Estimating Discrete Games with Many Firms and Many Decisions: An Application to Merger and Product Variety*

Ying Fan†
University of Michigan
CEPR and NBER

Chenyu Yang‡
University of Maryland

January 27, 2024

Abstract

This paper presents a method for estimating discrete games based on bounds of conditional choice probabilities. The bounds are probabilities that an action is dominant and that it is not dominated. Because the bounds are easy to compute, our method is scalable to models with many firms and discrete decisions. We apply the method to study the effects of a hypothetical merger on firm entry and product variety in local retail craft beer markets in California. We find that the merger induces firm entry. The net effect on product variety is ambiguous once a fixed-cost efficiency is taken into account.

JEL: D43, L13, L41, L66

Keywords: discrete games, incomplete models, entry, product choice, merger, beer

*We thank Zibin Huang, Sueyoul Kim, and Xinlu Yao for their excellent research assistance and participants at Barcelona Summer Forum, Boston College, Caltech, Drexel, CMU, FTC Microeconomics Conference, Georgetown, ITAM, Johns Hopkins, Mannheim IO Workshop, MIT, Northwestern, NYU IO Day, Penn State, Rice, Stanford, Stony Brook, SUNY, University of Montreal, UT Berlin, Warwick IO Workshop, Washington University, Yale, and Zhejiang University for their insightful comments. We also thank Ryan Lee, Marc Sorini, and Bart Watson for insights into the craft beer industry. The authors acknowledge the University of Maryland and the University of Michigan supercomputing resources (http://hpcc.umd.edu, https://arc.umich.edu/greatlakes/) made available for conducting the research reported in this paper. Researchers' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researchers and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

†Department of Economics, University of Michigan, Ann Arbor, MI 48109; yingfan@umich.edu.

‡Department of Economics, University of Maryland, College Park, MD 20740; cyang111@umd.edu.

1 Introduction

Discrete games of firm entry or product choice are often used to understand the effect of merger, divestiture, or industrial policy on market structure. In this paper, we consider the estimation of such models when there are many firms or when each firm makes a large set of discrete decisions. For example, a firm may need to choose a set of products to sell in a market from many potential products. In this case, the firm's choice can be represented as a long vector of binary decisions regarding each potential product's entry into the market. The estimation of such a model can quickly become challenging as the computational burden of solving the game increases exponentially with the number of firms and firm decisions. In this paper, we propose a computationally tractable estimation method and apply the method to study the effects of merger on firm entry, product variety, pricing, and welfare in the context of the craft beer market in California.

Our method is based on the bounds for conditional choice probabilities. Consider a binary action $a \in \{0, 1\}$. Assuming no equilibrium action is dominated, we can show that the equilibrium probability of $a = 1$ is larger than the probability that $a = 1$ is a dominant strategy and smaller than the probability that $a = 1$ is not a dominated strategy. These bounds hold when there is no pure-strategy equilibrium, when there are multiple equilibria, under any equilibrium selection rule, and when the selection rule varies across markets. More importantly, these bounds are easy to compute even with a large number of firms or firm decisions because the bounds can often be reduced to cumulative distribution functions evaluated at certain cutoffs. Using Monte Carlo experiments, we compare our method to existing methods for estimating discrete games. We show that as the number of firms increases, our method remains computationally feasible, while the computation time needed using existing methods increases exponentially.

We apply our method to study a merger's effects on firm entry and product variety. In antitrust litigation, merging parties often argue that the arrival of new entrants mitigates the increased market power resulting from a merger. One assumption behind this argument is that incumbent firms do not change their product offerings. In our paper, we study the effects of merger by addressing the following questions: Does a merger cause incumbents to add or drop products? Do new firms enter the market after a merger? What is the overall impact of product adjustments and firm entry on welfare? Do any changes in product variety offset or exacerbate the negative price effects on consumer welfare? How does a fixed cost merger efficiency influence the effect of merger on product variety?

The US craft beer industry provides an ideal empirical context to study the effects of merger on the market entry and product variety of multi-product firms. Firstly, craft breweries have recently become popular acquisition targets, and these transactions have drawn the attention of antitrust regulators (Codog, 2018). Secondly, consumer preferences for beer may vary widely, making product variety an important determinant of consumer welfare in this market. Lastly, there are rich demographic variations across geographical markets, which help to identify consumer tastes and firm costs. In our study, we focus on the state of California, which has the highest number of craft breweries and craft beer production among all US states, according to the Brewers Association, a trade group of the craft beer industry.

To address our research questions, we set up a model to describe consumer demand and firm decisions in the retail beer market in California. The demand side is a flexible random coefficient discrete choice model that allows for both observed and unobserved heterogeneity in consumer taste. The supply side is a static two-stage game. In the first stage, each firm is endowed with a set of potential products and chooses the set of products they will sell in a market. If a firm chooses not to enter the market, it chooses the empty set. In the second stage, firms observe demand and marginal cost shocks and simultaneously choose prices.

We use a newly compiled dataset to estimate our model. Our main data sources are the Nielsen Retail Scanner Data and Nielsen Consumer Panel from 2010 to 2016. We supplement these data with information on whether a beer is considered a craft beer based on the designation from the Brewers Association. We further augment our data by hand-collected information on owner identities and brewery locations.

Our demand estimates reveal substantial unobserved heterogeneity in consumer tastes and little substitution between craft and non-craft products. We obtain these estimates by combining standard macro moments with a new set of micro moments based on the panel structure of the consumer survey data. For example, to identify the dispersion in the unobservable heterogeneity in consumer tastes for craft products, we use the following intuition: if the standard deviation is large, then a household's taste for craft products is highly correlated over time. As a result, the expected total purchase of craft beers of a household in a year conditional on the household ever purchasing a craft beer in that year is large. We back out the marginal costs of beers based on the first-order conditions following the standard approach.

We apply our method based on bounds for conditional choice probabilities to estimate the fixed cost of product entry. This method is well suited to our empirical setting, which features many firms and many potential products. Our empirical setting also features rich market- and product/market-level variations resulting in variations in our bounds. Applying our method, we find higher fixed costs of entry for products by independent craft breweries. We also find that both the mean fixed cost of entry and the variance of the fixed cost shock increase with market size.

Using the estimated model, we conduct a counterfactual simulation where the largest macro brewery acquires three large craft breweries. This hypothetical merger case allows us to examine what would happen if the current acquisition trend (i.e., a so-called macro beer firm acquires small craft breweries) continues to the point where the craft beer market becomes as concentrated as the overall beer market. In our simulations, we consider the role of a fixed cost efficiency gain from the merger. Specifically, the merger allows craft breweries to use the acquirer's marketing networks and enjoy lowered fixed costs. Without the fixed cost efficiency, the merged firm drops more products than what other firms add; with the efficiency, the merged firm adds products, leading to an overall increase in product variety. In both scenarios, new firm and product entry are not enough to offset the negative effect of the merger on both consumer surplus and total surplus. The merger efficiency mitigates but does not reverse the overall welfare loss.

In our empirical implementation, we define a firm's set of potential products in a year as all products owned by the firm available in any market in the year. In other words, we focus on a firm's decision to sell an existing product in a market, a decision less costly than new brewery or brand creation. Our results show that even in such a setting that is favorable for firm or product entry, the merger decreases consumer welfare.

Contributions and Literature Review This paper makes two contributions to the literature. First, we develop a method for estimating discrete choice games with many firms and many decisions. Our method differs from existing methods for estimating discrete games, such as Aradillas-Lopez and Tamer (2008) and Ciliberto and Tamer (2009), in the construction of bounds.[1] These papers use bounds defined by the probability that a market-level outcome is a unique equilibrium. For example, the probability that an outcome is an equilibrium is larger than the probability that this outcome is a unique equilibrium. Computing such bounds, therefore, requires enumerating all possible outcomes (e.g., all possible entry outcomes regarding each potential entrant's entry decision) and checking whether each one of them is consistent with the behavioral assumption of the model (e.g., Nash equilibrium). Since the number of possible outcomes increases exponentially with the number of firms, the computational burden may become prohibitively high in settings with many firms. By contrast, our bounds are one-dimensional cumulative distribution functions evaluated at certain cutoffs. Computing these cutoffs does not require solving the full game. Therefore, our

[1] Our bounds and the bounds in these papers are not sharp. See Beresteanu, Molchanov and Molinari (2011), Galichon and Henry (2011), and Chesher and Rosen (2017) for characterizations of the sharp identification region. Other papers that also estimate an incomplete model and exploit the assumption of undominated strategies include, for example, Haile and Tamer (2003) and Barkley, Groeger and Miller (2021) in the auction literature.

method is scalable to settings with many firms and firm decisions. In a contemporaneous paper, Wang (2020) proposes a hybrid approach that replaces one side of the Ciliberto and Tamer (2009) bounds with probability bounds based on the concept of dominant strategies. Such an approach is less computationally intensive than Ciliberto and Tamer (2009), but still requires constructing bounds for all equilibrium outcomes and cannot scale.

