DOI: 10.6277/TER.202404_52(2).0001

The Effect of User-Generated Content on Sales: Evidence from the Toyota RAV4 Roof Leak Incident

Fang-Chang Kuo and Han-Sheng Chen*

This paper studies the effects of user-generated content on relative sales by exploiting the Toyota RAV4 roof leak incident as a natural experiment. Using a reduced-form difference-in-differences approach, our empirical results show that user-generated content from the roof leak incident caused sales of the Toyota RAV4 to fall by more than 30%. However, the estimates may capture not only the treatment effects but also other equilibrium effects on sales. Therefore, we estimate a structural nested logit demand model, and conduct counterfactual simulations showing that lost sales were approximately 9,000, equivalent to 25% of the total sales in the previous year. Cannibalization effects from Toyota Corolla Cross only led to an around 4% decrease in RAV4's sales.

Keywords: user-generated content, difference-in-differences, nested logit demand
JEL classification: L15, L62, M31

*The authors are Assistant Professor of Department of Economics at National Chung Cheng University and Master's student of Department of Economics at National Chung Cheng University. Correspondence: Fang-Chang Kuo, email: fckuo@ccu.edu.tw. We would like to thank the Editor, two anonymous referees, and participants at National Sun Yat-sen University and 2022 Taiwan Economic Association Annual Conference for their helpful comments and suggestions. We are also grateful for the support provided by the Ministry of Science and Technology of Taiwan (MOST110-2410-H-194-095-MY2). All remaining errors are our own.

經濟論文叢刊 (*Taiwan Economic Review*), 52:2 (2024), 121–153。
國立台灣大學經濟學系出版

臺大學術期刊資料庫

1 Introduction

In the past few decades, advances in information technology and prevalence of social media platforms have been facilitating the development in user-generated content (UGC), or online word-of-mouth (WOM). Facebook, Instagram, TikTok, TripAdvisor, Twitter, Yelp, and Youtube enable users to interact with one another by sharing and consuming posts, comments, links, photos, or videos. With the proliferation of UGC and online WOM, it is natural to ask whether these unvetted and seemingly unreliable information flow have substantive impacts on product sales.[1]

Answering the question faces several empirical challenges. First, popular products often draw more attentions and discussions from consumers. Therefore, user-generated content and online word-of-mouth reflect demand, and at the same time may also affect demand.[2] Second, firms could also advertise through promotional content on social medias. It is unclear whether sales increase because of UGC or because of advertising. Third, various social media platforms make it difficult to measure UGC across platforms, and sometimes the outcome variables are not available.

This paper investigates the effects of user-generated content on relative sales in the Taiwanese automobile market. Specifically, we use the Toyota RAV4 roof leak incident, which generated a sudden burst of online discussions about leak issues in October 2020, as a natural experiment. Consumers and content providers facilitated the online conversations through posts, comments, photos, and videos on various social media platforms, including mobile01, PTT, Facebook, and YouTube. Our analysis is of interest for several reasons. First, while automotive sales are mostly completed at brick-and-mortar stores where consumers can observe relevant characteristics by taking test-drives, it is unclear if results from previous literature focusing on experience goods, such as restaurants, movies, TV shows, or hotels, still hold in the automotive industry. Second, as firms increasingly engage in promotional content across social media platforms, mishandling of product information or marketing strategy could be amplified by UGC or online WOM. Third, our empirical analysis attempts to quantify the sales losses re-

[1] According to a survey in 2014, over 80% of consumers in Taiwan would consult online word-of-mouth before actual purchases. Source: https://mic.iii.org.tw/news.aspx?id=366

[2] For more details about identification issues in the reflection problem, please see Manski (1993).

臺大學術期刊資料庫

sulting from the incident. The estimates could be instrumental in assessing the potential impacts of public relation crises.

This article utilizes a unique product-level dataset on car sales to quantify the effects on relative sales. Initially, we adopt a difference-in-differences (DID) empirical strategy. The estimates suggest that Toyota RAV4's sales declined by more than 30% after the incident. However, the DID estimates may capture equilibrium effects from the market, rather than only direct effects of UGC on RAV4's sales, even with similar pre-trends. The DID estimates could include other changes in market structures if they coincide with the incident. Moreover, comparing the treated and untreated units becomes challenging when control group units are affected by the entries or exits of products. To disentangle the impacts of the roof leak incident from the equilibrium effects, we turn to a structural approach using logit demand models.

Following well-developed discrete choice framework for analyzing consumer demand in automotive markets, we use logit and nested-logit models to account for consumer's substitution patterns and introduction of new products (Berry, 1994). Aggregated market shares and product characteristics data are used to estimate underlying parameters in consumer's utility function. To address conventional price endogeneity, we construct two sets of instrumental variables and apply them to two-stage least squares (2SLS) estimation. With reasonable estimates on product characteristics and own-price elasticities, we then perform three counterfactual exercises to disentangle the sale losses due to RAV4 leak incident and those because of Toyota Corolla Cross. The first exercise investigates cannibalization effects from the 2021 best-selling car model, Toyota Corolla Cross, since its introduction coincided exactly with the roof leak incident in October 2020. Without Toyota Corolla Cross, RAV4 could have sold 1,450 more in a year, which is around 4% of the total sales in the previous year. We further get rid of the negative impacts from RAV4 roof leak incident as if the incident had never happened, letting consumers redo their purchase decisions. Counterfactual sales show that the lost sales are smaller than those implied by DID estimates. Had the incident not occurred, sales of RAV4 would have increased by around 9,000 in the following year. The third simulation focuses on the scenario in which both Toyota Corolla Cross and RAV4 roof leak incident do not exist. The simulated RAV4 sales essentially capture the losses from cannibalization effects and direct effects from UGSs.

The contribution of this paper is threefold. First, it adds to the growing

臺大學術期刊資料庫

literature on impacts of user-generated content and online word-of-mouth (Anderson and Magruder, 2012; Chevalier and Mayzlin, 2006; Chintagunta, Gopinath, and Venkataraman, 2010; Luca, 2016; Seiler, Yao, and Wang, 2017). Our paper differentiates itself from the literature by focusing on a durable good, whereas previous studies look at experience goods, such as books, movies, restaurants, and TV shows, and emphasize the informational roles of user-generated content. Despite the fact that consumers could test drive a car before buying it, our findings show that additional information revealing unobserved features or defects creates significant impacts on purchase decisions. While prior studies use data such ratings or reviews from particular review platforms, we incorporate UGC from various platforms, including posts, comments, photos, and videos.[3]

Second, in terms of the substantive questions addressed, this paper is related to a large literature using product recalls or industry scandals as natural experiments to quantify related effects on reputation (Ater and Yoseph, 2022; Bachmann et al., 2022; Bai, Gazze, and Wang, 2022; Freedman, Kearney, and Lederman, 2012; Hammond, 2013). The main distinction between this paper and prior studies is that our natural experiment has not been verified by the firm or any other government agencies. There was no product recall even though consumers and content providers were showing photos and videos in substantiation of the leak issues. The empirical results indicate that consumers could change their purchase decisions based on unverified but burgeoning online word-of-mouth. Motivating by Bachmann et al. (2022), we follow similar empirical strategy which combines difference-in-differences approach with discrete choice demand estimation to disentangle main effects of interests from equilibrium effects.[4]

Finally, to the best of our knowledge, this study is the first to estimate a structural demand system in the Taiwanese automotive industry and provide plausible estimates of own-price elasticities. The automotive market in

[3]There is a small but fast-growing literature focusing on how specific user-generated content like management responses (Proserpio and Zervas, 2017) and Q&As (Banerjee, Dellarocas, and Zervas, 2021) affect consumer reviews.

[4]Unlike Bachmann et al. (2022), we do not try to estimate the spillover effects, in which RAV4 leak incident may adversely impact other Toyota car models. This is because Toyota in Taiwan has the largest market shares of over 20% and Hammond (2013) finds that strong individual brand reputation could shield the adverse impacts from negative information. In addition, since we already consider removing Toyota CC as a counterfactual simulation, the spillover effects may not be well-identified in this case.

臺大學術期刊資料庫

Taiwan has been receiving little attention probably due to data limitations.[5] Our estimates of own-price elasticities may be of some value in terms of designing pricing strategies for firms, or evaluating tax policies for policy makers.

The remainder of the paper is structured as follows. In the next section, we introduce the Toyota RAV4 roof leak incident. Section 3 presents reduced-form analysis using DID empirical strategy. In Section 4, we estimate consumer demand in discrete choice framework, and simulate counterfactual sales to quantify direct effects of UGC and cannibalization effects. Finally, we provide some concluding remarks.

