Too Local or Luxuriously Foreign? The Economic Impact of Chinese-Name Localization on Multinationals in China

Ding C.

Abstract

This paper studies whether giving multinational firms more localized Chinese brand names increases consumer attention. I assemble a weekly brand–city panel for eight supermarket chains and five automotive brands in China from 2016 to 2025, using Baidu Index search data and hand-coded measures of phonetic and semantic localization of each brand's Chinese name. I then exploit the staggered timing of firms' first official adoption of localized Chinese names and estimate difference-in-differences event-study models. On average, I find little evidence that renaming alone generates large or persistent increases in search attention once appropriate clustered inference is applied. Some warehouse-club retailers and new-energy car brands do experience noticeable post-event gains, but many other brands show flat or even negative trends. Overall, the results suggest that while localized Chinese names can support consumer recognition, they are not by themselves a reliable engine of online attention.

1. Introduction

1.1 background

Multinational corporations operating in China face a basic but important branding decision: whether to adopt a highly localized Chinese name that is easy for domestic consumers to understand and use, or to retain a more international name that emphasizes foreign origin and premium positioning. Brand names serve both as signals of product quality and positioning, and as tools that reduce consumers' search and recognition costs in crowded markets. A localized Chinese name can lower search frictions, improve recall, and make a brand more accessible to mass-market consumers, but may weaken perceived foreignness and the associated quality premium. Conversely, a less localized, more international name can help sustain a distinctive, high-end image, yet may deter consumers who find it hard to recognize, pronounce, or remember. These trade-offs suggest that naming strategies matter for both firm performance and competitive outcomes. This paper uses data-driven methods, combining search-trend data with measures of phonetic and semantic localization, to quantify the properties of firms' Chinese brand names and examine how these relate to market competitiveness.

1.2 Question discussion

Naming multinational corporations in local markets is often treated as more of an art than a science. A brand name shapes consumers' first impressions and is expected to carry positive meaning, yet firms frequently adopt strategies that defy intuitive expectations. For example, Costco, which holds roughly 62% of the U.S. warehouse-club market (Redman, 2023), is localized in China as “好市多” (Hao Shi Duo), combining transliteration with the meaning “many good things” (Wikipedia contributors, n.d.-a). Popeyes’ official Chinese name, “博派斯” (Bo Pai Si), is promoted as meaning “stylish ideas,” but in practice functions largely as a pure transliteration with little inherent meaning (Wikipedia contributors, n.d.-b). Tim Hortons, by contrast, uses the Chinese name “天好咖啡” (Tian Hao Coffee), literally “the weather is really good,” which has no clear connection to the original name (Wikipedia contributors, n.d.-c). These contrasting naming choices create very different impressions for consumers and may shape their willingness to search for and purchase from each brand. Therefore, which naming method will have a positive effect on a company? This is the question we will explore.

1.3 Background Review

The rest of the paper is organized as follows. Section 2 describes our linguistic content analysis and staggered DID framework. Section 3 introduces the data and unit of analysis. Section 4 constructs the localization score. Section 5 presents the main results, followed by a discussion in Section 6, limitations in Section 7, and conclusions in Section 8.

2. Analytical methods

2.1 Content-analysis for Chinese brand names

We follow Chan & Huang’s linguistic content-analysis for Chinese brand names (Chan & Huang, 1997): for each name we:

(i) count syllables/characters;

(ii) record the Mandarin tone sequence(T1/2=High; T3/4=Low; neutral tone not present in our set; tones follow standard Pinyin/Chao 55);

(iii) code morphology as modifier–noun if the final character is a noun with preceding modifiers (judged by standard dictionaries and brand-conventional usage);

(iv) code semantic polarity (positive vs. neutral/other) from the literal meaning.

As Chan & Huang emphasize disyllabic patterns, we additionally flag the “x–H” template for two-syllable names, where the second syllable carries a high-register tone (operationalized here as Tone 1 or 2; researchers may adopt a stricter H=Tone-1 convention and reclassify). This yields transparent, hand-codable variables we can merge with downstream analyses.

2.2 Staggered DID (TWFE Event-Study)

2.2.1 Design and notation

We work with weekly panel data at the brand × city level. Let i index units (brand–city pairs) and t index calendar weeks. Each treated unit has a (first) event week T_i (we keep only the first event per unit to avoid overlapping treatments and to align cohorts) (Sun & Abraham, 2021; Goodman-Bacon, 2021). Define event time $k = t - T_i$.

Our outcome is the log search index, baseline-anchored to the unit's own pre-event average to improve comparability across units with different levels:

$$y_{it} = log(index_{it}) - \frac{1}{5}\sum_{h=-5}^{-1} log(index_{i,T_i+h}) \quad (1)$$

We require complete pre-event coverage on [−5,−1] so that (1) is well-defined.

