Online Appendix

"The Role of Digital Platforms in Data Markets: How Data Sharing through Advanced Analytics Empowers Small Business Innovation"

Table of Contents

A Adoption Rate of Advanced Data Analytics

In Figure A.1, we depict how the adoption rates of four types of analytics (i.e., descriptive data analytics, advanced data analytics, firm analytics, and market analytics) change over time for the original full sample (100,000 retailers). We observe that the adoption rate of advanced analytics will increase from 6.41% to 31.39% throughout 2021. In May 2021, there was a sharp increase in the adoption rate of advanced analytics, while the adoption rate of descriptive analytics grew smoothly. This illustrates that the free access policy significantly boosted the adoption of advanced analytics but not descriptive analytics. In addition, we also observe that the adoption rate of additional market analytics is significantly lower than that of firm analytics. This means that most retailers adopting advanced analytics only adopt firm analytics, while only a small portion are also adopting market analytics.

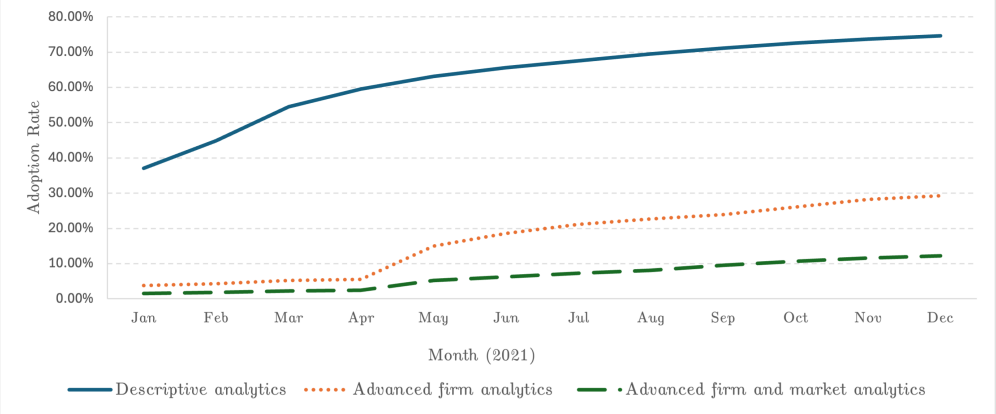


Figure A.1: The Adoption Rate of Data Analytics Over Time

B Dynamic Treatment Effects from SynthDiD Estimation

We retrieve the estimated unit and period weights from our main model to obtain treatment effects at each period. We estimate unit weights ω_i for each retailer in the control group and time weight λ_t for each pre-treatment period for equation (1) with the iterative method by Arkhangelsky et al. (2021). For each of the 12 periods in 2021, we estimate the following function:

$$\widehat{\tau}_t^{\text{adv}} = \left(\frac{\sum_{i \in N^{\text{tr}}} \log\left(Y_{it}\right)}{|N^{\text{tr}}|} - \left(\widehat{\omega}_0 + \sum_{i \in N^{\infty}} \widehat{\omega}_i \log\left(Y_{it}\right) \right) \right) \cdot \left(\widehat{\lambda}_t \cdot I(t \in \text{Pre-Periods}) \right). \tag{4}$$

Where ω_0 is the estimated intercept in the synthetic control procedure, and we use λ_t only for pre-treatment periods (Jan-Apr).

We estimate the heterogeneous effects for each period with equation (4) and the corresponding sub-cohorts described in section 4. The unit and time weights are fixed, and the outcomes and number of treated units are changed with the sub-cohorts. The standard error for each estimation is calculated using the jackknife method. Table B.1 presents the effects for each period. The name for each period is the distance relative to the policy change, and period 0 is May 2021.

Table B.1: Effect of Advanced Data Analytics at Each Period

	By Retailer Size			By Advanced Analytics Type	
Period	**Overall**	**Small**	**Large**	**Firm Only**	**Firm&Market**
-4	0.000	-0.003	-0.010**	0.002	-0.005
	(0.003)	(0.003)	(0.004)	(0.003)	(0.005)
-3	0.000	-0.001	0.002***	0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
-2	0.000	-0.001	0.003	-0.001	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
-1	0.000	0.005	-0.015***	-0.001	0.002
	(0.003)	(0.004)	(0.005)	(0.004)	(0.006)
0	0.183***	0.219***	0.077***	0.168***	0.226***
	(0.013)	(0.015)	(0.016)	(0.15)	(0.025)
1	0.268***	0.302***	0.166***	0.228***	0.307***
	(0.015)	(0.018)	(0.021)	(0.018)	(0.031)
2	0.274***	0.309***	0.169***	0.221***	0.303***
	(0.018)	(0.021)	(0.023)	(0.021)	(0.035)
3	0.277***	0.292***	0.232***	0.201***	0.301***
	(0.019)	(0.021)	(0.024)	(0.022)	(0.036)
4	0.274***	0.290***	0.225***	0.179***	0.311***
	(0.019)	(0.022)	(0.024)	(0.023)	(0.037)
5	0.272***	0.279***	0.250***	0.168***	0.281***
	(0.021)	(0.024)	(0.027)	(0.025)	(0.040)
6	0.282***	0.281***	0.287***	0.147***	0.315***
	(0.022)	(0.025)	(0.030)	(0.027)	(0.043)
7	0.277***	0.279***	0.270***	0.108***	0.321***
	(0.024)	(0.028)	(0.034)	(0.030)	(0.046)

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Robust standard errors in parentheses.

C Staggered Synthetic Difference-in-Differences Estimation

We further implement a staggered *SynthDiD* design following the logic of Callaway and Sant'Anna (2021) and the cohort-based event-time alignment strategy in Berman and Israeli (2022). This approach allows us to exploit variation in adoption timing after the policy change by incorporating multiple adoption cohorts, rather than focusing solely on the initial May adopters. For each cohort, we align treated and control retailers in event time relative to the month of adoption and restrict the control group to not-yet-treated retailers that remain untreated throughout a fixed six-month window around adoption. This design ensures that treated and control units are compared using outcomes from the same calendar months, while avoiding reliance on permanently untreated firms that may differ systematically from adopters. Aggregating across cohorts, we compute an overall treatment effect as a cohort-size-weighted average of the cohort-specific estimates, where both the point estimate and its standard error are constructed using weights proportional to the number of treated retailers in each adoption cohort. We find a positive and statistically significant effect of advanced analytics adoption on sales (ATT = 0.243, $p < 0.01$), which is largely similar to the method that using solely the cohort first adopted advanced analytics right after the policy change period (May 2021) (ATT = 0.276, $p < 0.01$). Table C.1 reports the cohort-specific estimates and corresponding sample sizes.

Table C.1: Estimation Results of Staggered SynthDiD

	log(sales)		
Cohort	**ATT**	Treated Retailers	Control Retailers
May 2021	0.186^{***}	5,566	4,979
	(0.0158)		
June 2021	0.245^{***}	2,059	3,977
	(0.0221)		
July 2021	0.257^{***}	1,607	2,989
	(0.0267)		
August 2021	0.415^{***}	1,002	1,640
	(0.0425)		
September 2021	0.365^{***}	988	494
	(0.0521)		
Weighted Average	$\mathbf{0.243}^{***}$	11,222	–
	(0.0241)		

Notes: This table reports cohort-specific and aggregated treatment effects estimated using a staggered Synthetic Difference-in-Differences design. For each cohort, treated retailers are compared to not-yet-treated retailers that remain untreated throughout a fixed six-month event window relative to adoption. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are reported in parentheses.

D Spillover Diagnostics: Peer Adoption and Non-Adopters' Outcomes

A central concern in our setting is whether adoption of advanced analytics by some retailers affects the outcomes of non-adopters within the same competitive environment. If such competitive spillovers exist, our estimated treatment effect may partially reflect losses among non-adopters rather than genuine improvements for adopters. In addition to the panel difference model in section 4.3 , we conduct a direct diagnostic test examining whether such competitive spillovers appear.

