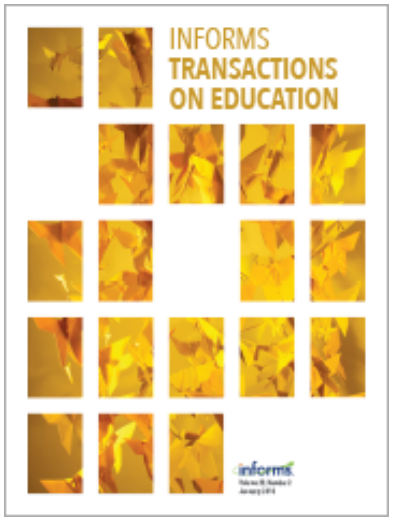
This article was downloaded by: [216.73.217.172] On: 17 November 2026, At: 00:00
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Case—The RealPro Customer Benefits Program (B): Implementing Covariate Balancing and Difference-in-Differences Analysis

Arnd Huchzermeier, Jannik Wolters, Marcel Uphues

To cite this article:
Arnd Huchzermeier, Jannik Wolters, Marcel Uphues (2023) Case—The RealPro Customer Benefits Program (B): Implementing Covariate Balancing and Difference-in-Differences Analysis. INFORMS Transactions on Education 24(1):33-34. https://doi.org/10.1287/ited.2021.0257csb

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. You are free to download this work and share with others, but cannot change in any way or use commercially without permission, and you must attribute this work as "*INFORMS Transactions on Education*. Copyright © 2022 The Author(s). https://doi.org/10.1287/ited.2021.0257csb, used under a Creative Commons Attribution License: https://creativecommons.org/licenses/by-nc-nd/4.0/."

Copyright © 2022 The Author(s)

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

https://pubsonline.informs.org/journal/ited

INFORMS TRANSACTIONS ON EDUCATION
Vol. 24, No. 1, September 2023, pp. 33–34
ISSN 1532-0545 (online)

Case

The RealPro Customer Benefits Program (B): Implementing Covariate Balancing and Difference-in-Differences Analysis

Arnd Huchzermeier,[a] Jannik Wolters,[a] Marcel Uphues[b]

[a] WHU – Otto Beisheim School of Management, 56179 Vallendar, Germany; [b] 40239 Düsseldorf, Germany

Contact: arnd.huchzermeier@whu.edu, https://orcid.org/0000-0003-2178-3254 (AH); jannik.wolters@whu.edu, https://orcid.org/0000-0002-1940-2291 (JW); marcel.uphues@web.de (MU)

Received: October 12, 2020
Revised: March 4, 2021; May 19, 2021
Accepted: May 19, 2021
Published Online in Articles in Advance: January 13, 2022

https://doi.org/10.1287/ited.2021.0257csb

Copyright: © 2022 The Author(s)

Open Access Statement: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. You are free to download this work and share with others, but cannot change in any way or use commercially without permission, and you must attribute this work as "*INFORMS Transactions on Education*. Copyright © 2022 The Author(s). https://doi.org/10.1287/ited.2021.0257csb, used under a Creative Commons Attribution License: https://creativecommons.org/licenses/by-nc-nd/4.0/."

Keywords: subscription program • customer loyalty • grocery retailing • covariate balancing • difference-in-differences analysis

1. Introduction

After reviewing the RealPro analysis based on the collected data, concerns were raised whether RealPro customers and the control group could be directly compared. Customers who were already spending a lot of money at Real in 2018 are more likely to join the RealPro program in 2019 as they receive large absolute discounts without any changes to their shopping behavior. This would distort a direct comparison between both groups, as it would be impossible to determine whether the difference in shopping behavior is the result of a RealPro membership or was already present in the data a priori.

The difference-in-differences (DiD) technique can help to remove biases in the comparison of the RealPro and the control group, such as the self-selection bias. Moreover, covariate balancing methods, which can be used in combination with the DiD methodology, can help negate the differences between RealPro and control group customers along characteristics referred to as covariates. The aim of this covariate balancing is to increase the validity of comparisons between both groups. Two methods appear particularly promising in this context: propensity score matching and entropy balancing.