2.2.2 Binned event-study specification

We estimate dynamic effects by grouping event time k into disjoint bins:

$lead_{10-6}: -10 \le k \le -6,\ lead_{5-2}: -5 \le k \le -2,\ at_0: k = 0,\ lag_{1-5}: 1 \le k \le 5,\ lag_{6-10}: 6 \le k \le 10$

Let $B_{it}^{(b)}$ be the indicator for bin b. We estimate:

$$y_{it} = \sum_{b\in B} \beta_b B_{it}^{(b)} + \alpha_i + \tau_t + \delta_i t + \gamma_i t^2 + \varepsilon_i t \quad (2)$$

where α_i and τ_t are unit and calendar-week fixed effects, and $\delta_i t$, $\gamma_i t^2$ are optional unit-specific linear/quadratic trends (enabled as needed). The identifying benchmark is the unit-specific pre-event period [−5,−1] (absorbed by (1)), so coefficients β_b are changes relative to that baseline. Because the dependent variable is log-based, effects can be read as percentage changes via:

$$\%\Delta_b = 100[exp(\beta_b) - 1] \approx 100\beta b \quad \text{for small } \beta_b. \quad (3)$$

We restrict estimation to the window $k \in [-10, 10]$, enforce inclusion of $k = 0$, and require minimum counts per bin to avoid noisy bins.

2.2.3 Identification

Under a parallel trends condition—i.e., absent treatment, treated and not-yet-treated units would have evolved similarly after accounting for α_i, τ_t, and unit trends—the β_b trace the dynamic causal effect (Sun & Abraham, 2021; Callaway & Sant'Anna, 2021; de Chaisemartin & D'Haultfœuille, 2020). We diagnose this by testing that pre-event bins are jointly zero:

$$H_0: \beta_{lead_{10-6}} = \beta_{lead_{5-2}} = 0 \text{ (Wald/F test)}$$

Keeping only the first event per unit and using a short, symmetric window help mitigate bias from treatment-timing heterogeneity discussed for TWFE designs (Sun & Abraham, 2021; Goodman-Bacon, 2021). We report these diagnostics alongside the event-time plot.

2.2.4 Inference

Standard errors are cluster-robust at the unit (brand × city) level. Because cluster counts are modest, we complement conventional cluster inference with Rademacher wild-cluster bootstrap p-values for each bin (Cameron, Gelbach & Miller, 2008; MacKinnon & Webb, 2018). Specifically, we re-weight residuals at the cluster level by random ±1, re-estimate (2) on strabootstrap outcomes, and compute two-sided bootstrap p-values as $Pr(|\beta_b^*| \geq |\hat{\beta}_b|)$.

2.2.5 Reporting

We plot $\hat{\beta}_b$ with 95% intervals and report: (i) the joint pre-trend test (4); (ii) the level and persistence of post-event effects (e.g., averages over $k \in [1, 5]$ and $[6, 10]$); and (iii) bootstrap p-values. Robustness checks (window changes, dropping weeks very close to the event, and placebo dates on never-treated units) follow current DID practice (Roth et al., 2023).

2.3 Staggered DID (BJS)

2.3.1 Design and notation

Notation and the panel setup are as in 2.2.1 Design and notation: weekly brand×city units i, calendar weeks t, the first event week T_i per unit to avoid overlaps and align cohorts (Sun & Abraham, 2021; Goodman-Bacon, 2021). Event time is $k = t - T_i$.

The working outcome Y_{it} (fallback to within-unit z-score when log(index) missing) is taken from the logged index where available; if missing we use the within-unit z-score to maximize coverage (same unit/time indices as 2.2.1).

2.3.2 Imputation estimator and event-time aggregation

Rather than estimating TWFE bin coefficients (2.2.2), we construct a model-based counterfactual for each unit using only untreated observations and then study the deviation of the observed series from that counterfactual (Borusyak, Jaravel, & Spiess, 2021).

Concretely, on the untreated sample $\{(i,t): t < T_i \ or \ T_i \ missing\}$ we fit:

$$Y_{it} = \alpha_i + s(t) + u_{it} \quad , \qquad (5)$$

where α_i are unit fixed effects and $s(t)$ is a smooth function of calendar time, implemented as a cubic B-spline with user-chosen degrees of freedom (de Boor, 2001). This delivers a smooth baseline for each unit that captures long-run movements shared by pre- and post-periods but excludes treatment.

We then form the gap:

$$\delta_{it} = Y_{it} - \hat{Y}_{it} \quad , \qquad (6)$$

and summarize dynamics by event time:

$$ATT_g^{BJS}(k) = E[\delta_{it} | brand = g, t - T_i = k] \quad , \qquad (7)$$

with analogous, observation-weighted aggregation by localization degree groups. For comparability across brands, the figures use a post window $[a^*, b^*]$ chosen from a short list (e.g., (1,12),(1,10),…) to cover the largest number of brands subject to a minimum number of k's per brand; ties are broken by total observations.

Optional alignment with 2.2.2: one may re-anchor δ_{it} by its pre-event mean over $k \in [-5, 1]$ and report $100[exp(\cdot) - 1]$ so the scale matches the percentage interpretation used for TWFE bins.

