Estimating new brand entry effects in plant-based beef alternative markets: a comparative study of (extended) two-way fixed effects and rolling approach

Jiayu Sun
Ph.D. Candidate
Michigan State University

Authors: Jiayu Sun, Vincenzina Caputo, Hanna Taylor

*This project was partially funded by the USDA, Economic Research Service. The findings and conclusions in this report are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.
*The analysis, findings, and conclusions expressed in this presentation should not be attributed to Circana.

Department of Agricultural, Food, and Resource Economics
MICHIGAN STATE UNIVERSITY

2024 AAEA Annual Meeting
New Orleans, LA

Motivation

- The rapid growth of the PBMAs market has attracted significant investments.
  - In 2022 alone, over twenty brands announced new plant-based facilities and product introductions, with most expected to launch by 2024 (GFI, 2022).

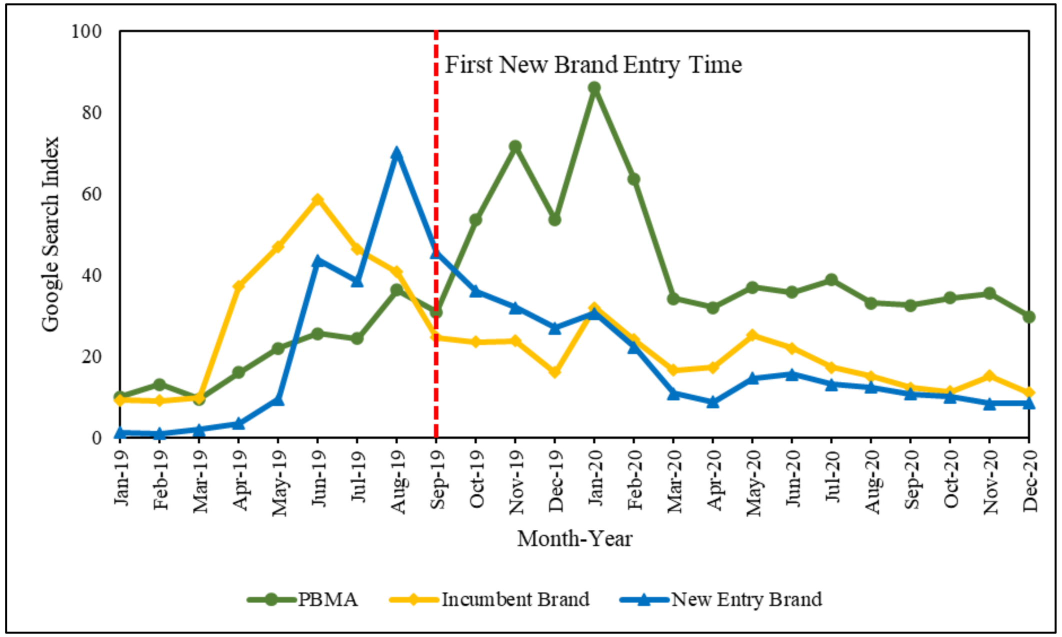
- New brands enter food markets every year, raising various questions including:
  - Can new brands replicate the early success of existing players in the PBMA market?
  - Will these new entrants compete with existing brands or attract new consumers?

- Addressing these questions is crucial for understanding the market dynamics of new PBMA entrants and their potential impact on consumer preferences and overall food industry.

Motivation and Objective 1

- However, the dynamics of market impacts resulting from new brand entries remain underexplored.
- Previous studies across various industries have shown mixed entry effects (Cao et al., 2021; Reshef, 2023):
    - New PBMA entrants may compete with incumbent brands without expanding the PBMA market.
    - New PBMA could stimulate market growth by attracting new consumers and potentially increasing overall demand for PBMAs.

Figure 2 Google Search Interest of PBMA, incumbent brand, and new entry brand in US.



Note: Data is from Google Trends.

Objective 1

Examine the impact of new PBMA brand entries on incumbent brands and their role in driving the overall market expansion of PBMAs

Methods for Objective 1

To reach Objective 1, I use …

Data: IRI Retail Scanner Dataset

- Store-month-brand level sales data (one incumbent brand and one entry brand)
- Timeframe: January 2019 to December 2020
- 6906 stores: 3018 control stores and 3888 treated stores
- Dependent variables: 1) Incumbent brand sales, 2) Incumbent brand price, and 3) total PBMA sales

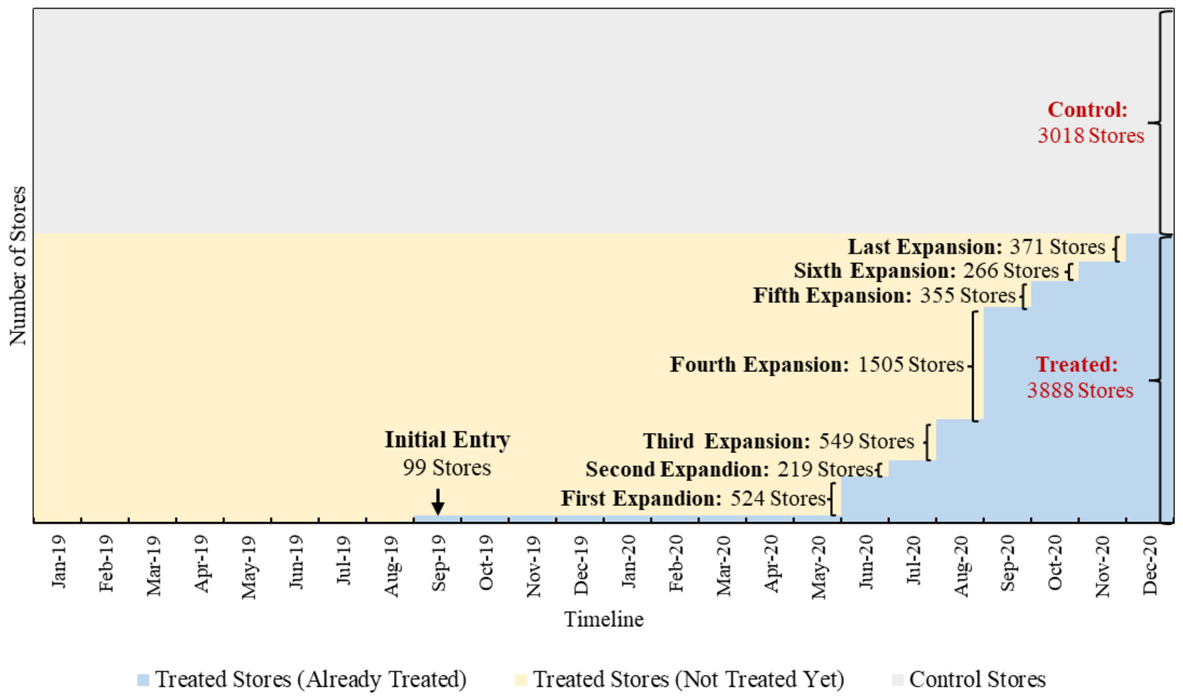
Model: Two-Way Fixed Effect (TWFE)

- Evaluates average entry effects.
- It has been widely used in the entry effect literatures (e.g., bike-sharing, transportation, and accommodation) (Zervas et al., 2017; Cao et al., 2021; Berger et al., 2018)

Modeling Issues I

- **The new brand entry is staggered.**
  - The new PBMA brand enters different stores at different times: initial entry + seven waves of expansion).
- **TWFE fails to capture …**
  - Heterogenous effects: The impact of the brand entry varies from one cohort to another.
  - Dynamic effects: The impact of the brand entry changes over time after the initial entry point.
- **TWFE strongly relies on …**
  - the homogeneous treatment effects across time and cohorts.

