# The Value of Fit Information in Online Retail: Evidence from a Randomized Field Experiment. Online Appendix.

Santiago Gallino
Tuck School of Business, Dartmouth College, Hanover, NH 03755, santiago.gallino@tuck.dartmouth.edu

Antonio Moreno
Harvard Business School, Boston, MA 02163, amoreno@hbs.edu

---

## Appendix. 1. Detailed data description.

### A. Study 1.

**Order level data.** We have an entry for each of the orders placed during the test after a first visit to one of the pages displaying digitized garments.

- Order id: Retailer order id
- Order time: Time stamp for the date and time the order was placed.
- Condition: 1 if the session was assigned to the Metail condition, 0 if assigned to the control condition.
- Order value: Total amount spent in the order.
- Number of products: Total number of products included in the order.
- Average price: Average price of products included in the order, calculated as Order value/Number of products.
- Indicator of digitized set: 1 if the order contained any items from the digitized set, 0 otherwise.
- Order return status: 1 if one or more items in the order was returned to the retailer, 0 otherwise. This variable was collected at the end of the experiment. Because the retailer accepted returns during 30 days, the variable is only accurate for the first month of the experiment.

### B. Study 2.

**Customer A/B assignment data.** We have an entry for each of the customers who clicked on one of the digitized garments and was assigned to one of the conditions, regardless of whether they purchased. We observe:

- Browser fingerprint

- IP address
- Condition: 1 if the customer is assigned to the Metail condition, 0 if assigned to the control condition.
- Order id: Retailer order id if that user made an order, null if not. If the user made multiple orders during the test, we see multiple entries. The orders may or may not contain digitized items.

**Order level data.** We have an entry for each of the orders placed during the test by customers who were assigned to one of the conditions after a first visit to one of the pages displaying digitized garments.

- Order id: Retailer order id
- Order time: Time stamp for the date and time the order was placed.
- Condition: 1 if the customer is assigned to the Metail condition, 0 if assigned to the control condition.
- Engagement indicator: 1 if the user used the Metail Technology in the corresponding order id, 0 if not
- Order level statistics (e.g., total order value, number of items, average price) is generated using the product level data

**Item level data.** We have an entry for each of the items included in orders placed during the test by customers who were assigned to one of the conditions after a first visit to one of the pages displaying digitized garments.

- Order id: Retailer order id for the order in which the item was purchased. This allows to inherit all the order level data described above.
- SKU id: Identifier of the item that was purchased
- Size: Size of the item that was purchased (e.g., L, M, etc.)
- Price: Price of the item that was purchased
- Quantity: Number of units of the item purchased.

**Product information.** For those SKUs that were digitized, we have the following additional information:

- Description: String that describes the product (e.g., "Saia Recorte Preta")
- Available sizes (e.g., P, M, G, etc.)
- Category: Type of garment (e.g., tops, dresses, trousers, etc.)

## Appendix. 2. Product-Level Analysis.

To confirm the robustness of our results, we run several additional analyses. There are two main goals to these analyses. First, we want to confirm that the results we find are not an artifact of some design decisions, such as the level at which we have conducted the main analysis. Second, we want to explore systematic differences in the effects observed. Since the products that were digitized were not chosen at random, it could be that the effects we find are very specific to those particular products. By focusing on different categories, we get a better sense of the potential generalizability of our main insights.

We conduct our robustness studies using detailed product-level sales, which we are able to track in Study 2. We conduct three main analyses. In the first analysis, we reproduce the analysis described in Section **??** but at the product level instead of the order level. In the second analysis, we analyze how sales evolve over time for customers with and without access to the fitting tool. Finally, we conduct a more granular analysis that considers sales in each of the available product categories.

For the first analysis, we break down the products sold during the trial to users who had access to the tool and those who did not have access to the tool. During the trial, 385,384 products were sold to customers who had been assigned to the treatment condition, while 315,050 products were sold to customers who had been assigned to the control condition. This suggests that sales for users with access to the tool were substantially higher overall (22.32 percent larger). If we restrict our attention to just products belonging to the digitized set, 14,142 of those products were sold to customers who had access to the tool, representing a 16.9 percent increase over the 12,095 products sold to customers who did not have access to the tool. We implement a binomial test (see Hollander et al. 2013, Chapter 2), to test whether share of sales that corresponds to the treatment condition exceeds 50% in a statistically significant way. The test allows us to reject the null in favor of the alternative hypothesis that the share of sales from the treatment condition exceeds 50% ($p<0.01$) and therefore we conclude that the positive effects of the tool documented in Section **??** persist even when we consider product-level sales.