2 Toyota RAV4 roof leak incident

Toyota RAV4 has been one of the top three besting-selling car models for almost a decade in Taiwan. With the introduction of the fifth generation RAV4 in 2019, total sales rose to the next level.[6] In 2020, it became the best-selling car model in Taiwan, with total sales of 31,870 accounting for 7% of the total market shares. Furthermore, the Toyota RAV4 is the first SUV and imported car model in the history of Taiwan's car market to become the top-selling car model, breaking the 19-year consecutive record of the domestically-made sedan, Toyota Corolla Altis.[7]

Because of RAV4's massive popularity, no wonder the roof leak incident has been receiving a great deal of attentions since mid October in 2020. Initially, some RAV4 owners shared their own experiences and photos of the leak issue on Mobile01. These posts soon reverberated across multiple social media platforms, including Mobile01, Facebook, PTT, and Dcard, as more owners inspected their own vehicles and many found similar issues. Hundreds of owners have organized self-help groups via various platforms and sought help from legislators.[8] Incentivized by this heated online discussion, it also drawn interests from various Youtubers. Videos about the

[5] One exception is Kuo and Shu (2023), in which the authors estimate the impact of VW Dieselgate on business buyers and self-use buyers in Taiwan and find heterogeneous responses to environmental fraud.

[6] Figure A1 plots the annual sales and market shares for Toyota RAV4 from 2014 to 2021.

[7] Since 2001, the best-selling car model in Taiwan has been the Toyota Corolla Altis. See: https://news.u-car.com.tw/news/article/64648

[8] During an interpellation in the Legislative Yuan, Democratic Progressive Party Legislator Kao, Chia-Yu brought up this issue for discussion with the Ministry of Trans-

臺大學術期刊資料庫

incidents like interviews with car owners, personal opinions, or discussions with mechanics, were made to get more views on their channels.[9]

To trace the dynamics of consumer attentions to the leak issue over time, we use aggregate search data from Google Trends and focus on the search term of "RAV4 leak" (漏水) in Taiwan. Figure 1 shows weekly search intensities on Google search and YouTube search respectively. In addition to keyword of "RAV4 leak", we also include "RAV4" search trend. Instead of the total number of searches, Google Trends only provides the relative volumes, ranging from 0 to 100, of the keywords with the highest search volume normalized to 100. There are few interesting patterns in the figure. First, search intensities of "RAV4 leak" before October 2020 are essentially non-existent except few tiny bumps on Google search. This may be due to below-average precipitation in 2020, which later contributed to serious drought in 2021.[10] RAV4's owners were less likely to detect roof leak issues when the weather conditions were mostly sunny.

Second, in October 2020, the search intensities of "RAV4 leak" and "RAV4" both peaked at 100, indicating surge of public interests at that time. Youtube search intensities, on the other hand, reached the highest in December. We believe that the highest point of Youtube search and the second peak of Google search were related to a video made by a used-car dealer, which had received more than half million views in a month.[11]

Finally, the whole incident seemed to come to the end as Toyota announced a 10-year special warranty in March 2021.[12] The search intensities also settled down. However, the low search intensities of "RAV4 leak" do not necessarily mean that the adverse impacts faded away since interested buyers might already factor in the potential concerns. Moreover, the search volume for the generic term, "RAV4", also dropped relative to that of 2020, potentially reflecting consumers' diminishing interests.

---

portation and Communications. See: https://www.facebook.com/ntufishfans/posts/10158861096412720/, https://www.youtube.com/watch?v=hplleqfiNM4&ab_channel=%E9%AB%98%E5%98%89%E7%91%9C

[9] See: https://www.youtube.com/results?search_query=rav4+%E6%BC%8F%E6%B0%B4

[10] https://www.cwb.gov.tw/Data/climate/Watch/twn/twn-monitor_2020-0.pdf

[11] See: https://www.youtube.com/watch?v=9jWDTtesfT8&t=36s&ab_channel=%E6%9D%B0%E9%81%8B%E6%B1%BD%E8%BB%8AAUTOSTAR Toyota also issued a official response to the video, see: https://autos.chinatimes.com/20201227001143-264301

[12] https://www.toyota.com.tw/owner_news_detail.aspx?id=587

臺大學術期刊資料庫

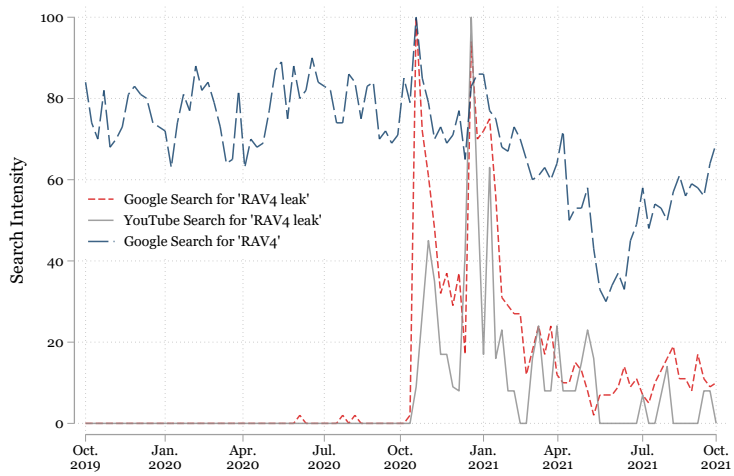


Figure 1: Search intensities of Google search and YouTube search

Notes: This figure plots weekly search intensities of "RAV4 leak" on Google search and YouTube search, and "RAV4" on Google search from October 2019 to October 2021. The search intensity ranges from 0 to 100. The intensities do not reflect actual numbers of searches as Google normalizes the highest search number of a keyword to 100 during the sample periods. Source: https://trends.google.com.tw/trends.

3 Reduced-form analysis: difference-in-differences

In order to quantify the effect of user-generated content, we use new vehicle registration data from Motor Vehicles Office.[13] The dataset contains monthly registration numbers, which would be a fairly accurate measurement for actual sales in that registration is mandatory.[14] As most studies on automotive industry, we limit our focus to passenger vehicles only since the segment accounts for the majority of household demand.[15] Each sales number is associated with specific vehicle model name, which is a combination of

[13]Various auto medias, like UCAR, 7Car, and Mobile01, regularly publish monthly new vehicle sales numbers to inform consumers of recent sales trends.

[14]Both Vehicle License Plate Tax and Fuel Tax are levied based on the registration records.

[15]The definition of a passenger car is a vehicle with no more than nine seats and vehicle weight less than 3,500 kilograms.

臺大學術期刊資料庫

car model and trim. The distinctions of trim levels are usually determined by engine displacements, or exterior designs. For example, Toyota RAV4 2.0 and Toyota RAV4 2.5 are two different models due to engine size differences. Mazda 3, a bestselling car model for Mazda, comes with two variants, 4-doors and 5-doors, in that Mazda 3 4D is a sedan and Mazda 3 5D is a hatchback.[16]

Following Kuo and Shu (2023), the monthly sales data is aggregated into quarterly levels since supplies of imported cars are subject to sea freight market conditions, which are extremely volatile during the COVID-19 pandemic in 2020 and 2021. As the incident came to public attention in October 2020, Q4 2020 is defined as the time of the incident. The sample periods span from from October 2019 to September 2021 with four quarters before and after Q4 2020.[17] The final sample consists of 1,456 model-trim-quarter combinations from 24 auto brands and 157 car models.[18]

In reduced-form analysis, we use a difference-in-differences (DID) event-study model to quantify the impact of user-generated content on auto sales. Conventional DID specification typically focuses on one single coefficient associated with interaction term between post and treated indicators to summarize treatment effects across all post-event periods.[19] Event-study specification, on the other hand, captures the dynamics of the treated before and after treatment. Specifically, the econometric model takes the following form:

$$\ln(\text{sales})_{jt} = \sum_{\substack{\tau=-4 \\ \tau\neq -1}}^{3} \beta_\tau \cdot \text{RAV4}_m \cdot \mathbf{1}\left(t - t^* = \tau\right) + \gamma_m + \delta_t + \epsilon_{jt}, \quad (1)$$

where $\ln(\text{sales})_{jt}$ is the natural logarithm of quarterly sales of trim-level

[16]Note that the trim levels in the data do not necessarily coincide with what consumers see while purchasing. For instance, Mazda 3 5D has four different trims at different prices (https://www.mazda.com.tw/globalassets/cars/mazda3/pdf/MAZDA-M3-Spec_0415.pdf). Our data uses coarser trim levels for each car model.

[17]We limit our sample periods to only two years since longer horizon may be associated changes in market structures, consumer tastes, and new environmental and fuel economy regulations. The impact of the leak incident is expected to be short-term.