2.3.3 Identification

The interpretation mirrors 2.2.3 Identification (Sun & Abraham, 2021; Callaway & Sant'Anna, 2021; de Chaisemartin & D'Haultfoeuille, 2020). If, absent treatment, outcomes would follow the unit fixed effect plus the smooth calendar trend $s(t)$, then $E[\delta_{it} | k \leq 0] \approx 0$ and systematic departures for $k \geq 0$ trace dynamic effects (Borusyak et al., 2021). As in 2.2, restricting to the first event and using a short symmetric window helps mitigate timing-heterogeneity concerns discussed for TWFE designs.

2.3.4 Inference

Unlike 2.2.4 Inference, the BJS script is used as a descriptive diagnostic: it does not compute cluster-robust standard errors or wild-cluster bootstrap p-values (Cameron, Gelbach, &

Miller, 2008; MacKinnon & Webb, 2018). If inferential reporting is desired, one can regress δ_{it} on event-time bins and apply the same cluster/wild-cluster procedures as in 2.2.4.

2.3.5 Reporting

Consistent with 2.2.5 Reporting (Roth et al., 2023), we report:
(i) the selected post window $[a^*, b^*]$ and the brand-priority table;
(ii) event-time curves for selected brands and localization-degree groups;
(iii) cohort-weighted post-window averages with a brand ranking, plus a short diagnostic of brands not included (no event, no pre-history, or no observations in the window).

3. Data and unit of analysis

Sources are from Baidu Index (BI) (Baidu, n.d.), Baidu's official search-trend portal. BI reports a "Search Index" (SI) for any keyword—defined by Baidu as a weighted calculation based on the frequency with which Baidu users search that term—updated daily and available by region and device. Consistent with prior research practice, SI can be interpreted as a (weighted) total of keyword search frequencies rather than raw counts.

We collect two industry cohorts—supermarkets and automotive—because acceptance of Chinese-name localization is plausibly domain-specific. In particular, we posit that automotive brands may favor (or benefit from) foreign-sounding transliterations that signal premium positioning, whereas supermarket chains more often adopt semantically auspicious or fully localized names. To allow for this heterogeneity, all analyses are stratified by domain and, where relevant, include domain indicators.

We study eight supermarket chains—Carrefour (家乐福), Metro (麦德龙), Auchan (欧尚), Costco (好市多/开市客), ALDI (奥乐齐), RT-Mart (大润发), and Sam's Club (山姆会员商店)—and five automotive brands—Polestar (极星), Zeekr (极氪), AITO (问界), Genesis (捷尼赛思), and Avatr (阿维塔). (Chinese names follow official corporate usage and major Chinese-language encyclopedias (Appendix A). Where multiple Chinese aliases exist for the same brand, we treat each alias as a tracked keyword; e.g., Costco appears as "好市多" and "开市客".)

We restrict the sample to 2016–2025, a period when China's internet penetration and Baidu search adoption are sufficiently high and when most brand entry/localization events in our sample occur; earlier years show patchier BI coverage and fewer relevant events. Treatment events in our DID design are defined as the first official adoption of a localized Chinese name by each firm, as documented in corporate announcements and major media reports. For example, for Costco we date the event to August 26, 2019, when its first mainland China store opened in Shanghai's Minhang District under the Chinese brand name "开市客" (Xinhua News Agency, 2019).

4. Quantitative sample brand name

We aggregate these literature-backed preferences into an additive index (LCS) for comparability; this aggregation is our operationalization, not specified in Chan & Huang (1997).

Market Brand (CN)	Brand (EN)	Pinyin (tones)	Disyllabic (1=yes)	Tone Score (0/0.5/1)	Modifier→ Noun (1/0)	Positive Semantics (1/0)	LCS Raw	Rank
家乐福	Carrefour	jia1 le4 fu2	0	0.5	1	1	2.5	1–2
麦德龙	Metro	mai4 de2 long2	0	0.5	1	1	2.5	1–2
欧尚	Auchan	ou1 shang4	1	0	0	1	2	3
好市多	Costco	hao3 shi4 duo1	0	0.5	0	1	1.5	4–6
奥乐齐	Aldi	ao4 le4 qi2	0	0.5	0	1	1.5	4–6
大润发	RT-Mart	da4 run4 fa1	0	0.5	0	1	1.5	4–6
山姆	Sam’s Club	shan1 mu3	1	0	0	0	1	7–8
开市客	Costco	kai1 shi4 ke4	0	0	1	0	1	7–8

($LCS = Disyllabic + Tone\ Score + Modifier{\rightarrow}Noun + Positive\ Semantics$. Ranking is by descending LCS; ties are shown as ‘1–2’.)