Figure 3 Data Structure



Potential Solutions to Solve Model Issue I

- Recent and emerging literatures suggested the cautious application of TWFE in staggered intervention framework and recommended alternative approaches.

	Average Effects	Dynamic Effects	Heterogenous Effects	Parallel Trend
Baseline				
TWFE	Biased	-	-	Relaxed
Alternative Approaches				
De Chausemartin and D'Haultfoeuille (2020)	Unbiased	-	-	Relaxed
Sun and Abraham (2021)	Unbiased	Y	Y	Relaxed
Callaway and Sant'Anna (2021)	Unbiased	Y	Y	Relaxed
Borusyak et al. (2021, 2024)	Unbiased	Y	Y	Strict
Wooldridge (2021) (ETWFE)	Unbiased	Y	Y	Hetero Linear

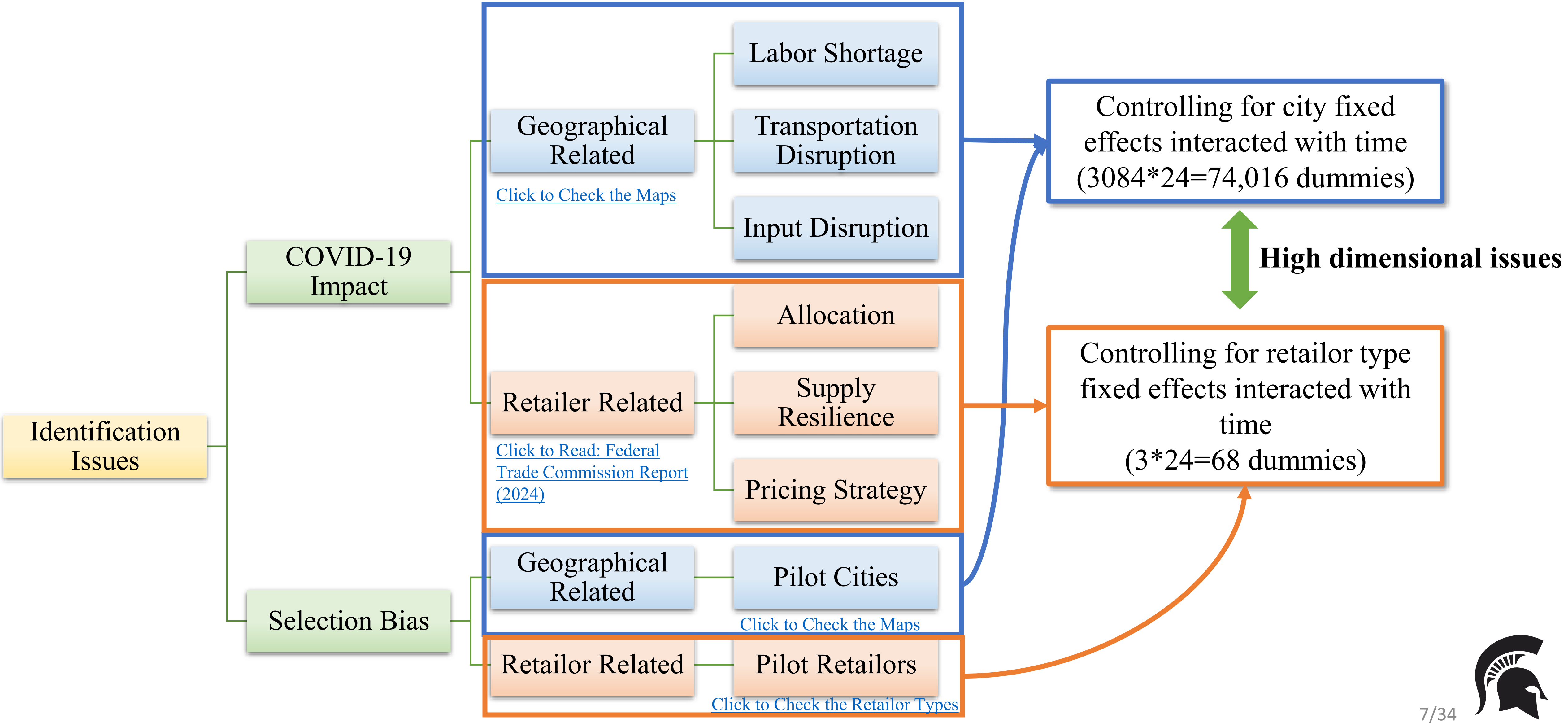
Objective 2

- ETWFE has been applied to study the effects of staggered adoption of new technologies and policy interventions (Berman and Israeli, 2022; Xiao et al., 2023).
- Its application in analyzing staggered entry effects in market scenarios remains underexplored.

Objective 2:

Extend the use of ETWFE in food economics to evaluate heterogeneous and dynamic effects associated with staggered entry of new PBMA brands.

Modeling Issues II



Combining Methods and Objective 3

Double Machine Learning (DML)

Chemozhukov et al. (2017, AER; 2018, The Econometrics Journal)

- DML provides doubly robust estimators when the covariates are high dimensional.
- It has been applied in conventional DID panel data structure but not staggered intervention situations.

Rolling Approach

Lee and Wooldridge (2023, WP)

- It is a unit-specific data transformation approach to estimate staggered treatment effects
- It allows the application of DML after data transformation.

Objective 3:

Extend the use of DML within staggered interventions in food economics by integrating it with the rolling approach to handle high dimensionality in estimation.

Contributions

Empirical Contribution

- Filling the literature gap by the empirical evidence of the impact of new PBMA brand entry on incumbent brand and its role in driving the overall market expansion of PBMAs.

Methodological Contributions

- Extending the use of ETWFE to evaluate heterogeneous and dynamic brand entry effects, providing a more accurate identification of these effects.
- Extending the use of DML within staggered intervention contexts, integrating it with the rolling approach in food economics.
- Comparing the performance of TWFE, ETWFE, and rolling approach with DML. Contributing to the discussion of the estimations in staggered intervention context.

Empirical Approach

Model 1: TWFE

- **Model Specification**:

$$Y_{it} = \beta PostEntry_{it} + \alpha_i + \gamma_t + \varepsilon_{it}$$

- Y_{it}: Dependent variables at store i in month t
    - Incumbent PBMA brand sales ($Ln(Sales)_{In,it}$)
    - Incumbent PBMA brand price ($Price_{In,it}$)
    - Total PBMA sales ($Ln(Sales)_{FPBBA,it}$)
- $PostEntry_{it}$: Dummy variable that equals one if month t is on or after the new brand began to be sold in store i;
- α_i: Store fixed effects;
- γ_t: Month fixed effects;
- ε_{it}: Error term.
- β measures the average impact of the new entry across time and cohorts.

