The Willingness to Pay for a Cooler Day: Evidence from 50 Years of Major League Baseball Games

Kevin Kuruc[*] Melissa LoPalo[†] Sean O'Connor[‡]

Abstract

The climate-economy literature has documented adverse effects of extreme temperatures on well-being through mechanisms such as mortality, productivity, and conflict. Impacts due simply to discomfort are less well understood. This paper investigates individuals' valuations of weather using a revealed preference approach. We first quantify the decline in attendance at Major League Baseball games on hot and cold days. Leveraging this finding coupled with the historically-informed assumption of a horizontal supply curve, we infer a monetized estimate of the disutility of extreme temperatures. We estimate a $1.53 utility loss per hour of exposure to high temperatures, implying non-trivial aggregate welfare effects.

JEL classification: Q51; Q54
Keywords: climate amenities, environmental valuation, climate change

[*]University of Texas at Austin; University of Oklahoma. Email: kevinkuruc@utexas.edu

[†]Montana State University. Email: melissa.lopalo@montana.edu

[‡]Federal Housing Finance Agency. This article was written by Sean T. O'Connor in his private capacity. No official support or endorsement by the Federal Housing Finance Agency is intended or should be inferred. Email: seantjoconnor@gmail.com

I. Introduction

A large literature in economics documents the impacts of extreme temperature on well-being: more frequent episodes of high heat are expected to harm human health (e.g. Deschenes and Moretti (2009); Barreca (2012)), make workers less productive (e.g. Adhvaryu et al. (2019); Zhang et al. (2018)), and lower the overall GDP of economies (e.g. Burke et al. (2015); Hsiang (2010)), for example. These impacts entail large losses of individual well-being. One impact that is relatively less well documented is loss of utility due to the simple fact that extreme temperatures are unpleasant. While utility losses from this discomfort may be relatively small at the individual level, on the aggregate level they could add up to large losses in welfare due to their universality.

In this paper, we identify the causal effects of weather on individuals' engagement in outdoor activities in a context where we can isolate and infer a monetized value for the disutility of extreme temperatures. Major League Baseball (MLB) games provide an ideal setting: attendance has been scrupulously documented for over a century, creating a rich dataset of over 80,000 games in our sample period with substantial variation in game-day weather and local climate. We use variation in hourly temperature at game time to identify the impacts of weather on attendance at games between 1950 and 2000, controlling for a variety of observable characteristics of the game, stadium-specific seasonality in attendance, and time fixed effects. The results indicate significant effects of both very hot and very cold days, with attendance declining by 16% at temperatures over 95 degrees.

We next calculate the decline in ticket prices that would reverse the decline in attendance on very hot or cold days to estimate the change in willingness to pay for baseball. To do this, we exploit the fixed pricing behavior of MLB teams that prevailed through the 20th century. Only over the last couple of decades has it become common for teams to adjust ticket prices for a given game according to expected or realized demand. Since fixed prices imply a horizontal supply curve, the decline in attendance on hot or cold days represents a leftward demand shift. Given the well-documented fact that MLB teams have historically priced tickets on the inelastic region of demand (Krautmann and Berri, 2007), we assume a price elasticity of demand of 0.7, although we show the results for a range of elasticity assumptions. This implies that a 22.9% decrease in price would offset the 16% decline in attendance when the temperature is over 95 degrees. The average of inflation-adjusted MLB ticket prices in our sample period is approximately \$18,[1] so the marginal consumer's willingness to pay for a ticket drops by roughly \$4.11 at temperatures over 95 degrees. Using this same method for unpleasantly cold days generates a slightly larger \$5.17-\$6.55 for games with temperatures under 55 degrees. This exercise provides information about individuals' preferences over weather, which encompass both the simple disutility from

[1] Source: https://eh.net/encyclopedia/the-economic-history-of-major-league-baseball/. This is in 2020 dollars.

the discomfort of extreme temperatures, as well as any concern the individuals may have about health consequences from exposure.

We complement this first approach with an alternative method for estimating how willingness to pay for baseball varies with the weather. During the 2021 season, we scraped game-level ticket price data from Seatgeek, a secondary market for tickets. Estimating the impacts of game-time weather directly on the prices of ticket listings on Seatgeek, we find that prices fall by 8.4% ($1.51) at temperatures over 90 degrees. The reason we use a less extreme temperature threshold—above 90 rather than 95—is that there are too few games with temperatures above 95 in the one-year sample. However, running the same regression, using the same stadiums, on our longer attendance data (and imputing price effects with our elasticity assumption) generates an estimate that is quantitatively similar. These results help validate our main results and generalize the findings to more recent games.

Translating these declines in willingness to pay for baseball to a monetized disutility of extreme temperature requires a few additional assumptions. In particular, the availability of alternative outdoor activities in general creates a wedge between the price declines necessary to reverse the attendance declines on hot and cold days and the monetized disutility we hope to estimate. We show that as long as individuals are choosing between baseball games, indoor activities, and other outdoor activities that become absolutely less enjoyable in the heat or cold, the willingness to pay estimate will be weakly smaller (in absolute value) than the disutility of extreme temperature. An exception would occur if individuals have the option to engage in an outdoor activity that becomes more enjoyable in the heat (or cold) relative to mild weather. In these cases, the decline in willingness to pay for baseball would be partially attributable to the availability of a better outdoor option. However, we show that attendance holds steady at baseball games in covered stadiums across all temperatures, indicating that declines in baseball attendance at outdoor stadiums are unlikely to be driven by an increasing desirability of alternative outdoor activities.

In order to translate our estimates to a per-hour disutility of extreme temperature exposure, we note that the average game we study lasts 161 minutes, so our $4.11 estimate for games over 95 degrees implies an hourly disutility of $1.53. For exposure to under 45 degrees the corresponding figure is $1.93. To understand the aggregate implications of these numbers, consider that individuals in the U.S. spend approximately 30 minutes outside on the types of days that tend to produce temperatures in excess of 95 degrees.[2] Applying our disutility estimate only to these 30 minutes—conservatively not accounting for utility lost due to individuals switching to less preferred indoor activities in the heat—means that each

[2] In our sample, games at over 95 degrees tend to occur on days with an average temperature between 85 and 89.5 degrees (this is the range between the 25th and 75th percentile of daily average temperature for game times in this bin). The estimate of time spent outdoors on these days is from the American Time Use Survey.

individual suffers welfare losses of $0.77 per day. The increase in such days under a business as usual climate change scenario will then result in additional annual total losses from very hot days on the order of $2.31 billion by 2080-2090. This number does not entail a significant revision to our understanding of the social cost of carbon, especially since the reduction of cold days in these scenarios provide benefits not counted in that $2.31 billion. However, the annual reduction in value from increased hot days *is* on par with the largest weather disasters experienced on a year-to-year basis, and so is not in general insignificant.

This paper contributes to the prior literature in several ways. First, we contribute to the literature on valuation of non-market climate amenities by exploiting unique aspects of the market for baseball tickets. Our paper is the first to provide an estimate of the valuation of mild weather based on consumer choice in a market setting, in a sufficient statistics framework that does not require strong structural assumptions. Settings where it is possible to use well-identified estimates of behavioral change to monetize revealed preferences are rare, in part because climate is closely tied to consumption of non-market goods and leisure which are themselves difficult to value. Recent work by Chan and Wichman (2022) derives a valuation method for the effects of climate on consumer surplus for any given activity, based on changes in participation in that activity and baseline consumer surplus. Our approach estimates the direct effect of temperature on utility through well-documented market behavior rather than relying on travel cost methods to obtain consumer surplus. A second advantage of the setting here is that the market for MLB tickets is better documented and studied than nearly any other form of outdoor recreation, allowing for more credible empirical evidence of behavior change and a more robust literature from which to draw external assumptions.

More generally, previous literature has aimed to identify individuals' monetary valuation of climate through several methods. One method is to identify preferences over long-term climate by observing individuals' choices over where to live (Albouy et al. (2016); Sinha et al. (2018)). These papers estimate these preferences using cross-sectional variation in climate; however, correlation between climate and other place-specific characteristics may bias these estimates. A second method is to use survey data on self-reported happiness or life satisfaction to examine the impacts of heat on utility (Denissen et al., 2008; Feddersen et al., 2012). Finally, Baylis (2020) finds evidence that expressed sentiment on Twitter becomes more negative, accompanied by an increase in profanities. Baylis then compares the impacts of heat on Twitter sentiment with the impacts of changes in quarterly local wages to back out a monetary valuation of temperature, finding that individuals would be willing to pay approximately $5 to $12 to exchange a 30-35 degree day with a 20-25 degree day (degrees in Celsius).[3]

[3] In addition, there is a related literature that seeks to estimate willingness to pay to avoid the health impacts of pollution, by valuing the sickness caused by air pollution in a value of statistical life framework and/or by estimating the willingness to

In addition, we add to the literature documenting changes in individuals' allocation of time according to the weather. Graff Zivin and Neidell (2014) find that individuals in the United States tend to reallocate leisure time indoors on hot days, indicating that outdoor leisure becomes relatively less valuable in the heat. Similarly, Connolly (2008) finds that rainy weather lowers the opportunity cost of labor by making leisure time less attractive, causing individuals to increase their labor supply. More recently, Chan and Wichman (2020) document the change in time spent cycling according to the weather, leveraging a relatively active outdoor leisure activity.[4] We contribute further evidence that individuals prefer to allocate time away from outdoor leisure on hot (and cold) days.

The rest of the paper proceeds as follows. Section II. provides background on MLB attendance, discussing the history of determinants of what fans pay both at the box office and on secondary markets such as StubHub and Seatgeek. Section III. describes the data we use on game attendance and game day weather. Section IV. introduces our regression framework, which uses stadium-specific seasonally unusual weather variation to causally identify the nonlinear impacts of temperature on game attendance. Section V. presents the results of the analysis of weather and attendance. Section VI. investigates whether the effects vary significantly by usual climate or stadium type. Section VII. translates the main results on attendance to an estimate of the disutility of extreme temperatures, and Section VIII. presents the results of a second exercise examining changes in Seatgeek ticket prices according to the weather. Section IX. discusses the implications of our estimates for hourly disutility and aggregate welfare losses on days with extreme temperatures, and Section X. concludes.

II. Background: the Market for Baseball Tickets

This section describes the market for baseball tickets, focusing on the determinants of game attendance and secondary market prices and providing the foundation for the structural assumptions made in Section VII.A.

II.A Attendance and the Primary Market for Tickets

Historically, the majority of fans bought tickets in the primary market, directly from teams. Teams sold tickets at a fixed price: prices varied only according to seat location and were otherwise kept constant

pay for exposure-reducing technologies (see, for example, Deschenes et al., 2017; Ito and Zhang, 2020). Staying inside on a hot day may be partially thought of as a defensive instrument against health impacts of high heat, but the willingness to pay parameter we estimate encompasses the pure utility impacts of high heat as well, aside from the health impacts.

[4] Chan and Wichman (2020) find more adjustments on cold days, leading them to conclude that climate change will, on net, increase outdoor recreation. Like this study, they do find a negative adjustment on very hot days as well.

throughout the season.[5] Furthermore, the sports literature has consistently found that teams have historically priced their tickets in the inelastic region of demand, below price levels that would be expected to maximize revenues.[6] Under this pricing model, observed variation in attendance from game to game is thus driven by non-price-related demand-side factors.

In recent years, each MLB team has moved to a variable pricing model and later adopted dynamic pricing models for tickets. Teams using variable pricing still set prices before the season begins, but they vary prices by additional game-level characteristics that predictably determine demand, such as day of the week, season, and popularity of the rival team. Teams using dynamic pricing, on the other hand, vary prices over time within the season according to observed changes in demand. Courty and Davey (2020) documents year of adoption of variable and dynamic pricing for each team in MLB, showing that most teams had adopted dynamic pricing by 2016, but none had before 2009. The earliest adopter of variable pricing was the Colorado Rockies in 1997. By definition, we would expect dynamic pricing to move pricing into a more elastic region of demand, which affects the assumptions of our willingness to pay exercise. Partially due to this concern, in our main specification we restrict our sample to 1950-2000, when fixed pricing was dominant.

II.B The Secondary Market for Tickets

Secondary sales have always been part of the market for MLB tickets: historically, scalpers purchased tickets in advance, especially for popular games, and then sold them on game day in front of the stadium. Scalpers attracted public distaste by creating unwanted competition for tickets in the primary market and selling them at exorbitant prices (Bhave and Budish, 2017). In the past, sports leagues aggressively moved to limit ticket scalping, even leading to anti-scalping state laws in many places, albeit with limited effectiveness (Drayer, 2011).

The market for secondary sales of baseball tickets has exploded in recent years with the creation of websites such as Ticketmaster (on the Internet starting 1996), StubHub (2000), and Seatgeek (2009).[7] These websites typically allow sellers to upload tickets in electronic format, and have proprietary algorithms that determine a recommended price for the listing. Sellers then have the ability to set their

[5] For example, season ticket holders and other fans were mailed pre-ordered tickets with fixed prices printed on them before the season began.

[6] The sports literature has proposed several explanations for this: one is that cheap tickets may draw fans through the gates to buy high-margin concessions goods (Krautmann and Berri, 2007). Another is that inexpensive ticket prices may "hook" new loyal baseball fans (Ahn and Lee, 2003). Fort (2004b) shows evidence that teams may keep ticket prices low due to the incentive to attract public subsidies. Overall, the literature suggests that pricing in the inelastic region may not be inconsistent with profit maximization (Fort, 2004a).

[7] Teams for the most part have also embraced the secondary market: In 2007, MLB formally signed an agreement with StubHub, making the website an official outlet for ticket sales complete with the use of official team logos. MLB in return shares in the revenue from sales of tickets on StubHub.

own price or take the recommended price, and can change their listing price as desired. Stubhub, for example, allows sellers to set up a notification that alerts them if market conditions have changed such that their price now falls outside of a recommended range.

Sellers on these websites are a mixture of casual fans, season pass holders, and professional ticket brokers. An estimate from one major primary ticket seller suggests that about 30 percent of major league sports tickets are sold to brokers in the primary market (GAO, 2018).[8] On the other hand, using data from e-Bay, Sweeting (2012) shows that 88% of MLB ticket sellers list only a single set of tickets to a particular game, suggesting that fans (especially season ticket holders) make up the majority of the market. Prices on secondary markets are typically significantly higher than primary market prices: Sweeting (2012) finds that tickets are listed on StubHub at about twice their face value, on average. These prices represent a much more elastic region of the demand curve for baseball tickets (Diehl et al., 2015). The same paper shows evidence that prices are similar across multiple secondary market sites, suggesting that they can be thought of as part of the same overall market. With fully dynamic prices on both the primary and second markets in recent years, and a fixed supply of seats in a stadium, these facts indicate that secondary market prices are very likely to track consumer willingness to pay for baseball.

Another major change in this market over the last 30 years has been the construction of new, well-equipped, and often smaller, stadiums. Combined with the ease of buying tickets online, this trend has meant that stadiums have steadily filled up over the past few decades. Figure A1 shows a marked increase in average monthly attendance as a percent of capacity across all teams since 1990.[9] We would expect fuller stadiums and a more robust secondary market for tickets to limit our ability to observe an impact of weather on attendance at MLB games, which we discuss further below.

III. Data

III.A MLB Attendance Data

Our data source for MLB game attendance and game characteristics is www.retrosheet.org, following Neidell (2009).[10] Retrosheet has collected game records from 1871 through present. We compile atten-

[8] With the use of software bots, brokers have a significant advantage over fans in purchasing low-price tickets on the primary market, and this rent-seeking behavior significantly eats away at the efficiency gains of secondary markets (Leslie and Sorensen, 2014).

[9] Note that one point shows average attendance over 100% of stadium capacity. This can happen because many stadiums sell standing room only tickets, which they count as exceeding their seating capacity. This is rare in practice, and should be rarer still for the monthly averages in this figure. That particular point is March 2014, when teams' opening days fell on March 30th and 31st, so that monthly average is really just an opening weekend average. This counter-intuitive capacity definition will not cause any issues for our analysis as we estimate raw attendance levels, not as a percent of reported capacity.

[10] Neidell (2009) uses Retrosheet data to analyze the effects of smog alerts on attendance. The context of MLB games has also been used to estimate the elasticity of labor supply, focusing on stadium vendors (Oettinger, 1999).

dance records from every game for each stadium in the U.S. from 1950-2000. Pre-1950 games are dropped because of the sporadic availability of weather data, and post-2000 games are dropped to eliminate the influence of secondary markets and more sophisticated pricing strategies by teams, which arose around the turn of the century. Before 2000, we rely on the fixed pricing behavior of teams to translate attendance declines into a monetary estimate of the disutility of extreme temperature under minimal assumptions. In all, this provides a sample of over 80,000 games.

The source of attendance records tends to vary by league in our sample period, with the National League reporting turnstile attendance (so no shows do not count) and the American League reporting ticket sales as attendance (so that no shows were included as attending). When the two leagues merged in the early 1990's, the National League switched to reporting ticket sales (Shaikin, 2005). Theoretically, this could affect our results to the extent that individuals mis-forecast weather when buying their tickets, decide to not go upon learning of unpleasant weather, and are unable to give away or sell their ticket. Using ticket sales to measure attendance could bias our estimates towards zero in these cases, since no shows generate an attendance decline not captured in ticket sales. We discuss this further below. Table A1 provides information on each team included in the sample.[11] The teams' home stadiums are well distributed throughout the U.S., and cool, temperate, and hot climates are well represented, as shown in Figure A2. In addition to attendance records, Retrosheet reports game-level information on day of the week, time of day, scores, and detailed play-by-play records of the progression of the game. This information allows us to control for a battery of game-level observable characteristics to improve the precision of our estimates. Retrosheet also reports game start time in the event files, but in practice this information is often missing. Therefore, we impute game start time in these cases to match with hourly weather. Specifically, we assign a start time of 1PM to afternoon games and 7PM to evening games without a start time. In a few cases, we have double-headers where both games took place in the same afternoon and are missing start times; in these instances, we assign a start time of 1PM to the first game and 4PM to the second game.

Table 1 reports summary statistics on game-level characteristics. On average, around 20,670 people attend an American League game and 22,130 people attend a National League game.[12] With average stadium capacity close to 50,000, this implies that stadiums are on average less than half full. The existence of substantial leftover capacity in the average baseball game allows for variation in attendance based on game-day factors such as temperature. The average game lasts about 159-163 minutes, and nearly two out of three take place in the evening, as opposed to in the afternoon. In a small percentage of cases,

[11] We restrict our attention to only American MLB teams, since some of the weather data is only available for the contiguous U.S. However, the Toronto Blue Jays and Montreal Expos do appear in our sample as visiting teams.

[12] These leagues have historically differed slightly in the rules they use to govern gameplay, but are both part of the umbrella organization of Major League Baseball and play against one another throughout the season.

more than one game takes place in the same stadium on the same day, usually due to postponement of a previous game.