[18]Table 3 shows the summary statistics of quarterly sales, and other product characteristics.

[19]In the estimation results, we also estimate the following model, $\ln(\text{sales})_{jt} = \beta \cdot \text{RAV4}_m \times \text{POST} + \gamma_m + \delta_t + \epsilon_{jt}$, and report estimates of $\hat{\beta}$.

臺大學術期刊資料庫

product j at quarter t.[20] RAV4_m is a dummy variable, which is equal to one if the car model is Toyota RAV4, and zero for other car models. $\mathbf{1}(t - t^* = \tau)$ is also a dummy variable indicating specific quarter relative to t^*, which is the fourth quarter of 2020.[21] Therefore, τ, ranging from -4, -3, -2, 0, 1, 2, and 3 indicates the number of quarters relative to Q4 2020. The omitted category is Q3 2020 with $\tau = -1$.

We are mainly interested in coefficients, β_τ, which measure the differentials, relative to the omitted quarter, in sales between treated and untreated car models at each quarter. For $\tau \geq 0$, β_τ assess the post-treatment impacts of roof leak incident on sales of Toyota RAV4 relative to other car models. For $\tau < -1$, the estimates of β_τ are expected to be close to zero and statistically insignificant if the sales for Toyota RAV4 and other car models were trending in similar patterns prior to the treatment. The examination of pre-trends is often used to assess the validity of the parallel trend assumption, which is required for identifying casual treatment effects in DID settings. However, as pointed out in Kahn-Lang and Lang (2020), pre-trend testing does not warrant the parallel trends in the post-periods.

Finally, γ_m and δ_t are model fixed effects and year-quarter fixed effects. The model fixed effects, γ_m, essentially absorb any time-invariant unobserved heterogeneity for car model m. Year-quarter fixed effects, δ_t, control prevailing market conditions that are common to all car models at time t. Robust standard errors are clustered at model level to allow for arbitrary correlations in ϵ_{jt} within the same car model.

Given Toyota RAV4 as the treated car model, we use four samples with different control groups in the estimation. The first one is a "naïve" sample which includes all available car models. Second sample uses only SUVs since RAV4 falls into the segment. Luxury cars with MSRP above 2.5 million NT$ are included as control group in the third sample. The last sample focuses on two different car segments, sedans and hatchbacks. The purpose of using various control groups is to assess the consistency of estimates as well as pre-trends similarities. Average sales and associated standard deviations

[20] For example, Toyota RAV4 2.0 and Toyota RAV4 2.5 are two different trim products, which will be indexed by two different j, since they have different engine sizes. However, they are the same car model, Toyota RAV4, which will be indexed by the same car model m. Similarly, sales of Mazda 3 4D and Mazda 3 5D are calculated separately while they share the same car model, Mazda 3.

[21] The sample periods include total 8 quarters from Q4 2019 to Q3 2021.

臺大學術期刊資料庫

Table 1: Sales before and after the incident

	Before		After	
Car Model	Mean	S.D.	Mean	S.D.
Toyota RAV4	4,533	2,080	2,931	1,522
Not RAV4	434	767	489	868
SUVs	496	814	664	1,164
Luxury Cars	352	485	326	418
Sedans and Hachbacks	457	829	425	647

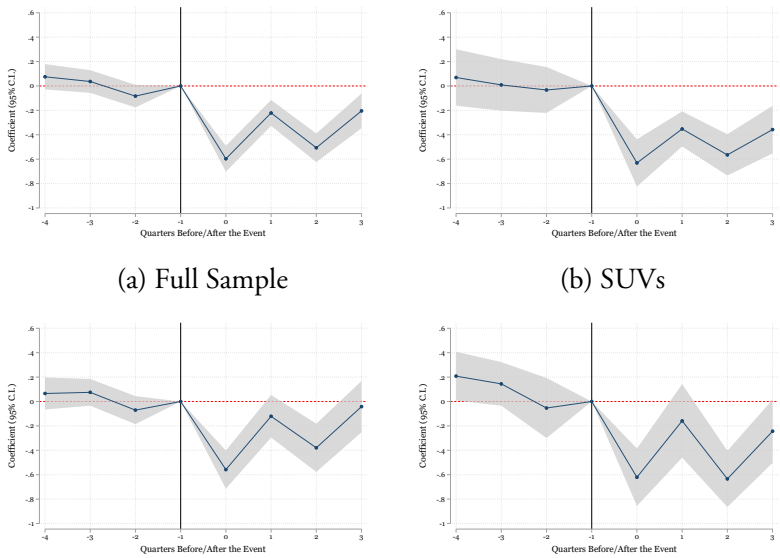
Note: This table shows average quarterly sales and associated standard deviations for five different sets of car models. We exclude products with quarterly sales less than 20. Each row refers a set of car models including Toyota RAV4, SUVs, luxury cars, and sedan and hatchback.

for treated and control groups are shown in Table 1. Toyota RAV4 has experienced a significant decrease in sales after the leak incident while other cars remained roughly the same levels of sales before and after Oct. 2020.

Estimation results of Equation 1 are presented in Figure 2 and Table 2. Figure 2 plots estimated coefficients β_τ along with associated 95% confidence intervals under four different control groups. The evolution of estimates under four specifications immediately provides several insights. First, estimates of β_τ in post-periods are all negative and mostly statistically significant. In particular, we find large decreases in quarterly sales for quarter zero, in which the roof leak incident took place, with drops between 45 to 47 percentage points relative to sales in the omitted quarter Q3 2020.[22] In later periods, the estimates are smaller in magnitudes and inconsistent across four different control groups in a way that some estimates are close to zero and statistically insignificant while others still show large decreases in sales. The inconsistency across estimates indicates that the dynamics of treatment effects may rely on choices of control groups as market conditions could change rapidly from quarter to quarter. In Table 2, the middle panel shows conventional DID estimates using one single coefficient to summarize all post-period effects. The estimates range from -0.327 to -0.489,

[22] When the estimated coefficient are -0.597 and -0.637, and the percentage changes are $\exp(-0.597) - 1 \approx -45\%$, and $\exp(-0.637) - 1 \approx -47\%$.

臺大學術期刊資料庫



(a) Full Sample (b) SUVs

(c) Luxury Cars (d) Sedans and Hatchbacks

Figure 2: Event-study results with different control groups

Notes: This figure plots estimated coefficient β_τ along with their 95% confidence intervals for Equation (1) using quarterly data spanning from October 2019 to September 2021. Outcome variable is the log of quarterly sales for product j. We exclude observations with quarterly sales less than 20. Figure 2a uses all available observations while Figure 2b limits to all SUVs, including RAV4. Figure 2c and Figure 2d change the control groups to luxury cars and cars within sedan and hatchback segments respectively. Detailed estimates and robust-clustered standard errors are reported in Table 2.

suggesting over 10% differentials in sales drop.

Second, the estimates before the incident from all four subfigures are close to zero and not statistically significant, indicating similar sales trends between Toyota RAV4 and other car models in various control groups. However, the presence of parallel trends in the pre-period does not mean that the same trends would still hold in the absence of treatment (Kahn-Lang and Lang, 2020). Specifically, in our setting, introduction of new car models would disrupt the trends in a way that both treated and untreated car mod-

臺大學術期刊資料庫

Table 2: Impact of UGCs from roof leak incident on RAV4's sales

	(a)	(b)	(c)	(d)
Event-Study Model:				
RAV4 × Q4 2019 ($\tau = -4$)	0.075	0.069	0.083	0.208**
	(0.052)	(0.116)	(0.080)	(0.099)
RAV4 × Q1 2020 ($\tau = -3$)	0.037	0.008	0.094	0.145
	(0.047)	(0.105)	(0.066)	(0.088)
RAV4 × Q2 2020 ($\tau = -2$)	−0.083*	−0.033	−0.033	−0.054
	(0.047)	(0.094)	(0.066)	(0.121)
RAV4 × Q3 2020 ($\tau = -1$, omitted)	0	0	0	0
RAV4 × Q4 2020 ($\tau = 0$)	−0.597***	−0.631***	−0.637***	−0.621***
	(0.054)	(0.097)	(0.091)	(0.116)
RAV4 × Q1 2021 ($\tau = 1$)	−0.221***	−0.352***	−0.091	−0.159
	(0.054)	(0.072)	(0.106)	(0.149)
RAV4 × Q2 2021 ($\tau = 2$)	−0.507***	−0.565***	−0.386***	−0.635***
	(0.059)	(0.084)	(0.121)	(0.114)
RAV4 × Q3 2021 ($\tau = 3$)	−0.204***	−0.358***	−0.051	−0.244*
	(0.072)	(0.097)	(0.121)	(0.128)
DID Model:				
RAV4 × POST	−0.389***	−0.488***	−0.327***	−0.489***
	(0.047)	(0.079)	(0.077)	(0.078)
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes
Model Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,456	564	524	246
R-squared	0.721	0.722	0.713	0.916

Note: This table shows the estimation results for Equation 1 in the top panel and DID specification in the bottom panel. Sample period includes 8 quarters spanning from October 2019 to September 2021 with four quarters for both baseline period and follow-up period. Outcome variable is the log of quarterly sales for product j. We exclude observations with quarterly sales less than 20. Columns (a) uses all available observations. Standard errors in parentheses are robust and clustered at model level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

els could deviate from pre-trends. For example, Toyota Corolla Cross, an indigenous compact SUV, was introduced to the market in October 2020.