Another strand of the literature on estimating discrete games exploits moment inequalities derived from a necessary equilibrium condition that no firm has an incentive to unilaterally deviate from the observed equilibrium. These papers typically rely on a mean-zero assumption of non-structural errors (Ho, 2009; Pakes, Porter, Ho and Ishii, 2015; Wollmann, 2018) or support restrictions (Eizenberg, 2014), and do not estimate the distribution of the structural errors associated with the discrete actions. Our approach estimates the structural error distribution and takes it into account in our counterfactual simulations.

Overall, our approach to estimating discrete games is scalable to large games and has significant advantages when solving for equilibria is costly and when it is important to consider shocks that are known to firms but unobservable to researchers.[2]

Second, we contribute to the literature on merger, entry response, and product variety. A number of papers study entry defense theoretically or through simulations (e.g., Werden and Froeb, 1998; Cabral, 2003; Spector, 2003; Gandhi, Froeb, Tschantz and Werden, 2008; Anderson, Erkal and Piccinin, 2020; Caradonna, Miller and Sheu, 2021). Ciliberto, Murry and Tamer, 2021 empirically analyze merger and entry in the airline industry. We contribute to this strand of the literature by expanding the examination to multi-product firms with endogenous product choice. In our model, because incumbents can reduce product offerings, it is possible for a merger to decrease product variety while inducing new entry.

Another strand within this literature studies how a merger affects product variety and welfare when there is no firm entry (e.g., Fan, 2013; Wollmann, 2018; Fan and Yang, 2020; Garrido, 2020; Li, Mazur, Park, Roberts, Sweeting and Zhang, 2022).[3] We contribute to

[2]There are three other alternative estimation approaches. First, one can obtain a unique equilibrium with additional assumptions and estimate the model via maximum likelihood (Reiss and Spiller, 1989; Garrido, 2020) or a simulated method of moments (Berry, 1992; Li, Mazur, Park, Roberts, Sweeting and Zhang, 2022). Second, Illanes (2017) estimates a dynamic discrete choice problem using a semi-parametric latent variable integration method (Schennach, 2014). This approach also avoids solving a game or an optimization problem, but depends on the availability of certain instruments and, in their absence, can result in relatively wide (and sometimes unbounded) confidence sets of parameters. Third, in their merger simulations, Fan and Yang (2020) make direct assumptions about the distribution of an unobserved fixed cost shock conditional on the observed equilibrium. In comparison, the approach in this paper estimates the distribution.

[3]Several papers (e.g., Berry and Waldfogel, 2001; Sweeting, 2010; Jeziorski, 2015) have studied the effects of merger on firm entry and product variety in the radio industry but do not quantify the impact of mergers on consumer welfare because radio stations do not set prices for their listeners. Mazzeo (2002), Seim (2006) and Draganska, Mazzeo and Seim (2009) also study entry with endogenous product choice but within the context of an incomplete information framework.

this strand of the literature by jointly studying firm entry responses and incumbent product adjustments after a merger to quantify the net changes in product variety.

We also contribute to understanding how merger efficiency shapes the effects of mergers. While Fan (2013) considers cost synergies in operation costs in the newspaper industry and Elliott, Houngbonon, Ivaldi and Scott (2021) study how economies of scale affect product quality and firm investment in the telecommunications industry, this paper examines a merger efficiency in reducing fixed costs and highlights the countervailing effects of such a merger efficiency: on the one hand, the merger efficiency can lead to an increase in product variety by the merging firms; on the other hand, it reduces new firm entry and new product entry because product choices tend to be strategic substitutes. In the end, merger efficiency only mitigates but does not reverse the merger's negative welfare effects.

The rest of the paper is organized as follows. Section 2 explains our estimation method and presents our Monte Carlo simulation results. Section 3 describes the craft beer market in California and our data. Section 4 presents the empirical model. Section 5 explains the estimation procedure and presents the estimation results. Section 6 discusses the counterfactual designs and results. Finally, Section 7 concludes.

2 Discrete Games and Our Estimation Strategy

The estimation of discrete games presents several challenges. First, since there might be multiple equilibria, the maximum likelihood approach may not apply without explicit equilibrium selection rules.[4] Second, a selection issue may complicate a moment inequality approach because the distributions of unobservables conditional on observed actions differ across these actions. We have discussed the existing methods dealing with these issues in the literature review part of the Introduction. In this section, we present our method by starting with a simple model to illustrate how we define our bounds. We then explain our estimation strategy for more general models. We present a set of Monte Carlo experiments to compare our approach to existing methods. We then provide a discussion on when our method is particularly useful. We conclude the section with an extension.

2.1 An Illustrative Model and Our Bounds

To illustrate our bounds, we start with a 2×2 model with two firms where each firm makes a single binary decision. We later extend the model to a setting with more firms where

[4] One exception is that Tamer (2003) considers a maximum likelihood estimator in the presence of multiple equilibria for bivariate games without specifying an equilibrium selection rule.

each firm makes a vector of binary decisions. In this bivariate model, firms 1 and 2 decide whether to enter market m. Let $Y_{nm} = 1$ indicate entry by firm n in market m. If firm n enters, its profit is $\pi_{nm}(Y_{-nm}) - C_{nm} - \zeta_{nm}$, where $\pi_{nm}(Y_{-nm})$ is a variable profit function that depends on the rival action Y_{-nm}, C_{nm} is the fixed cost of entry, and ζ_{nm} is a fixed cost shock observable to firms, which follows a distribution, F_{ζ}. Firm n enters market m if and only if its post-entry profit is positive, i.e.,

$$Y_{nm} = \mathbb{1}\left[\pi_{nm}(Y_{-nm}) - C_{nm} - \zeta_{nm} \geq 0\right]. \tag{1}$$

Our firm behavior assumption is as follows:

Assumption 1. *Y_{nm} is not a dominated strategy for n= 1 or 2.*

In other words, we assume that any observed Y_{nm} is not dominated. This level-1 rationality assumption implies the following bounds for $\Pr(Y_{nm} = 1)$:

$$\begin{aligned} & \Pr(Y_{nm} = 1 \text{ is a dominant strategy}) \\ \leq & \Pr(Y_{nm} = 1) \\ \leq & \Pr(Y_{nm} = 1 \text{ is not a dominated strategy}). \end{aligned} \tag{2}$$

Given that $Y_{nm} = 1$ is a dominant strategy if and only if $\zeta_{nm} < \min\{\pi_{nm}(0), \pi_{nm}(1)\} - C_{nm}$, and that $Y_{nm} = 1$ is not a dominated strategy if and only if $\zeta_{nm} < \max\{\pi_{nm}(0), \pi_{nm}(1)\} - C_{nm}$, it follows from (2) that

$$F_{\zeta}(\min\{\pi_{nm}(0), \pi_{nm}(1)\} - C_{nm}) \leq \Pr(Y_{nm} = 1) \leq F_{\zeta}(\max\{\pi_{nm}(0), \pi_{nm}(1)\} - C_{nm}).$$

Under the assumption that rival entry reduces a firm's profit, the inequality can be further reduced to

$$F_{\zeta}(\pi_{nm}(1) - C_{nm}) \leq \Pr(Y_{nm} = 1) \leq F_{\zeta}(\pi_{nm}(0) - C_{nm}).$$

Our estimation strategy builds on these inequalities. Before we discuss the general model and estimation, we first highlight the advantages of our bounds and compare our bounds with those in the literature.