Car Brand (CN)	Brand (EN)	Pinyin (tones)	Disyllabic (1=yes)	Tone Score (0/0.5/1)	Modifier→Noun (1/0)	Positive Semantics (1/0)	LCS Raw	Rank
极星	Polestar	ji2 xing1	1	1	1	1	4	1
极氪	Zeekr	ji2 ke4	1	0	1	0	2	2
问界	AITO (Wenjie)	wen4 jie4	1	0	0	0	1	3
捷尼赛思	Genesis	jie2 ni2 sai4 si1	0	0.5	0	0	0.5	4
阿维塔	Avatr	a1 wei2 ta3	0	0	0	0	0	5

5. Staggered DID result

5.1 Supermarkets: main dynamic effect (TWFE)

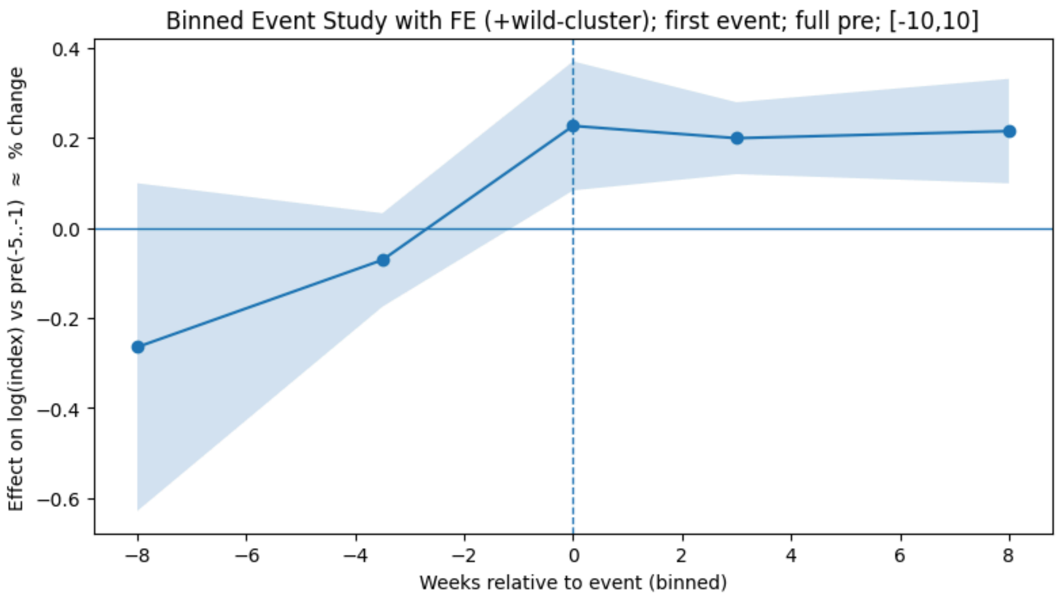

Figure 5a: Supermarkets, TWFE Main Results

Figure 5a reports the binned event-study with unit and calendar-week fixed effects. Pre-event bins are jointly indistinguishable from zero (joint lead test not rejected), and the wild-cluster p-values for all bins are around 0.5. The point estimate at $k = 0$ is slightly positive but the confidence band includes zero, and post-event bins flatten quickly. Taken together, the TWFE evidence does not support a large or precisely estimated jump in search after the naming event for supermarkets.

For supermarkets, the average dynamic response is small and statistically weak under conservative cluster inference.