A number of MLB teams play in covered stadiums, with some transitioning to such stadiums in recent years. These stadiums are either permanently enclosed (domed) or are temporally covered for bad weather (including hot days) using a retractable roof.[13] In our main specification, we drop all domed stadiums and stadiums with retractable roofs from the sample. We then later use these stadiums in a placebo test—we investigate the impacts of weather on baseball attendance in settings where baseball games are held in indoor or climate-controlled stadiums. This eliminates the influence of exposure to the weather on the decision to go to a baseball game.

III.B Weather Data

We use hourly weather station data downloaded from the National Oceanic and Atmospheric Administration (NOAA)'s Integrated Surface Database (ISD). We assign temperature readings for each game based on the average of station readings within 50km of the stadium, weighted by inverse distance. For each station, we take the average of the three hourly readings starting at the starting hour of each game, to approximate average weather during the duration of the game. One concern with using weather station data is that entry and exit of stations from the database during our sample period may cause variation in our measure of temperature for each stadium (Dell et al., 2014). To circumvent this, we only use readings from stations that reported continuously throughout the period that each stadium is operational.[14] Figure 1 displays summary statistics of the merged MLB-weather data, showing the number of MLB games in our sample whose game-time temperature fell into each of the 12 5-degree bins we use for our regression analysis. On average, game-time temperatures are warm but relatively mild (the baseball season runs from April through September), with 70-75 degrees being the most common temperature range. However, we observe thousands of games that occurred in weather over 90 degrees or under 50 degrees.

All MLB games in the contiguous U.S. occur close to a reporting weather station, but many of the stations have historically not consistently reported hourly, or sub-daily precipitation.[15] Therefore, we

---

13 The home stadiums of the Arizona Diamondbacks, Seattle Mariners, Texas Rangers, Houston Astros, Milwaukee Brewers, and Miami Marlins all now have retractable roofs, and were built in the past few decades.

14 Specifically, we drop any station that does not report every year between when each stadium opened and closed. Additionally, we drop any station that is missing more than 5% of all three-hourly readings in the time period that the stadium was open. This leaves a small number of missing station-game observations. We follow Park et al. (2020) in filling in these missing observations with the nearest station's reading. However, our results are invariant to dropping these instead. More details about the construction of our weather data can be found in Appendix B.

15 Precipitation is an "additional" reporting field in the ISD, not a mandatory one.

source our precipitation data from the PRISM weather data set from 1950-2000.[16] These data files give weather information on a 2.5 x 2.5 mile gridded basis for the contiguous United States. The data are adjusted by Wolfram Schlenker to provide a balanced panel of weather station data: missing daily station readings are filled in by the distance-weighted average of the cumulative density function of surrounding stations. We merge this precipitation information with baseball attendance data by taking the average of the daily weather readings from the four surrounding weather gridpoints for each stadium, weighted by inverse distance between each gridpoint and the stadium. In a robustness check, we use daily average temperature data derived from the same source, calculated as the average of minimum and maximum temperature for each gridpoint.

Figure 2 displays average weekly game attendance for each of five major teams as it progresses throughout the season after opening day. Attendance is normalized to each team's average attendance in the 13th and 14th week of the year, when opening day usually occurs, to net out the influence of stadium capacity and team popularity. There is a pronounced peak in attendance around the mid-summer months for all five teams, but attendance falls much more in the August heat in the home stadiums of the Texas Rangers and the Atlanta Braves. This figure provides suggestive evidence that weather may have a quantitatively important impact on attendance at MLB games.

IV. Empirical Strategy

We estimate the impacts of temperature on attendance at baseball games using a semi-parametric specification. This flexibly identifies the effect of game-time temperature falling into a certain bin on game day attendance, relative to a reference bin of 70-75 degrees Fahrenheit. We estimate the following equation:

$$y_{isdmv} = \sum_{j} \beta_j \cdot Exposure_{sdm}(T_j) + \theta_{sm} + \lambda_{my} + \eta precip_{sdm} + \nu X_{isdmv} + \phi_v + \epsilon_{isdmv} \tag{1}$$

where y_{isdmv} is logged total attendance at game i at stadium s on date d in month m against visiting team v. β_j is the coefficient of interest and gives the effect of game-time temperature falling in bin j on attendance, relative to the reference bin of 70-75 degrees Fahrenheit. We estimate the impact of temperature falling into 11 5-degree bins: $<$45 degrees, 45-50, 50-55, 55-60, 60-65, 70-75, 75-80, 80-85, 85-90, 90-95, and $>$95.[17] The omitted bin is 70-75 degrees Fahrenheit, so the thought experiment is to compare the impact of game-time temperature falling into the bin of interest with the impact of temperature instead falling between 70-75 degrees. θ_{sm} are stadium by month of year fixed effects (58 stadiums by 12 cal-

[16] The data can be downloaded from Wolfram Schlenker's website, at http://www.columbia.edu/ ws2162/links.html

[17] We code each bin to include the upper bound in case of an exact match; for example (45-50], (50-55], etc.

endar months), which net out the average popularity of games at a team's home location at a certain time of year and therefore control for usual place-specific monthly weather conditions. Additionally, controlling for stadium rather than home team fixed effects nets out changes in attendance within teams across stadiums due to differences in capacity or stadium amenities. λ_{my} are month by year fixed effects, which net out any universal time-varying determinants of baseball attendance. $precip_{sdm}$ refers to a linear control for daily total precipitation on game day.[18] X_{isdmv} are controls for observable characteristics of a particular game that are likely important determinants of attendance. In our main specification, we control for day-of-week fixed effects, the share of the home team's last 100 games that it won, and whether the game was an afternoon or evening game, which is an important control because afternoon games will tend to be both hotter and less popular.[19] Finally ϕ_v are fixed effects for the visiting team, netting out variation in game attendance due to the popularity of the rival team. Standard errors are clustered at the stadium level.

IV.A Identifying Assumption

Our identifying assumption is that any unobserved determinants of attendance at a given MLB game are uncorrelated with variation in game time weather after controlling for stadium by month and month by year fixed effects. That is, instances of weather that are unusual for a certain stadium in a certain month are not correlated with attendance for any reason other than the direct effects of weather on the enjoyability of attending a baseball game.

Our identifying assumption would not be violated, for instance, by attendance rising in later months in the season as excitement for playoffs builds in a way that is correlated with temperature falling in early fall. It would, however, be violated if the expected quality of the game were to be affected by the weather. For instance, if players' performance is affected by unexpected heat, and fans change their attendance behavior in anticipation of this, this would violate our identifying assumptions. However, as we show in section V., we don't see a significant negative impact of unexpected heat on indicators of game quality such as the number of runs in a game.

[18] Prior work has found that attendance at baseball games falls on rainy days, but that attendance subsequently bounces back, suggesting the presence of habit-formation in baseball attendance (Ge et al., 2020).

[19] If a team moved cities and changed names, we consider the team in the new city to be a new team. In our sample period, for example, the Kansas City Athletics moved to Oakland and became the Oakland Athletics, the Washington Senators moved to Minnesota and became the Minnesota Twins, and the Seattle Pilots moved to Milwaukee and became the Milwaukee Brewers. If a team changed names but stayed in the same city, we consider the teams to be the same. For example, the Florida Marlins became the Miami Marlins and the Anaheim Angels became the Los Angeles Angels. More information can be found in Appendix B.

V. Main Empirical Results

Figure 3 displays the results of Equation 1 for the full sample of games from 1950-2000. The results corroborate the suggestion from Figure 2 that extreme heat dissuades baseball fans from attending games. Attendance appears to be highest between 75-80 degrees, and it falls by 16 percent for games taking place at temperatures over 95 degrees Fahrenheit relative to games between 70-75 degrees. Attendance is also significantly lower on very cold days, with a 20-25 percent decline on days below 55 degrees.

Table 2 displays the results of several alternative specifications of Equation 1. The first column controls only for stadium by month and month by year fixed effects. The second column adds controls for daily precipitation, day of the week, and whether the game took place in the afternoon or evening. Finally, column 3 displays the results of the full specification, with visiting team fixed effects and controls for the home team's performance over its past 100 games. This is the same specification displayed in Figure 3. The results are stable across specifications, though the impacts of high temperature become larger and more precise as the additional controls are added in columns 2 and 3. Columns 4 and 5 show that the results are robust to extending the time period to 2019 or excluding sellout games (games where attendance exceeded 95% of capacity). As a final robustness check, column 6 estimates Equation 1, but using indicators for daily average temperature (from the Schlenker dataset) falling in each bin as the independent variables. These bins are all shifted down by 5 degrees to account for the fact that daily average temperatures are systematically lower than game-time temperatures. The results are strikingly similar, raising confidence in our main specification.

One mechanism for a decline in baseball attendance on unpleasant days could be that fans expect the quality of the game itself to decline. If this were true, then part of the effect we observe on attendance may not reflect valuation of weather conditions, but rather, valuation of high-quality game play. In Table A2 we investigate this possibility by taking advantage of the rich data on game play in the Retrosheet records. The table displays the results of Equation 1, but using indicators of the quality of the game as outcome variables. The results document that runs, and particularly home runs, become more frequent on particularly hot days. This is a well known phenomenon (see, for example, Florio and Shapiro, 2016; Koch and Panorska, 2013; Lindholm, 2014). An important physical mechanism is that the baseball flies farther on hot days due to lower air density (Bahill et al., 2009). Additionally, strike outs decline.

This could, in principle, bias our results in either direction. If fans prefer games with fewer runs and better pitching performances, that could account for a share of the decreased attendance for hot games. Our own understanding of this context is that the opposite is true—fans seem to enjoy increases in offensive performance. It is beyond the scope of this project to formally elicit fans' optimal game style, but recent anecdotal evidence corroborates our belief that the average fan prefers more offensive output. In one of its most substantive packages of rule changes in many decades, Major League Baseball banned an effective run-preventing strategy ('the shift') for the 2023 season. Whether or not its own first-order goal was to favor offenses, the organization explicitly advertised this to fans as having the effect of 'increased batting average on balls in play' (MLB, 2023).[20] At the very least, it suggests that the MLB seems to believe that fans prefer more offense. In short, while we cannot formally rule out that differences in expected game quality (from the fans' perspective) drives any of our result, it would be surprising if the increased run-scoring environment of hot days is the cause of a (positive) share of the observed decline in attendance. On the other hand, cold temperatures decrease runs and increase strike outs, so altered game play could theoretically cause us to overestimate the disutility of cold. In either case, we would expect this to be a second-order effect—the marginal fan plausibly does not consider the correlation between weather and run scoring when deciding whether to attend a game.

One limitation of the attendance data is that we are unable to observe any changes in the composition of which seats are purchased. One possibility is that some fans choose to still attend baseball games even if it's very hot or cold, but they buy seats in more sheltered areas of the stadium. To the extent that these seats are more expensive, this behavior reveals that the fans value avoiding the disutility of extreme temperatures in a way that we are unable to capture. If this occurs, we will derive an underestimate of the disutility. Another consideration is that many baseball fans do not stay for the entire duration of the game, and may be particularly likely to leave early on very hot or cold days. This is another margin of adjustment on hot days that theoretically reveals a disutility of heat or cold but is not captured by our empirical approach. On the other hand, it could be the case that some baseball fans choose not to attend on unpleasant days partially because they anticipate leaving the game early due to the temperature. There are scenarios in which some decline in willingness to pay could be attributed to this expected loss in baseball consumption. We cannot rule this out but expect it to be a second-order effect; if fans anticipated losing significant value via reduced baseball consumption, many would be in a position to re-optimize and instead stay.

Additionally, fans' experience at baseball games partially depends on other fans—it is more fun to attend a game in a full stadium with a lot of fan energy than in a half empty, quiet stadium. As a result, it is possible that attendance declines on very hot days or very cold days are self-reinforcing: fans may be further disincentivized to attend if they expect others to not attend. To the extent that this occurs, we would overestimate the disutility of unpleasant temperatures by erroneously including the expected loss of consumption value due to lower fan energy in the stadium. Our empirical strategy does not allow us to rule this out, though again it is plausibly a second order effect, as it would serve to amplify an existing

[20] And rather than offset this with other new run-preventing rules, the other critical rule change (the pitch-clock) made the game more difficult on pitchers. This change likely was done to speed the game up rather than increase run-scoring, but it suggests the package of rule changes was not intended to be run-neutral.

decline.

Finally, and as discussed previously, the leagues differed in how they reported attendance numbers during our sample period. Teams in the American League reported ticket sales rather than true 'turnstile' attendance data throughout the sample period. Teams in National League, in general, reported turnstile attendance until the early '90s, at which point they transitioned to reporting ticket sales. Effects estimated on attendance using turnstile vs. ticket sales could differ to the extent that unpleasant weather creates additional no shows. This could happen if individuals buy their tickets before an accurate weather forecast is available and decide not to go once the unpleasant weather is revealed. One reason to believe that this may be a relatively small effect is that once a fan has purchased a ticket, they are less likely to be dissuaded by an unpleasant weather forecast because going to the game is now (monetarily) costless. Fans who have not purchased a ticket yet are plausibly more likely to be swayed by the weather. However, to investigate whether measuring attendance by ticket sales is an important source of attenuation bias, in Figure A3 we limit the sample to teams in the National League, prior to 1992, to narrow in on a sample where this attenuation bias should not exist. The results are qualitatively unchanged, indicating that this variation in reporting standards is not affecting our estimates in an important way.

VI. Heterogeneity, Adaptation, and Placebo Tests

In this section, we examine heterogeneity in impacts by stadium type and usual climate. First, we leverage the subset of games played in domed or retractable stadiums for a placebo test. These games are climate-controlled, eliminating our proposed mechanism for lower attendance. Any remaining effect is interpretable as the impact of temperature on the enjoyability of other outdoor activities. For example, if baseball was equally enjoyable regardless of the heat, but visiting a water park or beach became much more pleasant, we would still pick up a negative attendance effect in the absence of heat-related disutility at the baseball game. Instead, Figure 4 demonstrates that baseball attendance is unaffected by outdoor weather in climate controlled environments. This suggests the decline in attendance for outdoor games is driven by spectatorship becoming less pleasant, rather than other activities becoming more pleasant. We discuss the impacts of temperature on alternative choices of outdoor activities further below.

As mentioned in Section III., our dataset also spans a variety of climates, from the hot and humid home of the Florida (now Miami) Marlins to frigid game days in Minneapolis. This variety allows us to examine whether fans' adaptation to their local climate affects their valuation of mild weather. To do so, we split the sample into thirds by the stadium's annual average temperature over its operational period (based on weather station readings). We then run our main regression, interacting the indicators for daily average temperature falling in the bins of interest flexibly with dummies for the home stadium falling in the coolest, hottest, and middle thirds of the sample. The results are displayed in Figure A4; there is little systematic evidence of the effects differing by usual climate. Thus, there is a lack of evidence that individuals' valuation of mild weather as opposed to very warm weather attenuates as they adapt and acclimatize to warmer climates.

One potential reason for the lack of differential effect here, which contrasts with evidence from the mortality literature (e.g. Barreca et al., 2016), is that we have eliminated domed stadiums from our analysis. Building a dome or retractable roof is the primary adaptation investment teams can (and do) make to cope with their respective climate. In fact, one way to view our placebo exercises on the sample that has made these investments (Figure 4) is in support of existing results in the climate-adaptation literature: cities like Phoenix, AZ and Houston, TX, do not see negative attendance responses on hot days precisely because they have paid the fixed adaptation costs meant to flatten the response curve. In cases where this investment has not been made, there are arguably few tools fans in hot climates could use to adjust to hot game days that are unavailable in cool climates on similarly hot days. The similarity of the experience of sitting outside for a baseball game in different regions of the country, conditional on that day's weather, is what makes this a useful context to study.

Figure A5 goes a step further by running the specification for each city independently and plotting the coefficient for games above 90 degrees.[21] This is meant to provide a richer understanding of the distribution of effect sizes across cities, which could be evidence for or against the generalizability of the finding. To summarize, the point estimate on attendance is negative in a majority of cities when games are above 90 degrees. The main result does not appear to be driven by a small number of extreme responses and many of the coefficients are not statistically distinguishable from one another. There are exceptions: the coefficients for Chicago and Cincinnati are significantly different, for example. However, these differences do not appear systematic. Recall that Figure A4 separates these cities by usual climate—a dimension along which we might expect heterogeneity—and does not show differences in point estimates. This might provide some reassurance that any differences observed in Figure A5 are at least partially due to noise.

VII. Calculating the Disutility of Extreme Temperatures

This section leverages the attendance results above to make inferences about how the willingness to pay for baseball games changes on hot and cold days. We then formalize under what conditions the will-

[21] Games above 95 are rare enough to prevent a sufficiently powered analysis when restricting regressions to one city at a time.

ingness to pay for baseball games informs us about the disutility of extreme temperature. We focus our theoretical exposition on the disutility of heat for simplicity and turn to the impacts of cold temperatures at the end of the section.

VII.A Declining Attendance and the Willingness to Pay for Baseball

To generate an estimate of the change in the willingness to pay to attend a baseball game according to the weather, we must make two assumptions, consistent with previously mentioned facts about the market for baseball tickets. First, we assume that our estimates constitute a change in demand, rather than a joint quantity-price equilibrium adjustment. Equivalently, we assume a perfectly elastic supply curve at the individual game level; prices do not adjust to demand declines for a given game. As mentioned previously, baseball box office ticket prices historically have not adjusted on a game-by-game basis. Only in the mid-2000s did teams begin using dynamic pricing, wherein ticket prices would change based on realized demand for respective games. We conservatively estimate the main results on only the pre-2000 sample to avoid the possibility of dynamic pricing practices biasing our results.[22]

In order to map these quantity changes into price-equivalent changes, we additionally need an assumption on the local price elasticity of demand. We take 0.7 as the central value. Theoretically, one would expect teams to price at unit elasticity. MLB organizations are usually local monopolists with extremely low marginal costs along the fan-per-game margin. However, as discussed in Section II., Krautmann and Berri (2007) summarizes a literature that is puzzled by elasticity estimates that consistently fall below unity for professional sports pricing. Taking an average of the estimates presented in this canonical reference gives an elasticity of 0.4.[23] However, this is a dormant literature lacking recent high-quality studies–once sub-unit-elasticity pricing became a robust empirical finding, the literature moved on to modeling why teams would behave this way. So we hesitate to fully update to this low value, especially because, as we show below, our main results are inversely related to the assumed elasticity and so this would inflate our monetization. The value of 0.7 represents a conservative implementation of the robust finding of inelastic pricing. We explore the implications of the use of alternative estimates of the price elasticity of demand below.

Figure 5 formalizes how these two assumptions allow us to translate our coefficient estimates into monetary values, focusing on the effects of hot temperatures. Note that this is a log-log plot in accordance with the log-quantity estimates from Section V.. If prices are fixed with respect to weather, it must

---

[22] Though we note that the bias would be downwards, as we would observe a smaller quantity effect if prices did in fact adjust where we assume they do not.