Advantages of Our Bounds

There are two advantages of using our bounds to estimate discrete games. First, our bounds do not rely on any equilibrium selection assumptions. Specifically, they hold when there are multiple equilibria, when the equilibrium selection mechanisms differ across markets, or when there is no pure strategy equilibrium for some values of fixed cost shocks. Moreover, because our bounds are constructed based on dominant and non-dominated strategies, they are valid under flexible information assumptions (Grieco (2014)). Second, since our bounds are one-dimensional CDFs, they are easy to compute. Therefore, the key advantage of using our bounds to estimate discrete games is that it is computationally feasible even in settings with many firms and when each firm makes multiple binary decisions simultaneously (such as product portfolio decisions).

Our bounds are also intuitive. In a single-agent binary choice model, the inequalities collapse into an equality used in the standard GMM estimator (McFadden, 1989). Thus, our approach can be considered an extension of the GMM estimation of binary choice models to a game setting.

Comparison to Bounds in the Literature

Ciliberto and Tamer (2009) Ciliberto and Tamer (2009) (henceforth, CT) assume the outcomes observed in the data are pure-strategy Nash equilibria and construct bounds for the probability of observing an outcome (Y_{1m}, Y_{2m}), denoted by $\Pr(Y_{1m}, Y_{2m})$, as follows:

$$\begin{aligned} &\Pr\left((Y_{1m}, Y_{2m}) \text{ is a unique pure-strategy Nash equilibrium}\right) \\ \leq &\Pr\left(Y_{1m}, Y_{2m}\right) \\ \leq &\Pr\left((Y_{1m}, Y_{2m}) \text{ is a pure-strategy Nash equilibrium}\right). \end{aligned} \tag{3}$$

The CT bounds are sharper than ours. The intuition is that $Y_{1m} = 1$ being a dominant strategy is a sufficient but not necessary condition for the event that either $(Y_{1m} = 1, Y_{2m} = 1)$ or $(Y_{1m} = 1, Y_{2m} = 0)$ is a unique Nash equilibrium. Therefore, we have

$$\begin{aligned} &\Pr\left(Y_{1m} = 1 \text{ is a dominant strategy}\right) \\ \leq &\Pr\left((Y_{1m} = 1, Y_{2m} = 1) \text{ is a unique pure-strategy Nash equilibrium}\right) \\ + &\Pr\left((Y_{1m} = 1, Y_{2m} = 0) \text{ is a unique pure-strategy Nash equilibrium}\right) \end{aligned}$$

In other words, the CT bounds imply a larger lower bound for $\Pr(Y_{1m} = 1)$ than our lower bound. Similarly, the CT bounds also imply a smaller upper bound than ours.

However, using the CT bounds for estimation in practice can be computationally challenging. First, one needs to obtain the bounds for all possible entry outcomes. With N firms making binary entry decisions, there are 2^N possible outcomes. Second, to simulate the lower bound for each outcome (i.e., the probability that this outcome is a unique equilibrium), one has to draw fixed-cost shocks (ζ_{1m}, ζ_{2m}) and, for each draw, find all equilibria by going over all possible outcomes and verifying whether each of them is an equilibrium. This procedure can be computationally costly, especially when there are many firms, as the number of possible outcomes grows exponentially with the number of firms.

Aradillas-Lopez and Tamer (2008) Similarly to CT, Aradillas-Lopez and Tamer (2008) (henceforth, AT) also construct bounds for $\Pr(Y_{1m}, Y_{2m})$ but consider weaker assumptions than Nash. Here we focus on the AT bounds based on the level-1 rationality assumption. Unlike the comparison between the CT bounds and our bounds, the AT bounds are not necessarily sharper than ours (and vice versa). For example, $Y_{1m} = 1$ being a dominant strategy does not necessarily imply that either $(Y_{1m} = 1, Y_{2m} = 1)$ or $(Y_{1m} = 1, Y_{2m} = 0)$ is a unique model implication according to level-1 rationality. There are scenarios where, although the model implication for (Y_{1m}, Y_{2m}) is not unique, firm 1 always chooses the dominant strategy $Y_{1m} = 1$. Our bounds exploit the uniqueness of a firm's action while the AT bounds do not.[5]

Computationally, one still has to construct bounds for all possible entry outcomes, making it challenging to estimate a discrete game in settings with many firms.

2.2 General Models and Estimation Using Our Bounds

We now describe a general model and explain how to estimate the model using our bounds.

2.2.1 General Model

We consider M markets and N firms, where each firm n makes a vector of binary decisions, $\boldsymbol{Y}_{nm}$, in each market m. For example, firms decide whether to enter a market and, if so, which subset of products from a potential set of products to sell. In this setting, $\boldsymbol{Y}_{nm} = (Y_{jm}, j \in \mathcal{J}_n)$, where $\mathcal{J}_n$ is the set of potential products for firm n, and $Y_{jm} \in \{0, 1\}$ indicates whether firm n sells product j in market m. Each product j is firm-specific.

We use $\boldsymbol{Y}_m = (\boldsymbol{Y}_{nm}, n = 1, ..., N)$ to denote all firm decisions in the market, and $\pi_n (\boldsymbol{Y}_m, X_{nm})$ to denote the variable profit function of firm n, which depends on $\boldsymbol{Y}_m$ as well as a set of observable covariates, X_{nm}. For example, X_{nm} may include the characteristics of all products in the market, features of firm n, and the demand conditions in market m. We further assume that there is a cost associated with choosing $Y_{jm} = 1$. This cost is $c (W_{jm}, \theta) + \zeta_{jm}$, where W_{jm} is a vector of exogenous covariates. The unobserved cost shock, ζ_{jm}, is assumed to be i.i.d. and follows the distribution $F_\zeta(\cdot, \sigma_\zeta)$.

[5]See Supplemental Appendix SA for a detailed comparison between our bounds and the AT bounds as well as for a graphic illustration of the comparison between our bounds and the CT bounds.

The parameters to be estimated include the coefficients θ and the distribution parameters σ_ζ in the fixed cost. Researchers observe (Y_{jm}, W_{jm}, X_{nm}), but not the fixed cost shock ζ_{jm}. The variable profit function $\pi_n(\boldsymbol{Y}_m, X_{nm})$ is either known or has been estimated. For example, one could follow the standard literature in Industrial Organization to estimate demand and marginal costs and then obtain the estimated variable profit at a Nash-Bertrand equilibrium for any given $\boldsymbol{Y}_m$.

We define the change in a firm's variable profit when Y_{jm} turns from 0 to 1:

$$\varDelta_j(\boldsymbol{Y}_{-jm}, X_{nm}) = \pi_n(Y_{jm} = 1, \boldsymbol{Y}_{-jm}, X_{nm}) - \pi_n(Y_{jm} = 0, \boldsymbol{Y}_{-jm}, X_{nm}), \tag{4}$$

where $\boldsymbol{Y}_{-jm} = (Y_{j'm}, j' \in \cup_{n'} \mathcal{J}_{n'}, j' \neq j)$. Given the discrete nature of $\boldsymbol{Y}_{-jm}$, the following minimum and maximum changes in variable profits exist: $\underline{\varDelta}_j(X_{nm}) = \min_{\boldsymbol{Y}_{-jm}} \varDelta_j(\boldsymbol{Y}_{-jm}, X_{nm})$ and $\overline{\varDelta}_j(X_{nm}) = \max_{\boldsymbol{Y}_{-jm}} \varDelta_j(\boldsymbol{Y}_{-jm}, X_{nm})$. Since $\pi_n(\boldsymbol{Y}_m, X_{nm})$ is either known or estimated, so are $\underline{\varDelta}_j(X_{nm})$ and $\overline{\varDelta}_j(X_{nm})$.