Model 2: ETWFE

- Following Wooldridge (2021), each specification was formulated as follows:

$$Y_{it} = \sum_{g=S}^{T} \delta_g \cdot D_{ig} + \sum_{g=S}^{T} \sum_{r=g}^{T} \tau_{gr} \cdot D_{ig} \cdot f_{rt} + \sum_{g=S}^{T} \varphi_g \cdot D_{ig} \cdot t + \alpha_i + \gamma_t + \varepsilon_{it}$$

- Y_{it}: Dependent variables at store i in month t;
- D_{ig}: Cohort dummy; = 1 if the new brand first enters store i in month g (referred to cohort g); and zero otherwise, meaning either the store was in control group, or the treatment occurred in a different month.
- f_{rt}: Binary indicator; =1 if the time t corresponds exactly to the post-entry time r, indicating a direct match in the timeline; otherwise, it is set to 0.
- φ_g captures the linear time trends of cohort g; the coefficients α_i and γ_t denote the store and time fixed effects, respectively; and ε_{it} is the error term.
- τ_{gr} is the coefficient that measures the entry effect of cohort g in post-entry month r.

Model 3: Rolling Approach with Double Machine Learning

Following Lee and Wooldridge (2023), we implemented four key steps:

Click for more details

Step 1 — Detrending the outcome variables → Run regressions at store level

Step 2 — Constructing the key independent variables → Cohort dummy (D_{ig}) as independent variable

Step 3 — Constructing sub-datasets → $SubData_{gr}$: each post-entry time r

- Treated: Observations of stores in entry cohort g
- Control: Observations of stores never treated

Step 4 — Assessing doubly robust estimators → Apply double machine learning on $SubData_{gr}$

- Controlling city and store-type fixed effect

Summary of Empirical Approaches and Outcomes

- **Classic Approach**
  - Model 1: TWFE → Average Effects
- **Advanced Approaches**
  - Model 2: ETWFE → Heterogenous and Dynamic Effects → Average Effects
  - Model 3: Rolling Approach with DML → Heterogenous and Dynamic Effects → Average Effects

- Comparison 1: Average Effects from three approaches
  - Disclose the biasness of TWFE estimates
- Comparison 2: Root Mean Squared Errors
  - Disclose the model precision

Results

Department of Agricultural,
Food, and Resource Economics
MICHIGAN STATE UNIVERSITY

Heterogenous and Dynamic Effects

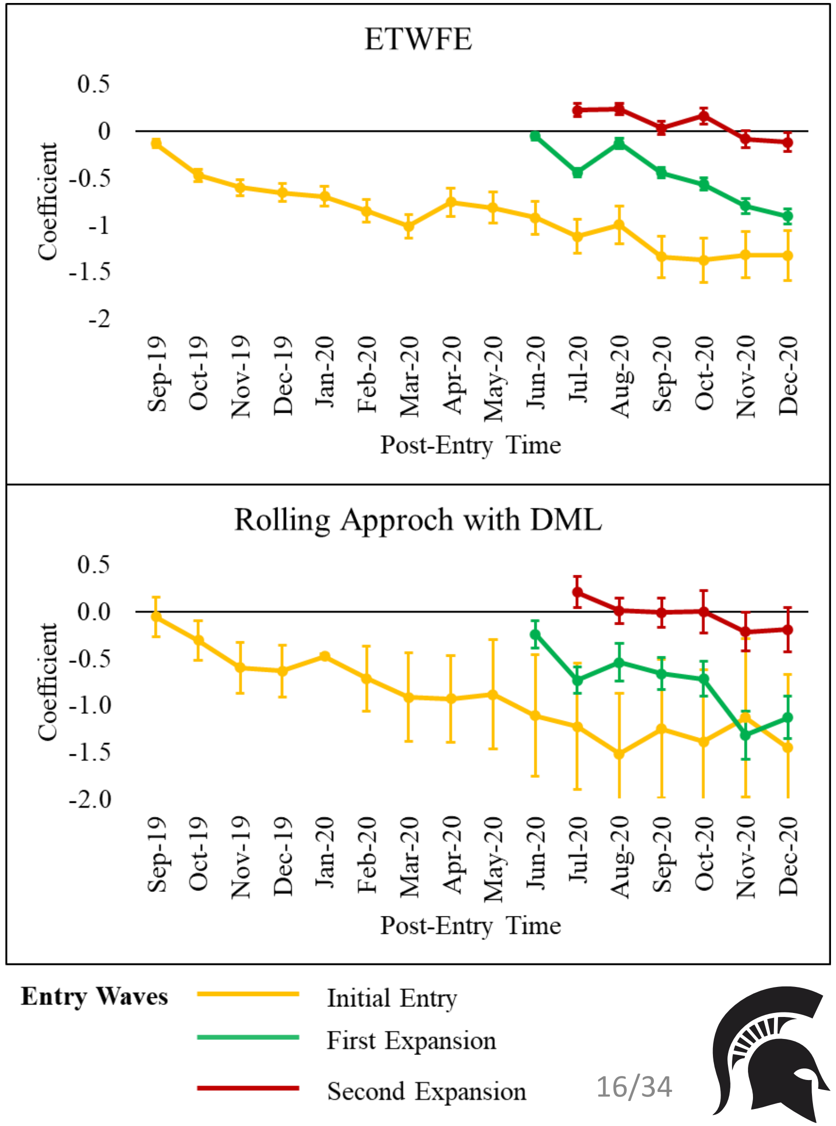
Heterogenous effects:

- The entry effects differ across entry cohorts: Initial Entry, First expansion, Second Expansion.
- When the new brand enters the market, the incumbent brand sales …
    - decrease in both initial entry and first expansion stores
    - increase in second expansion stores

Dynamic effects:

- The entry effects differ across different post-entry times, as shown in the x-axes of Figure 3.5.
- Effect increases with post-entry time periods in initial entry and first expansion, but diminishes with post-entry time periods in second wave of expansion

Figure 5 Heterogenous and dynamic entry effects on incumbent brand sales



Average Effects

- There are substantial differences between the TWFE and the ETWFE and the rolling approach with DML
- These differences have also been found in previous research in other topics
  - Callaway and Sant'Anna (2021)
  - de Chaisemartin and D'Haultfoeuille (2020)
  - Xiao et al. (2023)
  - Nagengast and Yotov (2023).
- This difference discloses the biasness of TWFE estimates and the improvement of the alternative approaches.

Dependent Variables	Average Entry Effects		
	TWFE	ETWFE	Rolling Approach with DML
$Ln(Sales)_{In,it}$	-0.015* (0.006)	-0.541* (0.050)	-0.682* (0.162)
$Price_{In,it}$	0.052* (0.011)	0.627* (0.050)	1.000* (0.268)
$Ln(Sales)_{FPBBA,it}$	0.358* (0.006)	-0.099* (0.048)	-0.227 (0.164)

Comparison 2:
ETWFE vs. Rolling Approach with DML

RMSE

The rolling approach with DML has **smaller** (24-45%) RMSE than ETWFE.