[23] This likewise sits squarely within the range presented in a more recent literature review [-.581,-.275] focused specifically on professional (American) football (Diehl et al., 2015).

be the case that the estimated quantity change β, from Table 2, represents a demand shift. Against a horizontal pricing curve any quantity change identifies a shift in demand. To translate this leftward shift in demand into an equivalent price change, we ask by how much (log) ticket prices would need to fall to reverse the decline in attendance. Formally, the (local) log-price-change-equivalent, x_p, for any log-quantity change, β, takes the following form, where e_p^d is the price elasticity of demand.

$$x_p = \frac{1}{e_p^d}\beta \tag{2}$$

Returning to Figure 5, let the equilibrium outcome at pleasant temperatures be point A; the equilibrium after a demand decrease from hot temperatures is point B. Point C determines the price decrease that would induce an offsetting increase in (log-)quantity. The inverse of the price elasticity of demand governs the slope of this curve, and accordingly how much of a price decrease, x_p, is necessary to get to point C.

Holding quantity fixed between points A and C ensures that for the same number of consumers the value of attending the game surpasses the outside option. Assuming homogenous preferences over temperature, the price decrease generating point C compensates consumers for their lost surplus due to the high heat.[24] If point A was generated by price-temperature bundle $(p_0, 70-75^{\circ})$ then point C would be generated by $(.77p_0,>95^{\circ})$.[25] Inflation-adjusted ticket prices over the sample are approximately \$18, and so this change in the willingness to pay for baseball is about \$4.11 per game. Using alternative elasticity estimates generates a range between \$2.88 and \$7.20 (Figure 6).

Under a more realistic assumption of heterogeneous dislike of high heat, the analysis is less straightforward. However, as long as the variance of the surplus from baseball games is large relative to the variance of the disutility of heat, the setting is reasonably approximated by one of homogeneous disutility.[26] As the surplus associated with attending baseball games varies tremendously (think of people's taste for sporting events, their time-cost of travel, etc) and human physiology only allows for so much variance over temperature related utility, we think this is a reasonable assumption. Thus, the \$4.11 estimate above approximately represents the average change in willingness to pay.

Further, if the supply curve has a non-zero slope in practice, our empirical quantity change estimates understate the true demand shift. In the limit, a perfectly inelastic supply curve would produce no

[24] To see this, consider that the marginal consumer receives 0 surplus at point A, and the price decreases enough to exactly offset the loss from temperature and return her to zero surplus. Assuming homogenous preferences for weather, this estimate of the decline in surplus due to temperature for the marginal consumer is representative.

[25] $0.77 = 1 - \frac{1}{e_p^d}\beta = 1 - 0.229$

[26] This is because the random variable determining attendance—baseball surplus plus heat disutility—will have a variance approaching the variance of baseball surplus in the case where this term is much larger. Summing two normal random variables, for example, results in a variance that will be very near the larger term if it is indeed much larger.

quantity change even for a large demand shift. We would erroneously conclude there is *no* preference for the cooler day. This is an additional way in which our conclusions err on the conservative side.

VII.B The Willingness to Pay for Baseball and the Disutility of Extreme Temperatures

The exercise above produces an estimate for the decline in willingness to pay for baseball due to extreme temperatures. In this section, we describe how we translate this to an estimate of the disutility of extreme temperature, again focusing first on heat. This conceptual leap is straightforward in the simple case where individuals have two possibilities for leisure time: indoor leisure and attending an outdoor baseball game. In this case, during hot temperatures, demand for indoor leisure shifts out by the exact same amount that demand for baseball games shifts inwards, and the magnitude of that shift, as measured by our calculation, clearly represents the disutility of heat. In the real world, individuals have a number of outdoor activities to choose from, and the "disutility of heat" depends on that choice. For instance, running a marathon or going for a bike ride would plausibly involve a higher disutility of heat, while going to the beach may involve a lower disutility of heat. Among truly heat-exposed outdoor activities, baseball may have a relatively low disutility of heat, due to its sedentary nature, the fact that many seats are shaded, and due to the availability of cold refreshments.

Another, related consideration is the relationship between our estimates, the disutility of heat, and the availability of alternative outdoor activities on hot days. We calculate the magnitude of the price drop that would reverse the decline in attendance at hot temperatures, which will in general be mediated by potential attendees' best available alternative activities. On a hot day, many individuals are likely to substitute from outdoor to indoor activities, but there may also be substitution towards other outdoor activities that are less demanding in the heat. This substitution across outdoor activities may lessen or exaggerate the decline in attendance at baseball games, affecting the size of our disutility estimate.

Here we formalize the impact of available alternatives on our disutility of heat estimate. Let $u_{bb,75}, u_{bb,95}$ be the monetized utility that a fan gets from baseball (*bb*) when it's 70-75 and over 95 degrees, respectively; $u_{other,75}, u_{other,95}$ is the utility they get from whatever activity is their best available alternative in the heat.

The following two equations describe the relationship between the monetized utility of baseball and the alternative activity in mild and hot weather. We focus on the individual who would be the marginal fan on the hot day if the price were decreased by the $4.11 we estimate would return attendance to its mild-temperature level. This fan can be thought of as the "last" attendee brought back into the stadium with the price drop and we can infer that they prefer their outside option by exactly $4.11.

$$u_{bb,75} - u_{other,75} = X \tag{3}$$

$$u_{bb,95} - u_{other,95} = -4.11 \tag{4}$$

The first equation is agnostic regarding the preference ordering between the alternative activity ('other') that they have chosen instead of baseball on the hot day and baseball at mild temperatures: it could be that the fan would have gone to the baseball game on a mild day, so that X is a positive number, or that the fan would have preferred this alternative activity if it were mild, such that X is a negative number. Note that we do not assume here that the best available alternative activity in mild temperatures is the same as the best available alternative activity in hot temperatures. Equation (4) holds following our empirical inference that $4.11 is the price drop necessary to return this fan to indifference (so they must prefer their other activity by exactly $4.11 on hot days).

Subtracting (4) from (3) yields:

$$(u_{bb,75} - u_{bb,95}) + (u_{other,95} - u_{other,75}) = X + 4.11 \tag{5}$$

and subsequently:

$$(u_{bb,75} - u_{bb,95}) = X + 4.11 - (u_{other,95} - u_{other,75}) \tag{6}$$

The lefthand side of this equation then describes the decline in the utility of attending the baseball game in the heat, which is the disutility of heat parameter that we seek to estimate empirically. The 4.11 on the righthand side is our empirical estimate. Therefore, our empirical estimate differs from the disutility of heat, in general, by $X - (u_{other,95} - u_{other,75})$. Does this imply an over- or underestimate the disutility of heat? Consider four cases. First, suppose the other activity is an indoor activity whose enjoyability is unaffected by the weather. Then $u_{other,75} - u_{other,95} = 0$, and our estimates differ from the disutility of heat only by X. For a fan that was just on the margin of going to the baseball game or going indoors in mild temperatures, so that X is 0, our estimates pick up exactly the monetary disamenity of the hotter day. To the extent that X is positive, so that the individual strictly prefers baseball to going indoors on a mild day, we underestimate the disutility of heat.

A second potential case is a fan that would have gone to the baseball game if it were mild (so that X is weakly positive), but due to the heat switched to an activity that got absolutely more pleasant due to hot temperatures (so that $u_{other,95} - u_{other,75}$ is positive, and we overestimate the disutility of heat). An example could be a water-based outdoor activity such as a water park. However, it seems to us that there are very few outdoor activities that are absolutely more pleasant at over 95 degrees than they are at mild temperatures. In suggestive support of this, the evidence shown in Figure 4 suggests that an indoor baseball game (in a domed stadium or one with a retractable roof) does not become less preferable on average compared to outside options on hot days. This is evidence that attendance declines on hot days in open stadiums are not driven by an increasing appeal of other outdoor activities - if that were the case, we would expect attendance declines for indoor stadiums as well. Furthermore, most summer baseball games take place in the evening hours, and the effects of heat on attendance persist even for these later games. While individuals may plausibly substitute towards more preferable outdoor activities such as swimming in the afternoon hours, this is unlikely to be the individual's outside option between 7:00-10:00 pm on a weekday. Therefore, we do not expect this second case to have an important influence on our estimates.

A third possibility is that a $4.11 decline in ticket prices would bring in a fan who would have attended a baseball game on a mild day, but absent any price decline would have chosen an alternative activity on a hot day that got less pleasant due to the heat, but to a lesser extent than baseball (in other words, $u_{other,75} - u_{other,95} < u_{bb,75} - u_{bb,95}$). In this case, $u_{other,75} - u_{other,95}$ is a positive number, and we underestimate the disutility of heat by $X + u_{other,75} - u_{other,95}$. Finally, a fourth possibility is a fan that would have strictly preferred a different activity (perhaps a more active one such as hiking or jogging) on a mild day to baseball, such that X is a negative number, but the disutility of heat for the alternative activity is so large that the $4.11 decline in ticket prices induces them to switch to baseball on a hot day.[27] In this case, it's theoretically ambiguous whether $X - (u_{other,95} - u_{other,75})$ is a positive or negative number and therefore whether we underestimate or overestimate the disutility of heat for baseball. However, if the strength of the individual's preference for the alternative activity were larger than the disutility of heat of that activity, it's unlikely the individual would be close enough to the margin of going to the baseball game on a hot day to be induced to switch by a $4.11 price drop. To see this, note that the decline in the level of attendance on a hot day means that there are enough people in the first three cases to induce to switch back to baseball with a price drop less than or equal to the disutility of heat without needing to compensate someone in this fourth category more than the disutility of heat to attend the baseball game.

The relationship between the empirical estimate we derive from our exercise and the true "disutility of heat" in this context will depend on what combination of these four cases is closest to the margin of attending or not attending a baseball game on a hot day. However, for the range of most common alternative activities at extreme temperatures, our empirical estimate is likely lower than the true disutility of heat. The amount by which we underestimate the disutility of heat will depend on the quantitative

[27] In the first case we discuss, we treat indoor activities as the outside option, assuming that the enjoyability of being indoors does not depend on the outdoor temperature. However, this is unlikely to be the case for individuals without air conditioning. It's even possible that these individuals will fall into this fourth category: they would have stayed inside on mild days but prefer a baseball game to the indoors on very hot days.

values for these terms, which are difficult to precisely ascertain. One possibility is that both terms are small, or close to 0. X is close to zero, as mentioned previously, if individuals are relatively similar in their disutility of heat, so that the marginal fan on a mild day is the same as the marginal fan on a hot day. It also seems plausible that the last individual brought back to the game with a $4.11 price decline had switched to an activity with a substantially smaller, or even zero, disutility of heat (such as simply going inside). If this is close to the real-world case, then our estimate aligns closely with the true disutility of heat.

We've focused thus far on the change in willingness to pay for baseball on a hot day, despite the fact that in Section V. we estimated even larger impacts of cold days on attendance. Following the same procedure, we would estimate a willingness to pay for warmer temperatures in the range of $5.17-$6.55. One reason for the relatively large disutility of cold estimate in this setting could be that baseball attendance is a sedentary activity, so attendees may feel cold in a temperature range where they may feel more comfortable participating in a more active outdoor activity. For example Greenstone et al. (2019) finds that the optimal temperature to run a marathon is in the 40's, which is a temperature range where we estimate very large negative effects on baseball attendance.

VIII. Seatgeek Ticket Prices and the Willingness to Pay for Baseball

In this section we exploit secondary market ticket price data directly for a much smaller sample of games to estimate the change in the willingness to pay for baseball under fewer structural assumptions. While box office ticket prices have historically been fixed with respect to variation in game-level characteristics, for the last 15+ years it has been possible to purchase tickets on online secondary markets, such as StubHub, Ticketmaster, and Seatgeek. On these secondary markets individuals update prices frequently and appear to optimize at the game level, rather than the season level (Sweeting, 2012). The existence of secondary markets, and generally more ways to buy and sell tickets with flexible prices, means tickets are allocated more efficiently. This amounts to making the market look more like one with perfectly inelastic supply—demand changes will then be reflected in price changes. Therefore, this setting is one where it makes more sense to translate a change in ticket prices directly into a measure of the change in willingness to pay for baseball.[28]

During the 2021 MLB season, from May to October, we scraped daily price data of listings on Seatgeek.com for each MLB game.[29] The data contain information on average, median, minimum, and

[28] To our knowledge, only Paul and Weinbach (2013) has previously examined the impacts of weather on baseball ticket prices, and we are the first to control for place-specific seasonality to separate out the causal effect.

[29] These sites do not make the necessary historical data available to perform this analysis for the entire period the secondary market has been active.

maximum prices for ticket offerings on Seatgeek for each game, as well as the overall count of listings. We pulled price information for listings left available on game day for each game, as well as for each day in the prior week. According to the Sports Business Journal, Seatgeek held 19.5% of the secondary market for Major League Baseball tickets in 2021 (McCormick, 2022). As mentioned in Section II., previous evidence finds that tickets on different websites tend to be comparable in price and move together (Sweeting, 2012).

The Seatgeek price data can be similarly merged with game-time temperature information using weather station data from the NOAA ISD.[30] For precipitation, we use the PRISM daily climate data, which is the same underlying precipitation data source as we used for the attendance data, but without the corrections to create a balanced panel undertaken by Schlenker (the corrected dataset as of the time of this writing runs only through 2019).[31] Table A3 displays summary statistics of the ticket prices. The ticket prices have a strong rightward skew: the average average price is around \$445, while the average median is only \$80. Accordingly, we drop the top 1% of prices in our main analysis to reduce the influence of outliers. A small number of games have very small numbers of reported listings, perhaps due to API error. We drop the lowest 1% of listings in our main analysis, but the results are invariant to retaining those observations or dropping larger percentages. Ticket prices decrease steadily in the days leading up to game day, while the overall number of listings declines. This matches with evidence from Sweeting (2012) that prices fall over time, as the opportunity cost of not holding the ticket in the next period declines for the seller.[32] Figure A6 shows how median prices and counts of listings evolved throughout the 2021 season. Prices tend to be higher at the beginning and end of the season, and on weekends. The count of listings increased throughout the season.

For our primary analysis, we restrict the sample to only game day listing prices so that each game contributes only one observation to the dataset and so that prices contain as much information about responses to game day weather as possible. Given that tickets are "perishable," most of the activity on the secondary market will have completed by this time, and the price will be a function of remaining inventory given the past realizations of demand. Weather forecasts will have been available for several days, presumably growing more precise as game day approaches.

To investigate the impacts of game day weather on prices, we run a slightly altered version of Equa-

[30] We use the same procedure here to ensure that we have a balanced panel of weather stations: we remove stations that were missing more than 5% of observations for the 3-hour average temperature during 2021.

[31] To conduct the merge, we interpolated the weather readings from the surrounding gridpoints in the weather data.

[32] As the paper mentions, theoretically consumers should rationally wait to buy tickets later, when ticket prices are lower, given this phenomenon. But customers might have costs of waiting, partially due to the investment required to attend a game, that could prevent this behavior from smoothing out the decline in prices as a game approaches.

tion 1:

$$y_{isdmv} = \sum_j \beta_j \cdot Exposure_{sdm}(T_j) + \theta_{sm} + \eta precip_{sdm} + \nu X_{isdmv} + \phi_v + \epsilon_{isdmv} \quad (7)$$

The fixed effects specification differs for this exercise due to the one-year sample period. y_{isdmv} is logged average listing price for game i at stadium s on date d in month m against visiting team v. β_j is the coefficient of interest and gives the effect of game-time temperature falling in bin j on prices, relative to the reference bin of 70-75 degrees Fahrenheit. For this data sample, we continue to use 5-degree bins, but the highest bin is >90 instead of >95. Due to the proliferation of domed stadiums and stadiums with retractable roofs in cities with warm climates in recent years, along with the limited sample period, there are too few observations above 95 degrees to use to estimate the effect separately. θ_{sm} are stadium by month of year fixed effects. X_{isdmv} are controls for observable characteristics of a particular game: day of week fixed effects, and whether the game was an afternoon or evening game. Finally ϕ_v are fixed effects for the visiting team. Standard errors are clustered at the stadium level. In this one year period, there were only 22 stadiums in the sample, raising concerns that there are too few clusters to accurately estimate clustered standard errors. To address this concern, we calculate wild cluster bootstrapped confidence intervals for the estimates (Cameron et al., 2008). We implement the calculations with 999 replications using Stata's boottest module (Roodman, 2021).

Figure 7a displays the results of Equation 7 on the full 2021 season sample. The results show declines in ticket prices at warmer and colder temperatures, with average ticket prices falling by about 8.4% on days over 90 degrees. Given the noise in these estimates, in Figure 7a we display the results of several alternative specifications, additionally regressing log average price on a quadratic in game time temperature and binned daily average temperature from PRISM, as well as estimating the impacts of binned game-time temperature on logged median price. Each of these specifications broadly shows the same U-shaped pattern.[33]

Using the same prevailing average ticket prices as with the attendance exercise, an 8.4% decline in ticket prices amounts to a \$1.51 estimate of willingness to pay for mild weather rather than weather $>$ 90 degrees. This is a smaller point estimate than derived from the attendance data but is not directly comparable given differences in sample and regression specification necessitated by the 2021 sample (for example, the upper most bin is above 90 degrees because of the small number of observations above 95 degrees).

Figure 7b documents that when a comparable regression is run on the attendance data, quantitatively similar results are obtained. Two adjustments need to be made to harmonize the analyses. First, we

[33] The U-shape relationship is shifted left for daily average temperature due to the fact that daily average temperature is usually lower than game-time temperature.

create a Retrosheet data sub-sample that restricts to only the stadiums that are observed in the 2021 Seatgeek data.[34] Second, we combine the bins above 90 and under 50 degrees in our attendance exercise to match the Seatgeek exercise. The results of this exercise are overlaid with the results of the exercise using the Seatgeek data in Figure 7b. The y-axis here is a common monetized disutility implied by the results for each temperature bin.[35] As might be expected, combining the >90-degree bins in the attendance analysis results in a smaller estimated effect on attendance, but the results show a similar quantitative pattern emerging from both exercises. And in fact, all of the confidence intervals for each coefficient in Figure 7b overlap.