Because retailers compete primarily within industries (e.g. fashion and nutrition products), we construct peer adoption measures at the industry–month level. For each retailer i in industry g and month t, we compute

$$\text{PeerAdoptRate}_{igt} = \frac{\#\{\text{other retailers in } g \text{ adopting at } t\}}{\#\{\text{other retailers in } g\}},$$

where the focal retailer is excluded from the numerator and denominator. We focus on the cohort of non-adopters, since their performance provides the cleanest evidence of whether adopters impose negative spillovers on untreated retailers. We estimate the following panel regression:

$$\log(1 + \text{Sales}_{igt}) = \beta\, \text{PeerAdoptRate}_{igt-1} + \mu_i + \lambda_{gt} + \varepsilon_{igt},$$

Where μ_i denotes retailer fixed effects and λ_{gt} denotes industry-by-month fixed effects. This specification absorbs all time-invariant retailer heterogeneity and all industry-level monthly shocks, including demand cycles, promotional events, and platform-wide seasonality. Standard errors are two-way clustered at the retailer and month levels. Identification therefore comes from within-retailer deviations relative to other firms in the same industry and month.

Table D.1 reports the estimates using both the lagged and contemporaneous peer adoption measures. Across all specifications, the coefficient on peer adoption is small in magnitude and statistically indistinguishable from zero. These patterns do not provide an indication of sizable spillovers in our setting.

This robustness test does not attempt to identify causal peer effects. Instead, it provides a falsification assessment of whether the main treatment effect may be driven by competitive losses among non-adopters rather than genuine improvements among adopters. The absence of any

detectable association between peer adoption intensity and non-adopters' performance suggests that competitive spillovers may not materially bias our main estimates. These results provide consistent evidence that violations of the stable unit treatment value assumption may not be a primary concern in this empirical setting.

Table D.1: Effect of Peer Adoption on Non-Adopters

	log(Sales)	
	Lagged Peer Adoption ($t-1$)	Current Peer Adoption (t)
	(1)	(2)
Peer adoption rate	-0.307	-0.236
	(0.540)	(0.448)
# observations	146,135	159,420
Adjusted R^2	0.8211	0.8100
RMSE	0.7937	0.8177
Retailer fixed effects	Yes	Yes
Industry fixed effects	Yes	Yes
Month fixed effects	Yes	Yes

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are two-way clustered at the retailer and month levels. Robust standard errors in parentheses.

E Panel Difference Estimator

To migrate the SUTVA concern, we also apply the same panel difference specification with the non-adopters following Goldberg, Johnson, and Shriver (2024) and Chiou and Tucker (2022). If spillovers were significant, the performance of non-adopters would decrease after the adoption window. Similarly with the design in the manuscript, we use outcomes from 2019 as the control group and outcomes from 2021 as the treatment group. Specifically, we estimate:

$$\begin{aligned}\log(Y_{it}+1) = \beta_0 \mathbb{1}_{\{2021\}_t} + \beta_1 \left(\mathbb{1}_{\{2021\}_t} \times \mathbb{1}_{\{\text{Post Change}\}_t}\right) \\ + \beta_2 \,\text{Month}_t + \beta_3 \,\text{Month}_t^2 + \alpha_i + \gamma_t + \varepsilon_{it}.\end{aligned} \tag{5}$$

Where $\mathbb{1}_{\{2021\}_t}$ identifies observations from the policy year, and $\mathbb{1}_{\{\text{Post Change}\}_t}$ indicates the months following the start of free access to the advanced analytics module. Duration_t denotes the number of months since the outbreak, and Duration_t^2 allows for nonlinear dynamics in this progression. This flexible time trend captures the accelerated and uneven expansion of global e–commerce following the outbreak. Seller fixed effects α_i absorb time–invariant heterogeneity across retailers, and calendar–month fixed effects γ_t account for platform–wide and seasonal shocks. As shown in column (2) in Table E.1, the estimates indicate no statistically significant change in sales outcomes, suggesting limited cross-unit interference along this margin.

Table E.1: Panel Difference Estimator

	log(Sales+1)	
	(1) Adopters	(2) Non-adopters
$1_{\{2021\}} \times \mathbb{1}_{\{\text{Post Change}\}}$	0.216***	-0.044
	(0.059)	(0.037)
$1_{\{2021\}}$	3.341***	2.212***
	(0.942)	(0.596)
Month	0.142***	0.144***
	(0.040)	(0.025)
Month^2	-0.004***	-0.003***
	(0.001)	(0.001)
Retailer fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
# observations	107,232	231,552
Adjusted R^2	0.311	0.230

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.. Standard errors are two-way clustered at the retailer and month levels. Robust standard errors in parentheses.

F Additional Robustness Checks on Main Effects

F.1 Difference-in-Differences Method

Self-selection is essential in causal inference studies; we use matching and fixed effects to enable a close comparison between the adopters and non-adopters of advanced analytics. For the matching part, we utilize the widely adopted propensity score matching method to pair each treated retailer with a comparable retailer that is very likely to move on to advanced analytics but, for some random reason, did not make the step forward in 2021.

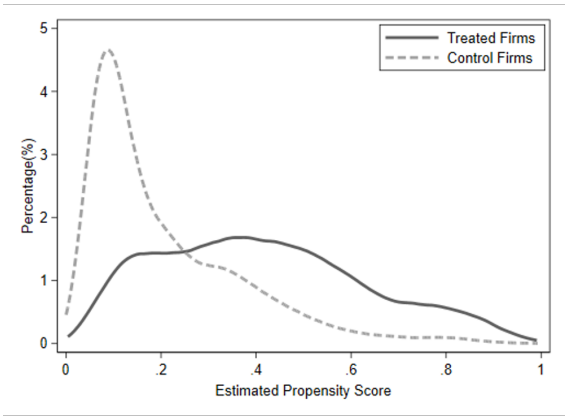
To estimate each retailer's propensity to adopt advanced analytics, we use the record for April 2021, one month before the policy change, and estimate the following function:

$$P(\text{Adopted}_i = 1) = P(\beta_0 + \gamma X_i + \epsilon_i),$$

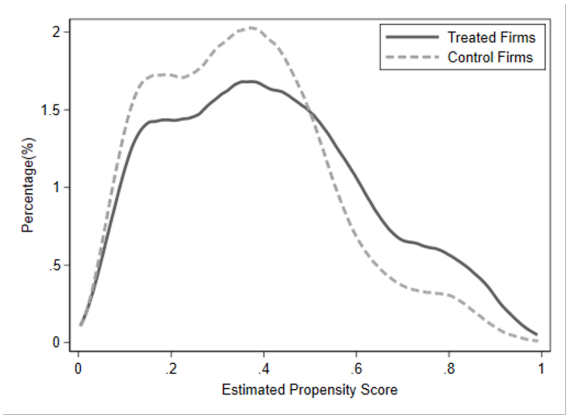
Where X_i is a vector of observable characteristics and performance measurements, including:

1. **Demographics:**
    - Store tenure
    - Store overall customer rating
    - Store grade (evaluated by the platform and shown to customers)
    - Industry
2. **Sales-related performance:**
    - Sales revenue
    - Sales from repeat customers
    - Sales from new customers
    - Sales of the best seller item
3. **Customer-related performance:**
    - Daily average page views
    - Daily average unique visitors count
    - Conversion rate

Figure F.1: Propensity Scores Distribution



(a) Prediction score differences distribution



(b) Prediction scores from both models

- Shopping cart size
- Average transaction value

4. **Product-related performance:**

- Number of items sold
- Median price of all listed products
- SKU count

These covariates all have minimal values for the variance inflation factor test. After fitting the logistic model with the April data, we assign pairs to compose the control and treatment groups. Since we have many more units that did not move on to advanced analytics, and the retailers in our sample are widely distributed across industries, we use a very small caliper (0.005) to find the match and eliminate the treated retailers without a match in the small radius. As a result, we yield 4,640 units left for each control and treatment group.