- The rolling approach with DML improves model precision over the ETWFE model.

Conclusion

- **Empirical:**
  - The results suggest that entry effects vary across geographical locations, entry waves, and post-entry times.

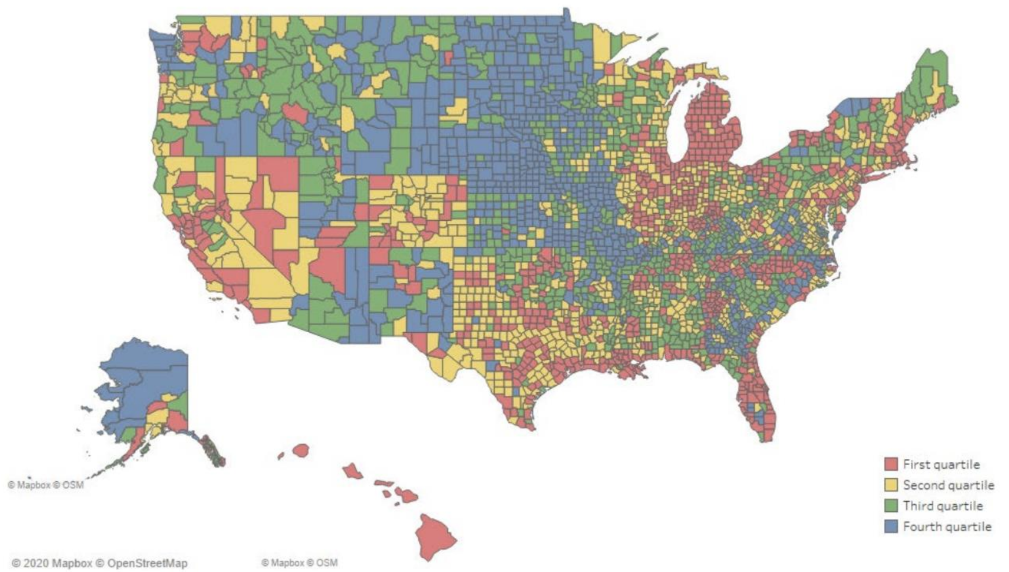
- **Methodological:**
  - The TWFE estimates could be biased when the staggered entry effects are not homogenous across entry waves and post-entry times, while ETWFE and the rolling approach with DML could produce less biased estimates.
  - Compared to the other models, the rolling approach integrated with DML controls for selection bias by including high-dimensional covariates, leading to an improved model precision ranging from 24.3% to 44.6%.

THANK YOU

Jiayu Sun
sunjiay5@msu.edu

Department of Agricultural, Food, and Resource Economics
MICHIGAN STATE UNIVERSITY

Projected change in the growth of all food sales because of COVID-19 recession



Source: USDA ERS (2021) https://www.ers.usda.gov/webdocs/publications/100426/ap-088.pdf?v=199

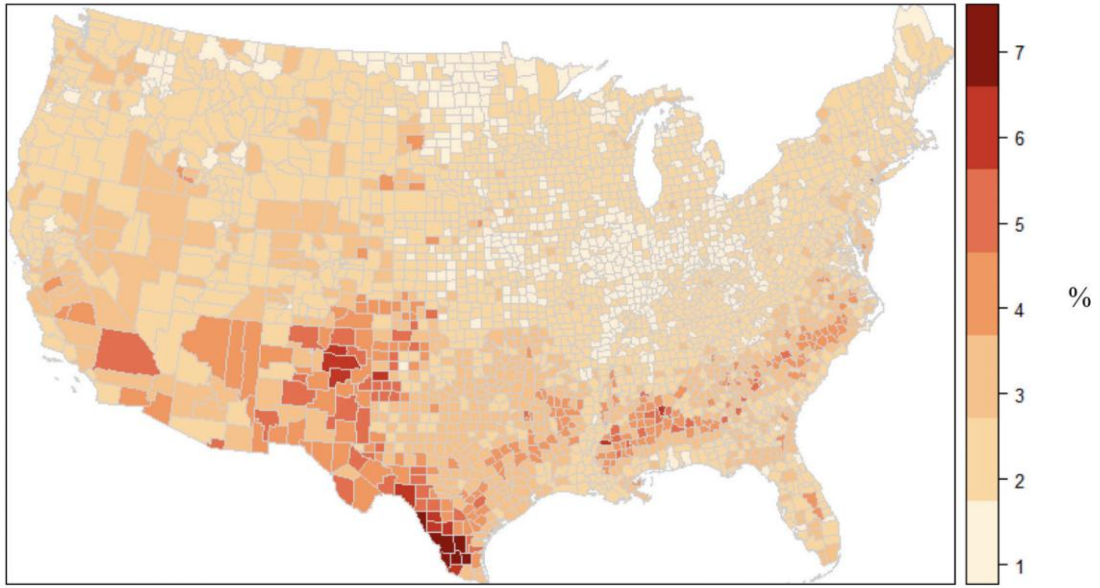
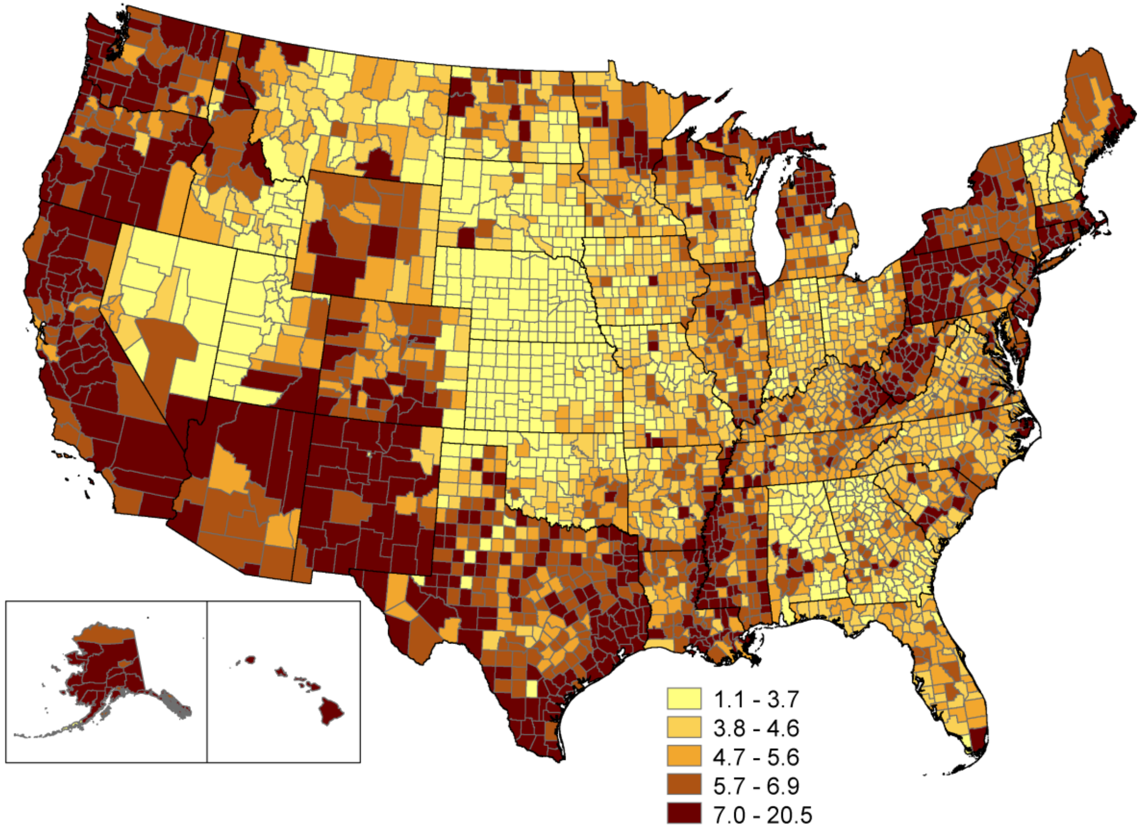


Fig. 5. Estimated share of production not realized due to COVID-19, by county. Source: simulation results.