VIII.A Robustness

Given that we are using game-level summary statistics of listing prices as our outcome variable, rather than individual seat-level data, one concern is that changing average ticket prices may partially reflect changing composition of seat offerings on Seatgeek. One example of this concern: if many fans that are just on the margin of purchasing or not purchasing a ticket to attend a game are likely to buy cheap, bleacher-type seats on Seatgeek, and these marginal fans are especially likely to be dissuaded by high game day temperatures, then a large negative impact on average prices of listings on Seatgeek could reflect these "nosebleed" seats being left unsold, as opposed to especially large changes in demand for baseball tickets. In this example, the underlying mechanism for the effect on prices is still reduced demand, but the magnitude of the effect does not necessarily correspond to the quantity of potential attendees dissuaded by heat. Sweeting (2012) shows evidence that the composition of available seats on StubHub does not change as game day approaches, reducing this concern.[36]

To investigate the potential contribution of these composition effects directly in our data, in Table 3, we examine the impacts of the weather on alternative summary statistics of listing prices. Column 1 displays the same specification as the blue line in Figure 7, with log average price as the outcome variable. Column 2 looks at the impact of weather on the minimum price among the Seatgeek offerings on game day. If cheap tickets are more likely to go un-purchased on hot days relative to average or expen-

[34] These stadiums are: Truist Park, Oriole Park at Camden Yards, Fenway Park, Wrigley Field, Guaranteed Rate Field, the Great American Ball Park, Progressive Field, Coors Field, Comerica Park, Kauffman Stadium, Angel Stadium of Anaheim, Dodger Stadium, Target Field, Citi Field, Yankee Stadium, RingCentral Coliseum, Citizens Bank Park, PNC Park, Petco Park, Oracle Park, Busch Stadium, and Nationals Park.

[35] For the attendance estimates, we multiply each coefficient by the inverse price elasticity of demand and then by the estimated average ticket price of $18, as described in Section VII.A. For the Seatgeek exercise, we multiply the estimates of the percent change in price by the average ticket price of $18 to back out a monetized disutility.

[36] Sweeting (2012) also tests the hypothesis that late buyers are more price sensitive by reasoning that sellers of more expensive tickets would have to slash their prices more dramatically than holders of cheap tickets to attract buyers in the days before a game. However, he finds no evidence of this, suggesting that the composition of buyers does not change leading up to a game.

sive tickets, this could drive down minimum prices to a greater degree than average prices. However, the effects in this specification are not statistically distinguishable and are qualitatively similar. Model 3 uses the maximum ticket price as the outcome variable: the price of the most expensive ticket offering seems little affected by temperature, but the coefficients are noisy. Column 4 uses the median price as the outcome variable, with similar albeit larger results to column 1. Finally, in column 5 we examine the impacts of temperature on the number of listings remaining on Seatgeek on game day. Unsold listings increase on very hot days and cold days, although the coefficients are very noisy. Any increase in listings may reflect a combination of more fans looking to sell their tickets and fewer fans willing to buy when the game day is forecasted to be very hot or cold.

The interpretation that both decreases in demand and increases in supply can be thought of as reductions in desire to attend baseball games may be complicated by the presence of professional ticket brokers on these secondary markets. Professional brokers may have an incentive to lower their price in anticipation of lower demand (willingness to pay) once very high temperatures make their way into the forecast, but the mechanism is no longer as direct since the brokers are not deciding whether to attend the game themselves. It is presumably the case that brokers play a bigger role in the market for tickets in high-demand games, so one way to gain suggestive evidence on the role of brokers is remove especially high-demand games from the sample. In column 6, we display the same specification from column 1, except we exclude games that had attendance over 95% of the stadium's capacity according to the Retrosheet data. The effects are very similar.

A final concern with using the 2021 Seatgeek data is that the Covid-19 pandemic had a major impact on operations for the season. This could impact prices on Seatgeek: many stadiums capped attendance at some percent of capacity, potentially restricting supply, and fans may have still been reluctant to attend crowded events, limiting demand. To the extent that Covid-related trends in attendance and ticket purchasing behavior interact with game day weather, this could create an external validity concern. For example, if games are regularly sold out to the limited capacity, where they would not have been sold out with full capacity, this could limit our ability to observe the impacts of heat on ticket prices, biasing our effects towards zero. On the other hand, if fans are on more "on the fence" about attending baseball games due to Covid, they may be more sensitive to weather impacts, causing us to estimate larger effects than in a typical season. Ultimately, this is an external validity issue, although the fact that we estimate similar effects using 2021 ticket prices as we get for 1950-2000 baseball attendance ameliorates this concern.

IX. Hourly Disutility of Extreme Temperature and Aggregate Welfare Losses

In Section VII.A, we discussed how our results translate into an estimate of the disutility of extreme temperature that a fan experiences at a baseball game. It is straightforward to convert the game-level estimates to per-hour estimates recognizing that the typical baseball game is about 161 minutes. The $4.11 per-game monetized cost for games at over 95 degrees then corresponds to $1.53 of utility loss per hour, or $1.97 for games below 45 degrees. These are our central estimates of the disutility of extreme temperatures: individuals would pay about $1.53 to avoid an hour spent outside in temperatures over 95 degrees or $1.97 to avoid an hour outside below 45 degrees. Recall that these values came from applying a price elasticity of demand of 0.7. Leaning fully on the empirical studies and applying a value of 0.4, we would estimate a higher $2.68 ($3.45) per hour for temperatures over 95 (under 45); the most conservative value of 1.0 produces a lower $1.07 per hour cost for over 95, or $1.38 for under 45. It is worth noting that in converting the utility costs of a three-hour baseball game to a per-hour cost, we assume that these utility costs scale linearly with time spent outdoors. If instead, the marginal costs are convex, this would cause our dollar-per-hour estimate to be an overstatement—one hour outdoors would be less unpleasant than the average of the three hours outdoors we observe. The opposite is true if marginal costs are concave.[37]

A natural extension of this exercise is to estimate the damages from climate change implied through this channel of disutility of exposure to extreme temperature. First setting aside the change in the distribution of cold days and only considering additional exposure to hot days, note that we estimated an hourly disutility of >95 degree temperatures. In our sample, these temperatures tend to occur on days with an average temperature between 85-90 degrees. We calculate that the average American will be exposed to about 29 additional days in this temperature range by 2080-90 under a business-as-usual (SSP5-8.5) warming scenario, compared with 1980-2000, or 17 additional days under a more moderate warming scenario (SSP2-4.5).[38] Estimates from the American Time Use Survey suggest that the average American spends about 30 minutes outside on these hot days, which implies utility loss per person

[37] While we typically think of costs as convex, decreasing marginal costs could arise if there are important fixed costs associated with experiencing unpleasant temperatures. For example, a bicycle commute on a very hot day requires two instances of becoming very sweaty (to and from work), relative to this happening only once if that time were spent outside in one continuous stretch.

[38] To derive this estimate, we calculate the population-weighted average number of days that produce temperatures >95 degrees under historical weather distributions. For this calculation, we take the average count of days in each gridpoint (using Wolfram Schlenker's version of the PRISM data) with daily average temperatures that tend to produce >95 degree temperatures (85-89.5 degree days) from 1980-2000. We then multiply that by the local population in 2020 using gridded population data from the Socioeconomic Data and Applications Center (SEDAC). Finally, we divide by the total population to get an estimated count of average days of exposure per person. We then compare that to the projected count of the same type of days in 2080-2090 that the average American will be exposed to, using projected population data for 2090. For the projected weather data, we use the average of the 35 global climate models (GCM's) from the Coupled Model Intercomparison Model 6 (CMIP6). The data are downloaded from the NASA Earth Exchange (NEX) Global Daily Downscaled Projections (GDDP).

per hot day is approximately $0.77. Therefore, these additional days of exposure will lead to additional annual welfare loss due to heat of $2.31 billion ($1.34 billion) by 2080-2090 under SSP5-8.5 (SSP2-4.5), according to our estimates.[39,40] On the one hand, these damages are small in the context of overall estimated damages from climate change. These are projected to be on the order of 5-20% of GDP annually in the U.S. by the end of this century in the business-as-usual scenario we study, or about 1 - 5 trillion dollars.[41] The estimates in this paper would not imply an important revision to optimal emissions trajectories or the overall costs of climate change. On the other hand, since 1980, the US has experienced an average of only 8 'Billion Dollar Weather Disasters' per year (NOAA, 2023). The annual costs of just the unpleasantness of 30 minutes per day in hot weather spread across the population would be large enough to be counted among this small class of extreme weather events.

A full accounting of the costs and benefits of climate change would incorporate the fact that fewer very cold days will occur, which will have benefits for outdoor recreation (see e.g., Chan and Wichman, 2020). If we incorporate the projected loss in exposure to cold days, the implied benefit using our estimates would entirely eliminate the aggregate costs of climate change via the mechanism of exposure disutility. We interpret this exercise to provide suggestive evidence that climate change will improve welfare for time spent outdoors in sedentary activities. However, more research is needed to understand the net effects of climate change on welfare from outdoor activities overall.

There are several caveats with aggregating our per-hour utility cost estimates to an estimate of population-level utility loss on days with extreme temperatures in this way. First and most importantly, it is likely the case that individuals optimize on hot and cold days by re-arranging their schedules to spend time outside during milder times of the day. For example, an individual might choose to go for a run in the early morning instead of during the afternoon on a hot day and thus avoid some of the utility costs of heat. This margin of adjustment (which is not available to baseball spectators, who face a set schedule of baseball games) would mitigate the aggregate welfare losses we estimate. In addition, any nonlinearity in the marginal costs of exposure to extreme temperatures will also affect our estimate of aggregate welfare loss. In light of these complications, our estimates of aggregate costs should be taken as suggestive predictions of the order of magnitude of these effects rather than precise forecasts.

[39] These 30 minutes do not include outdoor labor time, as the American Time Use Survey does not separate out whether the respondent's workplace is outside or inside. However, we can follow Graff Zivin and Neidell (2014) in defining "high risk" industries where a large proportion of work is likely to take place exposed to outdoor temperature: agriculture, forestry, fishing, and hunting; mining; construction; manufacturing; and transportation and utilities industries. If we include time spent at work in these industries as time spent outdoors, we get that the average American spends 61 minutes outdoors on hot days, so that damages from the additional 29 hot days will instead be $4.51B under a business as usual scenario.

[40] We conservatively estimate aggregate welfare losses only using time that individuals spend outside on hot days, not including welfare loss due to individuals spending less time outdoors on hot days.

[41] See, for example, Howard and Sterner (2017) or Barrage and Nordhaus (2023). These are 2022 dollars, based on a simplifying assumption of zero economic growth over this period. To the extent that the economy is much larger by the end of century the value of losses would be higher (but presumably so would the willingness to pay to avoid heat).

X. Conclusion

In this paper, we establish a new setting for examining the impacts of extreme temperatures on individual utility and leisure time. We first estimate the causal impacts of weather on MLB attendance, showing that preferences for engaging in this passive, outdoor leisure activity significantly decline at very hot and very cold temperatures. Next, we leverage these well-identified estimates to directly estimate fans' valuation of mild game-day weather using two methods. First, given fixed pricing of baseball tickets and findings from the literature that pricing is in the inelastic region of demand, we use the attendance estimates to back out an equivalent price change that would induce the same impact on quantity of tickets demanded, finding that fans would pay $4.11 to replace a game-time temperature over 95 degrees with one between 70 and 75 degrees. Second, we use secondary ticket pricing data from Seatgeek to directly estimate the impacts of heat on ticket prices, finding effects of a similar magnitude.

The results speak to an understudied implication of climate change. Heat waves are known from an established literature to pose significant risks to human health, productivity, and economic output, but they also cause a widespread increase in disutility due to the discomfort of hot weather. The estimates derived from this paper suggest that these effects are not large enough to alter our understanding of the costs of climate change, but are still economically meaningful.

References

Adhvaryu, Achyuta, Namrata Kala, and Nyshadham Nyshadham (2019) "The light and the heat: Productivity co-benefits of energy-saving technology," *Review of Economics and Statistics*, 1–36.

Ahn, Seung C and Young H Lee (2003) "Life-Cycle Demand for Major League Baseball," in *Western Economics Association International Conference, Denver, CO*.

Albouy, David, Walter Graf, Ryan Kellogg, and Hendrik Wolff (2016) "Climate amenities, climate change, and American quality of life," *Journal of the Association of Environmental and Resource Economists*, 3 (1), 205–246.

ATUS (2003-2018) "American Time Use Survey," https://www.bls.gov/tus/data.htm.

Bahill, A Terry, David G Baldwin, and John S Ramberg (2009) "Effects of altitude and atmospheric conditions on the flight of a baseball," *International Journal of Sports Science and Engineering*, 3 (2), 109–128.

Barrage, Lint and William D Nordhaus (2023) "Policies, Projections, and the Social Cost of Carbon: Results from the DICE-2023 Model,"Technical report, National Bureau of Economic Research.

Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S Shapiro (2016) "Adapting to climate change: The remarkable decline in the US temperature-mortality relationship over the twentieth century," *Journal of Political Economy*, 124 (1), 105–159.

Barreca, Alan I (2012) "Climate change, humidity, and mortality in the United States," *Journal of Environmental Economics and Management*, 63 (1), 19–34.

Baylis, Patrick (2020) "Temperature and temperament: Evidence from Twitter," *Journal of Public Economics*, 184, 104161.

Bhave, Aditya and Eric Budish (2017) "Primary-Market Auctions for Event Tickets: Eliminating the Rents of 'Bob the Broker'?"Technical report, National Bureau of Economic Research.

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel (2015) "Global non-linear effect of temperature on economic production," *Nature*, 527, 235–239.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller (2008) "Bootstrap-based improvements for inference with clustered errors," *The review of economics and statistics*, 90 (3), 414–427.

Chan, Nathan W and Casey J Wichman (2020) "Climate Change and Recreation: Evidence from North American Cycling," *Environmental and Resource Economics*, 76 (1), 119–151.

——— (2022) "Valuing Nonmarket Impacts of Climate Change: From Reduced Form to Welfare," *Environmental and Resource Economics*.

Connolly, Marie (2008) "Here comes the rain again: Weather and the intertemporal substitution of leisure," *Journal of Labor Economics*, 26 (1), 73–100.

Courty, Pascal and Luke Davey (2020) "The impact of variable pricing, dynamic pricing, and sponsored secondary markets in major league baseball," *Journal of Sports Economics*, 21 (2), 115–138.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken (2014) "What do we learn from the weather? The new climate-economy literature," *Journal of Economic literature*, 52 (3), 740–798.

Denissen, Jaap JA, Ligaya Butalid, Lars Penke, and Marcel AG Van Aken (2008) "The effects of weather on daily mood: a multilevel approach.," *Emotion*, 8 (5), 662.

Deschenes, Olivier, Michael Greenstone, and Joseph S Shapiro (2017) "Defensive investments and the demand for air quality: Evidence from the NOx budget program," *American Economic Review*, 107 (10), 2958–89.

Deschenes, Olivier and Enrico Moretti (2009) "Extreme weather events, mortality, and migration," *The Review of Economics and Statistics*, 91 (4), 659–681.

Diehl, Mark A, Joel G Maxcy, and Joris Drayer (2015) "Price elasticity of demand in the secondary market: Evidence from the National Football League," *Journal of Sports Economics*, 16 (6), 557–575.

Drayer, Joris (2011) "Examining the effectiveness of anti-scalping laws in a United States market," *Sport Management Review*, 14 (3), 226–236.

Feddersen, John, Robert Metcalfe, and Mark Wooden (2012) "Subjective well-being: Weather matters; climate doesn't."

Florio, John and Ouisie Shapiro (2016) "How Will Rising Temperatures Change Baseball?," https://www.theatlantic.com/science/archive/2016/10/how-will-rising-temperatures-change-to-baseball/502778/.

Fort, Rodney (2004a) "Inelastic sports pricing," *Managerial and decision economics*, 25 (2), 87–94.

——— (2004b) "Subsidies as incentive mechanisms in sports," *Managerial and Decision Economics*, 25 (2), 95–102.

GAO (2018) "Event Ticket Sales: Market Characteristics and Consumer Protection Issues,"Technical report, Government Accountability Office.

Ge, Qi, Brad R Humphreys, and Kun Zhou (2020) "Are fair weather fans affected by weather? Rainfall, habit formation, and live game attendance," *Journal of Sports Economics*, 21 (3), 304–322.

Graff Zivin, Joshua and Matthew Neidell (2014) "Temperature and the Allocation of Time: Implications for Climate Change," *Journal of Labor Economics*, 32 (1), 1–26.

Greenstone, Michael, Faraz Hayat, and Michael Galperin (2019) "Technical Support Document: The Marathon, The Climate and Your Race Against Time."

Howard, Peter H and Thomas Sterner (2017) "Few and not so far between: a meta-analysis of climate damage estimates," *Environmental and Resource Economics*, 68 (1), 197–225.

Hsiang, Solomon M. (2010) "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America," *Proceedings of the National Academy of Sciences USA*, 107 (5), 15367–72.

Ito, Koichiro and Shuang Zhang (2020) "Willingness to pay for clean air: Evidence from air purifier markets in China," *Journal of Political Economy*, 128 (5), 1627–1672.

Koch, Brandon Lee D and Anna K Panorska (2013) "The impact of temperature on major league baseball," *Weather, Climate, and Society*, 5 (4), 359–366.

Krautmann, Anthony C and David J Berri (2007) "Can we find it at the concessions? Understanding price elasticity in professional sports," *Journal of Sports Economics*, 8 (2), 183–191.

Lawrence, Jesse (2014) "Here's how Major League Baseball Teams are fighting the secondary ticket market," *Forbes*, https://www.forbes.com/sites/jesselawrence/2014/03/14/major-league-baseball-teams-are-combatting-the-secondary-market-by-selling-less-tickets/?sh=40d8ff88e999.

Leslie, Phillip and Alan Sorensen (2014) "Resale and rent-seeking: An application to ticket markets," *Review of Economic Studies*, 81 (1), 266–300.

Lindholm, Scott (2014) "Hitting and Temperature," https://www.beyondtheboxscore.com/2014/3/27/5546952/hitting-temperature-baseball.

McCormick, Bret (2022) "Dealmaking, the full return of live events, and league and team business up for grabs are positioning sports ticketing for years of growth," *Sports Business Journal*, https://www.sportsbusinessjournal.com/Journal/Issues/2022/02/28/In-Depth/Ticketing.aspx.

MLB (2023) "Pitch timer, shift restrictions among announced rule changes for '23," *mlb.com*, https://www.mlb.com/news/mlb-2023-rule-changes-pitch-timer-larger-bases-shifts.

Neidell, Matthew (2009) "Information, avoidance behavior, and health: The effect of ozone on asthma hospitalizations," *Journal of Human resources*, 44 (2), 450–478.

NEX-GDDP-CMIP6 (2080-2090) "NASA Earth Exchange Global Daily Downscaled Projections," https://www.nccs.nasa.gov/services/data-collections/land-based-products/nex-gddp-cmip6.

NOAA (1950-2019) "Global Hourly - Integrated Surface Database (ISD)," https://www.ncei.noaa.gov/products/land-based-station/integrated-surface-database.

——— (2023) "Billion-Dollar Weather and Climate Disasters," https://www.ncei.noaa.gov/access/billions/.

Oettinger, Gerald S (1999) "An empirical analysis of the daily labor supply of stadium vendors," *Journal of political Economy*, 107 (2), 360–392.

