Estimating Aging Curves: Using Multiple Imputation to Examine Career Trajectories of MLB Offensive Players

Quang Nguyen
Department of Statistics and Data Science
Carnegie Mellon University
Pittsburgh, PA 15213
`nmquang@cmu.edu`

Gregory J. Matthews
Department of Mathematics and Statistics
Loyola University Chicago
Chicago, IL 60660
`gmatthews1@luc.edu`

Abstract

In sports, an aging curve depicts the relationship between average performance and age in athletes' careers. This paper investigates the aging curves for offensive players in the Major League Baseball. We study this problem in a missing data context and account for different types of dropouts of baseball players during their careers. In particular, the performance metrics associated with the missing seasons are imputed using a multiple imputation model for multilevel data, and the aging curves are constructed based on the imputed datasets. We first perform a simulation study to evaluate the effects of different dropout mechanisms on the estimation of aging curves. Our method is then illustrated with analyses of MLB player data from past seasons. Results suggest an overestimation of the aging curves constructed without imputing the unobserved seasons, whereas a better estimate is achieved with our approach.

Keywords aging curve · baseball · multiple imputation · statistics in sports

1 Mitra and Reiter Paper

Mitra & Reiter (2016):

1.1 Introduction

The article discusses the use of propensity score matching in studies of causal effects and the use of multiple imputation to overcome the complication of missing covariate data. Propensity scores are the probability that a subject receives treatment given its vector of covariates, and by selecting control units with similar propensity scores to treated units, analysts can create a matched control group whose covariates are similar to the treated group's covariates. The article discusses different approaches to estimating propensity scores and the use of multiple imputation to fill in missing covariate data. It also explores two different approaches to matching treated and control units: the Within approach, where treated and control units are matched within each completed dataset, and the Across approach, where each unit's propensity scores are averaged across 'm' imputed datasets, and treated and control units are matched based on their averaged scores. The article uses simulations with artificial and actual data to investigate which approach is more effective in estimating the average treatment effect on those exposed.

1.2 Simulation study:

In this section, the authors compare the Across and Within approaches using simulations with artificial data. They generate two covariates X for n=1100 records, such that $x_i=(x_{i1},x_{i2})' \sim N(\mu,\text{sigma})$, where $\mu=(10,10)'$, and sigma has variances equal to 5 with correlation 0.5. The response Y is generated such that, for all i, $Y_i=x_{i1}+x_{i2}+e_i$, $e_i \sim N(0,1)$. The authors introduce missing data into x2 based on missing at random

mechanisms and consider three mechanisms for assigning treatment. They are (i) assignment depends only on x1, (ii) assignment depends only on x2, and (iii) assignment depends equally on x1 and x2.

1.3   Simulation Results:

The results suggest that both approaches perform similarly when treatment assignment is independent of the variables with missing data. However, the Across approach offers greater reductions in bias when assignment is dependent on the variables with missing data. This implies that the importance of variables with missing data in treatment assignment is a crucial factor in the differences observed between the two approaches.

When treatment assignment depends only on x1, which is fully observed, both methods perform similarly in terms of bias reduction. However, when treatment assignment depends on x2, biases increase for both approaches. The Within method results in very close balance on x1 and on the completed version of x2 in the treated and matched control set, but this balance does not imply balance on actual x2. In contrast, the Across method selects matched controls with missing x2 less frequently than the Within method does, thus mitigating the problems from inaccurate balance on true x2 in missing cases. The advantage of the Across method derives from lesser reliance on inaccurate imputed values.

In all simulations, the bias in the Within method does not change as m increases, while the bias in the Across method generally decreases with m. The variance for the Across method appears not to decrease with m, whereas the variance for the Within method does. This is because the Across method results in only one propensity score for each unit, while the Within method averages treatment effects over independently generated imputations.

1.4   Conclusion:

The authors conducted simulation studies to compare the performance of the two approaches in reducing bias and variance in treatment effect estimates. They found that the Across approach had greater potential for bias reduction when treatment assignment depended on missing covariates, but the Within approach resulted in smaller variances. However, the authors caution that these results may not generalize to all scenarios and recommend running simulation studies to evaluate the performance of the approaches in a specific context.

The Article also notes that the choice of imputation model and whether or not to condition on the response in the imputation models can affect treatment effect estimates. The authors recommend using the augmented Across approach as a default and suggest a heuristic for choosing values of m and r. Finally, the authors suggest that further research is needed to investigate and develop approaches to select m and r.

2   Jennifer Hill Paper

Hill (2004):

2.1   Overview:

In this paper, the author explained about propensity score matching and discussed about various methods for handling missing data. The methods discussed were Complete-cases, Complete variables, ECM-DR, and MI methods (MIPM-I, MIPM-I(Y), MIPM-II, MIPM-II(Y)).

The average bias for complete variables method is higher whereas for MIPM-I method has lower treatment effect bias. For MSE, MIPM-I(Y) performs the best, followed by MIPM-II(Y) and MIPM-I, with MIPM-II lagging, just as with balance diagnostics. All four MIPM methods outperform DR in terms of MSE.

The author also performed a simulation study for which the outcome variable also has missing data. For this simulation, MIPM-II(Y) method outperforms all other methods with respect to treatment effect bias and MSE.