Source: Haqiqi and Horeh (2021). https://doi.org/10.1016/j.agsy.2021.103132

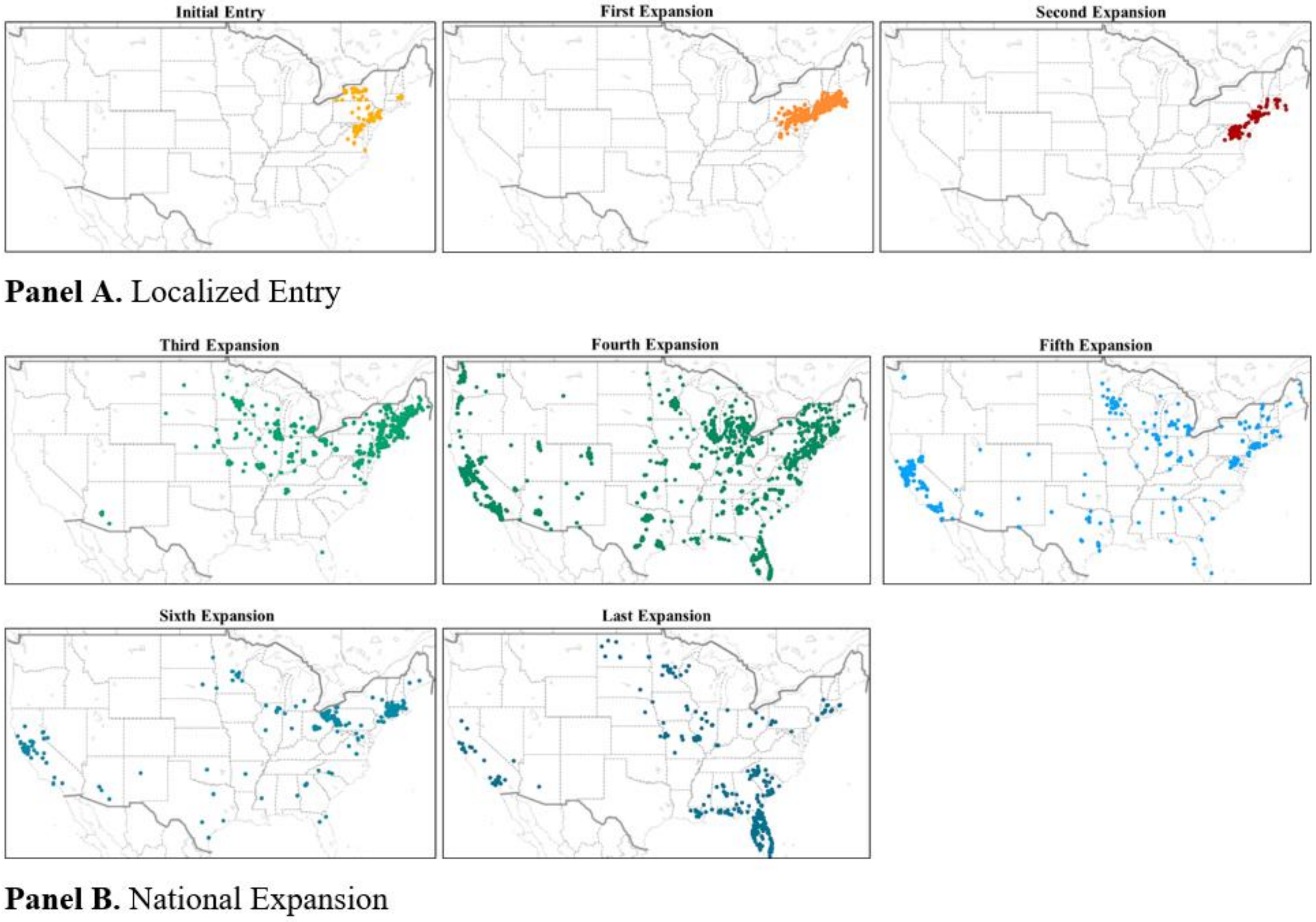
U.S. county unemployment rates during the week of March 12, 2021



Note: The ranges of unemployment rates shown in the map are quintiles of the distribution. Data are not seasonally adjusted.
Source: USDA, Economic Research Service using data from U.S. Department of Labor, Bureau of Labor Statistics, Local Area Unemployment Statistics program (accessed June 3, 2021).