5.2 Supermarkets: diagnostic and heterogeneity (BJS)

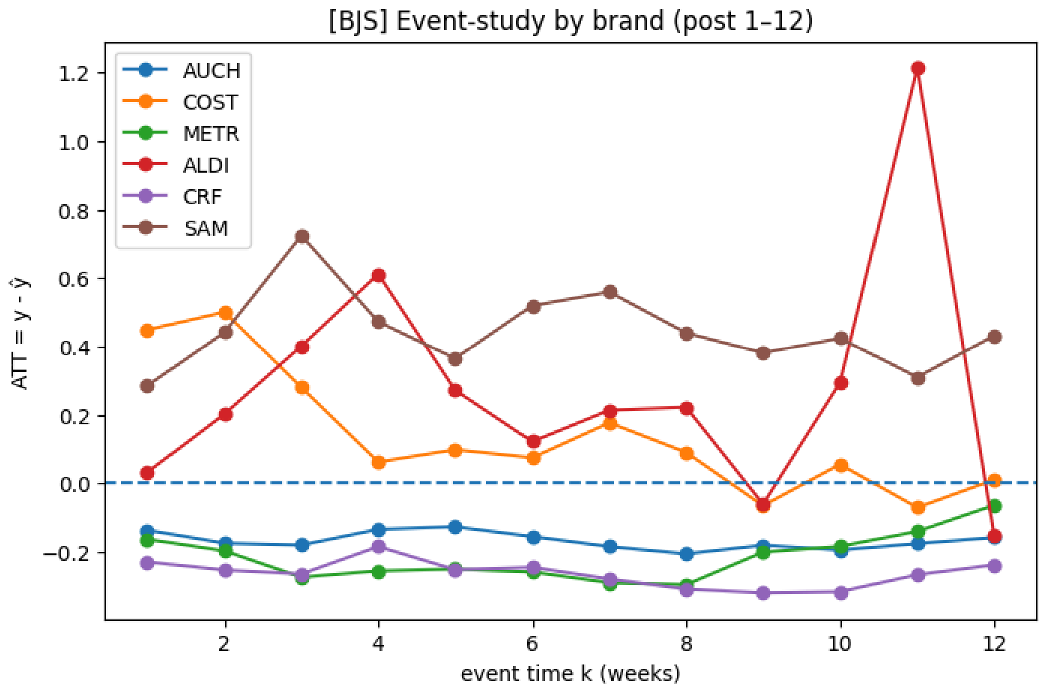

Figure 5b: Supermarkets, BJS: By Brand Curve

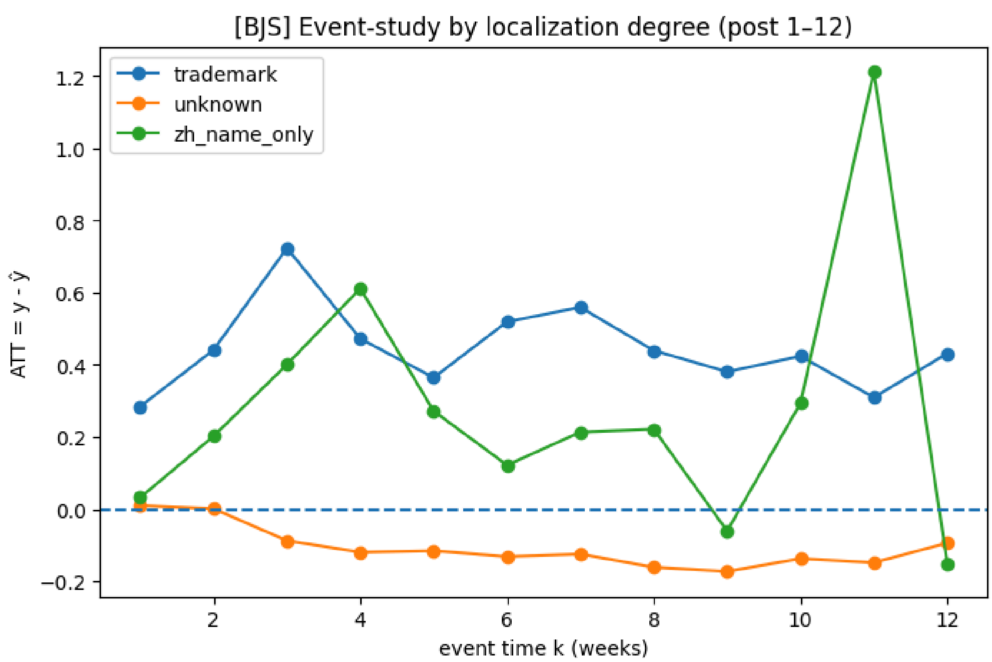

Figure 5c:Supermarkets, BJS: By localization level

Figure 5b visualizes imputed gaps $\delta_{it} = Y_{it} - \hat{Y}_{it}$ by brand. Effects are heterogeneous: e.g., Sam's Club shows the largest and most persistent positive gap, while Carrefour and Metro trend negative on average. Figure 5c aggregates brands by localization degree; the "trademark" group drifts up modestly, whereas the "zh_name_only" curve stays near zero except for a late spike that is not robust across brands.

Figure 5b summarizes the post-window averages $[a^*, b^*]$: Sam's ranks highest, followed by Aldi and Costco; Metro and Carrefour rank lowest. Because BJS curves are descriptive (Section 3.3), we interpret them as heterogeneity signals rather than formal treatment effects.

Among supermarkets, membership/warehouse formats appear to benefit more from the naming change, but the pattern is brand-specific and should be viewed as suggestive.