Our second analysis studies how sales evolve over time for customers with and without access to the fitting tool. To do that, we compute total number of sold products per week for both the treatment and control groups. If the effects are robust, we would expect sales for the treatment group to be consistently larger than sales for the control group. Figure 1 plots the evolution of both groups, and we can observe that sales to customers in the treatment group (i.e., those customers who have access to the fitting tool) are systematically higher. Table 1, column 1 presents the results in a regression format, where the coefficient of the virtual fitting tool is positive and significant.

**Table 1 Effect of Virtual Fitting Tool on Sales Over Time**

| | (1)<br>Log(Sales) |
|---|---|
| $METAIL$ | 0.205*** |
| | (0.015) |
| Time effects | Week |
| Observations | 20 |
| R-squared | 0.998 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

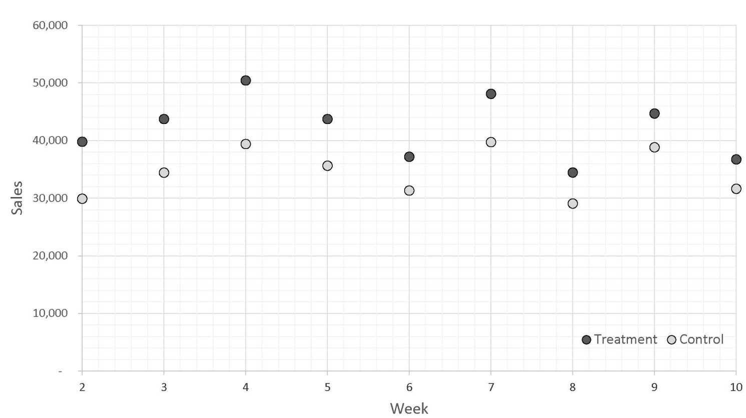


**Figure 1 Sales During the Test Weeks**

The aforementioned discussion considers the aggregate effects at the product level but does not consider the effects in the different product categories. Finally, we conduct a more granular analysis that considers sales in each of the available product categories. Table 2 shows the total number of units sold for products in the digitized set for both for treatment and control customers. If there was no effect in a category, we would expect a very similar share of treatment and control customers in the category. If the virtual fitting tool has an effect on sales in a given category, we would expect the share of treated customers among the purchased products of that category to exceed 50 percent. We formally test this using a binomial test (Hollander et al. 2013, Chapter 2) for each category. The test allows us to reject the null for all categories ($p<0.05$). This suggests that while the precise effects can be slightly different for different products, the benefits of virtual fit information are likely to be materialized for a wide range of products.

**Table 2 Sales by Category (Digitized Set)**

| | Dresses | Jackets & Coats | Knitwear | Shoes | Shorts | Skirts | Tops | Trousers | **Total** |
|---|---|---|---|---|---|---|---|---|---|
| Control | 4,262 | 1,532 | 180 | 334 | 393 | 510 | 3,724 | 1,160 | 12,095 |
| | 47.38 | 45.39 | 41.96 | 46.98 | 42.72 | 42.82 | 45.82 | 46.62 | 46.10 |
| Treatment | 4,734 | 1,843 | 249 | 377 | 527 | 681 | 4,403 | 1,328 | 14,142 |
| | 52.62 | 54.61 | 58.04 | 53.02 | 57.28 | 57.18 | 54.18 | 53.38 | 53.90 |
| **Total** | 8,996 | 3,375 | 429 | 711 | 920 | 1,191 | 8,127 | 2,488 | 26,237 |
| | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

## Appendix. 3. Additional Tables.

To make sure the weights are actually correcting for observable differences in treated- and control-unit characteristics, we follow Guo and Fraser (2009) and we run weighted regressions of the continuous covariates in the treatment. If the propensity score weighting effectively removes imbalances, the coefficient of the treatment variable should not be significant. The following table (Table 3) shows it is the case.

**Table 3 Evidence that Weighting Removes Imbalances**

| | *Dependent Variable:* | | |
|---|---|---|---|
| | order_amount | avg_price | n_items |
| | (1) | (2) | (3) |
| *METAIL* | 0.305 | −0.061 | 0.002 |
| | (0.921) | (0.400) | (0.007) |
| Constant | 187.324*** | 109.097*** | 1.931*** |
| | (0.651) | (0.283) | (0.005) |
| Observations | 211,116 | 211,116 | 211,116 |
| $R^2$ | 0.00000 | 0.00000 | 0.00000 |
| Adjusted $R^2$ | −0.00000 | −0.00000 | −0.00000 |
| Residual Std. Error (df = 211114) | 299.098 | 130.029 | 2.315 |
| F Statistic (df = 1; 211114) | 0.110 | 0.023 | 0.089 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | | |

### References

Guo, Shenyang, Mark W. Fraser. 2009. *Propensity Score Analysis: Statistical Methods and Applications*. 1st ed. SAGE Publications, Inc.

Hollander, Myles, Douglas A Wolfe, Eric Chicken. 2013. *Nonparametric statistical methods*. John Wiley & Sons.