Following the discussion in the previous section, we can see that the bounds of the conditional probability $Y_{jm} = 1$ given X_{nm} and W_{jm} are:

$$\begin{aligned} & F_\zeta\left(\underline{\varDelta}_j(X_{nm}) - c(W_{jm}, \theta), \sigma_\zeta\right) \\ & \leq \Pr(Y_{jm} = 1 \,|X_{nm}, W_{jm}) \\ & \leq F_\zeta\left(\overline{\varDelta}_j(X_{nm}) - c(W_{jm}, \theta), \sigma_\zeta\right). \end{aligned} \tag{5}$$

We define the following moment functions:

$$\begin{aligned} L(Y_{jm}, X_{nm}, W_{jm}, \theta, \sigma_\zeta) &= F_\zeta\left(\underline{\varDelta}_j(X_{nm}) - c(W_{jm}, \theta), \sigma_\zeta\right) - \mathbb{1}(Y_{jm} = 1), \\ H(Y_{jm}, X_{nm}, W_{jm}, \theta, \sigma_\zeta) &= \mathbb{1}(Y_{jm} = 1) - F_\zeta\left(\overline{\varDelta}_j(X_{nm}) - c(W_{jm}, \theta), \sigma_\zeta\right). \end{aligned} \tag{6}$$

The inequalities in (5) imply the following conditional moment inequalities:

$$\begin{aligned} E\left(L(Y_{jm}, X_{nm}, W_{jm}, \theta, \sigma_\zeta) \,|X_{nm}, W_{jm}\right) \leq 0, \\ E\left(H(Y_{jm}, X_{nm}, W_{jm}, \theta, \sigma_\zeta) \,|X_{nm}, W_{jm}\right) \leq 0. \end{aligned} \tag{7}$$

2.2.2 Identification

The identification of (θ, σ_ζ) based on the inequalities in (5) is similar to the idea of special regressors in entry games (Ciliberto and Tamer, 2009; Lewbel, 2019). To identify our parameters, we exploit exogenous variations in X_{nm} and W_{jm}. For example, to identify the coefficient in θ that corresponds to an indicator variable, we compare the entry probability conditional on this indicator being 1 versus 0, holding other covariates fixed. Similarly, to identify the coefficient of a continuous variable, we examine how entry probability varies across different ranges of the continuous variable. Exogenous variations in X_{nm} and W_{jm} are also helpful for identifying the distribution parameters σ_ζ. For example, consider a distribution of ζ_{jm} that is fully specified by its variance. If the variance is large, the upper and lower bounds will show little co-variance with the covariates. In the special case where ζ_{jm} follows a symmetric distribution, both bounds in (5) approach 0.5 as the variance increases. On the other hand, when the variance is close to 0, both bounds are close to 0 if $\overline{\Delta}_j(X_{nm}) - c(W_{jm}, \theta) < 0$ or close to 1 if $\underline{\Delta}_j(X_{nm}) - c(W_{jm}, \theta) > 0$. Therefore, if the variance is small, the model predicts large jumps in entry probabilities even with small changes in the covariates. Both the lack of sensitivity of entry probability to covariates (in the case of a large variance of ζ_{jm}) and the high sensitivity (in the case of a small variance) can be tested by data.

2.2.3 Estimation

Moment Inequalities Following the literature on conditional moment inequalities, we transform the conditional moment inequalities in (7) into unconditional ones:

$$
\begin{aligned}
&E\left(\frac{1}{\#\mathcal{J}}\sum_{j\in\mathcal{J}} L\left(Y_{jm}, X_{nm}, W_{jm}, \theta, \sigma_\zeta\right)\cdot g^{(k)}\left(X_{nm}, W_{jm}\right)\right) \le 0, \qquad (8)\\
&E\left(\frac{1}{\#\mathcal{J}}\sum_{j\in\mathcal{J}} H\left(Y_{jm}, X_{nm}, W_{jm}, \theta, \sigma_\zeta\right)\cdot g^{(k)}\left(X_{nm}, W_{jm}\right)\right) \le 0.
\end{aligned}
$$

Inside the expectation in (8), we average over potential products because even when the fixed cost shock ζ_{jm} is independent across j, the entry decisions Y_{jm} across j within the same market m may be correlated due to strategic interdependence among firms. However, the entry decisions $\boldsymbol{Y}_m$ are independent across markets.[6] The functions $g^{(k)}(X_{nm}, W_{jm}), k = 1, ..., K$ are non-negative and capture information contained in the conditioning variables. We define these functions below.

Approximate $\underline{\Delta}_j(X_{nm})$ and $\overline{\Delta}_j(X_{nm})$ To compute the moment functions, we need to compute $\underline{\Delta}_j(X_{nm}) = \min_{\boldsymbol{Y}_{-jm}} \Delta_j(\boldsymbol{Y}_{-jm}, X_{nm})$ and $\overline{\Delta}_j(X_{nm}) = \max_{\boldsymbol{Y}_{-jm}} \Delta_j(\boldsymbol{Y}_{-jm}, X_{nm})$, where $\Delta_j(\boldsymbol{Y}_{-jm}, X_{nm})$ is the change in firm n's expected variable profit when product j joins the market, as defined in equation (4). Directly solving for the minimum and maximum of

[6] Averaging over potential products is not strictly necessary. One could account for the correlation among observations indexed by jm and adjust the estimate of the covariance matrix of the moments accordingly.

the expected profit across all possible values of $\boldsymbol{Y}_{-jm}$ may be computationally prohibitive because there are $2^{(\text{length of } \boldsymbol{Y}_{-jm})}$ possible values of $\boldsymbol{Y}_{-jm}$ and for each value of $\boldsymbol{Y}_{-jm}$, one needs to solve a pricing game for multiple simulated draws of demand and marginal cost shocks in order to compute $\varDelta_j(\boldsymbol{Y}_{-jm}, X_{nm})$. Economic intuition suggests that, because products are substitutes, we can approximate the minimum and maximum by, respectively, the following:

$$\underline{\varDelta}_j(X_{nm}) \approx \varDelta_j((1,...,1), X_{nm}) \text{ and } \overline{\varDelta}_j(X_{nm}) \approx \varDelta_j((0,...,0), X_{nm}).$$

These approximate extrema are exact for entry games such as those in Berry (1992), Seim (2006), Ciliberto and Tamer (2009), Sweeting (2013), and Berry, Eizenberg and Waldfogel (2016). For more general demand and pricing models such as ours in the empirical part of the paper, we find that the approximate extrema coincide with the true values in all of our computational experiments.[7]

Non-negative Functions We define our non-negative functions $g^{(k)}(X_{nm}, W_{jm})$ in (8) as functions of $\underline{\varDelta}_j(X_{nm})$, $\overline{\varDelta}_j(X_{nm})$, and W_{jm}. In practice, we have found $\underline{\varDelta}_j(X_{nm})$ and $\overline{\varDelta}_j(X_{nm})$ to be informative summary statistics of the long vector of covariates capturing market structure and the characteristics of each potential product. The vector W_{jm} includes covariates in the fixed cost function. The extrema $\underline{\varDelta}_j(X_{nm})$ and $\overline{\varDelta}_j(X_{nm})$ are continuous variables while W_{jm} can include both continuous and discrete variables.

We first consider a scenario where W_{jm} consists of an indicator variable. We specify a series of cutoffs $\{b_l : l = 1, ..., L\}$. For each cutoff b_l, we define the following functions:

$$\begin{aligned} &\mathbb{1}\left(\underline{\varDelta}_j(X_{nm}) > b_l\right) \cdot W_{jm}, \\ &\mathbb{1}\left(\overline{\varDelta}_j(X_{nm}) < b_l\right) \cdot W_{jm}, \\ &\mathbb{1}\left(b_l < \underline{\varDelta}_j(X_{nm}) < \overline{\varDelta}_j(X_{nm}) < b_{l'}\right) \cdot W_{jm}, \end{aligned} \tag{9}$$

where $l' = l+1, ..., L$. We define another set of functions by replacing W_{jm} in (9) by $1-W_{jm}$. Our non-negative functions $g^{(k)}(X_{nm}, W_{jm})$ include all these functions.

When W_{jm} contains more than one indicator variables, we repeat the above process for each indicator variable and include these additional functions in $g^{(k)}(X_{nm}, W_{jm})$. When W_{jm}

[7] We randomly sample 100 markets and H potential products from our sample. For each selected potential product and each selected market, we hold fixed the entry outcomes of the products not in the H selected products, and go over all possible outcomes for the other $H-1$ products to find the actual extrema. For all sampled markets and H products when H takes the value of 6, 8, or 10, we find that the approximations coincide with the actual extrema.

contains a discrete variable that takes value from a finite set, we replace W_{jm} in (9) by an indicator of whether the value of this variable equals to each of the values in the finite set. When W_{jm} contains a continuous variable, we define a set of cutoffs for this continuous variable and replace W_{jm} in (9) by an indicator of whether the value of the variable falls into a bin defined by the cutoffs.

Inference We construct the confidence set for (θ, σ_ζ) by inverting the test in Andrews and Soares (2010). In Appendix A, we provide details on the construction of our confidence set and a step-by-step guide on the calculation of the confidence set.