Park, R Jisung, Joshua Goodman, Michael Hurwitz, and Jonathan Smith (2020) "Heat and learning," *American Economic Journal: Economic Policy*, 12 (2), 306–39.

Paul, Rodney J and Andrew P Weinbach (2013) "Determinants of dynamic pricing premiums in Major League Baseball," *Sport Marketing Quarterly*, 22 (3), 152.

PRISM Climate Group (2021) "PRISM Gridded Climate Data," https://prism.oregonstate.edu.

Retrosheet (1950-2000) "Game logs (various)," retrosheet.org.

Roodman, David (2021) "BOOTTEST: Stata module to provide fast execution of the wild bootstrap with null imposed."

Schlenker, Wolfram (1950-2019) "Daily Weather Data for Contiguous United States," http://www.columbia.edu/~ws2162/links.html.

Seatgeek (2021) "The SeatGeek Platform," https://platform.seatgeek.com/.

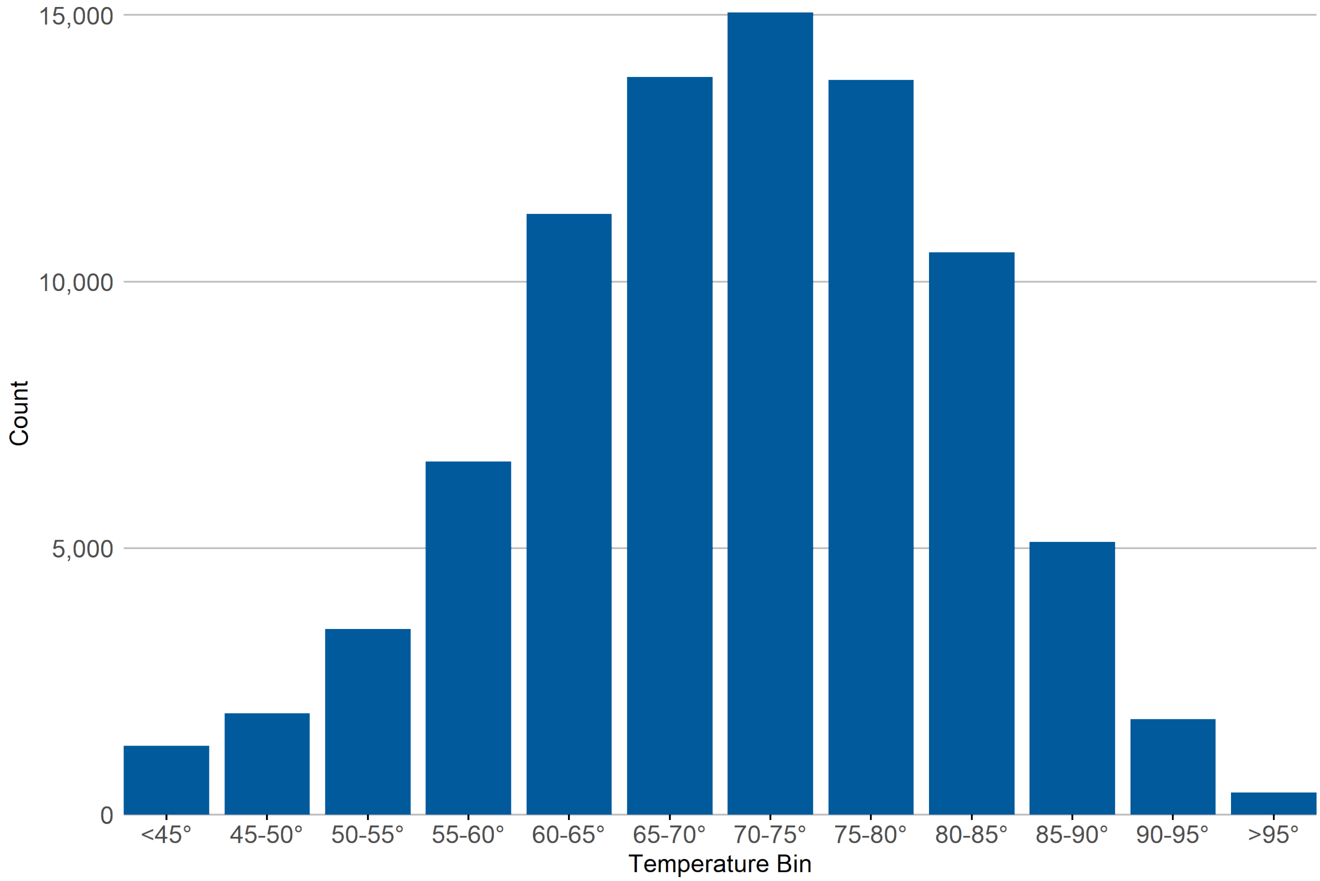
Shaikin, Bill (2005) "Attendance figures that count tickets sold, not turnstile clicks, make it hard for fans to reconcile what they hear with the empty seats they see," *Los Angeles Times*, https://www.latimes.com/la-sp-attendance-082305-story.html.

Sinha, Paramita, Martha L Caulkins, and Maureen L Cropper (2018) "Household location decisions and the value of climate amenities," *Journal of Environmental Economics and Management*, 92, 608–637.

Sweeting, Andrew (2012) "Dynamic pricing behavior in perishable goods markets: Evidence from secondary markets for major league baseball tickets," *Journal of Political Economy*, 120 (6), 1133–1172.

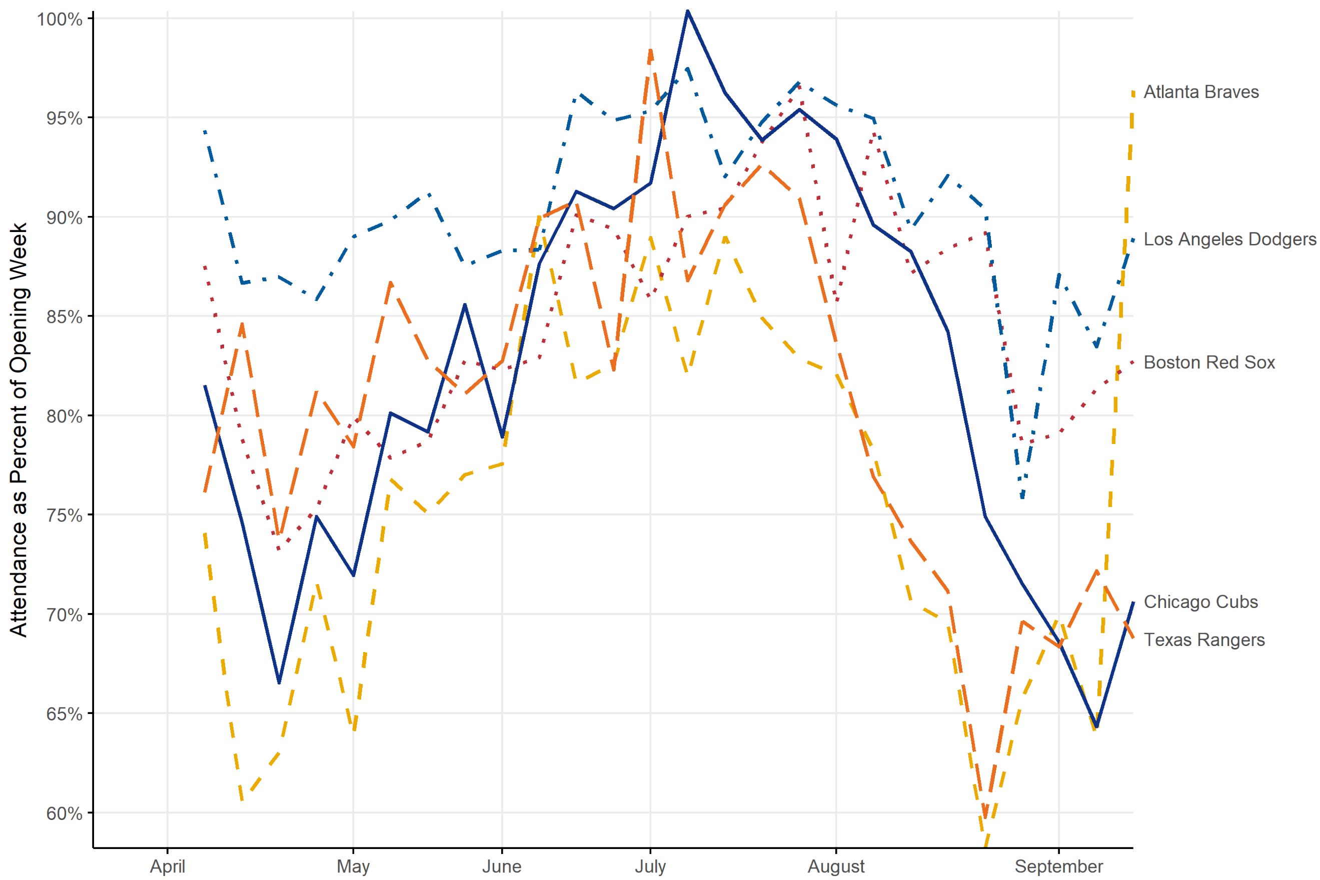
Zhang, Peng, Olivier Deschênes, Kyle Meng, and Junjie Zhang (2018) "Temperature Effects on Productivity and Factor Reallocation: Evidence from a Half Million Chinese Manufacturing Plants," *Journal of Environmental Economics and Management*, 88, 1–17.

Figure 1: Number of MLB games in each 5-degree temperature bin, 1950-2000



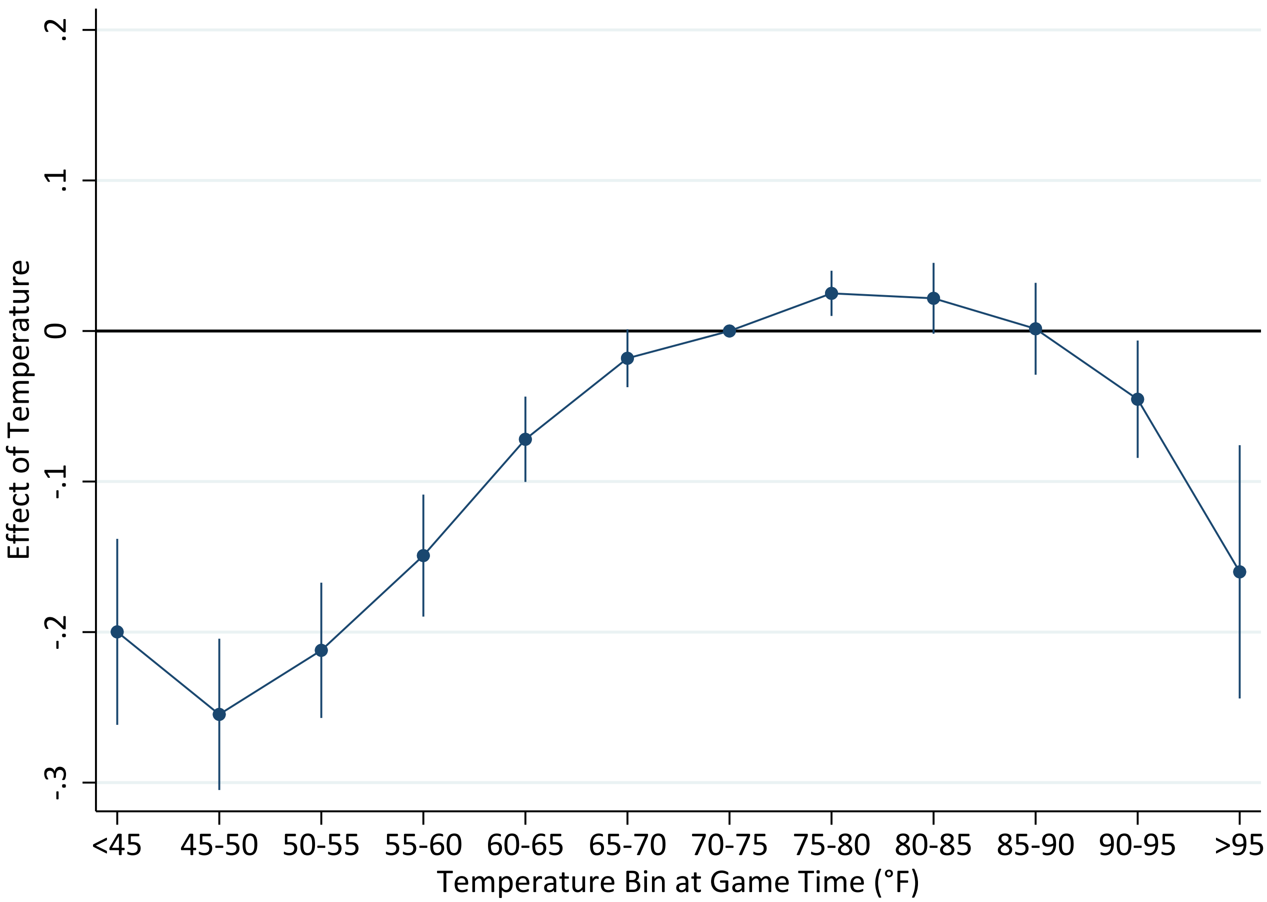
Note: This figure displays the count of MLB games that had game-time temperatures (the average of hourly temperature readings for the three hours after the start of the game) in each of the 12 bins we use in our regression analysis.

Figure 2: ATTENDANCE SLUMP IN AUGUST LARGER IN HOT PLACES



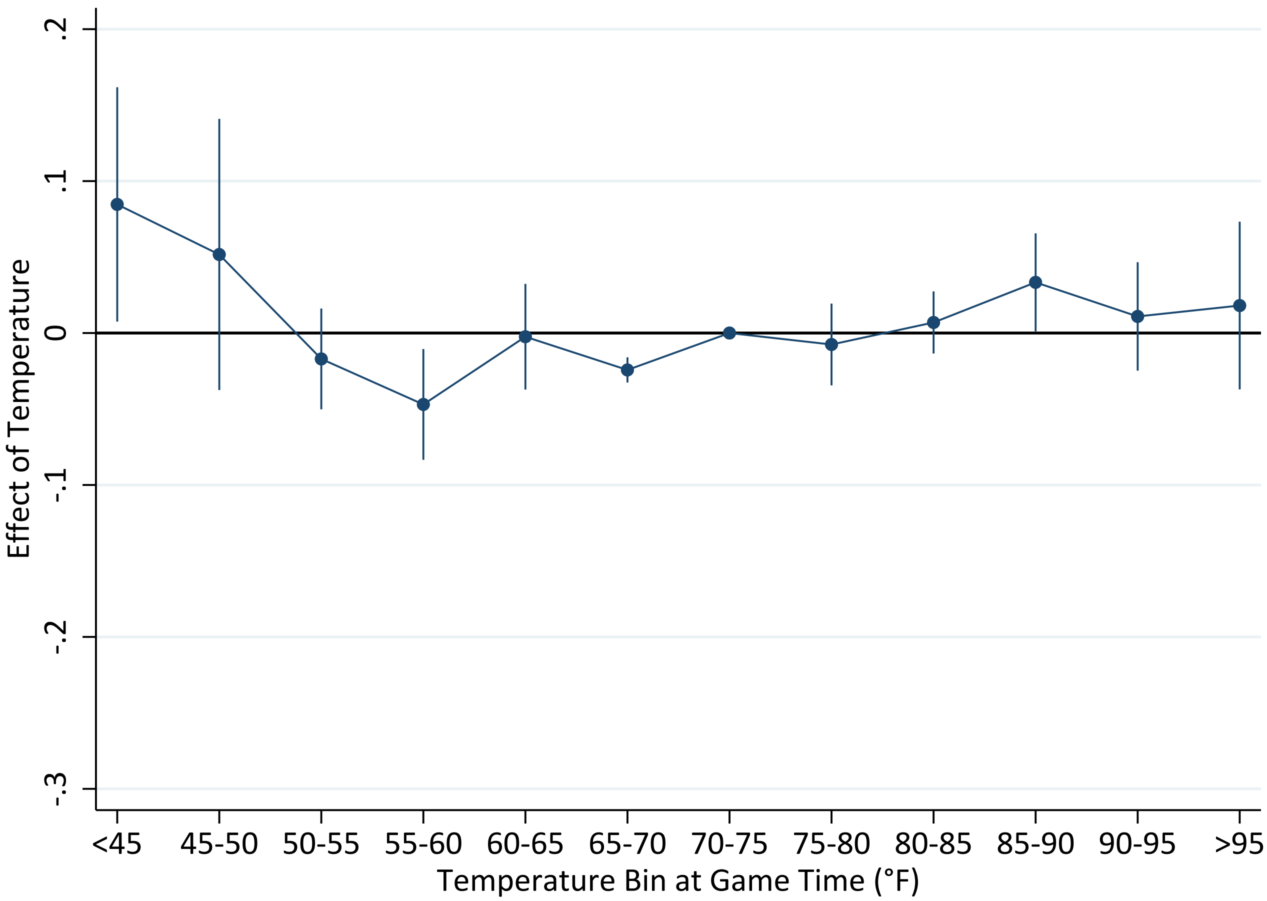
Note: This figure displays average weekly attendance as a percent of attendance in the 13th and 14th week of the year (when opening day usually falls) for home games for five major teams over the course of the average season from 1950 to 2019.