臺大學術期刊資料庫

Since then, its sales have been tremendous with sales of 40,675 in 2021, replacing Toyota RAV4 as the best-selling car model in 2021. Therefore, given parallel-trends in pre-periods, the shock from Toyota Corolla Cross may influence the post-trends, invalidating the common trend assumption required in DID.

Moreover, even with the same parallel trends before and after the incident, the DID estimates are essentially picking up equilibrium effects, including direct treatment effects, as well as cannibalization effects, because consumers may alter their purchase choices to other products in response to exogenous shocks or policies (Bachmann et al., 2022; Chen, Hu, and Knittel, 2021; Kuo and Shu, 2023). For instance, Figure 2b, in which all SUVs are included in the estimation, presents very similar pre-trends in sales as estimates before the incident are fairly close to zero. However, since RAV4 also belongs to the SUV segment, in addition to direct adverse effect on RAV4's sales, we also expect substitution from RAV4 to other SUVs, invalidating the role of control group since other SUVs were also affected by the incident. In addition, introductions of new car models, e.g. Toyota Corolla Cross, Nissan Sentra, and Mercedes-Benz GLB, in the fourth quarter of 2020 would impact both treated and untreated car models, adding another layer of equilibrium effect. Therefore, the reduced-form estimates are biased to some extent since the estimates reflect equilibrium outcomes in sales.

To disentangle the direct effect of user-generated content from equilibrium effects, we use similar empirical strategy in Bachmann et al. (2022) by complementing reduced-form analysis with a discrete-choice model of consumer demand. Counterfactual experiments based on the estimated demand system could isolate the direct effect of user-generated content by simulating market outcomes without the incident. Substitution effects and introduction of new products would be embodied in the demand system as consumers go through their decision process. However, the approach does not come without any cost. Additional behavioral and functional form assumptions are required to make the estimation feasible. It also needs more comprehensive data which include aggregated market share data, as well as product characteristics. Finally, we have to construct a set of valid instrumental variables to deal with price endogeneity in demand system.

臺大學術期刊資料庫

4 Structural analysis: discrete choice models

This section describes our empirical models used to estimate consumer demand in the Taiwanese automotive market, and discuss estimation results, as well as conterfactual experiments. In Section 4.1, we introduce a logit discrete choice model for consumer demand. Section 4.2 presents a nested-logit model which allows consumers to have heterogeneous preference over car segments. In Section 4.3, details about specifications, product characteristics, and instrumental variables used in the estimation are provided. Section 4.4 shows demand estimates and implied own-price elasticities. Finally, in Section 4.5, we conduct conterfactual experiments to disentangle equilibrium effects on sales.

4.1 Logit demand

Following logit discrete choice demand model developed in Berry (1994), we model consumer demand as a function of mean utility from product characteristics and prices. Specifically, consumer i's indirect utility u_{ij} from purchasing product j is

$$u_{ij} = \boldsymbol{x}_j\boldsymbol{\beta} - \alpha \cdot p_j + \xi_j + \epsilon_{ij} = \delta_j + \epsilon_{ij}, \tag{2}$$

where $j = 0, 1, \cdots, J$ is indexing different products in consumer's choice set in a market. $\boldsymbol{x}_j$ is a vector of observable product characteristics. p_j is the price of product j. ξ_j is the mean impact on consumer utility from unobserved product characteristics or demand shocks. δ_j is the meaning utility for product j. ϵ_{ijt} is a random taste shock. Finally, $\boldsymbol{\beta}$, and α are parameters to be estimated.

Consumers may choose not buy any product. Therefore, we introduce an outside good for not buying, which is indexed by $j = 0$. The indirect utility from this outside good is

$$u_{i0} = \delta_0 + \epsilon_{i0},$$

where δ_{0k} is the mean utility of outside option. Generally, δ_0 is normalized to zero since utility is ordinal and only the difference in utility can be estimated.

Consumers make their choices by comparing utilities across all alternatives, and select one which gives them the highest utility. We assume that ϵ_{ij} follows an i.i.d. Type I Extreme Value distribution. This assumption

臺大學術期刊資料庫

is mainly for computational purpose and it yields a clean functional form for market shares. Specifically, the choice probabilities for product j and outside option can be written as:

$$s_j = \frac{\exp\left(\delta_j\right)}{1 + \sum_{j'=1}^{J} \exp\left(\delta_{j'}\right)} = \frac{\exp\left(\boldsymbol{x}_j\boldsymbol{\beta} - \alpha \cdot p_j + \xi_j\right)}{1 + \sum_{j'=1}^{J} \exp\left(\boldsymbol{x}_{j'}\boldsymbol{\beta} - \alpha \cdot p_{j'} + \xi_{j'}\right)}, \tag{3}$$

$$s_0 = \frac{1}{1 + \sum_{j'=1}^{J} \exp\left(\delta_{j'}\right)} = \frac{1}{1 + \sum_{j'=1}^{J} \exp\left(\boldsymbol{x}_{j'}\boldsymbol{\beta} - \alpha \cdot p_{j'} + \xi_{j'}\right)}. \tag{4}$$

Taking logarithm and combining the above two equations yields:

$$\ln\left(s_j\right) - \ln\left(s_0\right) = \boldsymbol{x}_j\boldsymbol{\beta} - \alpha \cdot p_j + \xi_j. \tag{5}$$

As suggested in Berry (1994), Equation 5 can be estimated in a regression with aggregate market share data, prices, and associated product characteristics as observables while ξ_j is treated as unobserved error term in the regression. To construct s_j and s_0, we use total numbers of households in Taiwan to denote the potential market size in each quarter. Market share of outside option, s_0, is the fraction of consumers who choose not to purchase any of the products in the market.[23] s_j is the fraction of households that purchased product j, calculated as the sales of product j divided by the total number of households.

For logit model the price elasticities of market shares are derived as follows:

1. Own-price elasticity of product j

$$\Delta_{jj} = \frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j} = \alpha p_j\left(1 - s_j\right). \tag{6}$$

2. Cross-price elasticity between products j and k

$$\Delta_{jk} = \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = -\alpha p_k s_k. \tag{7}$$

[23] Specifically, for each quarter, the market share of the outside option is computed by subtracting the total sales from the total number of households and then dividing the result by the total number of households.

臺大學術期刊資料庫

Note that the cross-price elasticity of product j with respect to product k in Equation 7 is independent of product j. More specifically, if two car models have the same market share, then the cross-price elasticity of each with respect to a change in the price of Toyota RAV4 is the same regardless of prices, brands, or segments. This unattractive property of the logit model is the Independence of Irrelevant Alternatives (IIA) property, which is resulted from the independence of idiosyncratic tastes, ϵ_{ij}, across products. To alleviate the undesirable property, in the next section, we introduce nested logit model which allows for correlated individual taste shocks between alternative products in pre-specified groups.

4.2 Nested logit demand

Nested logit demand model differentiates from logit demand in that products are grouped into $G + 1$ exhaustive and mutually exclusive sets, $g = 0, 1, 2, \cdots, G$. For $g = 0$, the group contains the outside good. The set of products in group g is denoted as $\mathscr{J}_g$. For product j in group g, consumer i's indirect utility is

$$u_{ij} = \delta_j + \zeta_{ig} + (1 - \sigma)\epsilon_{ij}, \tag{8}$$

where $\delta_j = \boldsymbol{x}_j\boldsymbol{\beta} - \alpha \cdot p_j + \xi_j$, representing the mean utility for product j. ζ_{ig} is consumer i's idiosyncratic taste for products in group g.[24] ϵ_{ij} follows an i.i.d. Type I Extreme Value distribution and ζ_{ig} has the unique distribution such that $\zeta_{ig} + (1 - \sigma)\epsilon_{ij}$ is also distributed as Type I Extreme Value (Cardell, 1997). The distribution of ζ_{ig} depends on σ, with $0 \leq \sigma \leq 1$, which measures the correlation of utilities among products in the same group. When $\sigma = 0$, the model degenerates to standard logit model described in the previous section.