The inference procedure takes as input the variable profit function $\pi_n(\boldsymbol{Y}_m, X_{nm})$ and thus the extrema of the change in variable profits $\underline{\Delta}_j(X_{nm})$ and $\overline{\Delta}_j(X_{nm})$. In empirical studies, the variable profit function may be computed based on estimates of demand and marginal costs. In Appendix A, we also explain how to adjust the confidence set when $\pi_n(\boldsymbol{Y}_m, X_{nm})$ depends on these separately estimated parameters.

2.3 Monte Carlo Experiments

We use Monte Carlo experiments to compare both the performance and the computational burden of our method with those in the literature.

2.3.1 Monte Carlo Experiment Setup

In our Monte Carlo experiments, we consider an entry game similar to that in Ciliberto and Tamer (2009). Specifically, there are N potential entrants and each firm n makes a binary decision $Y_{nm} \in \{0,1\}$, where $Y_{nm} = 1$ represents entering market m. If firm n enters, its variable profit is:

$$\pi_n(\boldsymbol{Y}_{-nm}, X_{nm}) = O_m\Big(x_{nm} - \phi \log(1 + \sum_{n' \neq n} z_{n'm} Y_{n'm})\Big),$$

where $\boldsymbol{Y}_{-nm} = (Y_{n'm}, n' \neq n)$ denotes rival firms' entry decisions, O_m represents a market-level profit shifter, x_{nm} is a firm-level profit shifter, and z_{nm} captures the competitive effect of firm n on other firms. The logarithm functional form follows Berry (1992), where the competitive effects are homogeneous across firms (i.e., $z_{n'm} = 1$) and a firm's profit depends on $\log(1 + \sum_{n' \neq n} Y_{n'm})$. The parameter ϕ governs the magnitude of the competitive effects. We collect all covariates in $X_{nm} = (O_m, x_{nm}, \{z_{n'm}\}_{n' \neq n})$.

The fixed cost of entry is $C + \sigma\zeta_{nm}$, where the unobservable cost shock ζ_{nm} is assumed to be a standard normal random variable and i.i.d. across both firms and markets. The mean fixed cost parameter C and the standard deviation σ are the parameters to be estimated. We set the true values to be $C = \sigma = 1$.

For each Monte Carlo experiment, we simulate 300 data sets and each data set consists of 4000 markets. To simulate a data set, we draw X_{nm} for each firm and each market. Specifically, we draw O_m uniformly between 1 and 2, x_{nm} uniformly between 0 and 1, and z_{nm} uniformly between 0 and 0.5. We draw ζ_{nm} from the standard normal distribution. We compute the Nash equilibrium entry outcome for each market. In the case of multiple equilibria, an equilibrium is selected at random with equal probability.

Since the profit function π_n decreases in the entry decision of a firm's rivals, we have

$$\min_{\boldsymbol{Y}_{-nm}} \pi_n(\boldsymbol{Y}_{-nm}, X_{nm}) = \pi_n((1, ..., 1), X_{nm}) = O_m\Big(x_{nm} - \phi \log(1 + \textstyle\sum_{n' \neq n} z_{n'm})\Big),$$
$$\max_{\boldsymbol{Y}_{-nm}} \pi_n(\boldsymbol{Y}_{-nm}, X_{nm}) = \pi_n((0, ..., 0), X_{nm}) = O_m x_{nm}.$$

Based on (5), we use the following bounds in our estimation

$$\Phi\left(\frac{1}{\sigma}\Big(O_m\Big(x_{nm}-\phi \log(1+\sum_{n' \neq n} z_{n'm})\Big)-C\Big)\right) \leq \Pr(Y_{nm} = 1|X_{nm}) \leq \Phi\Big(\frac{1}{\sigma}(O_m x_{nm}-C)\Big), \quad (10)$$

where $\Phi(\cdot)$ is the standard normal distribution function.

2.3.2 Monte Carlo Experiment Results

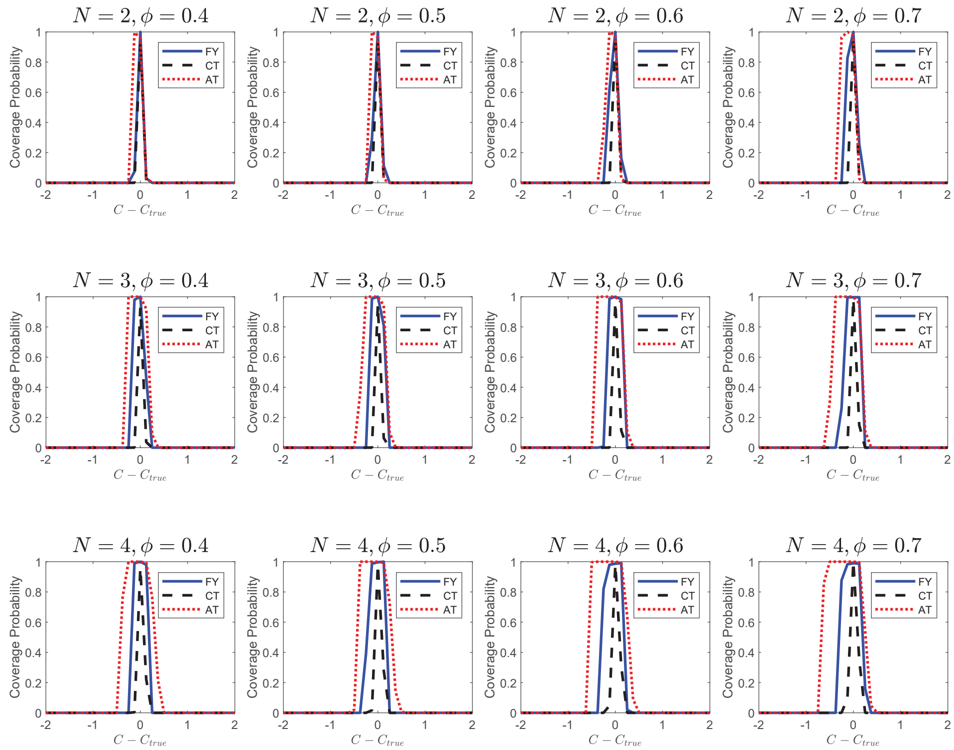
We first present the coverage probability of our 95% confidence set containing parameter values of C and σ in the neighborhood of the true parameter value (i.e., $(C, \sigma) = (1, 1)$). Specifically, for visibility, we plot the coverage probabilities for C and σ separately. For instance, for the coverage probability of $C = 0.8$ (i.e. $C - C_{\text{true}} = -0.2$), we report the fraction of the 500 data sets where the 95% confidence set corresponding to this data set contains $(0.8, \sigma)$ for some value of σ. We do so for candidate values of C from -1 to 3 and σ from $\exp(-1)$ to $\exp(3)$.

Figure 1 compares the coverage probabilities, for C in panel (A) and σ in panel (B), using our bounds versus the CT and AT bounds. Each row in these two panels correspond to a different number of potential entrants $N = 2, ..., 4$, and each column corresponds to a different value for the competitive-effect parameter $\phi = 0.4, 0.5, 0.6$, and 0.7. From the lower and upper bounds in (10), we can see that the gap between the bounds depends on both the number of firms N and the competitive parameter ϕ. Therefore, as we vary N and ϕ, we tighten or widen our bounds. Specifically, as N or ϕ increases, our bounds become wider.