Figure F.1 shows the distribution of propensity scores before and after the matching process. The matching procedure is valid for fixed effects estimation; we test the bias change and pre-treatment periods parallel trend on sales. Figure F.2 shows the matched and unmatched bias for selected covariates, and Figure F.3 presents the monthly average sales for the control and treatment groups. As expected, the sales in the first four periods have the same trend, and we test with a regression model to see if the adoption impacts sales in the pre-treatment periods.

We use the DiD approach to estimate the impact of adopting advanced analytics for retailers

Figure F.2: Covariates Bias After Propensity Scores Matching

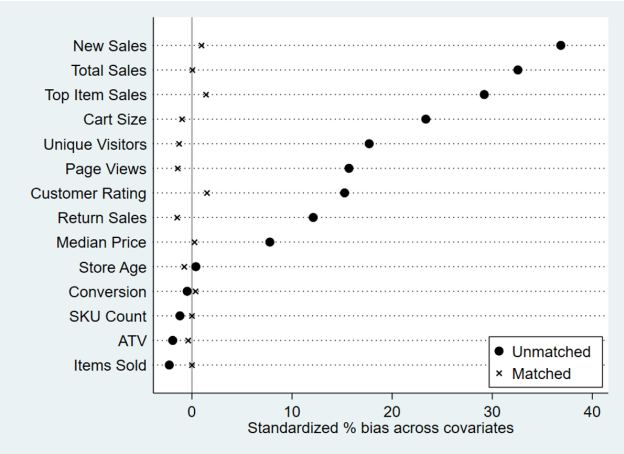
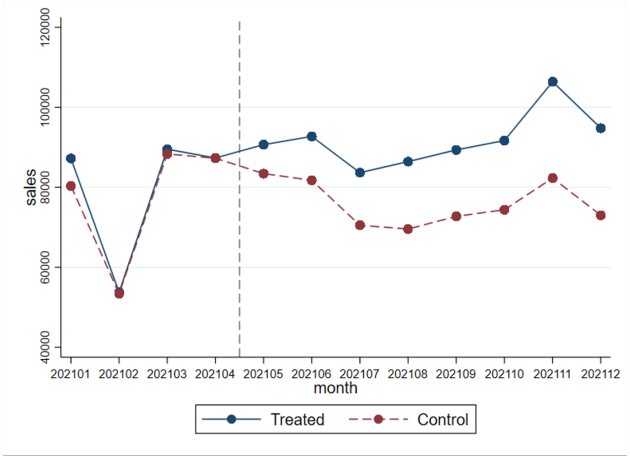


Figure F.3: Monthly Sales of Treatment and Control Group



with experience with descriptive analytics. We estimate the following two-way fixed effect model:

$$Y_{it} = \alpha_i + \lambda_t + \tau D_{it} + \varepsilon_{it}$$

Where α_i and λ_t are the retailer and time fixed effects, respectively, D_{it} indicates whether retailer i has adopted the advanced analytics at time t, and ε_{it} is the idiosyncratic error term. The coefficient τ measures the Average Treatment effect on the Treated (ATT).

Table F.1 presents the estimation results. The revenue growth attributable to advanced analytics is 34.85%. The median sales in pre-treatment periods for treated retailers left after matching is CNY 33,643.8; thus, the economic impact is CNY 11,724.8. This estimation result aligns with our main model.

Table F.1: Treatment Effect with Difference-in-Difference Method

	log(sales)
Treatment	0.299***
	(0.0184)
Constant	9.942***
	(0.0100)
AME (CNY)	11,724.8
AME (%)	34.85%
# observations	124,440
# retailers	10,370
R^2	0.057

Notes: Significance levels: $p < 0.1$ (*), $p < 0.05$ (**), $p < 0.01$ (***); Robust standard errors in parentheses.

Table F.2: Parallel Trend Test Results

	log(sales)
pre 4	0.0156
	(0.0234)
pre 3	-0.0237
	(0.0217)
pre 2	-0.00591
	(0.0139)
Policy Change	0.178***
	(0.0146)
post 1	0.271***
	(0.0181)
post 2	0.297***
	(0.0207)
post 3	0.299***
	(0.0222)
post 4	0.328***
	(0.0227)
post 5	0.314***
	(0.0249)
post 6	0.334***
	(0.0266)
post 7	0.339***
	(0.0292)
Constant	9.934***
	(0.0148)
# retailers	9,888
# observations	118,656
R^2	0.058

Notes: Significance levels: $p < 0.1$ (*), $p < 0.05$ (**), $p < 0.01$ (***); Robust standard errors in parentheses.

F.2 Alternative Control Group with Pioneer Retailers

A potential concern of the quasi-experiment is that there exist unobservable factors that keep the control group retailers from adopting advanced analytics. Therefore, we exclude those nonadopters and instead use retailers that had adopted prior to the policy change, and paid subscription fees. We construct the control group with these retailers, referred to as pioneer retailers. The pioneer retailers have had access to advanced analytics for the entire period of our sample, they would provide a more optimistic counterfactual for the treatment group than the original control group. In turn, the treatment in this analysis can be essentially viewed as the effect of filling the advanced analytics gap.

Following the concept of reverse-DiD framework (Kim and Lee 2019), we estimate synthetic weights for this alternative control group and estimate the SynthDiD model; the result is shown in Table F.3. Filling the advanced analytics gap significantly increases sales revenue by 16.3%,

equivalent to an economic benefit of CNY 6155.2 for the treated firms. We interpret this effect as follows; in this analysis, the treatment group adopted the analytics tools only after the "free access" policy, while the pioneer retailers had already adopted advanced analytics throughout the data period (i.e., both before and after the "free access" policy), and as a result, the difference in performance—post-policy minus pre-policy—for the treatment group should be larger than the corresponding difference for the pioneer retailers (control group). Note that pioneer retailers adopted the advanced analytics prior to the policy change, therefore, their sales trends in pre-treatment periods were expected to grow, setting a higher benchmark than the original control group. In turn, the impact is smaller here than in our main analysis.

Table F.3: Effect of Advanced Data Analytics Using Pioneer Retailers as Controls

	SynthDiD with Alternative Control Group
ATT	0.163^{***}
	(0.015)
# retailers	13,726
# observations	164,712

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Robust standard errors appear in parentheses.

The control group consists of retailers adopted advanced analytics prior to the policy change.

F.3 Alternative Treatment Group Excluding Outliers

To address the possibility that the outliers, particularly top performing retailers, might be driving the observed positive effects of advanced analytics, we implemented a trimming procedure on cohort G4. Specifically, we removed outliers defined by their mean sales falling outside the 1st to 99th percentile range. Following the trimming procedure, we re-estimate the synthetic weights for G3 and then implement the SynthDiD model. The results are presented in Table F.4, showing a similar outcome to our main analysis.

Table F.4: Estimation Results Using Alternative Treatment Group Excluding Outliers

	SynthDiD with Trimmed Treatment Group
ATT	0.275^{***}
	(0.014)
# retailers	20,706
# observations	248,472

Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Robust standard errors appear in parentheses.

F.4 Alternative Outcome Measurements

This robustness check examines whether the main effect of advanced analytics adoption is sensitive to the specific choice of performance outcome. In the main analysis, we use log sales revenue as the primary outcome because it captures both the scale and value of transactions and is the most commonly used metric of marketplace performance. As a robustness exercise, we re-estimate the main SynthDiD model using four alternative performance measures: (1) log(Paid Buyers), capturing the number of unique purchasing customers; (2) log(Sales Quantity), reflecting total items sold; (3) log(Repeat Sales), which isolates purchases from returning customers; and (4) log(New Sales), which measures sales attributable to newly acquired customers.

These outcomes provide complementary perspectives on firm performance. Paid buyers and new sales inform customer acquisition and conversion, sales quantity captures demand volume independent of price, and repeat sales reflect deeper engagement from existing customers. If the observed effect of advanced analytics is genuine rather than an artifact of a particular performance metric, then substituting sales revenue with these outcomes should yield consistent patterns.