Go Back

Geographical Entry Distribution

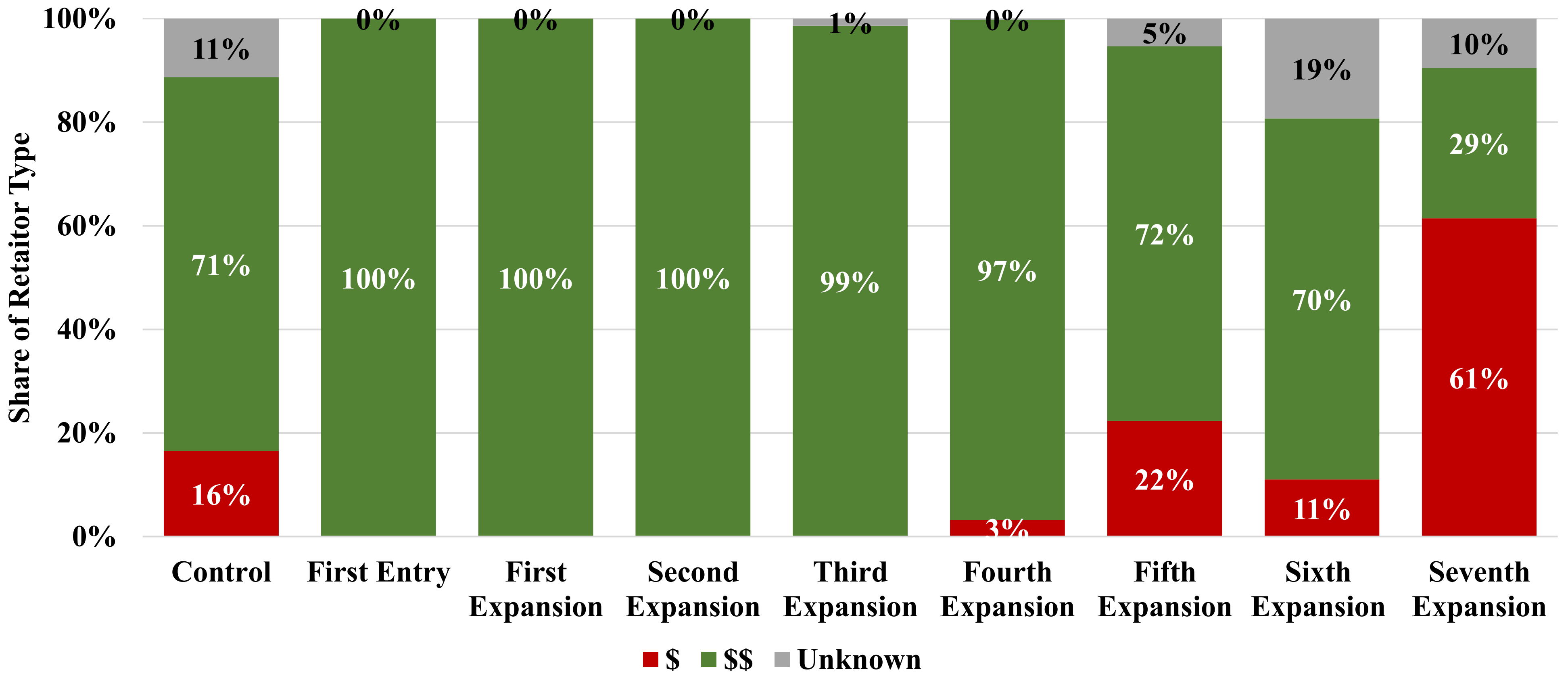


The entry of a new brand into specific geographical locations is not random, as its geographical distribution varies across entry waves

Go Back

Retailor Types by Entry Waves

- Stores are classified as $, $$, and unknown based on the price level in Google Map.
- New brand entered pricier stores ($$) first (from first entry to third expansion).



Go Back

Model 3: Rolling Approach with Double Machine Learning

Step 1

Detrending the outcome variables

For each store, i, in a treated cohort, g, we perform store-specific regressions for the pre-treatment period $t = 1, \dots, g-1$:

$$Y_{itg} = \alpha_i + \theta_i \cdot t$$

- α_i: Store fixed effect;
- θ_i: Store specific time trend.

Post-entry outcomes are adjusted based on these regressions to isolate the effects of new brand entries:

$$\dot{Y}_{irg} = Y_{irg} - \hat{Y}_{irg}$$

where $\hat{Y}_{irg}$ is the out-of-sample predicted value from equation

Step 2

Constructing the key independent variables

D_{ig}: Cohort dummy; = 1 if the new brand first enters store i in month g (referred to cohort g); and zero otherwise, meaning either the store was in control group, or the treatment occurred in a different month.

Go Back

Model 3: Rolling Approach with Double Machine Learning

Step 3

Constructing sub-datasets

- For each entry cohort g, there are $\mathrm{T} - g + 1$ sub-datasets.
- Each sub-dataset, denoted as $SubData_{gr}(r = g, \ldots T)$, includes:
  - Treated: observations of stores in entry cohort $g(D_{ig} = 1)$ in specific post-entry time r;
  - Control: observations of stores where the new brand never entered ($D_{i\infty} = 1$) in the same time r.

Step 4

Assessing doubly robust estimators

- Following Chemozhukov et al. (2017, 2018), the DML model is specified as follows:

$$\dot{Y}_{irg} = \theta_{rg} \cdot D_{ig} + g(\boldsymbol{X_i}) + U_{irg}$$

$$D_{ig} = m(\boldsymbol{X_i}) + V_{irg}$$

- The functions $g(\boldsymbol{X_i})$ and $m(\boldsymbol{X_i})$ represent unknown function of covariates $\boldsymbol{X_i}$ (city dummies, retailor dummies)
- θ_{rg} represents the new brand entry effect on the treatment group cohort g in post-entry month r.

[Click to check the estimation details of of θ_{rg}]

[Go Back]

DML Procedure

- To estimate the treatment effects, θ_{rg}, we followed Chemozhukov et al. (2017, 2018) and applied three additional steps.
  - First, we randomly and evenly split the data into K folds ($K = 5$) and each fold is represented by I_k ($k \in [K] = \{1, \dots, K\}$).
  - Second, for each fold I_k we estimated the nuisance functions ($\hat{g}(\boldsymbol{X_i})_{i \in I_{\neq k}}$ and $\widehat{m}(\boldsymbol{X_i})_{i \in I_{\neq k}}$) using the data from the remaining $K - 1$ folds ($I_{\neq k}$) as follows:

$$\hat{\theta}_{rg,k} = (\frac{1}{n}\sum_{i \in I_k}(D_{ig} - \widehat{m}(\boldsymbol{X_i})) \cdot D_{ig})^{-1} \cdot \frac{1}{n}\sum_{i \in I_k}(D_{ig} - \widehat{m}(\boldsymbol{X_i})) \cdot (\dot{Y}_{irg} - \hat{g}(\boldsymbol{X_i}))$$

  where the nuisance functions measure the relationships between covariates $\boldsymbol{X_i}$ and the treatment indicator D_{ig}.

  - Finally, we averaged the treatment effect estimates ($\hat{\theta}_{rg,k}$) across the 5 folds to obtain the overall estimation of $\hat{\theta}_{rg}$ for each entry cohort $g$and post-entry time, $\hat{\theta}_{rg} = \frac{1}{K}\sum_{k=1}^{K} \hat{\theta}_{rg,k}$.

RMSE Calculation

To compare the performance of the ETWFE model and the rolling approach integrated with DML, we used the Root Mean Squared Error (RMSE); the smaller out-of-sample RMSE represents more precise model estimation. Following Bajari et al. (2015), for the ETWFE method, the RMSE was calculated as the root mean squared differences between actual value of outcome variables and the predicted value of outcome variables on the out-of-sample data:

$$\sqrt{\frac{1}{n}\sum_{i=1,i\in I_k}^{n}(\hat{Y}_{irg} - Y_{irg})^2}.$$

For the method of rolling approach with DML, the RMSE was calculated by taking the square root of the average of the squared differences between the predicted values and the actual values of the outcome variables for each data point in the out-of-sample dataset:

$$\sqrt{\frac{1}{n}\sum_{i=1,i\in I_k}^{n}(\widehat{\dot{Y}_{rg}}_i - \dot{Y}_{rg_i})^2}.$$

It is important to note that the out-of-sample RMSEs for the ETWFE are based on the actual dataset, while those for the rolling approach with DML are derived from the detrended data. To make the RMSEs from these two methods comparable, we follow the normalization method described by Scherbakov et al. (2013).