Figure 3: BASEBALL ATTENDANCE FALLS SHARPLY AT EXTREMELY HOT AND COLD TEMPERATURES



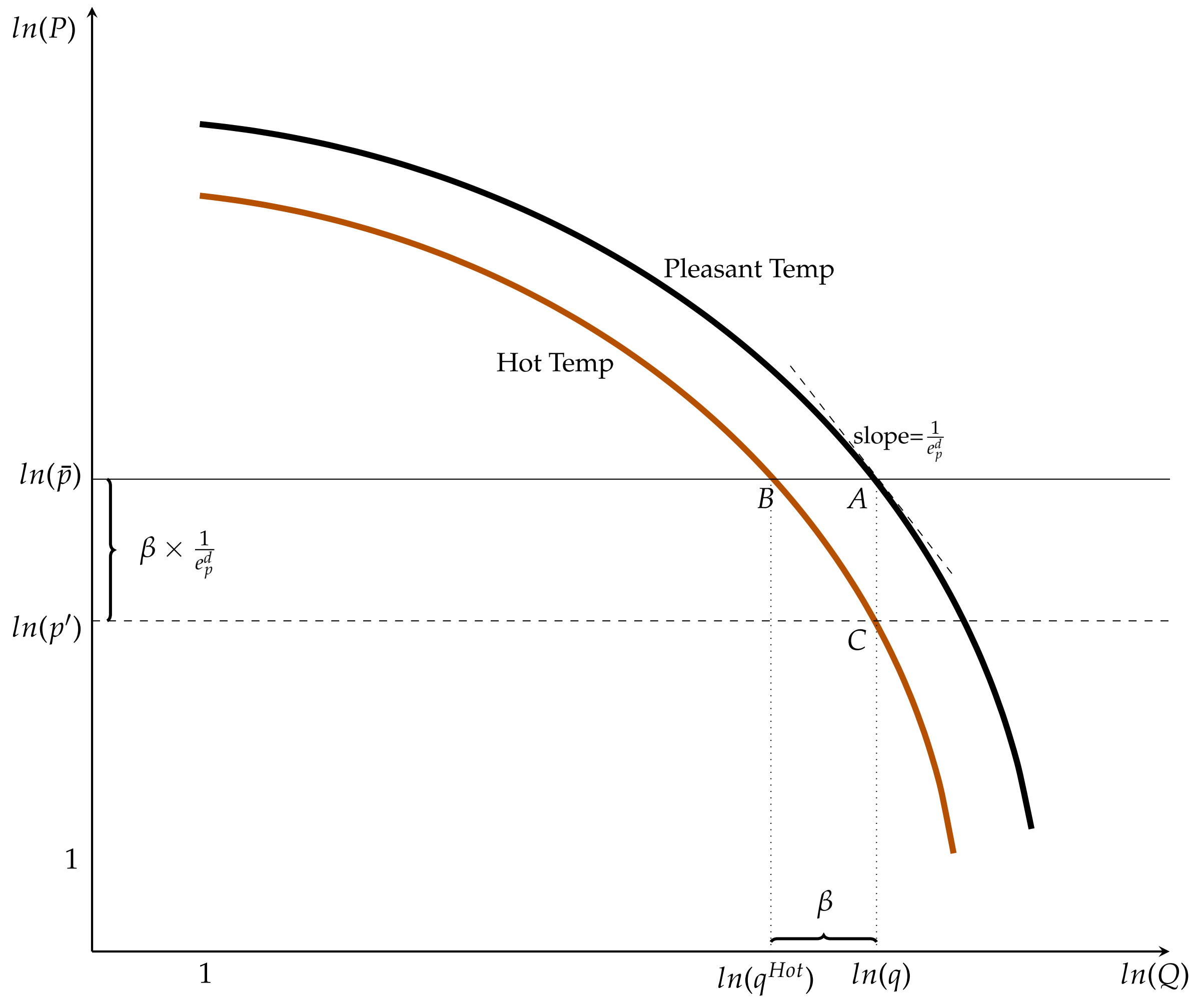
Note: This figure displays the results of Equation 1, run on the sample of MLB games from 1950 to 2000. The specification is the same as Column 3 in Table 2. The outcome variable is game-level log attendance, and the independent variables of interest are indicators for game-time temperature (the average of hourly temperature readings for the three hours after the start of the game) falling into the temperature bin of interest. The regressions also include stadium by month fixed effects, visiting team fixed effects, month by year fixed effects, and controls for daily precipitation, the share of the last 100 games the home team won, indicators for day of the week, and an indicator for whether the game took place in the day or evening. Standard errors are clustered by stadium. Point estimates and 95% confidence intervals are shown.

Figure 4: Placebo Exercise: Attendance unaffected by temperature at covered stadiums



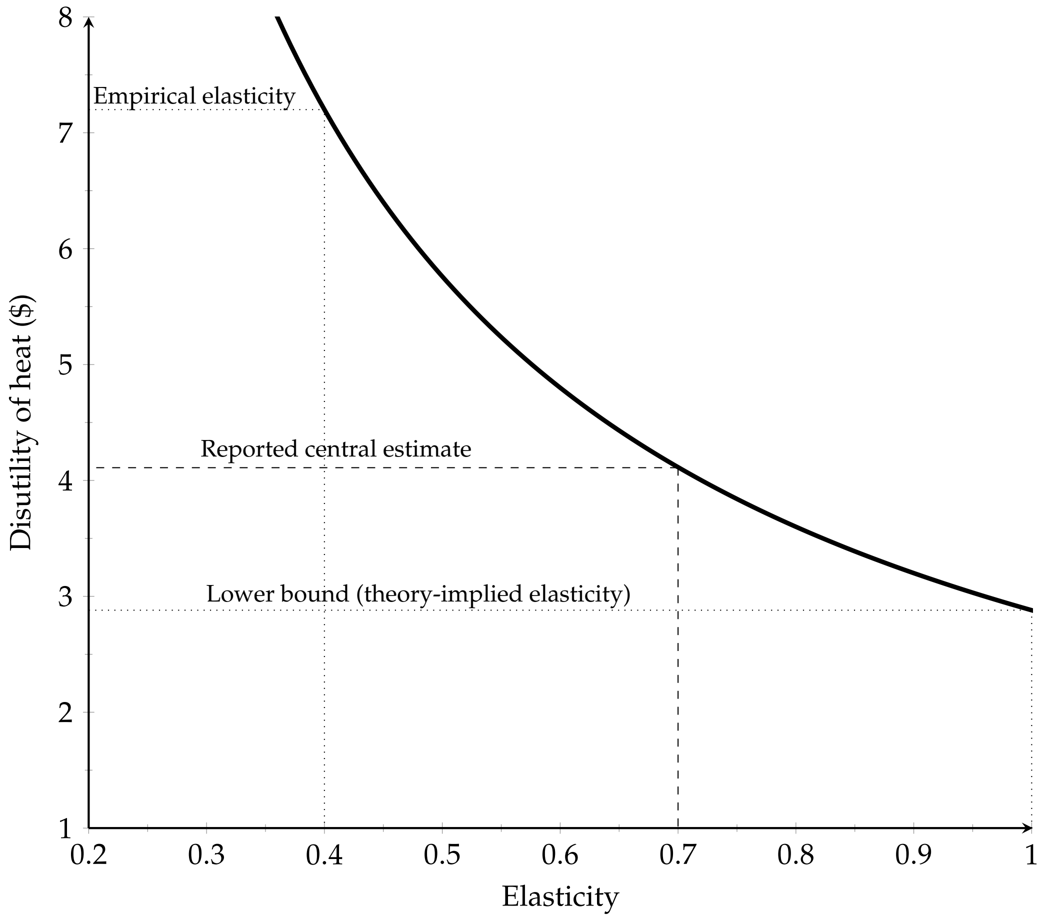
Note: This figure displays the results of Equation 1, run on the sample of MLB games from 1950 to 2000 that took place in stadiums that are either domed or have retractable roofs. The outcome variable is game-level log attendance, and the independent variables of interest are indicators for game-time temperature (the average of hourly temperature readings for the three hours after the start of the game) falling into the temperature bin of interest. The regression also includes stadium by month fixed effects, visiting team fixed effects, month by year fixed effects, and controls for daily precipitation, the share of the last 100 games the home team won, indicators for day of the week, and an indicator for whether the game took place in the day or evening. Standard errors are clustered by stadium. Point estimates and 95% confidence intervals are shown.

Figure 5: Elasticity Diagram



Note: Assume that the supply curve for tickets at the game level is perfectly elastic, i.e., that ticket prices are fixed at the game level (but may vary at the team or season level). Then the β we estimate as attendance falls from $ln(q) \rightarrow ln(q^{Hot})$ is demand-driven, as pictured. To compute the price-increase-equivalent, we need to multiply the demand change by the slope of the demand curve in the log-log space, i.e., $\frac{\partial ln(P)}{\partial ln(Q)}$. This is equal to the inverse of the price elasticity of demand, $\frac{1}{e_p^d}$. Most of the estimates from the sports literature are well below 1. This implies we should multiply β by *at least* 1 to obtain the price-increase-equivalent. We use a price elasticity of demand of 0.7 for our central estimate.

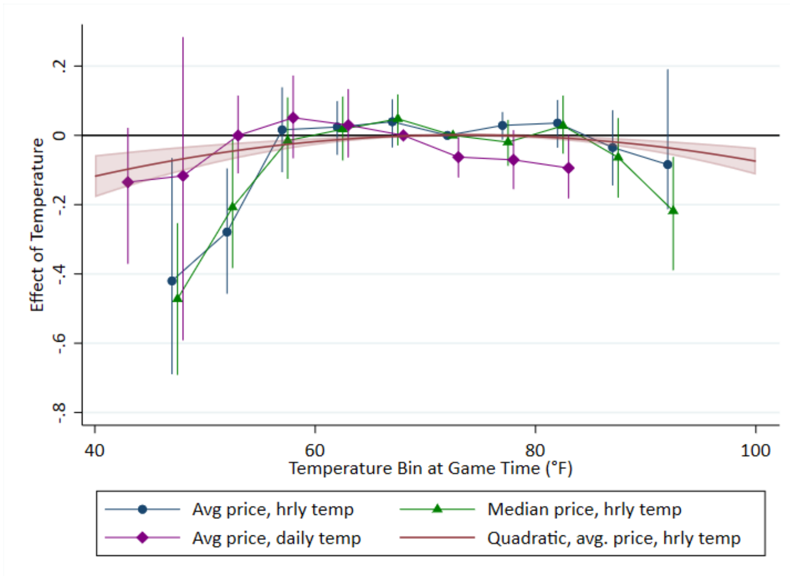
Figure 6: Disutility of heat increases with (in)elasticity



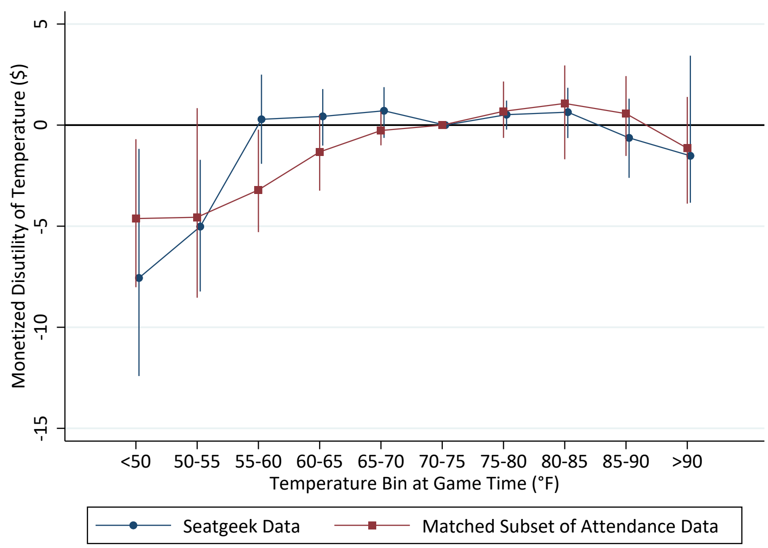
Note: Estimate of the disutility of heat (>95 degree weather) over the length of a baseball game as it changes with assumed price elasticity of demand. Baseline results use an elasticity of 0.7. Krautmann and Berri (2007) surveys a literature documenting that MLB teams price on the inelastic range of demand. If the elasticity is closer to those implied by the literature, the implied monetary willingness to pay is larger than the $4.11 in the baseline estimate. If the theoretically-predicted unit elasticity is correct, the disutility is only $2.88.

Figure 7: SEATGEEK TICKET PRICES FALL AT HOT AND COLD TEMPERATURES

(A) Main Results



(B) Comparison to Attendance



Note: Panel A displays the results of Equation 7, run on the sample of game-day Seatgeek listing prices from May-October 2021. The blue line shows the results of game-level log average listing price regressed on indicators for game-time temperature falling into the temperature bin of interest. The green line shows the results of game-level log median prices regressed on the same independent variables. The red line shows the results of a regression of logged average prices on a quadratic in game-time temperature, while the purple line shows estimates from a regression of logged average prices on indicators for daily average temperature on game day falling into the bin of interest. The regressions also include stadium by month fixed effects, away team fixed effects, controls for daily precipitation, indicators for day of the week, and indicators for whether the game took place during the day or evening. Standard errors are clustered by stadium and, except for the quadratic regression, calculated using the wild cluster bootstrap procedure from Cameron et al. (2008). Point estimates and 95% confidence intervals are shown. Panel B displays the results of the same regression as Panel A in the blue line, but the estimates have been multiplied by $18 to convert to a monetized disutility of each temperature bin. In the red line we display the monetized disutilities implied by a regression of log attendance on temperature similar to Equation 1. For this regression, we restrict to only the stadiums seen in the 2021 Seatgeek data and combine the bins for temperatures over 90 and below 50 degrees to match the Seatgeek regression. We then convert to a monetized utility by multiplying by the inverse price elasticity of demand and the estimated average ticket price of $18.

Table 1: SUMMARY STATISTICS: GAME AND STADIUM METRICS

	American League			National League		
	Mean	SD	N	Mean	SD	N
Attendance (Thousands)	20.67	12.66	44,242	22.13	12.96	42,462
Capacity (Thousands)	49.73	11.62	44,236	49.10	10.55	42,439
Percent Full	0.43	0.27	44,236	0.45	0.25	42,439
Stadium Age (Years)[1]	30.45	24.80	44,236	24.39	19.76	42,439
Game Duration (Minutes)	163.26	28.62	44,242	159.30	27.85	42,462
Night	0.64	0.48	44,242	0.62	0.49	42,462
Double Header[2]	0.08	0.26	44,242	0.07	0.25	42,462

[1]Calculated as the difference between the year a game takes place and the first year of stadium operation.

[2]Share of games that were played on days when more than one game occurred, usually due to prior inclement weather causing a game postponement.

Note: This table presents summary statistics for stadium and game-related control variables by league.

Table 2: MAIN RESULTS: BASEBALL ATTENDANCE FALLS SHARPLY AT EXTREMELY HOT AND COLD TEMPERATURES

	(1)	(2)	(3)	(4)	(5)	(6)
<45	-0.206	-0.186	-0.200	-0.144	-0.244	-0.127
	(0.028)	(0.032)	(0.032)	(0.023)	(0.034)	(0.031)
45-50	-0.272	-0.252	-0.255	-0.191	-0.269	-0.204
	(0.028)	(0.027)	(0.026)	(0.022)	(0.024)	(0.026)
50-55	-0.225	-0.210	-0.212	-0.158	-0.226	-0.171
	(0.022)	(0.022)	(0.023)	(0.020)	(0.021)	(0.016)
55-60	-0.162	-0.149	-0.149	-0.108	-0.158	-0.108
	(0.019)	(0.020)	(0.021)	(0.017)	(0.022)	(0.012)
60-65	-0.081	-0.070	-0.072	-0.053	-0.077	-0.042
	(0.012)	(0.014)	(0.014)	(0.011)	(0.015)	(0.009)
65-70	-0.022	-0.019	-0.018	-0.011	-0.022	
	(0.010)	(0.011)	(0.010)	(0.007)	(0.010)	
70-75						0.019
						(0.007)
75-80	0.025	0.022	0.025	0.018	0.027	0.026
	(0.008)	(0.009)	(0.008)	(0.006)	(0.008)	(0.007)
80-85	0.037	0.019	0.022	0.020	0.025	0.003
	(0.012)	(0.013)	(0.012)	(0.009)	(0.012)	(0.011)
85-90	0.032	0.002	0.001	0.001	0.002	0.006
	(0.020)	(0.017)	(0.016)	(0.012)	(0.016)	(0.022)
90-95	-0.001	-0.044	-0.045	-0.035	-0.045	-0.140
	(0.024)	(0.022)	(0.020)	(0.016)	(0.022)	(0.031)
>95	-0.128	-0.167	-0.160	-0.153	-0.147	
	(0.048)	(0.047)	(0.043)	(0.032)	(0.042)	
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Game Day Controls	No	Yes	Yes	Yes	Yes	Yes
Team Controls	No	No	Yes	Yes	Yes	No
Years	Pre-2000	Pre-2000	Pre-2000	1950-2019	Pre-2000	Pre-2000
Exclude Sellouts	No	No	No	No	Yes	No
Weather Data	Hourly	Hourly	Hourly	Hourly	Hourly	Daily Avg
Obs.	75,757	75,757	74,004	108,949	70,116	74,014

Standard errors in parentheses

Note: This table displays the results of Equation 1, run on the sample of MLB games from 1950 to 2000. The outcome variable is game-level log attendance, and the independent variables of interest, in columns 1-5, are indicators for game-time temperature (the average of hourly temperature readings for the three hours after the start of the game) falling into the temperature bin of interest. The regression in the first column includes stadium by month and month by year fixed effects. Column 2 adds controls for daily precipitation, indicators for day of the week, and an indicator for whether the game took place in the afternoon or evening. Column 3 adds visiting team fixed effects and controls for the share of the last 100 games the home team won. This is the specification displayed in Figure 3. Columns 4 and 5 display the same specification as column 3, except column 4 contains the sample from 1950-2019, and column 5 excludes games where attendance was greater than 95 percent of stadium capacity. Column 6 replaces the hourly weather station data with daily average temperature data. The omitted and highest bin are moved down by 5 degrees since daily average temperatures are cooler on average in our sample. All standard errors are clustered by stadium.