The results, reported in Table F.5, closely track the main estimates. In Panel (A), adopting advanced analytics increases paid buyers by 22.4%, sales quantity by 25.4%, repeat sales by 25.6%, and new sales by 34.3%, demonstrating broad and substantial improvements across both acquisition-driven and loyalty-driven outcomes. Panel (B) further shows parallel patterns across analytics types. Retailers relying solely on firm-level analytics experience meaningful gains (14.5–17.4% across acquisition and quantity, 18.0% in repeat sales, and 18.6% in new sales), while those adopting both firm- and market-level analytics achieve markedly larger improvements (29.2–29.9% for acquisition and quantity outcomes, 24.6% in repeat sales, and 33.4% in new sales). These

consistent increases across four alternative performance measures reinforce that the main effect is robust and that market-level data materially amplify the benefits of advanced analytics.

Table F.5: Impact of Advanced Analytics on Paid Buyers and Sales Quantity

Panel (A): Overall Impacts				
Dependent Variable	**log(Paid Buyers)**	**log(Sales Quantity)**	**log(Repeat Sales)**	**log(New Sales)**
ATT	0.224***	0.254***	0.256***	0.343***
	(0.013)	(0.014)	(0.025)	(0.016)
AME	25.1%	28.9%	29.2%	40.9%
# treated retailers	5,566	5,566	5,566	5,566
# control retailers	13,285	13,285	13,285	13,285
# observations	226,212	226,212	226,212	226,212

Panel (B): Impacts by Data Source								
Dependent Variable	**log(Paid Buyers)**		**log(Sales Quantity)**		**log(Repeat Sales)**		**log(New Sales)**	
Data Source	Firm Only	Firm&Market	Firm Only	Firm&Market	Firm Only	Firm&Market	Firm Only	Firm&Market
ATT	0.145***	0.292***	0.174***	0.299***	0.180***	0.246***	0.186***	0.334***
	(0.015)	(0.027)	(0.017)	(0.029)	(0.038)	(0.061)	(0.019)	(0.031)
AME	15.6%	33.9%	19.0%	34.9%	19.7%	27.9%	20.4%	39.7%
# treated retailers	3,166	910	3,166	910	3,166	910	3,166	910
# control retailers	13,285	13,285	13,285	13,285	13,285	13,285	13,285	13,285
# observations	197,412	170,340	197,412	170,340	197,412	170,340	197,412	170,340

Notes: Significance levels: $p < 0.1$ (*), $p < 0.05$ (**), $p < 0.01$ (***). Robust standard errors in parentheses. All dependent variables are transformed via $\log(x + 1)$

F.5 Addressing the Selection Issue on Unobservable for the Advanced Analytics Adoption: Heckman Two-step Approach

One might concern that unobserved factors might affect the retailers' decisions to adopt advanced analytics at the policy change period (May 2021). To address this, we also use the Heckman two-step approach to mitigate this concern.

In the Heckman model, we use the policy change of free access to advanced analytics as the instrumental variable (IV). We employ the advanced analytics service access cost as an instrumental variable (IV). The cost remained stable during the 12 months preceding May 2021, after which it was reduced to zero. Consequently, we assign the indicator a value of 0 for all firms before the policy change (May 2021) and 0 thereafter.

We argue that this instrument is valid, satisfying both the relevance and exclusive restriction conditions. First, the free access policy positively affects the retailers' adoption decision because it significantly reduces the monetary adoption costs for the retailers. Since the service is delivered via web pages, no additional effort is required from retailers. The first stage of estimation results of the Heckman model, presented in Table F.6, empirically confirms that the free access indicator is positively associated with the retailers' adoption decision. Thus, the IV satisfies the relevance condition. Second, because the free access policy is an exogenous event—stemming from a decision

made by the Alibaba Group's management board to inclusively support their retailers—it does not directly affect the platform retailers' performance except through their adoption decisions.[15] Thus, the IV satisfies the exclusive restriction condition.

In the Heckman analysis, we include all retailers in our sample, except those who never adopt descriptive analytics, ensuring that the estimated impact reflects the influence of advanced analytics as an addition to descriptive analytics. In this analysis, retailers that adopted advanced analytics before the free access policy change ("pioneers") must be included to ensure variation in advanced analytics adoption before and after the policy change, allowing the IV to be estimable in the first stage.

In the first step of the Heckman two-step correction model, we address the issue of sample selection bias by estimating a selection equation. This equation models the probability of adopting the advanced analytics service, referred to as the selection mechanism. We specify the selection equation as:

$$P(\text{Adopted}_{it} = 1) = \alpha + \beta X_{it} + \gamma Z_t + \varepsilon_{it} \tag{6}$$

Where:

- $P(\text{Adopted}_{it} = 1)$ indicates whether retailer i has adopted advanced analytics at period t
- Z_t is the instrumental variable: 0 for periods before the cost policy change and 1 thereafter
- X_{it} represents the vector of variables that may influence both the selection process and the outcome:
  - Daily pageviews and unique visitor counts
  - Overall ratings
  - Advertising cost levels
  - SKU counts
  - Number of free descriptive analytics logins

[15] At the time of the "free access" policy launch (May 2021), the platform did not introduce any functional changes to the analytics tool. Additionally, no other significant events occurred on the platform alongside the "free access" policy launch that could have potentially impacted retailers' performance. This helps ensure the validity of the instrumental variable.

Observations after adoption time t are not used to estimate Equation (6) because the decision to adopt advanced analytics is made once. Estimation results in the first step are shown in Table F.6 Panel(A). We validate that the instrument is not weak using:

- Kleibergen-Paap test
- Anderson-Rubin test
- Stock-Wright test

The IV has a significant positive impact on the adoption of advanced analytics. Then, we estimate the inverse Mills ratio from the predicted values of the probit model.

In the second step of the Heckman two-step model, we implement a two-way fixed effect model and estimate the impact of the treatment on the adopted firms while accounting for selection bias corrected in the first step. We estimate the following equation:

$$\log(Y_{it}) = \alpha_i + \beta_t + \lambda \cdot \text{Adopted}_{it} + \gamma \cdot IMR_{it} + \omega X_{it} + \varepsilon_{it} \tag{7}$$

Where:

- α_i and β_t are the retailer entity and time-fixed effects
- IMR_{it} is the inverse Mills ratio, serving as a correction term adjusting for the non-random selection of adoption
- X_{it} represents a vector of time-varying control variables (same as in Equation (6))

Estimation results of the second step are shown in Table F.6 Panel(B). The impact of adopting advanced analytics aligns with our main analysis. Meanwhile, the coefficient for the IMR is significant and negative, which underscores the importance of correcting for selection bias in our analysis. Failing to account for selection bias would likely lead to an overestimation of the treatment effect.

Table F.6: Heckman Two-Step Estimation Results

Panel A: Selection Equation (First Step)		Panel B: Outcome Equation (Second Step)	
Variable	Coefficients	Variable	Coefficients
IV_{it}	1.932***	$Adopted_{it}$	0.240***
	(0.024)		(0.007)
$SKU\ Count_{it}$	-0.000***	IMR_{it}	-0.045***
	(0.000)		(0.001)
$Advertising\ Budget_{it}$	0.093***	$SKU\ Count_{it}$	0.000***
	(0.001)		(0.000)
$Pageviews_{it}$	0.000***	$Advertising\ Budget_{it}$	-0.086***
	(0.000)		(0.001)
$Unique\ Visitors_{it}$	0.000***	$Pageviews_{it}$	0.000***
	(0.000)		(0.000)
$Rating_{it}$	0.354***	$Unique\ Visitors_{it}$	0.000***
	(0.030)		(0.000)
Constant	-6.198***	$Rating_{it}$	0.169***
	(0.157)		(0.003)
		Constant	8.964***
			(0.019)
Retailer fixed effects	Yes	Retailer fixed effects	Yes
Month fixed effects	Yes	Month fixed effects	Yes
# observations	516,349	# observations	516,349
# retailers	57,669	# retailers	57,669

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

G Robustness Check on Estimating Market Analytics

G.1 Estimation of Market Data Impact on Pioneer Firms

As a robustness check isolating the incremental value of market analytics beyond firm analytics, we restrict attention to pioneer retailers that had already adopted firm analytics (with subscription fees) in the pre-treatment period. Within this group, retailers that never adopted market analytics throughout the pre- and post-treatment periods serve as the control group, while retailers that adopted market analytics at the time of the policy change constitute the treatment group. By construction, the two groups differ only in their adoption of market analytics. We then estimate a SynthDiD model on this restricted sample.