For product j in group g, the choice probability of product j conditional on group g equals

$$\bar{s}_{j|g} = \frac{\exp\left(\frac{\delta_j}{1-\sigma}\right)}{D_g}, \tag{9}$$

[24] Another way to think about ζ_{ig}, as mentioned in Berry (1994), is to interpret ζ_{ig} as a random coefficients on group-specific dummies.

臺大學術期刊資料庫

where

$$D_g = \sum_{k \in \mathcal{J}_g} \exp\left(\frac{\delta_k}{1-\sigma}\right). \tag{10}$$

As group 0 contains only one product, the outside good with $j = 0$ and $\delta_0 = 0$, the choice probability for the outside is

$$s_0 = \frac{1}{\sum_g D_g^{(1-\sigma)}}. \tag{11}$$

The probability of selecting group g, i.e. the group share, is

$$\bar{s}_g = \frac{D_g^{(1-\sigma)}}{\sum_g D_g^{(1-\sigma)}} \Rightarrow \ln\left(D_g\right) = \frac{\ln\left(\bar{s}_g\right) - \ln\left(s_0\right)}{(1-\sigma)}. \tag{12}$$

Using the conditional probability formula, the unconditional choice probability for product j in group g is

$$s_j = \bar{s}_{j|g} \cdot \bar{s}_g = \frac{\exp\left(\frac{\delta_j}{1-\sigma}\right)}{D_g^{\sigma}\left[\sum_g D_g^{(1-\sigma)}\right]} \Rightarrow \ln\left(s_j/s_0\right) = \frac{\delta_j}{(1-\sigma)} - \sigma \ln\left(D_g\right). \tag{13}$$

Substituting Equation 12 into Equation 13 and making some arrangements give us similar logit equation in the previous section

$$\ln\left(s_j\right) - \ln\left(s_0\right) = \delta_j + \sigma \ln\left(\bar{s}_{j|g}\right) = \boldsymbol{x}_j\boldsymbol{\beta} - \alpha \cdot p_j + \sigma \ln\left(\bar{s}_{j|g}\right) + \xi_j. \tag{14}$$

Unknown parameters $\boldsymbol{\beta}$, α, and σ can be estimated via an instrumental variables regression.

For nested logit model the price elasticities of market shares are derived as follows:

1. Own-price elasticity of product j in group g

$$\Delta_{jj} = \frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j} = \frac{\alpha p_j}{1-\sigma}\left(1 - \sigma\bar{s}_{j|g} - (1-\sigma)s_j\right), \tag{15}$$

2. Cross-price elasticity between products j and k when two products are in the same group g

$$\Delta_{jk} = \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = -\frac{\alpha p_k}{1-\sigma}\left(\sigma\bar{s}_{k|g} + (1-\sigma)s_k\right). \tag{16}$$

臺大學術期刊資料庫

Table 3: Summary statistics

	Mean	S.D.	Median	Min	Max
Sales	496.19	903.27	184.50	20.00	8289.00
Price (10,000NT$)	163.79	112.98	124.80	52.90	840.00
HP/Weight (kW/1,000kg)	121.06	30.73	113.99	78.71	284.86
Displacement (liter)	1.82	0.57	1.80	0.00	3.98
Size (m^3)	13.06	1.97	12.74	8.52	18.79
Fuel costs (NT$/100km)	1.74	0.52	1.72	0.00	4.06

Note: Sample period spans from October 2019 to September 2021, total 8 quarters. We exclude observations with quarterly sales less than 20. In order to include electric vehicles, we set both engine displacement and fuel costs to zero for EVs since engine size and fuel costs are not applicable to EVs. For other vehicles, fuel costs are adjusted by quarterly average prices for gasoline and diesel. Total number of observation, model-trim-quarter combinations, is 1,456.

3. Cross-price elasticity between products j and k when two products are in different groups

$$\Delta_{jk} = \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = -\alpha p_k s_k. \tag{17}$$

4.3 Specification and instrumental variables

We model consumer tastes over price and several measures of automobile performance, including horsepower divided by weight, engine displacement, exterior dimensions, and fuel costs. These product characteristics are scraped from 8891, an online auto media website. Summary statistics for the variables included in our empirical demand model are shown in Table 3.[25]

Adding fixed effects in the structural demand estimation could improve precision of estimates by reducing variations in structural errors (Nevo, 2000). Here, we include two sets of fixed effects in the specification. Model fixed-effects could control for unobserved time-invariant heterogeneity within

[25] In the Appendix, we report a correlation matrix of these variables in Table A1. In generally, cars with lower prices are associated with more sales, as these vehicles tend to be smaller, and more fuel efficient.

臺大學術期刊資料庫

specific car model, and time fixed-effects absorb countervailing market conditions that are common to all cars. We model Toyota RAV4 roof leak incident as demand shocks to Toyota RAV4 only and use interaction terms between quarterly indicator variables and RAV4 dummy to capture the post-event dynamics of user-generated content's impacts. Specifically, we model ξ as follow:

$$\xi_{jt} = \sum_{l=1}^{4} \lambda_l \cdot RAV4_m \cdot D_l + \phi_m + \rho_t + \Delta\xi_{jt}, \tag{18}$$

where $RAV4_m$ is a dummy variable equal to 1 if the car model is Toyota RAV4, and 0 otherwise. D_l is also a dummy variable indicating the number of quarter, l, starting from Q4 2020. When $l = 4$, D_l refers Q3 2021. λ_l as the associated coefficients of the interaction terms between $RAV4_m$ and D_l captures demand shocks specific to Toyota RAV4 from Q4 2020 to Q3 2021. ϕ_m controls for unobserved characteristics which are persistent over time while ρ_t absorbs time-specific demand shocks common to all products. Finally, the specification still leaves us with unobserved heterogeneity $\Delta\xi_{jt}$.

Since the model is linear, it is natural to use OLS in the estimation. However, OLS estimation will result in a biased price coefficient since price is likely to be correlated with unobserved structural errors $\Delta\xi_{jt}$ as firms could adjust price in response to unobserved product characteristics. In addition, in the nested-logit model, $\ln(\bar{s}_{j|g})$ is also endogenous since $\Delta\xi_{jt}$ would affect market shares. To deal with the endogeneity issues, Berry (1994) suggests instrumental variable approach via two-stage least square (2SLS) method. Following Berry, Levinsohn, and Pakes (1995), I use product characteristics from competitors as instruments. This type of instruments relies on the assumption of exogeneity of observed product characteristics space and unobserved structural errors. Specifically,

$$\mathbb{E}\left(\Delta\xi_{jt}|\boldsymbol{x}_t\right) = 0, \tag{19}$$

where $\Delta\xi_{jt}$ is unobserved time and product specific structural errors after controlling for fixed effects and demand shocks from leak incident. $\boldsymbol{x}_t$ is a matrix which include the whole characteristics space at time t.

The rationale of BLP-style instruments is based on oligopoly pricing and horizontal competition in characteristics space. Products with close proximity in characteristics space to other products implies that they are competing with good substitutes and thus tend to have lower markups, which limit

臺大學術期刊資料庫

their abilities to raise prices and result in lower prices. The economics intuition establishes relevance condition, which can also be checked empirically. In contrast, the validity of the exclusion restriction hinges on conditional mean independent of unobserved characteristics, Equation 19, which implies that product characteristics $\boldsymbol{x}_t$ are determined before $\Delta\xi_{jt}$ is known. In our empirical setting, which allows consumers to make decisions in every quarter, the assumption is likely to hold since auto makers cannot respond to quarter-product specific unobserved demand shocks or characteristics immediately by changing their product designs in a quarter. More often than not, it will take at least one year to change product characteristics on a car model.

BLP-style instruments are popular in the literature for two reasons. First, product characteristics information is already included in the dataset. No additional information is needed.[26] Second, the identifying assumption essentially allow for flexible constructions of product characteristics for instrumental variables. Therefore, our first set of instrumental variables, $Z_{jt}^{k,1}$ follows Gandhi and Houde (2019) by constructing "differentiation IVs" to approximate optimal instruments of Chamberlain (1987).[27] The second set of instrumental variables, $Z_{jt}^{k,2}$, are average product characteristics from competing products within the same segment. The period t instrument for product j and characteristic k is

$$Z_{jt}^{k,1} = \sum_{j \neq j'} \mathbf{1}\left(\left|x_{j't}^k - x_{jt}^k\right| < SD\left(x_{\cdot t}^k\right)\right) \cdot x_{j't}^k, \tag{20a}$$

$$Z_{jt}^{k,2} = \frac{1}{\left|\mathscr{J}_g\right| - 1} \sum_{\substack{j \neq j' \\ j, j' \in \mathscr{J}_g}} x_{j't}^k, \tag{20b}$$

where $SD(x_{\cdot t}^k)$ is the standard deviation of product characteristics k in period t, and $|\mathscr{J}_g|$ is number of products in group g.