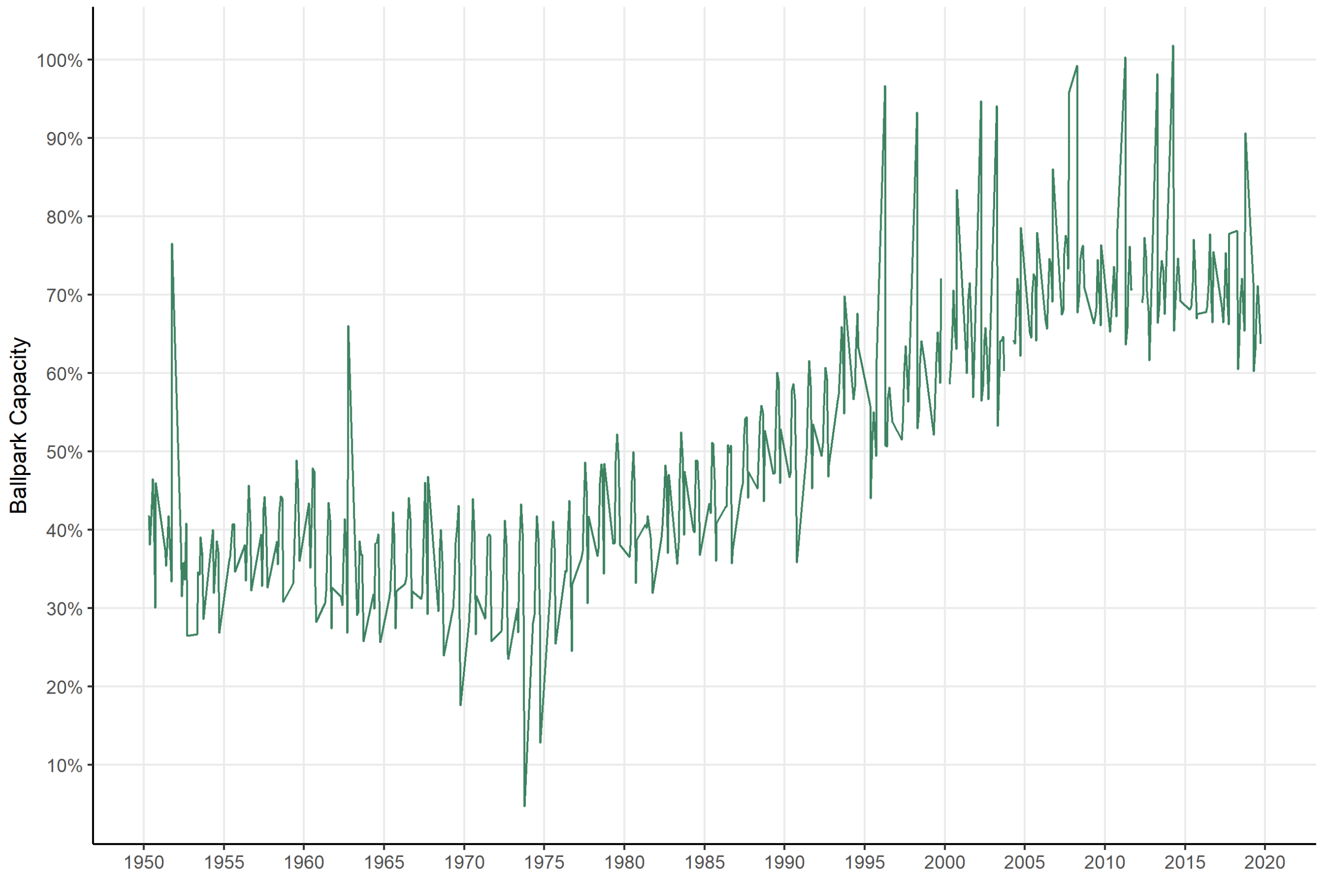
Table 3: Main Results: Ticket Prices

	(1) Avg. Price	(2) Min Price	(3) Max Price	(4) Median Price	(5) Listings	(6) No Sellouts
<50	-0.420	-0.585	0.023	-0.472	0.531	-0.428
	(0.03)	(0.01)	(0.95)	(0.00)	(0.08)	(0.07)
	[-0.690, -0.065]	[-0.873, -0.262]	[-0.802, 0.748]	[-0.692, -0.253]	[-0.072, 1.109]	[-0.704, 0.058]
50-55	-0.279	-0.260	-0.319	-0.208	0.136	-0.294
	(0.00)	(0.08)	(0.16)	(0.03)	(0.44)	(0.01)
	[-0.453, -0.100]	[-0.499, 0.025]	[-0.705, 0.190]	[-0.379, -0.022]	[-0.226, 0.478]	[-0.537, -0.075]
55-60	0.016	0.012	0.047	-0.015	0.031	0.045
	(0.80)	(0.84)	(0.69)	(0.78)	(0.72)	(0.38)
	[-0.102, 0.134]	[-0.090, 0.117]	[-0.241, 0.292]	[-0.121, 0.109]	[-0.150, 0.226]	[-0.066, 0.154]
60-65	0.024	0.049	-0.033	0.018	-0.037	0.011
	(0.54)	(0.39)	(0.62)	(0.66)	(0.49)	(0.74)
	[-0.059, 0.105]	[-0.074, 0.163]	[-0.172, 0.102]	[-0.074, 0.108]	[-0.140, 0.070]	[-0.063, 0.078]
65-70	0.039	0.042	0.025	0.048	-0.020	0.043
	(0.25)	(0.25)	(0.78)	(0.20)	(0.66)	(0.25)
	[-0.031, 0.107]	[-0.035, 0.110]	[-0.169, 0.212]	[-0.029, 0.119]	[-0.116, 0.081]	[-0.035, 0.117]
75-80	0.029	-0.017	0.196	-0.020	0.092	0.018
	(0.17)	(0.57)	(0.00)	(0.59)	(0.04)	(0.41)
	[-0.014, 0.069]	[-0.084, 0.042]	[0.108, 0.292]	[-0.089, 0.046]	[0.003, 0.179]	[-0.028, 0.061]
80-85	0.036	0.042	0.150	0.028	-0.042	0.012
	(0.27)	(0.30)	(0.10)	(0.54)	(0.68)	(0.74)
	[-0.033, 0.103]	[-0.040, 0.123]	[-0.033, 0.336]	[-0.056, 0.123]	[-0.248, 0.173]	[-0.054, 0.084]
85-90	-0.035	-0.031	-0.058	-0.064	0.044	-0.051
	(0.52)	(0.55)	(0.58)	(0.28)	(0.62)	(0.33)
	[-0.144, 0.076]	[-0.133, 0.072]	[-0.269, 0.173]	[-0.184, 0.052]	[-0.145, 0.233]	[-0.155, 0.056]
>90	-0.084	-0.098	0.140	-0.218	0.077	-0.110
	(0.30)	(0.28)	(0.50)	(0.01)	(0.75)	(0.14)
	[-0.215, 0.191]	[-0.225, 0.127]	[-0.348, 0.701]	[-0.404, -0.059]	[-0.503, 0.541]	[-0.233, 0.102]
N	1457	1457	1457	1457	1457	1311

Note: This table displays the results of Equation 7, run on the sample of game-day Seatgeek listing prices from May-October 2021. The outcome variables are game-day log average listing price (columns 1 and 6), log minimum price (column 2), log maximum price (column 3), log median price (column 4), and logged listings count (column 5). The independent variables of interest are indicators for game-time temperature (the average of hourly temperature readings for the three hours after the start of the game) falling into the temperature bin of interest. The regressions also include stadium by month fixed effects, away team fixed effects, and controls for daily precipitation, indicators for day of the week, and indicators for whether the game took place during the day or evening. Standard errors are clustered by stadium and calculated using the wild cluster bootstrap procedure from Cameron et al. (2008). The table shows the coefficient, p value, and 95% confidence interval for each estimate.

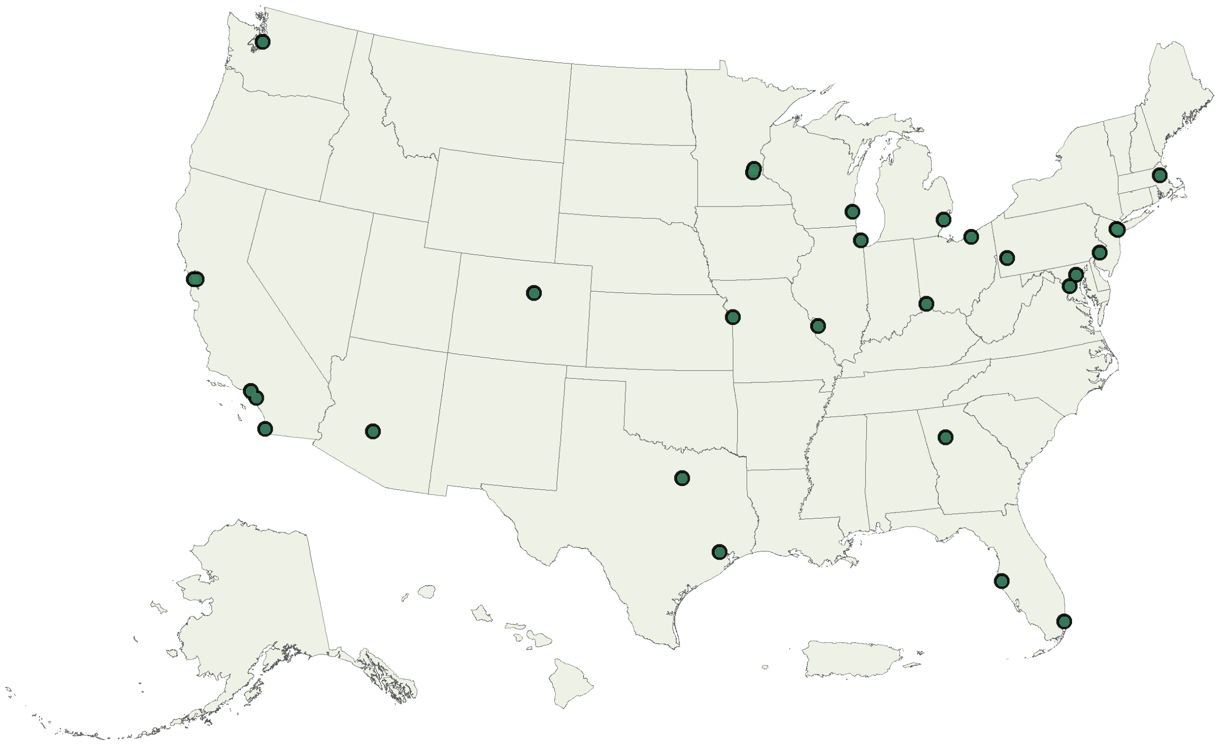
A Appendix A

Figure A1: SUMMARY STATISTICS: STADIUMS ARE MORE FULL THAN IN PAST



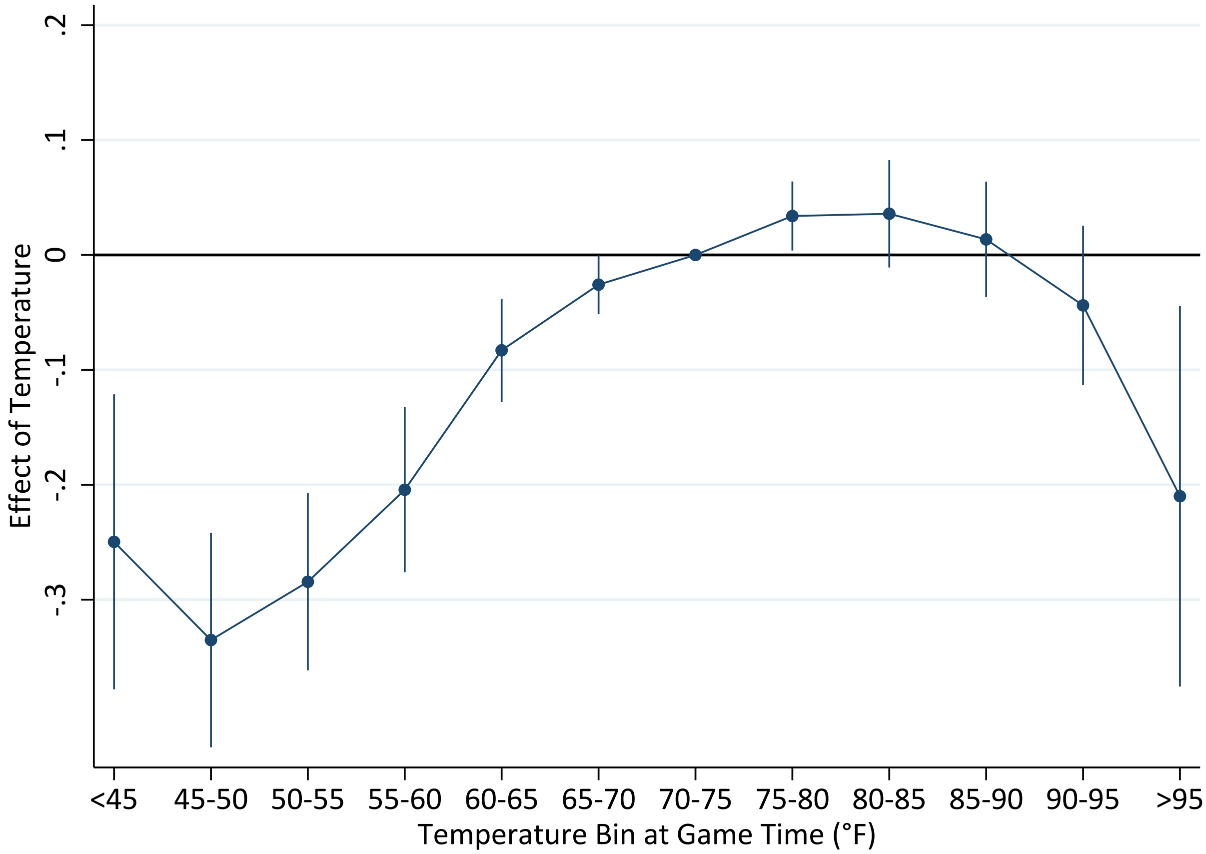
Note: This figure displays attendance as a percent of baseball stadium capacity for each month of each season from 1950 through 2019 using data from Retrosheet.org. The large spikes are generally associated with the first game of each season, which is the most well-attended game of the year for most teams.

Figure A2: CITIES WITH MLB STADIUMS



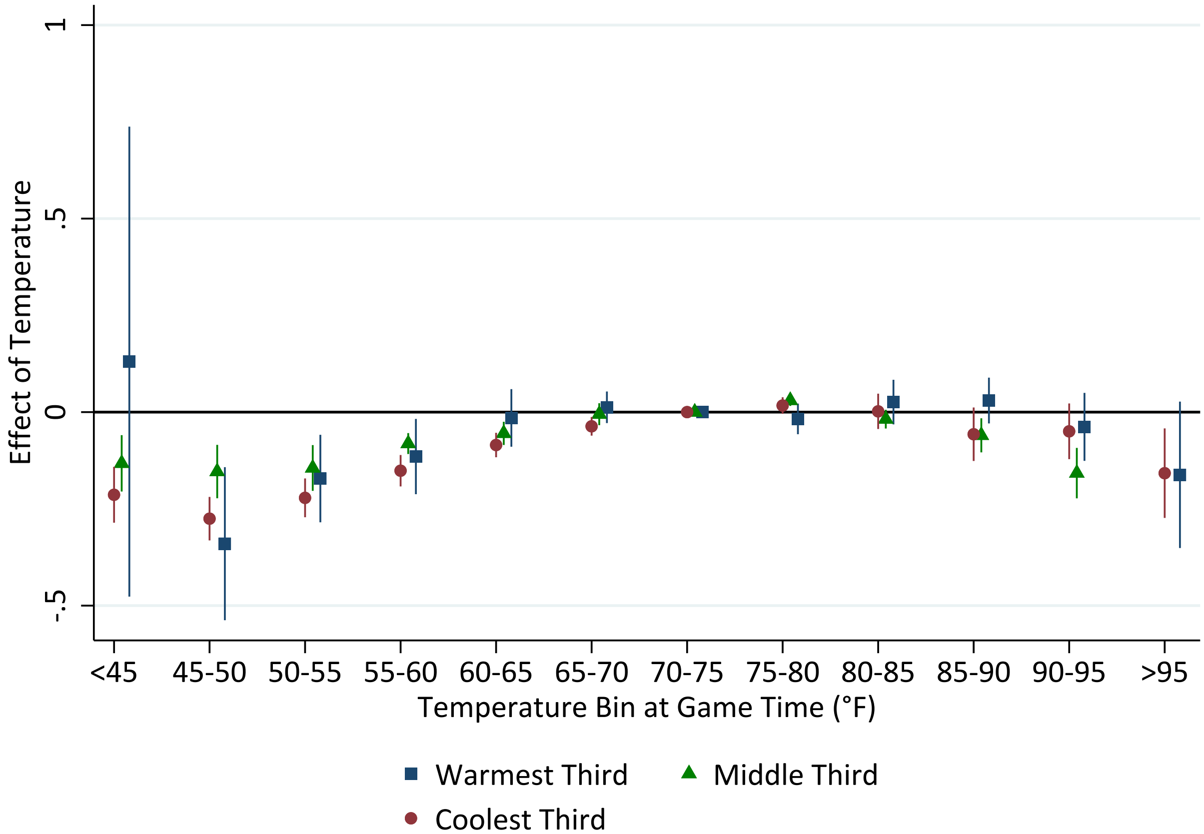
Note: This figure displays the location of the MLB stadiums in our sample. Some large cities, such as New York City and Chicago, have multiple stadiums operating at the same time. For a complete list of teams in the sample and their years of operation, see Table A1.

Figure A3: ESTIMATES ROBUST TO LIMITING SAMPLE TO TEAMS REPORTING TURNSTILE ATTENDANCE



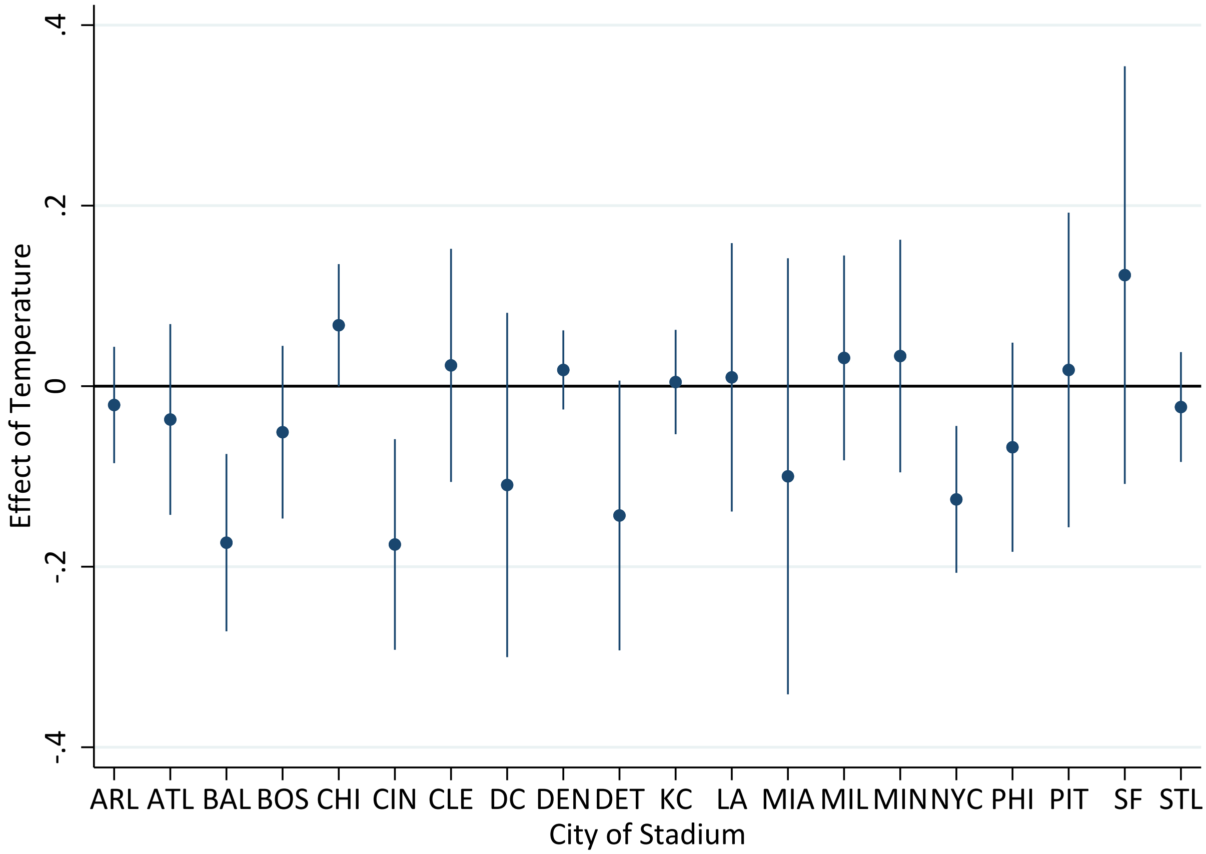
Note: This figure displays the results of Equation 1, run on the sample of National League games from 1950 to 1992. The outcome variable is game-level log attendance, and the independent variables of interest are indicators for game-time temperature (the average of hourly temperature readings for the three hours after the start of the game) falling into the temperature bin of interest. The regression also includes stadium by month fixed effects, visiting team fixed effects, month by year fixed effects, and controls for daily precipitation, the share of the last 100 games the home team won, indicators for day of the week, and an indicator for whether the game took place in the day or evening. Standard errors are clustered by stadium. Point estimates and 95% confidence intervals are shown.