As shown in Table G.1, the impact of market analytics, in addition to firm analytics, is significantly positive and generates a 34.3% increase in sales revenue for pioneer retailers.We also estimated the dynamic impact of analytics with additional market data, as shown in Figure G.1. The effects persist over time and align with the main findings, consistent with market data functioning as an innovation input.

Table G.1: The Impact of Market Analytics on Sales Revenue for Pioneer Retailers

	log(Sales)
ATT	0.295^{***}
	(0.057)
Retailer fixed effects	Yes
Month fixed effects	Yes
Control group	Pioneer retailers only adopted firm analytics in both pre- and post-treatment periods (N=2,209)
Treatment group	Pioneer retailers only adopted firm analytics in the pre-treatment period and adopted market analytics in May 2021 (N=336)

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Robust standard errors in parentheses

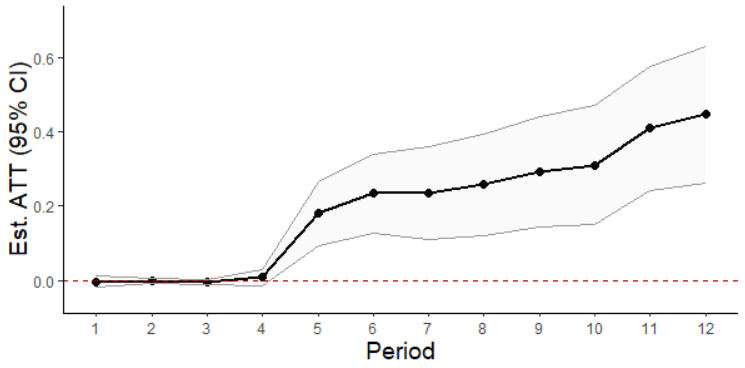


Figure G.1: Dynamic Effects of Advanced Analytics with Market Data Source among Pioneer Retailers

This figure is estimated using pioneer retailers that had already adopted advanced firm-level analytics before the policy change; the treatment group adopted advanced analytics with market data source at the policy change, while the control group never adopted market-data-based analytics.

G.2 Instrumental Variable Approach

Among adopters of advanced analytics, some chose to adopt market-level data analytics in addition to firm-level data analytics, while others adopted only firm-level data analytics. The adopters choosing to adopt market-level data analytics might be affected by some unobserved factors. To address this issue, we follow Berman and Israeli (2022) and use market analytics adoption rates among retailers within the same industry as an instrumental variable for the market adoption decision.

In this analysis, we first focus on retailers in the treatment group, using industry-level market analytics adoption rate as the instrumental variable, following Berman and Israeli (2022). Theoretically, the industry-level market analytics adoption rate will influence the focal retailer's market-level analytics adoption decision within the same industry, satisfying the relevance condition. However, it will not directly affect the focal retailer's sales performance, satisfying the conditions of exclusive restrictions. Thus, it is a valid instrumental variable. We limit periods of study to be June to December 2021 because, first, the adoption cost has to be uniform across all units, and second, the first adopters in May are very likely to be those who always willing to adopt but restrained by cost. For each month between June 2021 and December 2021, we calculate the market analytics adoption rate for each of the major industries. Figure G.2 illustrates the variation in this instrument.

We use the Heckman two-step selection method similarly, as used in Online Appendix D4, to estimate equations (6) and (7) using the new sample and IV. In the first IV model, we solely use

retailers in the treatment group, as they all upgrade to advanced analytics at the first period of policy change (May 2021), but exclude those adopted market analytics at the first period. In the second IV model, we extend to all retailers that upgrade to advanced analytics in the post-treatment periods (after May 2021) and exclude those adopted market analytics at the first period as well. We only use retailers that adopted market analytics after the first period as the treatment units, to minimize the interference of cost change, and ensure the predictive power of the instrumental variable. Again, observations after adoption time t are not used because the decision to adopt market analytics was also made once. Table G.2. A presents the results of the first step. We also validate that the instruments are not weak in either model using Kleibergen-Paap, Anderson-Rubin, and Stock-Wright tests.

In the second step of the Heckman two-step model, we implement a two-way fixed effect model and estimate the impact of the treatment on the adopted firms while accounting for selection bias corrected. The covariates used in the first and second steps are the same as the Heckman model implemented in Online Appendix D4.

The estimation results for the second step are shown in Table G.2. As shown in column (1) of Table G.2, firms that adopt market analytics have a significant positive impact on sales revenue (9.3% increase), indicating the value of market analytics in addition to those that adopt firm-level data analytics only. We then expand the IV model to include all firms that upgraded from descriptive analytics to advanced analytics in the post-treatment period (after May 2021), and the impact is similar, 16.8%, which is shown in column (2) of Table G.2. These results align with our finding that market analytics contribute more to the value of advanced analytics to firms.

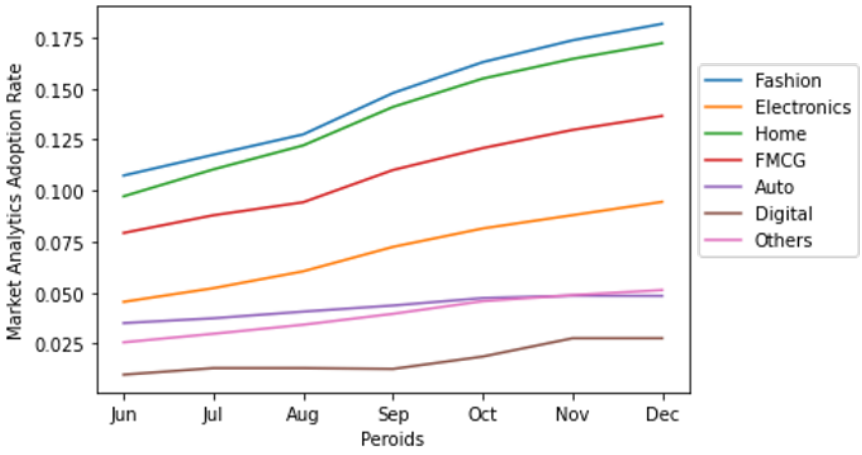


Figure G.2: Market Analytics Adoption Rate by Industry

Table G.2: First Step of Heckman Method

	log(sales)	
	IV Model 1	**IV Model 2**
IV_{it}	11.319***	16.683***
	(2.272)	(1.361)
SKU $Count_{it}$	-0.000	-0.000
	(0.000)	(0.000)
Descriptive Analytics $Usage_{it}$	0.029***	0.035***
	(0.003)	(0.002)
Advertising $Budget_{it}$	-0.046***	-0.045***
	(0.017)	(0.009)
$Pageviews_{it}$	0.000	0.000
	(0.000)	(0.000)
Unique $Visitors_{it}$	-0.000	0.000
	(0.000)	(0.000)
$Rating_{it}$	0.073	0.116*
	(0.149)	(0.064)
Constant	-3.759***	-4.640***
	(0.833)	(0.385)
Retailer fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
# observations	27,460	100,440
# retailers	4,656	16,537

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

Table G.3: Second Step of Heckman Method

	log(sales)	
	IV Model 1	**IV Model 2**
Adopted$_{it}$	0.089***	0.155***
	(0.024)	(0.014)
IMR$_{it}$	-0.213***	-0.271***
	(0.016)	(0.007)
SKU Count$_{it}$	0.000*	0.000***
	(0.000)	(0.000)
Advertising Budget$_{it}$	-0.036***	-0.047***
	(0.006)	(0.003)
Pageviews$_{it}$	0.000***	0.000***
	(0.000)	(0.000)
Unique Visitors$_{it}$	-0.000***	0.000***
	(0.000)	(0.000)
Rating$_{it}$	0.423***	0.214***
	(0.048)	(0.016)
Constant	8.739***	9.821***
	(0.238)	(0.081)
Retailer fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
# observations	27,460	100,440
# retailers	4,656	16,537

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

H Robustness for Heterogeneous Effect across Firm Size

We conducted a robustness check by dividing the control groups into control groups for large retailers (top 25%) and control groups for small retailers (bottom 75%). Then, we use the bottom 75% as the control group for the treatment group of small retailers (bottom 75%) and the top 25% as the control group for the treatment group of large retailers (top 25%). The results, shown in the below table, remain consistent with those in Table 4.

