Homework 7
Question 3

Name: Romulo Jimenez
Current Choice for Final Project: MLB Econometrics

Each person in the group should find 2 more academic articles related to your current choice of final project. Write a short paragraph on each, concentrating on what data is used (and whether it is accessible), what econometric techniques, and what questions are addressed.

Evidence From 50 Years of Major League Baseball Games (K. Kuruc, 2025)
This paper leverages game-level data over ~50 years of MLB seasons to explore fans' revealed preferences for weather at the ballpark. The dataset comprises attendance figures for MLB games, matched with temperature and weather conditions, allowing the author to estimate how shifts in extreme heat or cold affect crowd size. The econometric technique is a revealed-preference approach: using the variation in attendance as a function of weather, controlling for other factors (team-fixed effects, season-fixed effects, maybe game-level controls) and assuming a horizontal supply curve to infer welfare (utility) losses from weather. They estimate a per-hour utility loss (e.g., ~$1.53) when exposures to high temperatures occur. Data are publicly accessible (attendance by game, weather records) though the exact merged dataset may require effort. The question addressed is: How much do fans value comfort/temperature when deciding to attend MLB games, and what is the welfare cost of extreme weather conditions on sporting events? This is more economics of sport than player performance, but uses econometric modelling and a rich dataset.

Empirical Determination of Baseball Eras: Multivariate Changepoint Analysis in Major League Baseball (Whalen, Matthews & Mills, 2024)
This study uses a long-span dataset of MLB team and player statistics from the late 1800s through 2020 to empirically detect structural shifts ("eras") in baseball via multivariate change-point analysis. The data include multiple time-series of performance metrics (batting, pitching, fielding) aggregated by season across teams. The econometric/statistical technique is a multivariate changepoint detection method (identifying shifts in mean and/or variance across series) rather than traditional regression; the authors cluster multiple time series simultaneously to detect when baseball entered e.g. the "Steroid Era" or "Post-Steroid Era". The data are somewhat accessible (historical MLB stats are publicly or commercially available), though the authors' exact cleaned panel may not be open. The research question is: At what points did structural breaks occur in the evolution of MLB performance metrics, and can we statistically identify distinct "eras" without relying solely on historical narrative? Their findings support known era boundaries but add nuance via multivariate modelling.