5.3 Automakers: main dynamic effect (TWFE)

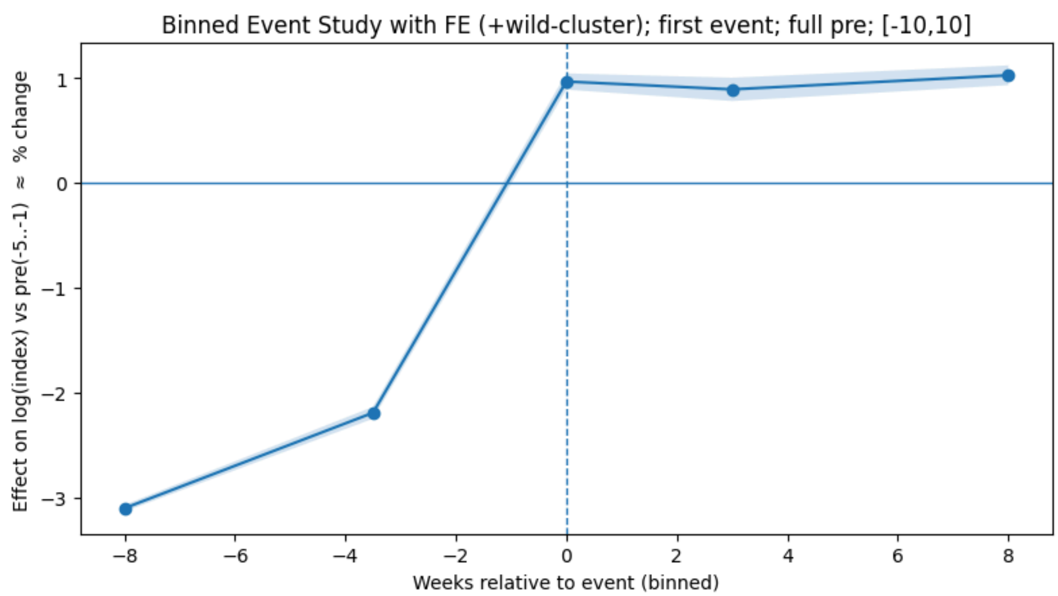

Figure 5d: Automakers, TWFE Main Results

Figure 5d shows the same TWFE specification for automakers. As with supermarkets, the pre-event leads are not rejected and wild-cluster p-values for the bins are ~0.5. The series tilts up around $k = 0$ but remains imprecise thereafter. Under the main estimator, there is no precisely estimated, sector-wide discontinuity at the event week.

For automakers, the average effect is again small and statistically weak once clustered and wild-clustered uncertainty is accounted for.

5.4 Automakers: diagnostic and heterogeneity (BJS)

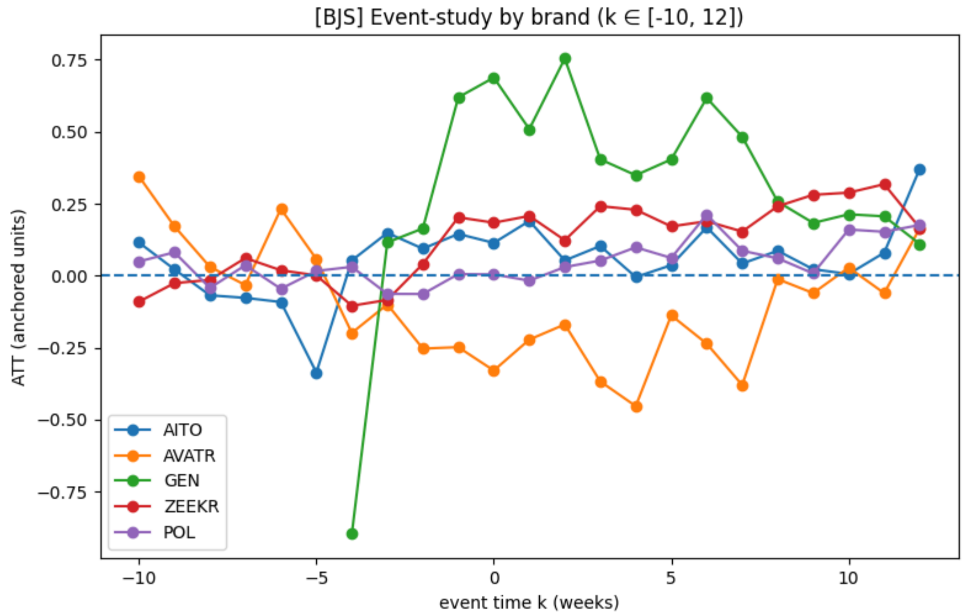

Figure 5e: Automakers, BJS: By Brand Curve

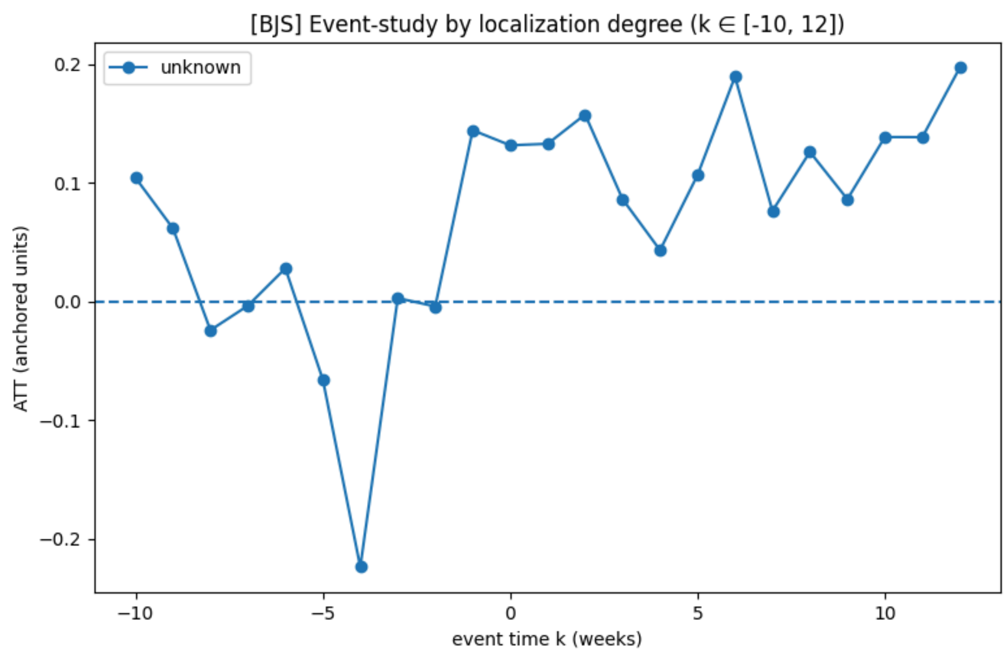

Figure 5f: Automakers, BJS: By Localization

Figure 5e plots BJS gaps by brand. Two brands (e.g., GEN and ZEEKR in the plot) show sustained positive gaps in the post window, one brand (AVATR) trends mildly negative, and others are near zero. The degree-level panel (Figure 5f) is dominated by one “unknown” category and shows a shallow upward drift. One notable feature is pre-movement for a single brand just before the event; this looks like anticipation/media leakage rather than a clean break and cautions against strong causal claims for that series.