We have three findings from these comparisons in Figure 1. First, using all three bounds

Figure 1: Coverage Probabilities: FY vs. CT vs. AT Bounds

(A) C



(B) σ

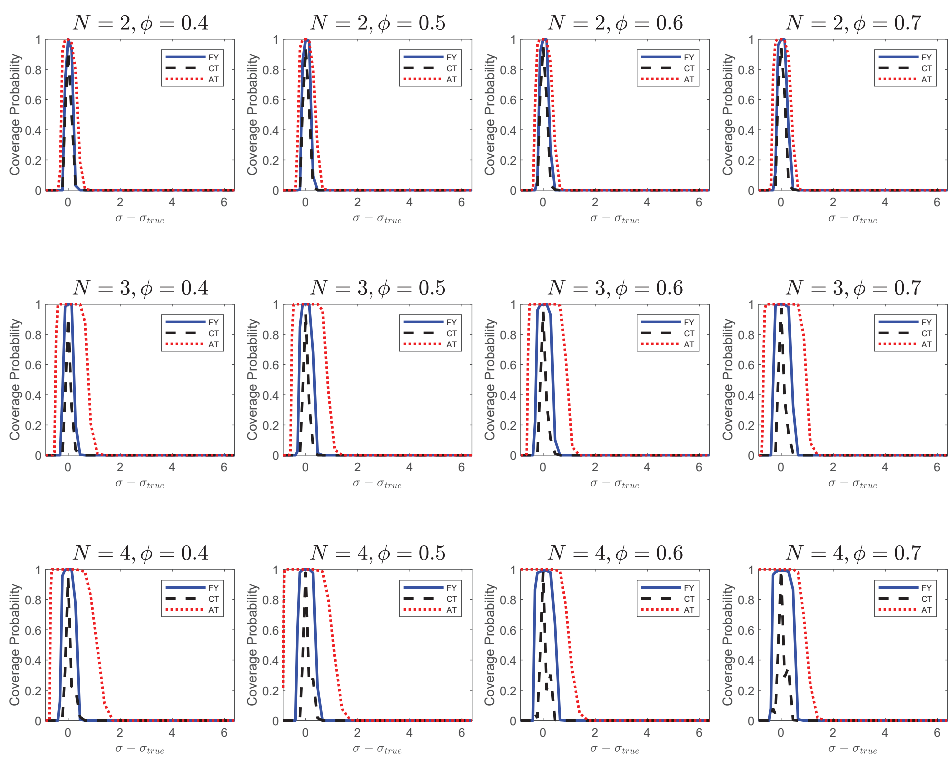


Table 1: Computation Time: FY vs. CT vs. AT Bounds

N (#Potential Entrants)	FY (s)	CT (s)	AT (s)
2	0.017	0.341	0.017
3	0.014	0.739	0.084
4	0.022	1.930	0.452
5	0.023	5.580	1.682
6	0.021	21.743	13.891
7	0.016	151.646	132.086
8	0.126	1459.528	1421.882

Note: the table reports the computation time required to evaluate the test statistic used for constructing the confidence set once. The three columns correspond to our bounds (FY), the Ciliberto and Tamer (2009) (CT) bounds, and the Aradillas-Lopez and Tamer (2008) (AT) bounds.

(i.e., our bounds – labeled FY, the CT bounds, and the AT bounds), the coverage probability decreases for parameter values further away from the true value. Second, FY's coverage probability is higher than CT's, but lower than AT's. This finding is consistent with our discussion on the comparison of our bounds to the CT and AT bounds in Section 2.1, i.e., while the CT bounds are sharper than ours, there is no clear ranking between our bounds and the AT bounds. Third, FY's and AT's coverage probabilities increase as N or ϕ increases. This result is expected because the gap between bounds in both estimators increases in N and ϕ. However, even for the largest N and ϕ, FY's coverage probability decreases quickly for parameters away from the true values.

Turning to the computational burden of the three methods, we report the time needed to evaluate the test statistic used for constructing the confidence set. Specifically, we compare the time needed to compute the test statistic once using our bounds versus the CT and AT bounds in Table 1.[8]

Table 1 shows that our method's computational advantage grows exponentially with the number of potential entrants in a game. Across the three columns of the table, the computation time of FY bounds is consistently smaller than that of either CT or AT bounds. Across the rows, the computation time of FY bounds is relatively stable as the number of potential entrants (N) increases. In contrast, the computation time using the CT and AT bounds increases from less than 1 second to more than 23 minutes. The FY bounds are scalable because they only require evaluating one-dimensional CDFs. In contrast, CT and AT bounds require enumerating all possible market outcomes, the number of which grows exponentially as the number of potential entrants increases.

[8] The results are computed on a cluster using Intel Xeon Gold 6154 processors (2x 3.0 GHz). We use 500 simulation draws of (ζ_{1m}, ζ_{2m}) to simulate the CT bounds.

2.4 Discussion

In this section, we provide a further discussion on the computational advantage of our method and the tightness of our bounds. With many firms, FY bounds' computational advantage is greater, but so is the gap between the bounds. We discuss the trade-off between the computational savings and the identifying power of our bounds below.

First, it is worth noting that as the number of firms increases, while the computational advantage increases exponentially, we expect the gap between our bounds to increase at a lower, possibly diminishing, rate. For example, the profit difference between a duopolist and a triopolist is typically smaller than that between a monopolist and a duopolist.[9] As a result, having a large number of firms does not necessarily mean our bounds are too wide to be useful.

Moreover, the tightness of the bounds also depends on the nature of competition between firms and is ultimately an empirical question. Generally, when products are more heterogeneous, firm profit is less sensitive to the entry of a competitor. As a result, our bounds are tighter. Therefore, regardless of the number of firms and products, we expect the identifying power of our bounds to be stronger in settings where there exist a sizable group of products that are quite differentiated from the rest of the products.

2.5 Extension

So far, we have assumed that the unobserved cost shock is i.i.d. In this section, we extend the model to allow for correlated shocks. We first explain how an additional set of bounds can be constructed and used to identify the correlation using the illustrative model in Section 2.1. We then extend the general model in Section 2.2 and explain the estimation details. Finally, we show the results of the Monte Carlo experiment for this extension.

2.5.1 Extension to the Illustrative Model and Additional Bounds

We first extend the illustrative model in Section 2.1 and allow unobserved cost shocks ζ_{1m} and ζ_{2m} to be correlated. To estimate the correlation in unobserved cost shocks, we consider the bounds for the probability of $\Pr(Y_{1m} = 1, Y_{2m} = 1)$. The level-1 rationality assumption

[9] See Bresnahan and Reiss (1991) for an early work establishing this result. Berry (1992) specifies and estimates a profit function linear in the logarithm of the number of firms to capture the diminishing competition effect as the number of firms increases.

implies the following:

$$
\begin{aligned}
&\Pr\left(Y_{nm}=1 \text{ is a dominant strategy for both } n=1 \text{ and } n=2\right)\\
\leq &\Pr\left(Y_{1m}=1, Y_{2m}=1\right)\\
\leq &\Pr\left(Y_{nm}=1 \text{ is not a dominant strategy for either } n=1 \text{ or } n=2\right).
\end{aligned}
$$

To save notation, we define $\overline{\pi}_{nm}=\max\left(\pi_{nm}\left(0\right),\pi_{nm}\left(1\right)\right)$ and $\underline{\pi}_{nm}=\min\left(\pi_{nm}\left(0\right),\pi_{nm}\left(1\right)\right)$. The bounds above can be expressed as

$$
\begin{aligned}
&\Pr\left(\zeta_{1m}<\underline{\pi}_{1m}-C_{1m},\ \zeta_{2m}<\underline{\pi}_{2m}-C_{2m}\right)\\
\leq &\Pr\left(Y_{1m}=1, Y_{2m}=1\right)\\
\leq &\Pr\left(\zeta_{1m}<\overline{\pi}_{1m}-C_{1m},\ \zeta_{2m}<\overline{\pi}_{2m}-C_{2m}\right),
\end{aligned}
\tag{11}
$$

which can be further rewritten as

$$
\begin{aligned}
&\Pr\left(\zeta_{1m}<\underline{\pi}_{1m}-C_{1m}\left|\zeta_{2m}<\underline{\pi}_{2m}-C_{2m}\right.\right)\cdot\Pr\left(\zeta_{2m}<\underline{\pi}_{2m}-C_{2m}\right)\\
\leq &\Pr\left(Y_{1m}=1, Y_{2m}=1\right)\\
\leq &\Pr\left(\zeta_{1m}<\overline{\pi}_{1m}-C_{1m}\left|\zeta_{2m}<\overline{\pi}_{2m}-C_{2m}\right.\right)\cdot\Pr\left(\zeta_{2m}<\overline{\pi}_{2m}-C_{2m}\right).
\end{aligned}
$$

As the correlation between ζ_{1m} and ζ_{2m} increases, the conditional probability $\Pr(\zeta_{1m}<\underline{\pi}_{1m}-C_{1m}\left|\zeta_{2m}<\underline{\pi}_{2m}-C_{2m}\right.)$ increases, making the lower bound more likely to be violated. Conversely, as the correlation decreases, $\Pr\left(\zeta_{1m}<\overline{\pi}_{1m}-C_{1m}\left|\zeta_{2m}<\overline{\pi}_{2m}-C_{2m}\right.\right)$ decreases, making the upper bound more likely to be violated. Therefore, these bounds are informative about the correlation.