Finally, in the nested-logit model, pre-determined groups are required to calculate within group market shares, $\bar{s}_{j|g}$. We use total 9 car segments, including SUV, full-size SUV, compact sedan, mid-size sedan, large sedan,

[26] Cost information is another viable candidate for instrumental variable. However, cost data with cross-sectional variations are rarely observed in practice.

[27] Miravete, Moral, and Thurk (2018); Bachmann et al. (2022) also use various differentiation IVs to identify parameters in discrete choice demand models.

臺大學術期刊資料庫

compact hatchback, mid-size hatchback, wagon, and MPV.[28] Toyota RAV4 and Toyota Corolla Cross both belong to the SUV segment.

4.4 Empirical results

Before presenting the demand estimation results, our first-stage regressions in 2SLS estimation are shown in Table 4. The estimates show that our constructions of IVs are correlated with endogenous prices and within group shares. One concern of the first stage regression results is that the F-statistics are not bigger than 10, a typical test for weak IVs. The weak IV issue is coming from the fact that the demand system is based on quarterly market shares and thus variations in both prices and product characteristics are limited since auto makers do not change their product designs or pricing significantly within two years. While shortening the time horizon in the demand system seems necessary in our empirical setting, it actually makes the identifying assumption more credible. Firms do not respond to quarter-product specific unobserved demand shocks and unobserved characteristics immediately by changing their product designs in a quarter. Longer time horizons not only provide more variations in prices and characteristics spaces, but also allow firms to strategically adjust the product characteristics. Therefore, it is essentially a trade-off between credibility of identifying assumption and statistical power of IVs. Our choice of quarterly demand system is motivated by the empirical setting, and comes with both costs and benefits. This deviation from existing literature strengthens the argument of identifying assumption but brings weak IVs. To assess the potential bias from weak IVs, we will look into implied price elasticities more closely.

Table 5 displays demand estimates in Logit-OLS, Logit-IV, and Nested Logit-IV. The main difference in estimates between the first and the other two columns is the estimated price coefficients. In OLS estimation, the price coefficient is close to zero and insignificant, which is indicative of potential biases resulted from positive correlation between unobserved structural errors and prices. With two sets of instrumental variables, price estimates in Logit-IV and Nested Logit-IV become significant with larger magnitudes. The estimates of product characteristics in the last two columns show that in general consumer utility decreases with price, displacements, and fuel costs and increases with a car's size, and its horsepower relative to weight. The

[28] The classification is based on Full-size SUVs usually have 7 seats while SUVs only have 5.

臺大學術期刊資料庫

Table 4: First-stage regression results

	Price	$\ln(s_{j\bar{\mid}g})$
IV1-HP	−0.00042	0.00005*
	(0.00031)	(0.00003)
IV1-Cost	−0.06429***	0.00203***
	(0.01634)	(0.00078)
IV1-Dis	0.05642**	−0.00139
	(0.02525)	(0.00157)
IV1-Size	0.00638	−0.00033
	(0.00397)	(0.00035)
IV2-HP	−0.17713	−0.01149
	(0.11406)	(0.00780)
IV2-Cost	−1.99413	0.38510
	(7.66356)	(0.45028)
IV2-Dis	4.69638	−0.40832
	(9.22553)	(0.44840)
IV2-Size	0.02943	0.00894***
	(0.02452)	(0.00173)
HP/Weight (kW/1,000kg)	0.65304***	0.00264
	(0.09053)	(0.00451)
Displacement (liter)	6.14575	−1.45045***
	(7.28489)	(0.47219)
Size (m^3)	9.12706**	−0.22401
	(3.62919)	(0.21103)
Fuel costs (NT$/100km)	−7.84373	0.05100
	(5.07818)	(0.14201)
RAV4 $\times$**1**(Q4 2020)	0.27373	−0.78182***
	(0.42331)	(0.05492)
RAV4 $\times$**1**(Q1 2021)	−0.55469	−0.48829***
	(0.81160)	(0.05864)
RAV4 $\times$**1**(Q2 2021)	−0.28307	−0.58530***
	(0.88298)	(0.06847)
RAV4 $\times$**1**(Q3 2021)	−0.33588	−0.31674***
	(1.58744)	(0.09235)

Note: IV1s and IV2s are constructed according to the Equation 20 as x_{jt}^{k} refers to one of the product characteristics. Standard errors in parentheses are robust and clustered at model level. F-statistics are 5.65 and 7.48 respectively. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

signs of these parameters are intuitive while some of them are statistically insignificant.

臺大學術期刊資料庫

Table 5: Logit demand estimates

	Logit-OLS	Logit-IV	Nested Logit-IV
Price (10,000NT$)	−0.008	−0.031*	−0.024**
	(0.007)	(0.016)	(0.012)
HP/Weight (kW/1,000kg)	0.005	0.021*	0.016*
	(0.006)	(0.012)	(0.009)
Displacement (liter)	−1.598***	−1.236**	−0.952**
	(0.402)	(0.486)	(0.395)
Size (m^3)	−0.190	0.024	0.012
	(0.167)	(0.279)	(0.219)
Fuel costs (NT$/100km)	0.059	−0.211	−0.188
	(0.208)	(0.312)	(0.230)
$\ln(\bar{s}_{j\|g})$			0.268**
			(0.126)
RAV4 ×**1**(Q4 2020)	−0.604***	−0.595***	−0.371***
	(0.053)	(0.054)	(0.109)
RAV4 ×**1**(Q1 2021)	−0.248***	−0.239***	−0.105
	(0.051)	(0.054)	(0.073)
RAV4 ×**1**(Q2 2021)	−0.530***	−0.539***	−0.367***
	(0.057)	(0.059)	(0.093)
RAV4 ×**1**(Q3 2021)	−0.243***	−0.267***	−0.143*
	(0.072)	(0.070)	(0.084)
# IVs	—	8	8
Model FE	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes
Observations	1456	1456	1456

Note: Overidentification test statistics are 7.602 for logit model with a p-value of 0.369, and 6.945 for nested logit model with a p-value of 0.326. First-stage regression results are shown in Table 4. Standard errors in parentheses are robust and clustered at model level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

The adverse effects of Toyota RAV4 roof leak incident are captured by estimates in the middle panel of Table 5. They are all negative as the largest impact on consumer utility emerges in the first quarter of the incident. Note that these estimates cannot be interpreted in the same way as DID event-

臺大學術期刊資料庫

Table 6: Summary statistics for own-price elasticities

	Mean	Min	P25	P50	P75	Max
Logit-OLS	−1.23	−6.32	−1.53	−0.94	−0.67	−0.40
Logit-IV	−5.05	−25.88	−6.28	−3.84	−2.76	−1.63
Nested Logit-IV	−5.41	−27.94	−6.81	−4.08	−2.96	−1.44

Note: Calculations of price-elasticities in the first two rows are based on Equation 6. The last row follows Equation 15. Price coefficients, α, and nested parameter, σ, are from Table 5.

study model since the outcome variable in discrete choice model is market share relative to outside option not sales. Negative estimates on the interaction terms imply that consumer utilities toward Toyota RAV4 have dropped since the fourth quarter of 2020. In Logit-IV, the estimated coefficient, −0.595, is equivalent to an average decrease of NT$192,000 in consumer valuation, or willingness to pay, of Toyota RAV4.[29] For Nested Logit-IV, the incident has reduced consumer valuations of RAV4 by an average of 155,000NT$ in Q4 2020. Given the MSRP of Toyota RAV4 is around 1 to 1.2 million NT$, the impacts on consumer valuations are both statistically and economically meaningful.

Using the estimates in Table 5, we can calculate the own-price elasticities according to Equation 6 and 15. Distribution of own-price elasticities implied by nested logit demand is plotted in Figure 3 while Table 6 compares summary statistics of own-price elasticities across three specifications.[30] The estimated price elasticities under Logit-IV and Nested Logit-IV are less than −1, with median elasticities being −3.84 and −4.08, suggesting that all vehicles are reasonably priced at the elastic part of the demand curve. Figure 3 shows that majorities of elasticites range from −1 to −10 with a fairly long left tail.

[29]The estimated price coefficient measures marginal utility of additional one thousand NT$. Dividing other estimates by the price coefficient yields monetary valuations in mean consumer utilities. Here, $-0.595/0.031 = -19.2$, equivalent to 192,000NT$.