Mr. Uphues and Mr. Laenge were wondering whether analyzing the market test data with these techniques will give different results as compared with a direct comparison between the RealPro and the control group and what impact this will have on the overall assessment of the RealPro program.

Exhibit 1

The data analysis in this case extension is based on one of two available RealPro market test data sets. When running the DiD analysis without prior covariate balancing, the prebalanced data set with 75 thousand transactions, based on 572 RealPro and control group customers each, should be used. This data set is the same as the one used in the main case. When combining covariate balancing and the DiD analysis, an unbalanced data set with 83 thousand transactions, 572 RealPro customers, and 963 control group customers should be employed.

For both the RealPro group and the control group, transaction data are reported from May 1, 2018, through November 30, 2018, and also for the same time period in 2019. This approach allows one to examine changes in the purchasing behavior of RealPro customers after joining the program. The reported test market data are purposely limited to only seven months because, during the program's first two months (i.e., March and April 2019), many customers were still in the process of joining the program; therefore, including these months in the data set would result in transaction histories of unequal length. Note also that only those RealPro customers who became members before April 1, 2019, are included. This restriction ensures less volatile demand patterns, since the initial analysis established that many customers increased their purchase volume abnormally in their first month of membership but exhibited a more level demand pattern in subsequent months.

Downloaded from informs.org by [216.73.217.172] on 17 November 2026, at 00:00 . For personal use only, all rights reserved.

Table 1. Data Set Description

Feature	Description
Transaction_ID	The unique transaction ID associated with the purchase
Customer_Group	Indicates whether the customer/household has signed up for the RealPro program ("Pro") or is part of the control group ("Control")
Customer_ID	The ID associated with the given customer/household
Store_ID	The ID of the store where the purchase was made
Date	The date of the purchase (yyyy-mm-dd format)
Year	The year of the purchase
Month	The month of the purchase
Week	The week of the purchase
Day	The day of the week of the purchase
Revenue_Transaction	The revenue generated from the purchase[a,b]
Num_Items	The total number of items bought in the purchase
Revenue_Transaction_Pro	The revenue generated from the purchase of products discounted as part of RealPro[a]
Revenue_Transaction_NonPro	The revenue generated from the purchase of products not discounted as part of RealPro[b]
Revenue_Transaction_Promo	The revenue generated from the purchase of products that were price promoted as part of the high–low pricing strategy[b]

[a]RealPro price discount has not been deducted.
[b]High–low price discount has been deducted.

Table 1 describes all features of the two data sets. For each transaction, which is given a unique ID, multiple features are reported. In this data set, the *Customer_Group* column indicates whether the customer/household is a RealPro member or part of the control group. To match all transactions to their respective customers/households, the customer ID is reported in each case. The store ID is similarly reported to enable identification of which transactions occurred at each of the seven stores used in the market test. The *Date* column reports the time of purchase (in yyyy-mm-dd format), and there are several other date-related columns. *Revenue_Transaction* gives the total amount spent by the customer/household on the respective transaction. Although the discount due to high–low promotions has already been deducted from the reported revenue figures, the RealPro discount has not been deducted. The Real team suggests that you do not deduct the RealPro discount when analyzing changes in or associated with the revenue, but only for calculations associated with the profitability of the program. The *Num_Items* variable captures the number of items purchased in the transaction. This number represents the sum of all items bought, not the number of unique products. In addition to the total revenue generated by each transaction, the data set reports the revenue from the purchase of products discounted under the RealPro program separately from that from products that were discounted as part of Real's high–low pricing strategy. There is no overlap between these two revenue figures because, in the RealPro program, already promoted products do not receive an additional program discount.

Downloaded from informs.org by [216.73.217.172] on 17 November 2026, at 00:00 . For personal use only, all rights reserved.