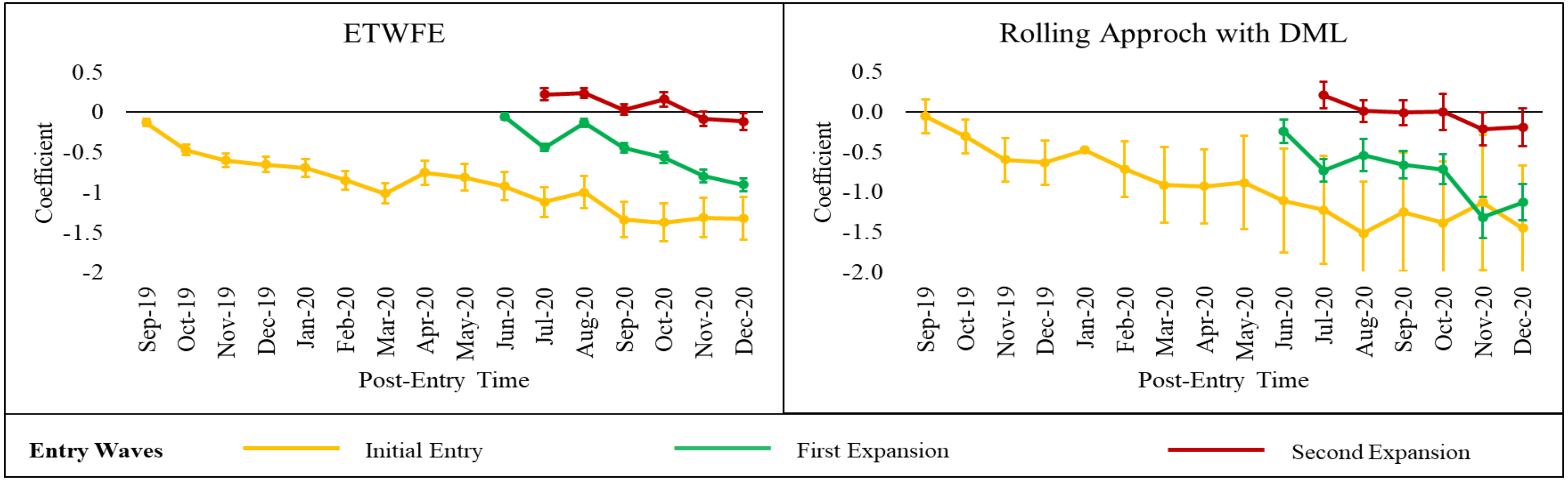
Results from the ETWFE Rolling Approach with DML: Incumbent Brand Sales

Heterogenous effects:

- When the new brand enters the market, the sales of incumbent PBMA brand …
  - decrease in both initial entry and first expansion stores
  - increase in second expansion stores

Dynamic effects:

- (-) Negative effect of new brand entry increases with post-entry time periods in initial entry and first expansion
- (+) Positive effect of new brand entry diminishes with post-entry time periods in second wave of expansion



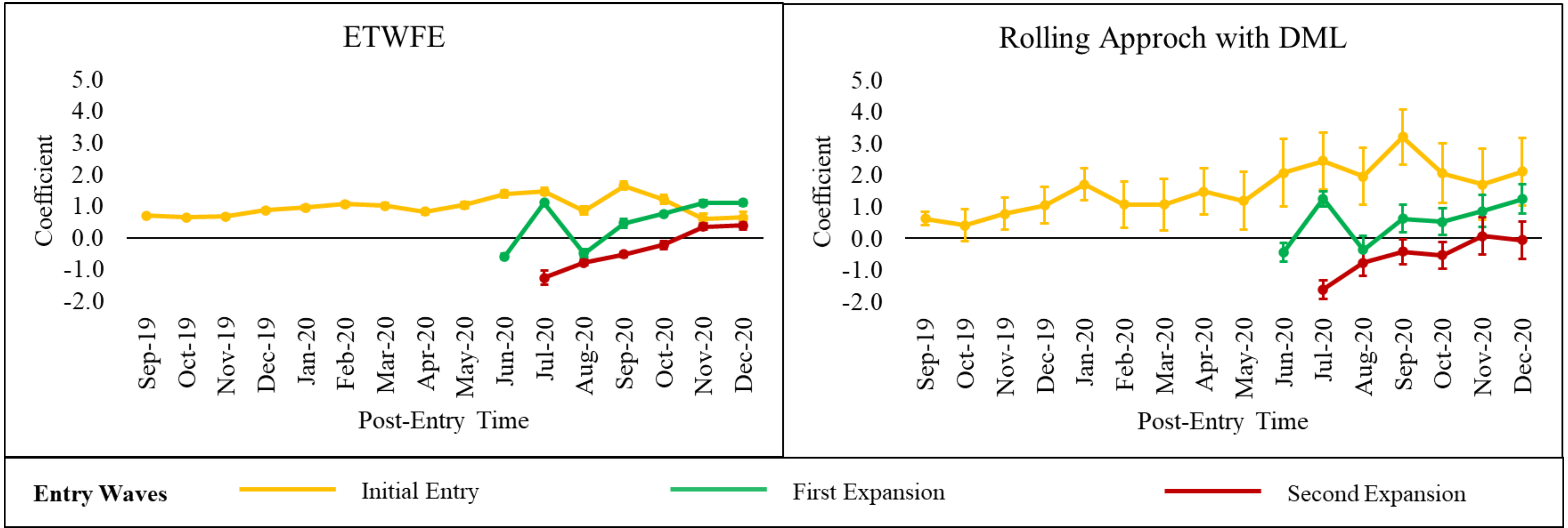
Results from the ETWFE Rolling Approach with DML: Incumbent Brand Price

Heterogenous effects:

- When the new brand enters the market, the incumbent PBMA brand price …
    - increase in both initial entry and first expansion stores
    - decrease in second expansion stores

Dynamic effects:

- (+) Positive effects are statistically significant across post-entry time periods in initial entry and first expansion
- (-) Negative effects are only statistically significant in early post-entry times in the second expansion



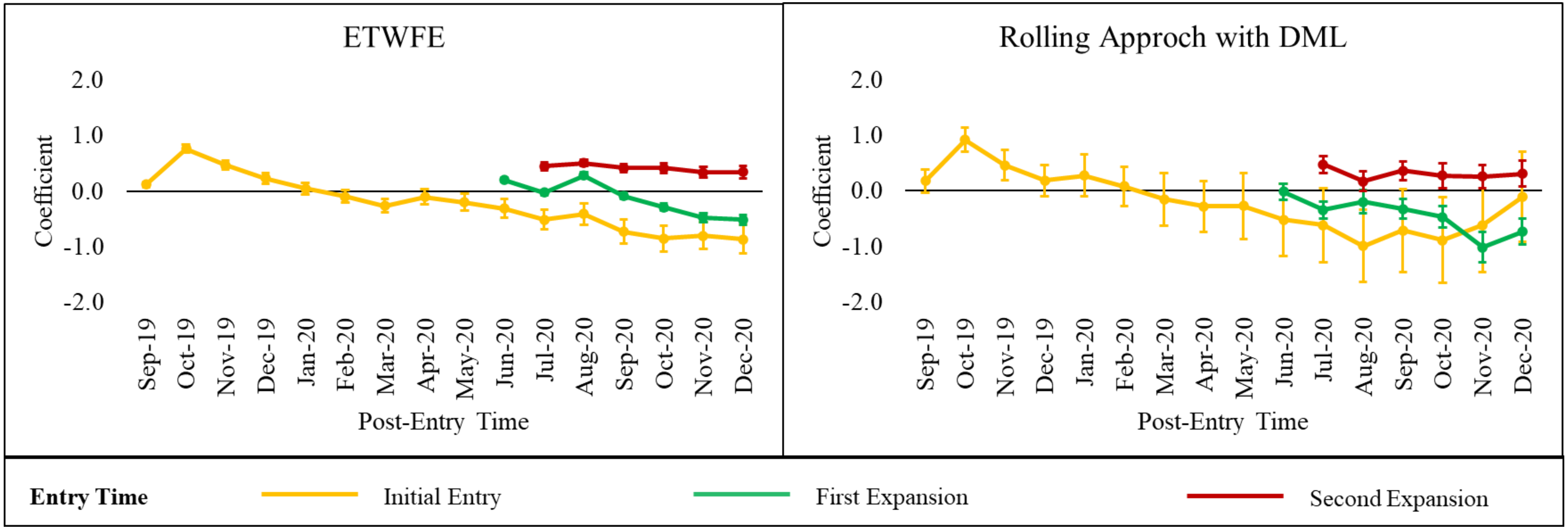
Results from the ETWFE Rolling Approach with DML: Total PBMA Sales