[30]For OLS and Logit-IV, we plot the histograms in Figure A2 of Appendix.

臺大學術期刊資料庫

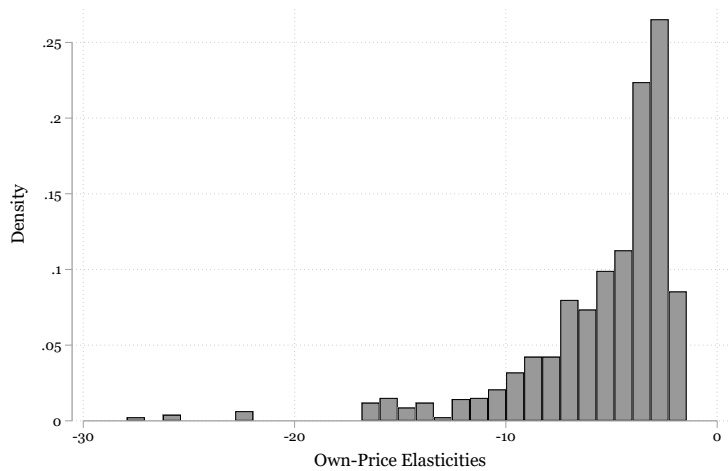


Figure 3: Histogram of own-price elasticities

Notes: The above figure shows the distribution of own-price elasticity of over 1,400 vehicles in the Taiwanese automotive market. The calculation is based on Equation 15 and estimates from Nested Logit-IV specification.

4.5 Counterfactual exercises

Given a reasonably-estimated consumer demand system, the structural model can be used to simulate the impacts of user-generated content on equilibrium sales of Toyota RAV4. Specifically, using Toyota Corolla Cross and the RAV4 leak incident as two treatments, we perform three counterfactual exercises.[31] The first counterfactual simply investigates how sales of RAV4 would have changed if Toyota Corolla Cross was removed from consumer's choice set. This exercise allow us to quantify cannibalization effects given the leak incident. In the second counterfactual, we impose a hypothetical case in which roof leak incident did not happen. Essentially, we recompute market shares after removing the adverse effects of RAV4 after October 2020 on consumer's mean utilities, i.e. setting the estimates of interaction terms in Table 5 to zero. The third scenario combines the previous two cases and allows consumers to make purchase decisions with no Toyota Corolla Cross

[31] Along with the actual scenario, it becomes a total of four different cases.

臺大學術期刊資料庫

and leak incident. The first two simulations provide us with two separated effects which took place exactly at the same time while reduced-form analysis can only estimate a combination of these two effects.

Figure 4 presents the actual sales and counterfactual RAV4 sales in three simulations.[32] First, we investigate the cannibalization effects from the introduction of Toyota Corolla Cross, the best-selling car model in 2021, on RAV4's sales. The purpose of this exercise is to show that counterfactual experiments in structural approach could disentangle the equilibrium effects. In Figure 4, the light orange bars indicate the counterfactual sales if Toyota Corolla Cross is no longer available to consumers. As consumers are remaking their purchase decisions, some of them may decide to go with Toyota RAV4 since both car models are both within the SUV segment, resulting in changes in sales.[33] From our estimates, there are only modest increases in RAV4's sales, 369, 451, 243, and 386 in each quarter, approximately 4% of the total sales in the previous year.

Second, the light blue bars in Figure 4 are the simulated sales when the roof leak incident was removed. These simulated sales are significantly higher than actual sales especially in the first quarter of the incident. The numbers of lost sales in each quarter are 3,714, 955, 2,956, and 1,310.[34] While the sales gaps seem to be smaller in later periods, the number of total lost sales is around 9,000, which is equivalent to 25% of the total sales in a year before the incident. Recall that the DID estimates Table 2 suggest that the impact of roof leak incident is around -32%.[35] Smaller estimated impact of user-generated content in the structural model is internally-consistent since simulated equilibrium accounts for the demand shocks from new products by allowing consumers to substitute to Toyota Corolla Cross while observing the leak incident. Apparently, the direct impacts of leak incident are much larger than those caused by removing Toyota Corolla Cross, suggesting that the cannibalization effects are not major drivers for the decreased sales of RAV4.

Finally, the third simulation includes the two scenarios and re-calculate

[32] Detailed sales numbers are reported in Table A2 of Appendix.

[33] Of course, consumers can choose other car models when Toyota Corolla Cross is not available. But the changes in sales in other car models are not the main focus of this article.

[34] These numbers correspond to 39%, 13%, 38%, and 17% decreases relative to hypothetical sales when leak incident did not happen.

[35] We use the estimate in Column (a) from Table 2. $\exp(-0.389) - 1 \approx -32\%$

臺大學術期刊資料庫

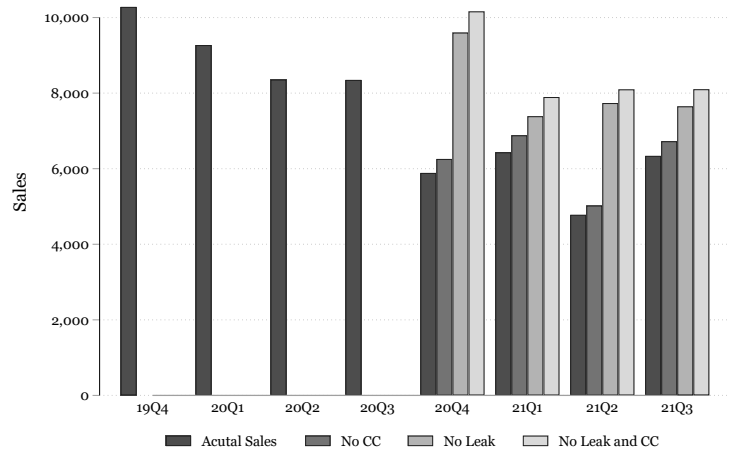


Figure 4: Counterfactual sales of Toyota RAV4

Notes: Simulations are based on nested logit model and estimates in the last column of Table 5. We first calculate implied market shares of RAV4 in each quarter and multiply the market shares with market sizes, total number of households, to get the counterfactual sales. "No Leak" refers to the scenario in which the adverse effects of roof leak incident were removed. "No CC" indicates that Toyota Corolla Cross was taken out of consumer's choice set. Finally, "No CC & Leak" includes the previous two scenarios.

equilibrium sales for RAV4 when consumers cannot choose Toyota Corolla Cross and do not observe leak incident. The sales, as expected, are the highest among three simulated sales. The differences between actual sales and simulated sales in the third simulations are essentially picking up the combined effects of cannibalization effects from Toyota Corolla Cross and direct effects from leak incident.

5 Concluding remarks

This paper uses both reduced-form and structural approaches to investigate the effect of user-generated content on relative sales by leveraging Toyota RAV4 roof leak incident as a natural experiment in the Taiwanese automo-

臺大學術期刊資料庫

tive market. In DID analysis, similar pre-trends do not necessarily ensure identification of treatment effects from leak incident as DID estimates are unable to disentangle the equilibrium effects, suggesting over 30% decrease in sales relative to those in the previous year. Structural approach, on the other hand, could isolate the treatment effects via counterfactual simulations, showing that the incident has led to around 9,000 lost sales, approximately 25%, of the total sales in the baseline periods.

Note that we are not implying that structural approach is a superior empirical strategy. In general, the structural approach demands a fully specified model, behavioral and computational assumptions, comprehensive market share data from all products, and valid instruments to address price endogeneity. While many empirical applications would find it infeasible to meet these requirements, we are fortunate enough to have the context in automobile industry in that it is the most widely studied industry in empirical industrial organization literature. Guidance from prior studies and data availability considerably lowers the costs of implementing the structural analysis. In contrast, DID empirical strategy would be viable with minimal requirements on data and identifying assumption. For example, instead of data from all available products, DID only needs a sample consisting with a valid control group and a treatment group. Besides, no behavioral model is required for reduced-form approaches, avoiding misspecification risk. In fact, as two approaches excel in different ways, they could complement each other in impactful ways (Todd and Wolpin, 2023).

Finally, our estimates highlight the power of user-generated content and online word-of-mouth in a context of negative information reverberated across various social media platforms. Nevertheless, Seiler, Yao, and Wang (2017) finds low elasticity of TV viewership with respect to micro-blogging activity, suggesting that practitioners might overestimate the effectiveness of online WOM as a marketing channel. The discrepancy could be resulted from asymmetric responses to positive or negative user-generated content since consumers understand, to some extent, that a large portion of UGC are in fact promotional content. Besides, consumers may become risk-averse while making major purchases. Future researches are necessary to further the understanding of user-generated content and online word-of-mouth.