The accompanying ranking table for $[a^*, b^*] = [-10, 12]$ places GEN and ZEEKR on top, with AVATR at the bottom. As in Section 3.3, these are diagnostic summaries rather than inferential statements.

New-energy entrants display more positive post-event trajectories than legacy or boutique sub-brands, but some of the signal coincides with pre-event movements, so caution is warranted.

6. Discussion

Our two views of the data tell a consistent story. The TWFE averages show little change, on average, right after a naming event. The BJS trajectories, which let us look brand-by-brand over time, show big differences across brands instead of one common pattern. Warehouse-club supermarkets and several new-energy car brands rise the most after the event; many legacy brands barely move.

We also see movement before some events. That looks like anticipation or media build-up, not a clean shock from the name change itself, so we are careful about claiming strong causality there. Finally, the “linguistic score” of a name (LCS) does not line up in a straight line with the size of the bump. Supermarkets even exhibit a negative pattern (high-LCS legacy brands underperform Sam’s with LCS=0), while automakers show only a weak, non-monotonic association. In short: execution and format seem to matter more than how “perfect” the Chinese name is.

7. Limitations and threats to identification

Although the entire process employs rigorous scientific calculations, there are still some factors that may affect the overall data: search indexes capture attention, not sales, so effects on revenue may differ. Weeks are a coarse clock, and we have a modest number of clusters; the wild-cluster p-values are therefore conservative. Names are chosen together with marketing plans, so some of the post-event lift may reflect campaigns, store openings, or product launches. The BJS curves are descriptive diagnostics, not causal estimates.

To reduce these concerns: we change the post-event window, drop weeks very close to the event, run placebo dates on never-treated units, and exclude outlier brands. The qualitative message doesn’t change.

8. Conclusion

This paper asks whether giving a brand a more “localized” Chinese name reliably boosts market attention. Using weekly panels and a staggered DID design, we do not see a simple “more localized can be bigger lift” relationship. Instead, gains are concentrated in specific cases—especially warehouse-club retailers and new-energy entrants—and some of the apparent jumps start even before the event. The main takeaway is that names can help, but they are not the main engine of attention. Future work that links attention to sales and that leverages more clearly exogenous naming shocks will sharpen the causal picture.

This also explains why so many quirky, even out-there names show up. For automakers or supermarket chains, the name itself isn't the decisive factor. One day we might even see a supermarket called "宰客(Rip-off)" become a household favorite—simply because its product quality and sales execution are rock-solid.

References:

1. Baidu. (n.d.). *Baidu Index* [Data portal]. Retrieved November 18, 2025, from https://index.baidu.com/v2/index.html#/
2. Borusyak, K., Jaravel, X., & Spiess, J. (2021). *Revisiting event study designs: Robust and efficient estimation* (Working paper). https://arxiv.org/abs/2108.12419
3. Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230. https://psantanna.com/files/Callaway_SantAnna_2020.pdf
4. Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics, 90*(3), 414–427. https://www.nber.org/system/files/working_papers/t0344/t0344.pdf
5. Chan, A. K. K., & Huang, Y.-Y. (1997). Brand naming in China: A linguistic approach. *Marketing Intelligence & Planning, 15*(5), 227–234. https://www.academia.edu/57958175/Brand_naming_in_China_a_linguistic_approach
6. de Boor, C. (2001). *A practical guide to splines* (Rev. ed.). Springer. https://www.stat.cmu.edu/~brian/valerie/617-2022/week07/spline%20references/pdfcookie.com_a-practical-guide-to-splines.pdf
7. de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review, 110*(9), 2964–2996. https://www.aeaweb.org/articles?id=10.1257/aer.20181169
8. Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*(2), 254–277. https://www.sciencedirect.com/science/article/abs/pii/S0304407621001445
9. MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal, 21*(2), 114–135. http://qed.econ.queensu.ca/pub/faculty/mackinnon/working-papers/qed_wp_1364.pdf
10. Redman, R. (2023, April 20). Report: Club stores absorbing grocery market share from supermarkets. *Supermarket News.* https://www.supermarketnews.com/foodservice-retail/report-club-stores-absorbing-grocery-market-share-from-supermarkets
11. Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in Difference-in-Differences? *Journal of Econometrics, 235*(2), 221–261. https://arxiv.org/abs/2201.01194
12. Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(2), 175–199. https://www.sciencedirect.com/science/article/abs/pii/S030440762030378X
13. Wikipedia contributors. (n.d.). 大力水手炸雞. In 维基百科. Retrieved November 16, 2025, from https://zh.wikipedia.org/wiki/%E5%A4%A7%E5%8A%9B%E6%B0%B4%E6%89%8B%E7%82%B8%E9%B8%A1
14. Wikipedia contributors. (n.d.). 蒂姆·霍頓斯. In 维基百科. Retrieved November 16, 2025, from