Table H.1: Alternative Control Group for Large vs. Small Retailers

	log(Sales)	
ATT	0.289^{***}	0.253^{***}
	(0.017)	(0.026)
Treatment group	Small retailers in in main treatment group	Large retailers in main treatment group
Control group	Small retailers in main control group	Large retailers in main control group

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

I Interviews with Retailers

I.1 A Haircare Brand

> *"In the rapidly changing market, I will always maintain a high degree of sensitivity to both my firm data and the industry data, allowing me to anticipate industry trends in advance. The Business Advisor gives me more data support combined with my industry experience and insight so that I can find opportunities in the industry one step ahead."*
>
> — Mr. Li, head of operation at TIGI

TIGI, a haircare brand in the United States, entered the online retailing market in China in 2016. Initially, it offered only four types of curling and perm products to female consumers. However, as competition in the female market intensified, TIGI faced challenges such as declining sales, a limited product range, and difficulty in attracting new customers.

A midsize retailer on the Taobao platform, TIGI competed with large brands such as Sassoon, Schwarzkopf, and L'Oreal and experienced a decline in sales in 2018. By using the "Market Overview" function in the "Market Insights" module of Business Advisor, TIGI found that men have been increasingly focusing on personal care in recent years, with both their conversion and repurchase rates surpassing those of women. Recognizing the potential in the men's market, TIGI decided to expand into this segment. Additionally, by analyzing their own SKU (Stock Keeping Unit) data from Business Advisor (firm-level analytics), they found that female consumers primarily purchase 100ml vials, typically buying one bottle at a time, whereas male consumers prefer 385ml bottles and tend to buy two to three at a time. Based on these insights, TIGI designed different SKU sizes, tailored usage scenarios, and distinct packaging names to better cater to each customer segment.

After deciding to enter the men's market, TIGI used the "Market Ranking" function of the "Market Insights" module in Business Advisor to identify potential markets for hair spray. They set out to analyze which market gaps existed. Through market-level data analytics, they found the top-selling products are primarily low-ticket items. Based on these insights, TIGI determined they had a greater opportunity to differentiate by offering high-end hair spray in the male market. By the end of 2021, TIGI had increased its investment in targeting male customers to promote hair spray.

Since entering the men's market, TIGI has experienced continuous growth, and by June 2022,

its men's hair spray category accounted for 24.74% of total store sales. These results show the impact of market-level data analytics on retailers' performance and demonstrate how the tool helps them successfully launch new products.

I.2 An Office Supplies Specialty Store

Prior to 2020, the small office-supplies specialty retailer had focused primarily on the poster category for five years. The onset of the pandemic, however, caused a sharp contraction in this segment, making category expansion an urgent strategic priority.

To identify viable opportunities, the retailer turned to the Market Overview feature within the Market Insights module of Business Advisor. By tracking cross-category search volume, transaction scale, and conversion trends, the retailer pinpointed categories with stronger and more resilient market potential. This analysis revealed that the signage category had been growing at an average annual rate of 27% over the previous three years and offered a substantially larger market size than the store's existing offerings. Guided by this evidence, the retailer selected signage as its expansion direction in 2021.

Once the category was chosen, the retailer needed to select specific SKUs. Using the Search Keywords Analysis tool in Market Insights, the retailer examined market-level search behavior to identify products with high search volume, low competitive density, and strong conversion performance. These insights informed the SKU-level assortment for the signage expansion. For instance, as illustrated in Figure I.1, the retailer originally sold signage products with white backgrounds. Analysis of search rankings and associated conversion metrics revealed that yellow backgrounds performed significantly better due to higher visibility. The retailer adopted this finding and introduced a yellow-background product line, which both retained existing customers and drove notable sales increases.

Following product launch, the retailer further optimized traffic acquisition through search-based advertising. Using Market Insights, the retailer benchmarked competitors' high-performing traffic-driving and conversion keywords and refined its own keyword strategy accordingly to enhance exposure and paid traffic efficiency.

The impact of this data-driven expansion was substantial. From 2020 to 2021, customer visits to the Taobao store increased by 27% following the signage expansion. From 2021 to 2022, visits grew an additional 31%. These sustained gains highlight how market-level analytics can effectively

Figure I.1: Screenshot of Signage with White and Yellow Backgrounds

guide category and SKU expansion and support renewed growth for small retailers.

J Robustness Checks on Product Assortment Innovation

J.1 Controlling for the Influence of Category Expansion

We find that retailers that adopted only advanced data analytics introduced more SKUs in our data settings. An alternative explanation might be that retailers launched new categories and then expanded their SKUs. We conducted a robustness check by removing the effect of category expansion.

Specifically, we excluded the retailers which launched new categories and compared the remained treated retailers with retailers in the control group. The results are shown as follows. In Table J.1, we found that the treatment effect is still statistically significant for aftering removing the effect of category expansion. This is consistent with the result in the manuscript that treatment group retailers launched more SKUs after they adopt the advanced data analytics.

Table J.1: Impact on SKU for Firms without Category Expansion

Panel (A): Overall Impacts	
ATT	0.0615^{***}
	(0.011)

Panel (B): Impacts Across Data Types		
	Firm Only	**Firm & Market**
ATT	0.0426^{***}	0.0882^{***}
	(0.0130)	(0.0266)

Panel (C): Impacts Across Data Types and Firm Size		
	Firm Only	**Firm & Market**
Small Retailers	0.0286^{**}	0.0671^{**}
	(0.0146)	(0.0326)
Large Retailers	0.1050^{***}	0.1408^{**}
	(0.0214)	(0.0430)

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

J.2 Validating the Category Expansion Mechanism

We find that large retailers conduct less category expansion than small retailers do after adopting market-level analytics. An alternative explanation might be that large retailers already have a large category coverage, so there is less room for expansion.

We conducted additional analysis to rule out this alternative explanation. Since the synthetic DID method used in our main analysis does not allow for the inclusion of control variables, we divided the firms into two groups with large and small numbers of categories based on the median of all firms' category counts. We estimated the synthetic DID for eight groups in total, respectively ($2 \times 2 \times 2$): large retailers vs. small retailers, firm analytics only vs. firm and market analytics, and a large number of categories vs. a small number of categories. The estimation results are shown in Table J.2.

In our data, the mean number of categories is 2.27 for large retailers adopting market analytics with a small number of categories, which is smaller than the mean number of categories (9.88) for small retailers adopting market analytics with a large number of categories. The estimation results in Table J.2 show that adopting market analytics for large retailers even with a small number of categories does not significantly positively affect category expansion (0.181). However, adopting market analytics for small retailers with a large number of categories significantly still positively affects their category expansion (0.838^{***}). These findings do not support the alternative explanation that large retailers already have a large category coverage, so there is less room for expansion.