Heterogenous effects:

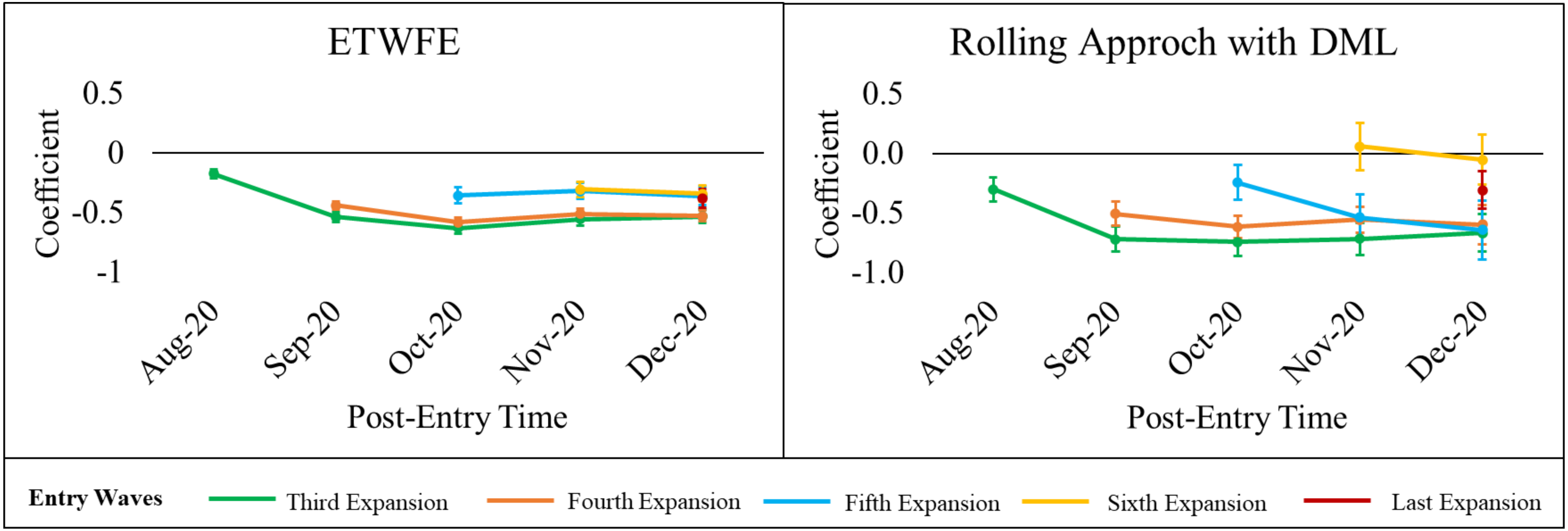
- When the new brand enters the market, the total PBMA sales …
  - decrease in both initial entry and first expansion stores (except in early post-entry times)
  - increase in second expansion stores

Dynamic effects:

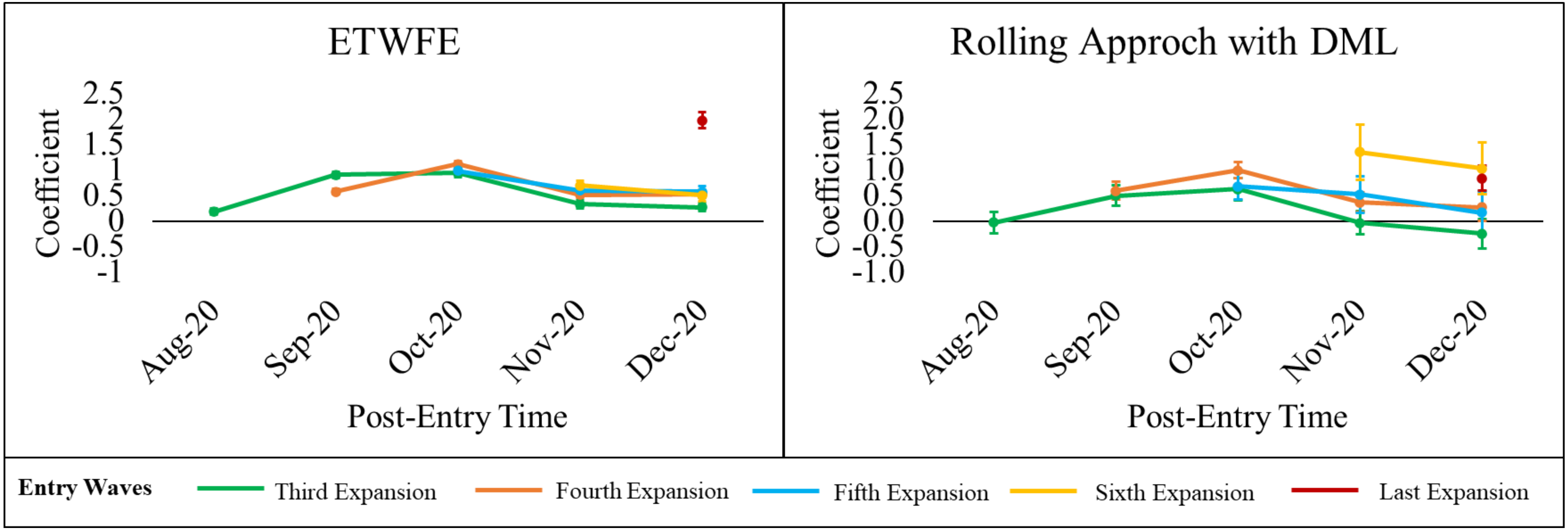
- In initial entry stores, the new brand entry increases total PBMA sales in early post-entry times but decreases it in later post-entry times.



Results 3: Heterogenous Entry Effects on Incumbent PBMA Brand Sales



Results 3: Entry Effects on Incumbent PBMA Brand Price



Results 2: Entry Effects on Total PBMA Sales