臺大學術期刊資料庫

Appendix

A Additional figures and tables

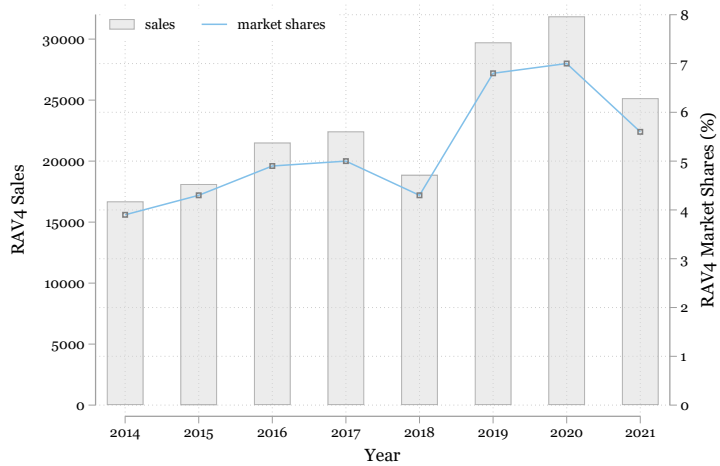
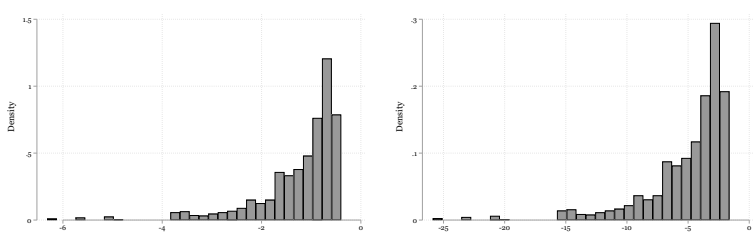


Figure A1: Annual Sales and Market Shares for Toyota RAV4



(a) Logit-OLS (b) Logit-IV

Figure A2: Histogram of own-price elasticities

Notes: The above figures plot the distributions of own-price elasticity of over 1,400 vehicles in the Taiwanese automotive market. The calculation is based on Equation 6 and estimates from Logit-OLS and Logit-IV specification in Table 5.

臺大學術期刊資料庫

Table A1: Correlation matrix

	Sales	Price	HP/ Weight	Displace-ment	Size	Fuel costs
Sales	1.000					
Price	−0.177	1.000				
HP/Weight	−0.125	0.640	1.000			
Displacement	−0.022	0.459	0.217	1.000		
Size	−0.028	0.559	0.289	0.531	1.000	
Fuel costs	−0.072	0.255	0.146	0.658	0.390	1.000

Note: We exclude observations with quarterly sales less than 20. In order to include electric vehicles, we set both engine displacement and fuel costs to zero for EVs since engine size and fuel costs are not applicable to EVs. For other vehicles, fuel costs are adjusted by quarterly average prices for gasoline and diesel. Total number of observation, model-trim-quarter combinations, is 1,456.

Table A2: Actual sales and counterfactual sales for Toyota RAV4

	Actual Sales	No CC	No Leak	No CC & Leak
Q4 2019	10,283	—	—	
Q1 2020	9,272	—	—	
Q2 2020	8,359	—	—	
Q3 2020	8,351	—	—	
Q4 2020	5,889	6,258	9,603	10,165
Q1 2021	6,434	6,885	7,389	7,897
Q2 2021	4,782	5,025	7,738	8,101
Q3 2021	6,341	6,728	7,651	8,104

Note: Simulations are based on nested logit model and estimates in the last column of Table 5. We first calculate implied market shares of RAV4 in each quarter and multiply the market shares with market sizes, total number of households, to get the counterfactual sales. "No Leak" refers to the scenario in which the adverse effects of roof leak incident were removed. "No CC" indicates that Toyota Corolla Cross was taken out of consumer's choice set. Finally, "No CC & Leak" combines the previous two scenarios.

臺大學術期刊資料庫

References

Anderson, Michael and Jeremy Magruder (2012), "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *Economic Journal*, 122(563), 957–989.

Ater, Itai and Nir S. Yoseph (2022), "The Impact of Environmental Fraud on the Used Car Market: Evidence from Dieselgate," *Journal of Industrial Economics*, 70(2), 463–491.

Bachmann, Rüdiger, Gabriel Ehrlich, Ying Fan, Dimitrije Ruzic, and Benjamin Leard (2022), "Firms and Collective Reputation: A Study of the Volkswagen Emissions Scandal," *Journal of the European Economic Association*, 21(2), 484–525.

Bai, Jie, Ludovica Gazze, and Yukun Wang (2022), "Collective Reputation in Trade: Evidence from the Chinese Dairy Industry," *Review of Economics and Statistics*, 104(6), 1121–1137.

Banerjee, Shrabastee, Chrysanthos Dellarocas, and Georgios Zervas (2021), "Interacting User-Generated Content Technologies: How Questions and Answers Affect Consumer Reviews," *Journal of Marketing Research*, 58(4), 742–761.

Berry, Steven (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25(2), 242–262.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Cardell, N. Scott (1997), "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13(2), 185–213.

Chamberlain, Gary (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34(3), 305–334.

Chen, Chia-Wen, Wei-Min Hu, and Christopher R. Knittel (2021), "Subsidizing Fuel-Efficient Cars: Evidence from China's Automobile Industry," *American Economic Journal: Economic Policy*, 13(4), 152–184.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43(3), 345–354.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29(5), 944–957.

臺大學術期刊資料庫

Freedman, Seth, Melissa Kearney, and Mara Lederman (2012), "Product Recalls, Imperfect Information, and Spillover Effects: Lessons from the Consumer Response to the 2007 Toy Recalls," *Review of Economics and Statistics*, 94(2), 499–516.

Gandhi, Amit and Jean-François Houde (2019), "Measuring Substitution Patterns in Differentiated Products Industries," *Working Paper*.

Hammond, Robert G. (2013), "Sudden Unintended Used-Price Deceleration? The 2009–2010 Toyota Recalls," *Journal of Economics & Management Strategy*, 22(1), 78–100.

Kahn-Lang, Ariella and Kevin Lang (2020), "The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications," *Journal of Business & Economic Statistics*, 38(3), 613–620.

Kuo, Fang-Chang and Liang-Hong Shu (2023), "The Impact of Volkswagen Dieselgate on the Taiwanese Automotive Market," *Journal of Social Sciences and Philosophy*.

Luca, Michael (2016), "Reviews, Reputation, and Revenue: The Case of Yelp. Com," Harvard Business School NOM Unit Working Paper.

Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60(3), 531–542.

Miravete, Eugenio J., María J. Moral, and Jeff Thurk (2018), "Fuel Taxation, Emissions Policy, and Competitive Advantage in the Diffusion of European Diesel Automobiles," *RAND Journal of Economics*, 49(3), 504–540.

Nevo, Aviv (2000), "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9(4), 513–548.

Proserpio, Davide and Georgios Zervas (2017), "Online Reputation Management: Estimating the Impact of Management Responses on Consumer Reviews," *Marketing Science*, 36(5), 645–665.

Seiler, Stephan, Song Yao, and Wenbo Wang (2017), "Does Online Word of Mouth Increase Demand? (And How?) Evidence from a Natural Experiment," *Marketing Science*, 36(6), 838–861.

Todd, Petra E. and Kenneth I. Wolpin (2023), "The Best of Both Worlds: Combining RCTs with Structural Modeling," *Journal of Economic Literature*, 61(1), 41–85.

投稿日期: 2022年7月4日, 接受日期: 2023年4月27日

臺大學術期刊資料庫

使用者生成內容對銷售之影響: 以豐田 RAV4 車頂漏水事件為例

郭芳彰
國立中正大學經濟系
陳翰陞
國立中正大學經濟系

本文利用豐田 RAV4 車頂漏水事件作爲自然實驗, 研究使用者生成內容對相關銷售的影響。透過差異中差異計量方法, 我們的實證結果顯示, 車頂漏水事件的使用者生成內容導致豐田 RAV4 的銷售下降超過 30%。但是, 估計結果可能不僅捕捉到處理效果, 還捕捉到銷售的其他均衡效應。因此, 我們估計了一個結構化需求模型, 並進行了反事實模擬, 顯示失去的銷售約爲 9,000 輛, 相當於上一年總銷量的 25%。而另一款同品牌的新車推出僅對 RAV4 的銷售造成大約 4%的下降。

關鍵詞: 使用者生成內容, 差異中差異, 巢狀對數需求
JEL 分類代號: L15, L62, M31

臺大學術期刊資料庫

臺大學術期刊資料庫