Table J.2: Heterogeneous Effects on Category Expansion by Firm and Market Analytics

	Category Expansion	
	Upper	**Lower**
Large Retailers		
Firm&Market	-0.123	0.181
	(0.295)	(0.139)
Firm Only	0.589^{***}	0.289^{**}
	(0.222)	(0.130)
Small Retailers		
Firm&Market	0.838^{***}	0.416^{***}
	(0.301)	(0.125)
Firm Only	-0.143	0.348^{***}
	(0.245)	(0.132)

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

J.3 Weighted Category Expansion Index

Mere expansion into additional categories, however, does not reveal whether firms are using platform-shared data to target underexplored areas of demand. If platform-shared data truly act as informational capital, we would expect retailers to be more likely to enter less crowded or emerging categories, where external demand signals provide a clearer advantage beyond what can be inferred from their own histories. To shed light on the extensive margin behind this change, we also track the number of categories that appear in a retailer's assortment for the first time in that period. Specifically, for each new category k introduced by retailer i in period t, we assign a weight $\left(1 - \frac{N_{kt}}{N_{\mathcal{I}t}}\right)^2$, where N_{kt} is the number of retailers offering category k in period t, and $N_{\mathcal{I}t}$ is the total number of retailers in the industry in period t. The weight increases when fewer firms in the industry (such as fashion or consumer electronics) serve the category, and decreases when the category is widely covered, so the resulting index reflects the extent to which retailers expand into potentially under-served, categories.

Table J.3 shows that the treatment effect on this competition-weighted expansion measure is substantial and statistically significant when retailers have access to market-level information, but is not significant when they rely only on firm-level data. This pattern is intuitive: retailers are less likely to experiment with unfamiliar categories without boarder market knowledge, and are more likely to do so when platform-level signals reveal demand patterns beyond their existing footprint. The heterogeneity by firm size echoes the unweighted counts but with magnified differences, reinforcing the view that external market information is particularly important for smaller firms when they consider entering new and under-served categories.

Table J.3: Impact of Advanced Analytics on Weighted Category Expansion

Panel (A): Overall Impacts		
ATT	0.369^{***}	
	(0.132)	
AME	44.6%	
# treated retailers	5,566	
# control retailers	13,285	
# observations	226,212	
Panel (B): Impacts by Data Source		
	Firm Only	**Firm&Market**
ATT	0.223	0.423^{***}
	(0.142)	(0.156)
AME	25.0%	52.7%
# treated retailers	3,166	910
# control retailers	13,285	13,285
# observations	197,412	170,340
Panel (C): Cross Effects with Firm Size and Data Source		
	Firm Only	**Firm&Market**
Small Retailers	0.141	0.608^{***}
	(0.153)	(0.171)
AME	15.1%	83.7%
# treated retailers	2,423	641
# control retailers	13,285	13,285
# observations	188,496	167,112
Large Retailers	0.490^{***}	-0.019
	(0.170)	(0.218)
AME	63.2%	-1.9%
# treated retailers	743	269
# control retailers	13,285	13,285
# observations	168,336	162,648

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

J.4 Causal Mediation Analysis

This appendix provides additional evidence that product-assortment innovation serves as a mediating channel through which advanced analytics adoption affects retailers' sales revenue. Because mediation analysis requires explicit modeling of intermediate variables and heterogeneous treatment timing, we implement the analysis using a staggered difference-in-differences (DiD) framework with propensity-score matching, rather than the Synthetic DiD estimator used in the main analysis. The DiD design is well suited for mediation: it accommodates staggered adoption across cohorts, allows mediators and time-varying covariates to enter flexibly, and yields interpretable decompositions of total, direct, and indirect effects. In contrast, the Synthetic DiD framework is optimized for constructing aggregate counterfactuals but does not naturally incorporate mediation variables or sequential regression structures. Therefore, the DiD-with-matching design serves as an appropriate complementary estimator that preserves causal comparability while enabling path decomposition. The estimation procedure follows Habel, Alavi, and Linsenmayer (2021) and Wu, Chen, and Naik (2024).

The analysis proceeds in three steps. First, we estimate the total effect of advanced analytics adoption on sales revenue using a staggered DiD specification identical to that in Online Appendix F.1. Second, we estimate DiD models in which each assortment-innovation measurement, is treated as the dependent variable, yielding the effect of adoption on the mediators. Third, we estimate full mediation models in which sales revenue is regressed on both the treatment indicator and the mediators, allowing decomposition of the treatment effect into direct and indirect components.

Table J.4 reports the results of the full mediation analysis. Consistent with the mechanism evidence in Section 5.2, the adoption of advanced analytics significantly increases all three assortment-innovation measures. In turn, these measures positively and significantly predict sales revenue when included jointly with the treatment indicator. The residual direct effect of adoption is attenuated when the mediators enter the model, indicating that a meaningful share of the total impact operates through the assortment-innovation channel. Taken together, these results provide formal causal evidence that product-assortment innovation mediates the effect of advanced analytics adoption on retailers' sales revenue.

Table J.4: Causal Mediation Analysis

		Product Assortment Innovation			**log(Sales)**		
	log(Sales)	**log (SKU count)**	**log (Category Count)**	**Distinction Score**		**log(Sales)**	
ATT	0.280^{***}	0.070^{***}	0.069^{***}	0.018^{***}	0.262^{***}	0.210^{***}	0.267^{***}
	(0.022)	(0.004)	(0.003)	(0.004)	(0.022)	(0.009)	(0.022)
Mediating effects							
log (SKU count)					0.256^{***}		
					(0.004)		
log (Category Count)						1.292^{***}	
						(0.009)	
Distinction Score							0.697^{***}
							(0.019)
Retailer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# observations	124,440	124,440	124,440	124,440	124,440	124,440	124,440
$AdjustR^2$	0.033	0.001	0.001	0.045	0.072	0.128	0.043

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.Robust standard errors in parentheses. Category Expansion is the weighted number of new categories at each period.

J.5 Usage-Based Falsification Test

To further assess whether the observed assortment-innovation effects are driven by active use of advanced analytics rather than inherent differences between adopters and non-adopters, we conduct a falsification test based on retailers' actual usage intensity of the analytics tool, following the usage-data construction in Section 4.3. The intuition is simple: if assortment innovation arises from analytics-informed decision-making, then such effects should be present only among retailers that actively use the tool. Retailers that activate the analytics service but do not engage with it should exhibit no systematic changes in innovation outcomes.

We classify retailers into three groups based on their observed usage behavior. Non-users are those who activated the analytics account but never logged into the system. Among retailers who logged in at least once, we split them at the median (or 75th percentile, as noted) of usage frequency into high-intensity and low-intensity users. Holding the control group and synthetic weights fixed, we re-estimate the SynthDiD model for each usage group to examine how treatment effects vary with actual utilization.

The results, presented in Table J.5, show a clear monotonic pattern that supports the mechanism. Non-users display no statistically significant changes in SKU Count, Category Expansion Index, or Category Distinction Score, indicating that mere activation of the analytics tool—without meaningful usage—does not lead to product-assortment adjustments. In contrast, both low-intensity and high-intensity users exhibit positive treatment effects on assortment-innovation outcomes, with substantially stronger effects for the high-intensity group. These patterns reinforce a causal interpretation: the innovation responses documented in the main analysis arise from active engagement with the analytics tool than from unobserved differences in retailer characteristics.

Table J.5: Estimation with Analytics Usage Levels

Usage Group by Cutoff Level	*50th Percentile*		*75th Percentil*		
	High Usage	**Low Usage**	**High Usage**	**Low Usage**	**None-users**
log(SKU Count)	0.170^{***}	0.081^{***}	0.153^{***}	0.055^{***}	0.031
	(0.011)	(0.007)	(0.009)	(0.008)	(0.024)
log(Category Count)	0.10^{*}	0.038	0.181^{***}	0.032	0.074
	(0.056)	(0.049)	(0.065)	(0.047)	(0.085)
Distinction Score	0.006	-0.004	0.000	-0.003	-0.001
	(0.005)	(0.003)	(0.004)	(0.004)	(0.001)
# treated retailers	2,756	2,810	1,348	4,182	135
# control retailers	13,285	13,285	13,285	13,285	13,285
# observations	192,492	193,140	176,026	209,604	161,040

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses. Category Expansion is the weighted number of new categories at each period.