2.5.2 Extension to the General Model and Estimation with the Additional Bounds

We now extend the general model in 2.2 and allow the unobservable fixed cost shock ζ_{jm} to be correlated. We continue to use σ_ζ to denote the parameters governing their joint distribution.

For each pair of potential products (j,j'), the analogy of the inequalities in (11) is:

$$
\begin{aligned}
&\Pr\left(\zeta_{jm}<\underline{\Delta}_j(X_{nm})-c(W_{jm},\theta),\ \zeta_{j'm}<\underline{\Delta}_{j'}(X_{nm})-c(W_{j'm},\theta);\sigma_\zeta\right)\\
\leq &\Pr\left(Y_{jm}=1, Y_{j'm}=1\left|X_{nm},W_{jm},W_{j'm}\right.\right)\\
\leq &\Pr\left(\zeta_{jm}<\overline{\Delta}_j(X_{nm})-c(W_{jm},\theta),\ \zeta_{j'm}<\overline{\Delta}_{j'}(X_{nm})-c(W_{j'm},\theta);\sigma_\zeta\right).
\end{aligned}
$$

The resulting new moment functions are

$$
\begin{aligned}
& L\left(Y_{jm}, Y_{j'm}, X_{nm}, W_{jm}, W_{j'm}, \theta, \sigma_\zeta\right) \\
& = \Pr\left(\zeta_{jm} < \underline{\Delta}_j(X_{nm}) - c(W_{jm}, \theta),\ \zeta_{j'm} < \underline{\Delta}_{j'}(X_{nm}) - c(W_{j'm}, \theta); \sigma_\zeta\right) \\
& - \mathbb{1}\left(Y_{jm} = 1, Y_{j'm} = 1\right), \\
& H\left(Y_{jm}, Y_{j'm}, X_{nm}, W_{jm}, W_{j'm}, \theta, \sigma_\zeta\right) \\
& = \mathbb{1}\left(Y_{jm} = 1, Y_{j'm} = 1\right) \\
& - \Pr\left(\zeta_{jm} < \overline{\Delta}_j(X_{nm}) - c(W_{jm}, \theta),\ \zeta_{j'm} < \overline{\Delta}_{j'}(X_{nm}) - c(W_{j'm}, \theta); \sigma_\zeta\right).
\end{aligned}
$$

In estimation, we combine the moments based on a single product's entry probability in (8) with the following additional moments:

$$
E\left[\frac{1}{\frac{1}{2}\#\mathcal{J}\left(\#\mathcal{J}-1\right)} \sum_{\{(j,j'):j,j'\in\mathcal{J}, j\neq j'\}} L\left(Y_{jm}, Y_{j'm}, X_{nm}, W_{jm}, W_{j'm}, \theta, \sigma_\zeta\right) \cdot g^{(k)}\left(X_{nm}, W_{jm}\right)\right] \leq 0,
$$

$$
E\left[\frac{1}{\frac{1}{2}\#\mathcal{J}\left(\#\mathcal{J}-1\right)} \sum_{\{(j,j'):j,j'\in\mathcal{J}, j\neq j'\}} H\left(Y_{jm}, Y_{j'm}, X_{nm}, W_{jm}, W_{j'm}, \theta, \sigma_\zeta\right) \cdot g^{(k)}\left(X_{nm}, W_{jm}\right)\right] \leq 0,
$$

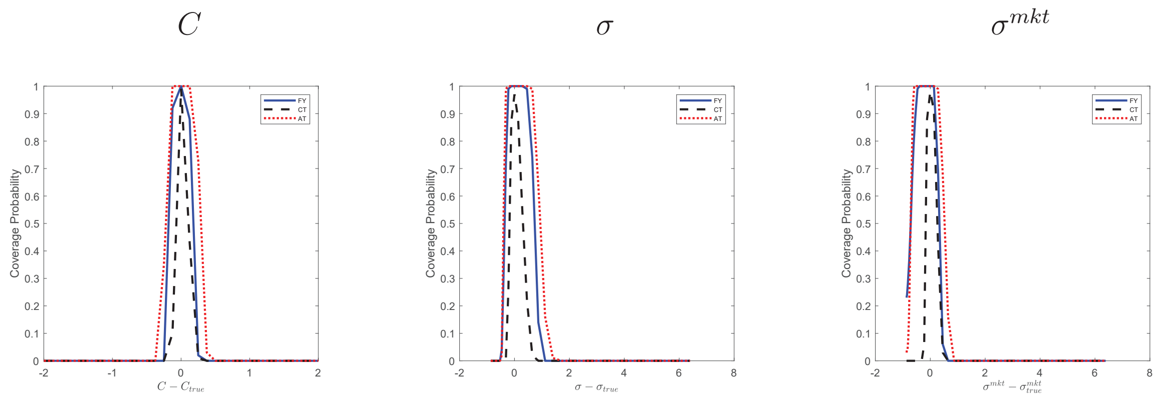
as well as the above moments where $g^{(k)}\left(X_{nm}, W_{jm}\right)$ is replaced by $g^{(k)}\left(X_{nm}, W_{j'm}\right)$.

One could also additionally consider analogous inequalities for triplets or quadruplets of potential products.

2.5.3 Extension to the Monte Carlo Experiment

We extend the setup of the Monte Carlo in Section 2.3 by adding a market-level fixed cost shock. Specifically, the fixed cost of entry is now $C + \sigma\zeta_{nm} + \sigma^{mkt}\zeta_m$, where ζ_m represents a common market-level fixed cost shock. We set the true values to be $\sigma^{mkt} = 1$. Figure 2 shows the coverage probability for parameter values in the neighborhood of the true parameter values for C, σ, and σ^{mkt}. Again, FY's coverage probability is slightly larger than CT's but smaller than AT's.

Figure 2: Coverage Probabilities: FY vs. CT vs. AT Bounds, Allowing for a Market-level Fixed Cost Shock



3 Empirical Background and Data

We apply our method to study how a merger can affect firm entry and product variety in the retail craft beer market in California.[10] This setting features many firms and many products. For example, there are 26 breweries producing 95 craft products in the sample in 2016. As a result, while existing methods may be computationally prohibitive, our method is well suited.

Our analysis is based on a new dataset that we compiled from various sources. The primary datasets are the market-level data in the Nielsen Retail Scanner Data and the micro-level panel data in the Nielsen Consumer Panel between 2010 and 2016. We define a product to be a brand in the Nielsen data (e.g., Samuel Adams Boston Lager). We aggregate the Nielsen scanner data from its original UPC/week level to a product/month-level dataset by homogenizing the size of a product (a unit represents a 12-ounce-12-pack equivalent), adding quantities across weeks within a month, and using the quantity-weighted average price across weeks within a month as a given product's price in a given month. We then supplement the dataset with information on whether a beer is considered a craft beer based on the designation by the Brewers Association. We also add hand-collected data on the identities of the corporate owner and the brewery as well as the location of the production facility for each product in our dataset. For example, Samuel Adams Boston Lager is produced at the Samuel Adams Boston Brewery in Boston and owned by the Boston Beer Company. We

[10]There has been growing interest in the market structure of the craft beer industry. For example, Tremblay, Iwasaki and Tremblay (2005) document the entry of microbreweries in the US. Elzinga and McGlothlin (2021) analyze a macro brewery's acquisition of a craft brewery. Bronnenberg, Dubé and Joo (2022) study the formation of preferences for craft beer and its implications for the future market structure of the industry. California accounted for 18% of craft beer volume and 12% of craft breweries in the nation, the highest among all US states according to the 2015 Brewers Association estimates.

define a firm as a corporate owner (e.g., Boston Beer Company). A firm can own multiple breweries and products. Finally, we merge the data with county demographics obtained from the US Census.

In our analysis, we define a market as a retailer-county pair. The Nielsen consumer panel data suggest that cross-retailer shopping is rare. For example, we find that over 80% of households in our study purchase all of their beer from one retailer-county combination in 2016. This finding is consistent with those of Illanes and Moshary (2020) and Huang, Ellickson and Lovett (2022), who find little evidence of retailer competition in the spirits category. In our estimation, we define market size as the average monthly alcohol sales in a market (in the unit of a 12-ounce-12-pack equivalent) multiplied by 8, which is the average number of household trips per month in the panel data.[11]

We consider a product to be available in a market in a calendar year if the product's monthly sales are more than 20 units for more than 6 months in the market in the year. Moreover, for craft products, we keep those produced by the top 60 craft breweries according to their national volumes in the 2015 Brewers Association production data. We thus focus on breweries established in the 1990s or earlier. In the end, our sample covers 91% of California's craft beer quantity in the Nielsen Scanner Data across our sample periods.[12]

We define a firm's set of potential products in a year as all products owned by the firm that are available in any market in the year. We do not consider the creation of new breweries or brands. We focus on a firm's decision to sell an existing product in a market, a decision that is less costly than *de novo* entry.[13] Therefore, our setting can be considered favorable for firm or product entry. As we see later, even in this favorable setting, a merger can reduce consumer welfare.