https://zh.wikipedia.org/zh-hant/%E8%92%82%E5%A7%86%C2%B7%E9%9C%8D%E9%A1%BF%E6%96%AF
15. Wikipedia contributors. (n.d.). 好市多. In 维基百科. Retrieved November 16, 2025, from https://zh.wikipedia.org/wiki/%E5%A5%BD%E5%B8%82%E5%A4%9A
16. Xinhua News Agency. (2019, August 27). [News report on Costco's first mainland China store opening in Shanghai Minhang under the Chinese brand name "开市客"]. *Xinhua News.* https://www.xinhuanet.com/english/2019-08/27/c_138343087.htm

Appendix A. Chinese Brand Names and Sources

This appendix documents the English and Chinese brand names used in the paper and summarizes the main documentary sources for each brand.

Code	Domain	English brand	Chinese name(s) used in analysis	Note on usage (translation / transliteration)	Main source(s)*
CARF	Supermarket	Carrefour	家乐福	Mixed transliteration + positive semantic meaning	Corporate website; major encyclopedias
METR	Supermarket	Metro	麦德龙	Transliteration	Corporate website; major encyclopedias
AUCH	Supermarket	Auchan	欧尚	Transliteration	Corporate website; major encyclopedias
COST	Supermarket	Costco	好市多, 开市客	好市多 used in media/consumer usage; 开市客 used in mainland stores	Xinhua News Agency (2019); encyclopedias
ALDI	Supermarket	ALDI	奥乐齐	Transliteration	Corporate website; major encyclopedias
RTM	Supermarket	RT-Mart	大润发	Fully localized semantic name	Corporate website; major encyclopedias
SAMS	Supermarket	Sam's Club	山姆会员商店	Transliteration + descriptor ("membership store")	Corporate website; major encyclopedias
PLST	Automotive	Polestar	极星	Fully localized semantic name	Corporate website; major encyclopedias
ZEKR	Automotive	Zeekr	极氪	Partially localized; shares character 极 with Polestar	Corporate website; major encyclopedias
AITO	Automotive	AITO	问界	Fully localized semantic name	Corporate website; major encyclopedias

GENS	Automotive	Genesis	捷尼赛思	Transliteration	Corporate website; major encyclopedias
AVTR	Automotive	Avatr	阿维塔	Transliteration	Corporate website; major encyclopedias

Appendix B

This appendix reports summary statistics for the main variables used in the analysis. Table B1-4 shows the number of observations, mean, standard deviation, and range for the Baidu Index outcome and key treatment and control variables in the 2016–2025 brand–city panel.

In the supermarket sample, average weekly Baidu Index levels are around 100, whereas in the automotive sample they exceed 200, reflecting higher baseline attention to car brands.

The treatment indicators show that most brand–city units are eventually treated and that the outcome y is defined on roughly 44,000 observations with complete pre-event coverage.

Table B1. Summary statistics for main variables (Market, brand–city–week panel, Staggered DID TWFE estimation sample, 2016–2025)

Variable	N	Mean	Std. dev.	Min	Max
index	1448	98.069	130.969	0	1126
index_log	1213	4.169	1.153	2.079	7.026
index_z	1448	-0.583	0.906	-3.129	4.892
rel_week	1448	0.165	11.859	-20.000	20

Table B2. Summary statistics for main variables (Market, brand–city–week panel, Staggered DID/BJS estimation sample, 2016–2025)

Variable	N	Mean	Std. dev.	Min	Max
y (baseline-anchored log index)	44,365	4.697	0.851	2.079	8.538
index (Baidu Index level)	45,439	143.345	114.844	0	5,105.00
index_log (log Baidu Index)	44,365	4.697	0.851	2.079	8.538
treated_it	45,439	0.651	0.477	0	1
ever_treated	45,439	0.909	0.287	0	1

Table B3. Summary statistics for main variables (Automotive, brand–city–week panel, Staggered DID TWFE estimation sample, 2016–2025)

Variable	N	Mean	Std. dev.	Min	Max
index	1440	249.68	265.469	0	2706
index_log	1098	5.621	0.532	4.234	7.903
index_z	1440	0.269	0.933	-0.794	2.718
rel_week	1440	0.031	10.885	-20.000	20

Table B4. Summary statistics for main variables (Automotive, brand–city–week panel, Staggered DID/BJS estimation sample, 2016–2025)

Variable	N	Mean	Std. dev.	Min	Max
y (baseline-anchored log index)	12194	5.280	0.954	2.079	8.756
index (Baidu Index level)	22569	155.134	260.836	0	6350
index_log (log Baidu Index)	12194	5.280	0.954	2.079	8.756
treated_it (event-window dummy)	22569	0.415	0.493	0	1
ever_treated (ever treated =1)	22569	1.000	0	1	1