K Sub-Group Estimation on Price and Advertising Budget Effects

Table K.1: Impact of Advanced Analytics by Data Source and Firm Size

Panel (A): Impacts by Analytics Data Source										
Dependent Variable	**log(Average Price)**		**log(Median Price)**		**Increased Price Count**		**Decreased Price Count**		**Advertising Budget Change**	
	Firm Only	Firm & Market	Firm Only	Firm & Market	Firm Only	Firm & Market	Firm Only	Firm & Market	Firm Only	Firm & Market
ATT	-0.002	-0.002	-0.001	-0.005	1.219	-0.264	-1.084	-1.307	0.012	0.011
	(0.012)	(0.008)	(0.008)	(0.013)	(1.080)	(0.480)	(1.958)	(1.037)	(0.013)	(0.021)
# treated retailers	3,166	910	3,166	910	3,166	910	3,166	910	3,166	910
# observations	197,412	170,340	197,412	170,340	197,412	170,340	197,412	170,340	197,412	170,340
Panel (B): Impacts Across Firm Size										
Dependent Variable	**log(Average Price)**		**log(Median Price)**		**Increased Price Count**		**Decreased Price Count**		**Advertising Budget Change**	
	Firm Only	Firm & Market	Firm Only	Firm & Market	Firm Only	Firm & Market	Firm Only	Firm & Market	Firm Only	Firm & Market
Small Retailers	0.002	-0.002	0.003	-0.008	0.983	0.023	-0.908	-0.259	0.018	0.031
	(0.009)	(0.015)	(0.009)	(0.015)	(0.830)	(0.309)	(1.491)	(0.251)	(0.015)	(0.027)
# treated retailers	2,423	641	2,423	641	2,423	641	2,423	641	2,423	641
# observations	212,016	239,412	188,496	167,112	209,508	209,508	209,508	209,508	188,496	167,112
Large Retailers	-0.016	-0.001	-0.015	0.000	-0.172	-0.947	-0.918	-3.802	-0.014	-0.043
	(0.012)	(0.021)	(0.013)	(0.021)	(0.392)	(1.373)	(0.699)	(3.431)	(0.017)	(0.027)
# treated retailers	743	269	743	269	743	269	743	269	743	269
# observations	204,300	229,620	168,336	162,648	176,124	176,124	176,124	176,124	168,336	162,648

Notes: Significance level: $p < 0.1$ (*), $p < 0.05$ (**), $p < 0.01$ (***). Robust standard errors in parentheses.

L Marginal Effects under Synthetic Difference-in-Differences

Our main analysis employs the synthetic difference-in-differences estimator, which combines features of both synthetic control and traditional two-way fixed effects (TWFE) to estimate treatment effects in panel data. In this appendix, we describe the construction of marginal effects from SynthDiD estimates for interpretation on the original outcome scale.

Let Y_{it} denote the observed outcome (in our case, log(sales)) for unit i at time t, and let $W_{it} \in \{0, 1\}$ indicate treatment status. We assume a panel of N units and T time periods, with T_{pre} pre-treatment periods.

The SDID estimator first uses the pre-treatment outcomes to estimate:

- A vector of time weights $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_{T_{\text{pre}}})$ to balance pre- and post-treatment periods;
- A vector of unit weights $\hat{\omega} = (\hat{\omega}_1, \ldots, \hat{\omega}_{N_{\text{co}}})$ to construct a synthetic control group that best matches the treated units in pre-trends.

Using these weights, the estimator for the average treatment effect on the treated (ATT) is:

$$\hat{\tau} = \frac{1}{N_{\text{tr}}} \sum_{i > N_{\text{co}}} \hat{\delta}_i - \sum_{i \leq N_{\text{co}}} \hat{\omega}_i \hat{\delta}_i, \tag{8}$$

where the unit-level pseudo-treatment effects $\hat{\delta}_i$ are given by:

$$\hat{\delta}_i = \frac{1}{T_{\text{post}}} \sum_{t=T_{\text{pre}}+1}^{T} Y_{it} - \sum_{t=1}^{T_{\text{pre}}} \hat{\lambda}_t Y_{it}. \tag{9}$$

To obtain marginal effect interpretation, we follow the regression-based representation of SDID in Arkhangelsky et al. (2021), where the ATT can be equivalently obtained by solving a weighted two-way fixed effects regression:

$$Y_{it} = \mu + \alpha_i + \beta_t + \tau W_{it} + \varepsilon_{it}, \tag{10}$$

where $(\mu, \alpha_i, \beta_t, \tau)$ are estimated via weighted least squares using the estimated SDID weights: $\hat{\omega}_i$ over units and $\hat{\lambda}_t$ over time. This regression-based view provides an interpretable decomposition of the predicted potential outcomes.

For each treated unit $i > N_{\text{co}}$ and post-treatment period $t > T_{\text{pre}}$, we construct predicted outcomes under both treatment ($W_{it} = 1$) and counterfactual non-treatment ($W_{it} = 0$) as follows:

$$\hat{Y}_{it}^{(1)} = \mu + \hat{\alpha}_i + \hat{\beta}_t + \hat{\tau}, \tag{11}$$

$$\hat{Y}_{it}^{(0)} = \mu + \hat{\alpha}_i + \hat{\beta}_t. \tag{12}$$

Given that the original outcome is in logs, we back-transform the predicted log-scale outcomes to the level scale, correcting for the log-normal bias using the residual variance estimate $\hat{\sigma}^2$:

$$\hat{y}_{it}^{(1)} = \exp\left(\hat{Y}_{it}^{(1)} + \frac{\hat{\sigma}^2}{2}\right), \tag{13}$$

$$\hat{y}_{it}^{(0)} = \exp\left(\hat{Y}_{it}^{(0)} + \frac{\hat{\sigma}^2}{2}\right). \tag{14}$$

We compute the average marginal effect (AME) across all treated units and post-treatment periods as:

$$\text{AME} = \frac{1}{N_{\text{tr}} T_{\text{post}}} \sum_{i > N_{\text{co}}} \sum_{t=T_{\text{pre}}+1}^{T} \frac{\hat{y}_{it}^{(1)} - \hat{y}_{it}^{(0)}}{\hat{y}_{it}^{(0)}}. \tag{15}$$

This AME provides an interpretable, level-scale measure of the relative change in outcomes due to treatment, correcting for the non-linearity of the log transformation.

M The Impact of Descriptive Analytics

We focus on those who adopted the descriptive analytics between March and September 2021 for the staggered synthetic DID analysis. While our complete data period spans from January to December 2021, we left two months in the pre-period (January and February 2021) to construct synthetic weights in the DID model and reserved three months in the post-period (October to December 2021) to examine post-treatment effects.

The average treatment effect (τ) from the SynthDiD model is 0.198 ($p < 0.001$). Thus, the ATT for the impact of descriptive analytics is 21.9% after adoption. This impact translates to an economic effect equivalent to CNY 7366.73 per month for a median retailer in our sample. Our results show a significant positive effect of descriptive analytics on retailer sales, consistent with the main findings of Berman and Israeli (2022) and Bar-Gill et al. (2024).

Table M.1: The Impact of Descriptive Analytics Adoption

Cohort	**ATT**	**Std. Err**	**Number of Treated Retailers**
March 2021	0.174^{***}	0.018	6,169
April 2021	0.182^{***}	0.019	3,589
May 2021	0.231^{***}	0.022	2,320
June 2021	0.184^{***}	0.022	1,803
July 2021	0.213^{***}	0.024	1,516
August 2021	0.232^{***}	0.024	1,634
September 2021	0.253^{***}	0.025	1,354
Weighted average	0.198^{***}	0.020	

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors in parentheses.

References in the Online Appendix (not in the Manuscript)

Habel, J., Alavi, S., and Linsenmayer, K. (2021). Variable compensation and salesperson health. Journal of Marketing, 85(3), 130-149.

Wu, B., Chen, Y., and Naik, P. A. (2024). How own delivery services influence customer behavior and sales in online retail? Building trust and improving delivery quality in digital economy. Journal of Marketing, 88(5), 131-152.