3 Leyrat et al paper

Clemence Leyrat & Williamson (2019):

3.1 Introduction:

The article discusses the use of propensity scores (PS) to estimate treatment effects in observational studies. PS is the individual's probability of receiving treatment given their baseline characteristics. The article focuses on the use of inverse probability of treatment weighting (IPTW) to balance covariates between treatment groups. The main issue with estimating PS is the presence of partially observed covariates. The article discusses four unresolved questions regarding the use of multiple imputation (MI) for PS analysis.

The first question is whether to combine treatment effects estimated on each imputed dataset (MIte) or combine PS values across imputed datasets (MIps). The second question is whether to include the outcome in the imputation model for missing covariates. The third question is how to estimate the variance of the IPTW estimator after MI. The final question is how well MI performs compared to other popular approaches to handle missing data, namely complete case analysis (CC) or the missingness pattern (MP) approach.

The article shows that the best method when using MI with IPTW is to include the outcome in the PS model and use Rubin's rules to combine treatment effect estimates from imputed datasets (MIte). The article highlights that suboptimal methods are commonly used in recent studies, such as using MIps instead of MIte, and imputing data without including the outcome in the imputation model. The article concludes by providing new recommendations to address the unresolved questions regarding the use of MI for PS analysis.

3.2 Simulation study results:

The IPTW estimator on the full data is found to be approximately unbiased while the CC estimator is strongly biased in all scenarios except when the outcome is not associated with missingness and there is no treatment effect. The missingness pattern (MP) approach is also evaluated, and it is found to be always biased, even more so than the CC approach. The reason for the bias in the MP approach is due to an incorrect specification of the propensity score model in each missingness pattern.

For Mi, the study found that the inclusion of the outcome variable in the imputation model is crucial in all three MI approaches to reduce bias. When the outcome is not included in the imputation model there is a large bias but when the outcome is include Bias has reduced. However, only one of the MI approaches (MIte) led to an unbiased estimate. Combining the propensity score (PS) parameters to estimate the PS of the average covariates (MIpar) performed better than combining the PS themselves (MIps), but both approaches were slightly biased.

3.3 Discussion:

The Article discusses the results of a study that aimed to answer three main questions about multiple imputation (MI) in the context of inverse probability of treatment weighting (IPTW). The first question was whether the outcome must be included in the imputation model. The results showed that the outcome must be included in the imputation model, even if it is not a predictor of missingness. When the outcome is not included in the imputation, the results showed bias for all MI imputation methods. The second question was whether Rubin's rules should be applied on the IPTW treatment effect estimates or on the PS estimates themselves. The results showed that combining the treatment effects after MI(MIte approach) is the preferred MI strategy in terms of bias reduction under a missing at random (MAR) mechanism. The third question was how to estimate the variance of the IPTW estimator after MI. The results showed that Rubin's rules perform well for the MIte approach and the proposed variance approximation showed good performance in the simulation study. The Article also discusses the balancing properties of the different MI approaches and the need for further investigation in assessing the balancing properties of the pooled PS on the average covariate values across the imputed dataset or on each dataset.

The authors' recommended method is to pool treatment effects after imputation, including the outcome in the imputation model, and using inverse probability of treatment weighting (IPTW) to estimate the treatment effect. They compared their approach to other methods, including pooling PSs across imputed datasets (MIps), the matching propensity (MP) approach, and complete case (CC) analysis. The authors found that the MP approach showed poor performance under a missing at random (MAR) mechanism, while MI followed by pooling treatment effect estimates was the preferred approach when data are MAR and the outcome must be included in the imputation model. However, the authors note that further investigation is needed to understand the usefulness of the MP approach in practice.

3.4 Conclusion:

The Article discuss the limitations of a study on imputation-based methods for estimating treatment effects, particularly focusing on the use of the inverse probability of treatment weighting (IPTW) estimator only. The study generated only three covariates in their simulation, whereas in reality, propensity scores (PS) are often calculated from a large number of covariates. The study also assumed a PS model with only main effects and did not consider interactions or quadratic terms, which could complicate the imputation model.

The article highlights the importance of using IPTW after multiple imputation (MI) correctly, particularly when handling time-dependent confounding. However, some issues may arise when some estimated weights are too extreme, in which case PS matching could be a better solution.

Overall, MI followed by pooling of treatment effect estimates is the preferred approach among those studied when data are missing at random (MAR), and the outcome must be included in the imputation model. The Authors expected that PS-based methods (matching, stratification, covariate adjustment, and IPTW) perform well in estimating the treatment effect within each imputed dataset as long as the underlying assumptions are fulfilled.

Supplementary Material

All code for reproducing the analyses in this paper is publicly available at

References

Clemence Leyrat, I. R. W., Shaun R Seaman, & Williamson, E. J. (2019). Propensity score analysis with partially observed covariates: How should multiple imputation be used? *Statistical Methods in Medical Research*, *28*(1), 3–19. `https://doi.org/10.1177/0962280217713032`

Hill, J. (2004). REDUCING BIAS IN TREATMENT EFFECT ESTIMATION IN OBSERVATIONAL STUDIES SUFFERING FROM MISSING DATA. *Institute for Social and Economic Research and Policy*, *25*(1), 188–204.

Mitra, R., & Reiter, J. (2016). A comparison of two methods of estimating propensity scores after multiple imputation. *Statistical Methods in Medical Research*, *25*(1), 188–204. `https://doi.org/10.1177/0962280212445945`