Figure A4: No Evidence of Systematic Differences by Usual Climate



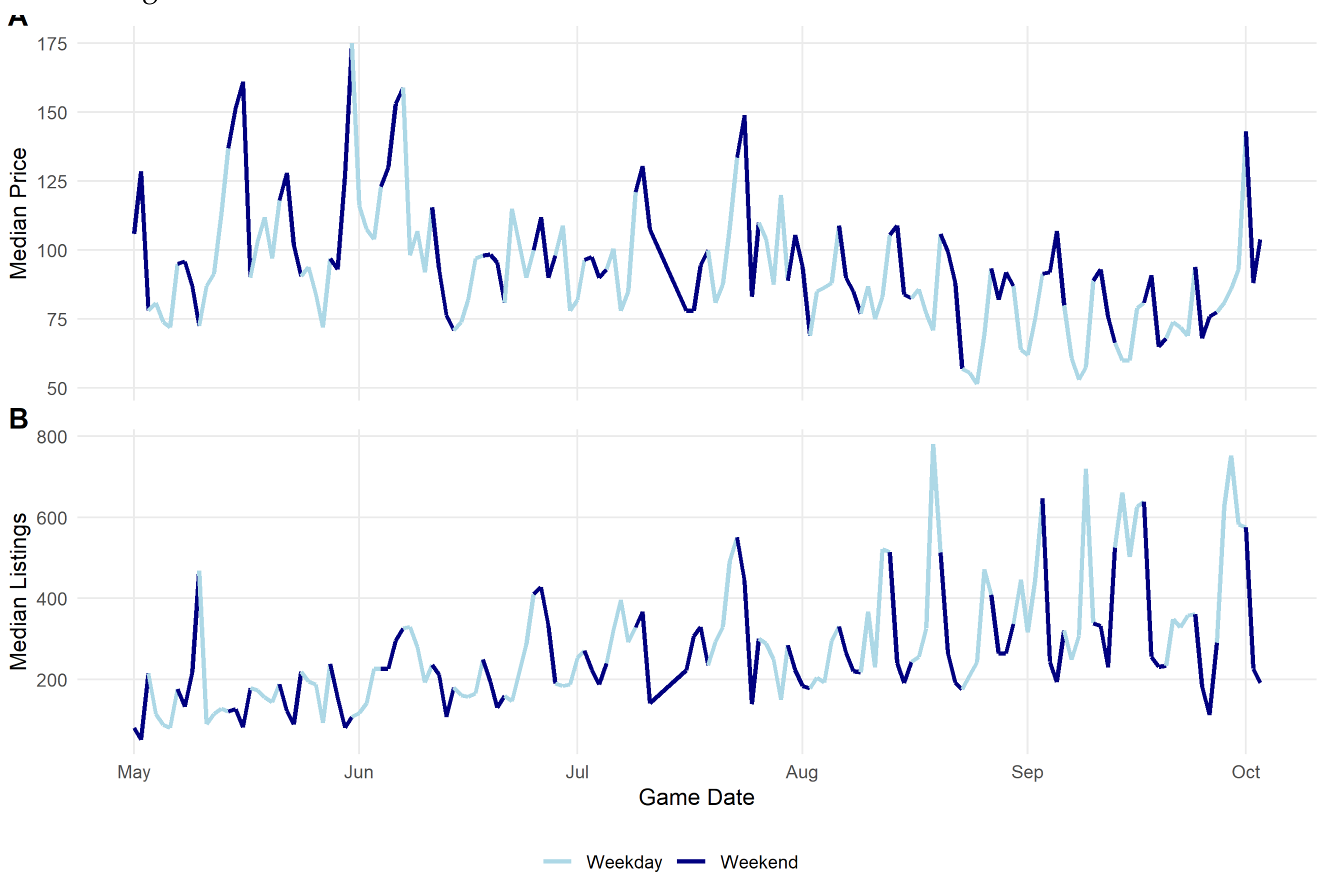
Note: This figure displays the results of Equation 1, run on the sample of MLB games from 1950 to 2000. The outcome variable is game-level log attendance, and the independent variables of interest are indicators for game-time temperature (the average of hourly temperature readings for the three hours after the start of the game) falling into the temperature bin of interest. These indicators are interacted flexibly with indicators for the stadium being in the warmest, coolest, or medium third of stadiums in the sample based on annual average temperature throughout the operation of each stadium. The regressions also include stadium by month fixed effects, visiting team fixed effects, month by year fixed effects, and controls for daily precipitation, the share of the last 100 games the home team won, indicators for day of the week, and an indicator for whether the game took place in the day or evening. Standard errors are clustered by stadium. Point estimates and 95% confidence intervals are shown.

Figure A5: Little Evidence of Heterogeneity by City



Note: This figure displays the results of Equation 1 on the sample of MLB games from 1950 to 2000, run separately for each city in the sample. The outcome variable is game-level log attendance, and the independent variables of interest are indicators for game-time temperature (the average of hourly temperature readings for the three hours after the start of the game) falling into the temperature bin of interest. Coefficients on indicators for game time temperature being above 90 degrees are displayed (the 90-95 and over 95 degree bins have been combined in this specification to increase statistical power). The regressions also include visiting team fixed effects, month by year fixed effects, and controls for daily precipitation, the share of the last 100 games the home team won, indicators for day of the week, and an indicator for whether the game took place in the day or evening. Point estimates and 95% confidence intervals are shown.

Figure A6: Seatgeek Summary Statistics: Median Price and Listings Count



Note: This figure displays median prices (Panel A) and median listing counts (Panel B) on Seatgeek from May through October of 2021, averaged by day of the season.

Table A1: Teams Included in Sample

	N. Home Games	First Year	Last Year
West			
Arizona Diamondbacks	1,780	1998	-
Colorado Rockies	2,146	1993	-
Los Angeles Angels	4,636	1961	-
Los Angeles Dodgers	4,872	1958	-
Oakland Athletics	4,029	1968	-
San Diego Padres	3,976	1969	-
San Francisco Giants	4,786	1958	-
Seattle Mariners	3,406	1977	-
Seattle Pilots	74	1969	1969
South			
Atlanta Braves	4,172	1966	-
Baltimore Orioles	4,922	1954	-
Houston Astros	4,578	1962	-
Miami Marlins	2,125	1993	-
Tampa Bay Devil Rays	1,778	1998	-
Texas Rangers	3,754	1972	-
Washington Nationals	1,207	2005	-
Washington Senators	734	1950	1960
Washington Senators	778	1961	1971
Northeast			
Boston Braves	193	1950	1952
Boston Red Sox	5,320	1950	-
Brooklyn Dodgers	532	1950	1957
New York Giants	520	1950	1957
New York Mets	4,341	1962	-
New York Yankees	5,171	1950	-
Philadelphia Athletics	304	1950	1954
Philadelphia Phillies	5,199	1950	-
Pittsburgh Pirates	5,200	1950	-
Midwest			
Chicago Cubs	5,238	1950	-
Chicago White Sox	5,137	1950	-
Cincinnati Reds	5,253	1950	-
Cleveland Guardians	5,122	1950	-
Detroit Tigers	5,282	1950	-
Kansas City Athletics	942	1955	1967
Kansas City Royals	3,944	1969	-
Milwaukee Braves	926	1953	1965
Milwaukee Brewers	3,867	1970	-
Minnesota Twins	4,608	1961	-
St. Louis Browns	251	1950	-
St. Louis Cardinals	5,382	1950	-

Table A2: BASEBALL GAME QUALITY AND TEMPERATURE

	(1) Duration	(2) Runs	(3) Home Runs	(4) Strike Outs	(5) Pitchers
<45	1.269	-1.017***	-0.692***	0.428**	-0.255**
	(1.084)	(0.269)	(0.113)	(0.139)	(0.088)
45-50	0.396	-0.872***	-0.558***	0.539***	-0.264***
	(0.860)	(0.212)	(0.080)	(0.123)	(0.064)
50-55	0.417	-0.656***	-0.416***	0.375***	-0.129**
	(0.604)	(0.164)	(0.062)	(0.105)	(0.047)
55-60	-0.465	-0.546***	-0.311***	0.228*	-0.157**
	(0.426)	(0.111)	(0.044)	(0.097)	(0.045)
60-65	-1.208**	-0.414***	-0.215***	0.128	-0.133***
	(0.352)	(0.085)	(0.034)	(0.069)	(0.030)
65-70	-0.566	-0.240**	-0.092***	0.117**	-0.111**
	(0.394)	(0.081)	(0.023)	(0.040)	(0.037)
75-80	1.122***	0.272***	0.116***	-0.154**	0.097***
	(0.309)	(0.073)	(0.023)	(0.048)	(0.023)
80-85	1.385**	0.404***	0.177***	-0.296***	0.127***
	(0.465)	(0.114)	(0.044)	(0.066)	(0.035)
85-90	2.831***	0.798***	0.305***	-0.276**	0.271***
	(0.550)	(0.171)	(0.059)	(0.089)	(0.048)
90-95	3.114**	0.881***	0.363***	-0.446***	0.262***
	(0.887)	(0.248)	(0.091)	(0.110)	(0.073)
>95	1.576	1.544***	0.513***	-0.253	0.457**
	(1.235)	(0.346)	(0.098)	(0.252)	(0.138)
Obs.	74,004	74,004	74,004	74,004	74,004
Mean of Dep.	161	9	2	11	5

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This figure displays the results of Equation 1, run on the sample of MLB games from 1950 to 2000. The dependent variables are indicators of the quality of game play, and the independent variables of interest are indicators for game-time temperature (the average of hourly temperature readings for the three hours after the start of the game) falling into the temperature bin of interest. The regressions also include stadium by month fixed effects, visiting team fixed effects, month by year fixed effects, and controls for daily precipitation, the share of the last 100 games the home team won, indicators for day of the week, and an indicator for whether the game took place in the day or evening. Standard errors are clustered by stadium.

Table A3: SEATGEEK PRICE AND LISTING SUMMARY STATISTICS

	Game Day			One Day Before			Two Days Before		
	Mean	P50	N	Mean	P50	N	Mean	P50	N
Average Price	445.70	90.50	1,992	278.23	94.00	1991	264.25	98.00	1980
Median Price	79.20	69.00	1,992	79.91	71.00	1991	81.13	73.00	1980
Lowest Price	32.52	25.00	1,992	30.18	24.00	1991	30.36	24.50	1980
Highest Price	2626.66	466.50	1,992	2755.85	590.00	1991	3232.35	683.50	1980
N. Listings	349.99	235.00	1,992	488.42	360.00	1991	537.72	402.50	1980

Note: This table presents summary statistics for SeatGeek ticket listing data from the 2021 MLB season. Data was accessed using SeatGeek's API, beginning on May 1, 2021 and ending on the final day of the season, October 3.

B Data Appendix

In this Appendix, we provide more information on the construction and contents of the weather, attendance, Seatgeek ticket price, and other data used in the main analysis.

BA Weather Data

NOAA's Integrated Surface Database We downloaded hourly temperature data from weather stations in the contiguous U.S. from the National Oceanic and Atmospheric Administration (NOAA's) Global Hourly - Integrated Surface Database (ISD). The ISD combines digitized weather station data from many data sources on over 20,000 stations into a single database, providing information on parameters such as temperature, wind speed, pressure, and precipitation at various time scales. We downloaded the ISD data using the worldmet package in R, and compiled a list of stations that reported continuously throughout the tenure of each baseball stadium in our sample. To ensure that a station reported continuously, we first remove any stations that stopped reporting for an entire year during the tenure of the stadium, and then further remove any stations that were missing more than 5% of 3-hourly temperature readings. These steps left 1,435 missing station-game observations in our weather dataset, which we fill in with the closest reporting weather station, if available. This results in a dataset with only 229 missing station-game observations, out of over 300,000 total. We then took the average of weather station temperature readings for the three hours starting at the start of each game, and then took the average for each weather station within 50km of the stadium, weighted by inverse distance. The worldmet package returns the weather data in UTC, so we convert to the local time zone before merging with the Retrosheet data.

Every stadium in the contiguous U.S. had at least one weather station nearby that reported continuously throughout its tenure. However, in many cases these stations reported only once per three hours for a period of time, usually in the late 1960's and early 1970's. In these cases, the station's identifying code often changed. A station's code consists of a USAF (U.S. Air Force) code and a WBAN (Weather-Bureau-Army-Navy) code. In cases where the code changed, one of the USAF or WBAN code always stayed the same, so we combine stations with the same USAF or WBAN code when judging the continuity of a station's reporting. NOAA has confirmed that these represent the same station, and we confirm that these stations are in the same location (within 1km in most cases, although it can be further and still be the same airport, for example). During the years that a station reported only once per three hours, if it's the only station available, the game-time assigned weather becomes the one reading in the three hour period, rather than an average of three readings.

Schlenker's Fine-Scaled Weather Data Set For the analysis of the effects of weather on attendance, we source our daily precipitation data from Wolfram Schlenker's Daily Weather for Contiguous United States dataset, which is derived from the PRISM dataset. This dataset provides total precipitation at the 2.5 mile resolution. This dataset creates a balanced panel of monitoring stations by interpolating missing station observations with the distance-weighted average of the cumulative density function of surrounding stations. This dataset is available from 1900-2019.

PRISM Data For the analysis of the effects of weather on Seatgeek ticket prices, we use the PRISM weather dataset, developed by Oregon State University's PRISM Climate Group, for our daily precipitation data. These data are compiled from monitoring stations across the U.S. and are spatially interpolated to a 4km resolution using the PRISM model, which accounts especially for the effects of elevation on the spatial distribution of temperature and precipitation. The dataset provides daily total precipitation, which is the sum of rain and melted snow for the day. We download PRISM data using the prism package in R.

Weather Projections from CMIP6 We use the average of the 35 global climate models (GCM's) from the Coupled Model Intercomparison Model 6 (CMIP6). The data are downloaded from the NASA Earth Exchange (NEX) Global Daily Downscaled Projections (GDDP). The NEX-GDDP is downscaled to improve the resolution of output from the GCM's, allowing for analysis at the 0.25 degree latitude/longitude resolution. Daily maximum and minimum near surface air temperatures are provided at the daily level, which we average to form a measure of daily average temperature. We then calculate a count of days between 85-89.5 degrees, which corresponds to the 25th and 75th percentile for game-time temperatures over 95 in our sample, for each grid cell for each year from 2080-2090, and then create a decadal average of the yearly count. We calculate these averages for two climate change scenarios from the new scenario framework introduced in the IPCC Sixth Assessment Report. SSP5-8.5 is a high-emissions scenario with no mitigation policy, in which economic growth is fueled by fossil fuels. This scenario produces warming of 4.4 degrees Celsius by end of century. SSP2-4.5 is a middle-of-the-road scenario in which emissions begin to fall by the mid 21st century, and socioeconomic factors are stable. This scenario produces warming of 2.7 degrees Celsius by 2100.

BB Baseball Data

Retrosheet Data We download data on game-level attendance and game play from retrosheet.org using the retrosheet package in R. We make a few adjustments to the retrosheet data to derive a consistent set of team-stadium observations. First, we code any team that changed its name but remained in the same city to be a single team.

- The Florida Marlins became the Miami Marlins; we code them as the same team.
- The California Angels became the Anaheim Angels in 1996 and subsequently the Los Angeles Angels of Anaheim in 2005; we code them as the same team.

If a team moved cities and changed names, we consider the team in the new city to be a new team.

- The Kansas City Athletics moved to Oakland and became the Oakland Athletics in 1968; we code these as different teams.
- The Washington Senators moved to Minnesota and became the Minnesota Twins in 1961. A new Washington Senators team sprung up in its place, from 1960-1971, before moving to Dallas-Fort Worth and becoming the Texas Rangers in 1972. The original Washington Senators, replacement Washington Senators, Minnesota Twins, and Texas Rangers are each coded as different teams in our sample.
- The Seattle Pilots moved to Milwaukee and became the Milwaukee Brewers in 1970; they are coded as different teams.

We combine a list of stadiums from Retrosheet with park configuration data from Seamheads. Seamheads.com provides data on capacity and type of cover for each stadium by year. We use this information in combination with Retrosheet to form our measure of attendance as a share of capacity as well as for our sample restriction to only open-air stadiums. We drop some one-off venues from our dataset: the Oakland Athletics played 6 games in Las Vegas in 1996, the Atlanta Braves played 1 game at Fort Bragg in 2016, the Tampa Bay Rays played 3 games at the Ballpark at Disney's Wide World in 2007, and the Pittsburgh Pirates played 1 game at BB&T Ballpark at Bowman Field in 2017. Our results are robust to excluding or including these venues from the sample.

Seatgeek Data We accessed the SeatGeek.com API on a daily basis at 8:00A.M. Central Time starting on May 1, 2021 and ending on October 3, 2021, the final day of the 2021 regular season. Each access of the API collected data on that day and the following two days' MLB games. The variables collected include:

- Home and away teams
- Date of the API request
- Game date
- Game start time
- Geographic information, including the city, state, latitude, and longitude of where the game was to take place
- The number of ticket listings
- Mean, median, minimum, and maximum ticket price for seats available during the game

The data are at the game-date level, not the individual ticket level. It should also be noted that the number of ticket listings does not equal the number of tickets available; for example, a listing for four seats next to each other would be counted as one listing.

BC Other Data

American Time Use Survey Data We use American Time Use Survey (ATUS) to calculate approximately how much time the average American spends outside at hot and cold temperatures. To create this statistic, we download the 2003-2018 multi-year ATUS data and merge it with daily average temperature data from Wolfram Schlenker's version of the PRISM data, by county and date of the diary entry. We then classify activities as taking place outdoors vs. indoors to calculate how much time is spent outside each day. We follow Graff Zivin and Neidell (2014) in defining time outdoors as time spent outdoors away from the home, or doing specific home activities such as walking, exterior maintenance, and lawn, garden, and houseplants. Outdoor labor time is not included in the main measure, as the American Time Use Survey does not distinguish whether the "respondent's workplace" is indoors or outdoors. However, in a secondary estimate we follow Graff Zivin and Neidell (2014) in defining a set of climate-exposed industries, where much of the work is likely done exposed to outdoor temperatures: agriculture, forestry, fishing, and hunting; mining; construction; manufacturing; and transportation and utilities industries. We then count time at work in these industries as additional time outdoors.