Table 2 reports summary statistics based on 110 markets present in the data every year from 2010 to 2016. These markets account for 82% of the total quantity from all markets and years. The table shows that the annual craft beer sales in the sample are, on average, about 5 million units, which accounts for roughly 10% of total beer sales for a given year. The average price for craft beer is around 17 dollars per unit in 2016 dollars, which is higher than the average beer price of 11 dollars per unit. Although craft beer sales account for only

[11] Our results are robust to alternative scaling factors.

[12] Although our retail data precludes a direct comparison of the retail beer market with the "on-premises" market (such as taprooms, bars, and restaurants), the Brewers Association estimates that the retail channel accounts for 65% of craft beer volume (Watson, 2016). Likely due to similar data limitations, existing research on the beer industry has also focused on the retail segment (Ashenfelter, Hosken and Weinberg, 2015; Asker, 2016; Miller and Weinberg, 2017; Miller, Sheu and Weinberg, 2021; Döpper, MacKay, Miller and Stiebale, 2022; Hidalgo, 2023).

[13] We have limited information on product characteristics. Moreover, many features of a beer are difficult to quantify (e.g., aroma and mouthfeel). As a result, we include product fixed effects in our demand and marginal cost specifications.

Table 2: Annual Total Quantity, Prices, and Numbers of Firms and Products

	Total Quantity (12-oz 12-pk equiv)	Avg. Price (2016 $)	# Firms Per Year	# Products Per Year
Craft	4,710,634	17	35	129
All	53,155,825	11	54	263

Note: for each year from 2010 to 2016, we first calculate a year's total and craft beer quantities, quantity-weighted average prices, number of firms and number of products, and then take the average across years.

Table 3: Shares of Total Quantity and Number of Products by Beer Types

		Ale	Lager	Light
Quantity				
	Craft	69.10%	29.34%	0.71%
	All	11.95%	46.07%	40.94%
Number of products				
	Craft	65.82%	27.44%	0.50%
	All	43.11%	40.38%	7.60%

Note: the shares reflect the respective proportions of the total quantity from 2010 to 2016, or of the total number of unique products in these years.

10% of the total market, the number of craft firms and products make up more than half of the market.

Table 3 provides a breakdown of the sales and number of products by beer types. Among craft products, ales constitute 66% of the product counts and 69% of sales. Lagers account for 29% and 46% of the craft and overall beer market share, respectively. While light beers account for 41% of the overall beer sales, their market share within craft products is only 0.71%, and this pattern has remained stable over time.

A key primitive in the product variety decisions is the fixed cost of product entry. According to our interviews with industry experts,[14] the main cost of product entry is a flow cost of the marketing support that a firm needs to provide to a retailer in a local market. By contrast, the sunk cost of convincing a retailer to carry a brewery's products or contracting with a distributor seems negligible compared to the fixed cost of marketing support. In our setting, it is illegal at both the federal and state level and extremely rare for grocers or distributors to charge slotting fees. California has its own tied-house laws that expand on federal statues prohibiting "tied-houses", which refer to vertical relationships between manufacturers and retailers that exclude small manufacturers such as craft breweries from placing their products with retailers. In addition, California passed competition laws that further prohibit payments for stocking products (Croxall, 2019).

[14]They are the Chief Economist, Bart Watson, and the General Counsel, Marc Sorini, at the Brewers Association.

4 Empirical Model

4.1 Demand

We use a random coefficient discrete choice model to describe consumer demand for beer. A product's characteristics include its flavor type (ale, lager, light, and others),[15] whether it is imported from outside North America, and whether it is designated as a craft product. These characteristics can overlap. For example, Bud Light is a light, North American, non-craft beer, while Samuel Adams Lager is a lager, North American, craft beer. These characteristics of product j are captured by a vector of indicator variables $\boldsymbol{x}_j = \left(x^{\text{ale}}, x_j^{\text{lager}}, x_j^{\text{light}}, x_j^{\text{import}}, x_j^{\text{craft}}\right)$. We allow both household income and unobservable heterogeneity to affect preferences. We specify the utility function of household i in market m from product j in month t as

$$\begin{aligned} u_{ijmt} = {} & (\sigma_0 \nu_i + \kappa_0 y_i) + (\alpha + \kappa_\alpha y_i)\, p_{jmt} \\ & + \sigma^{\text{ale}} \nu_i^{\text{ale}} x_j^{\text{ale}} + \sigma^{\text{lager}} \nu_i^{\text{lager}} x_j^{\text{lager}} + \sigma^{\text{light}} \nu_i^{\text{light}} x_j^{\text{light}} \\ & + \sigma^{\text{import}} \nu_i^{\text{import}} x_j^{\text{import}} + (\sigma^{\text{craft}} \nu_i^{\text{craft}} + \kappa^{\text{craft}} y_i) x_j^{\text{craft}} \\ & + \boldsymbol{d}_{jm}\beta + FE_j^{\text{demand}} + FE_m^{\text{demand}} + FE_t^{\text{demand}} \\ & + FE_m^{\text{demand, craft}} + FE_t^{\text{demand, craft}} + \xi_{jmt} + \varepsilon_{ijmt}, \end{aligned} \tag{12}$$

where y_i is the natural logarithm of household i's annual income and $\nu_i^{(\cdot)}$ is the household-specific unobserved taste shock, which follows a normal distribution and is independent across households. Therefore, the $\sigma^{(\cdot)}$ parameters capture the dispersion in unobserved household tastes while the κ parameters measure the effect of household income on tastes. We do not include mean coefficients for $\boldsymbol{x}_j$ because they are absorbed in product fixed effects. The covariates $\boldsymbol{d}_{jm}$ represent the interaction between the craft indicator and a set of indicator functions for whether the distance from the brewery's nearest production facility to the market falls within a certain distance range. Distance potentially plays an important role in demand as a local beer may lack name recognition outside its local market (see, for example, Tamayo, 2009). Moreover, its importance may be different for craft vs. non-craft products. We also include in our model product fixed effects (FE_j^{demand}) as well as market and month fixed effects to capture unobserved factors that may vary at these levels. We allow both the market and month fixed effects to be different for craft and non-craft products and denote these fixed effects by ($FE_m^{\text{demand}}, FE_m^{\text{demand, craft}}$) and ($FE_t^{\text{demand}}, FE_t^{\text{demand, craft}}$). The error term ξ_{jmt}, therefore, captures the transient, month-to-month variations of demand shocks

[15] Some examples in the category of "others" include stout, porter, and near beers, which collectively account for 1% of the craft quantities.

specific to a product, market, and month combination. Finally, the last term in (12), ε_{ijmt}, is a household's idiosyncratic taste, which is assumed to be i.i.d. and follows a type-1 extreme value distribution.

Overall, our demand specification gives us the market share $s_{jmt}(p_{jmt}, p_{-jmt})$ of product j in month t and market m, where p_{-jmt} is a vector of the prices of all other products in market m and month t. Other determinants of demand (product characteristics, fixed effects, and demand shocks of all products in the market) are absorbed by the subscript jmt of the function $s_{jmt}(\cdot,\cdot)$. Multiplying the market share by the corresponding market size then gives us the demand for product j, $D_{jmt}(p_{jmt}, p_{-jmt})$.

4.2 Supply

The supply side describes firms' product and price decisions. In each market, firms simultaneously choose which beer products, if any, to sell.[16] This product choice is made at the beginning of each year τ and is fixed throughout the year. Then, in each month t, firms simultaneously choose the retail prices for their products.

Specifically, we model the supply side as a two-stage static game. In the first stage, at the beginning of year τ, firms observe fixed costs for all potential products and simultaneously decide on a set of products to offer in market m. In the second stage, at the beginning of month t, the demand and marginal cost shocks $(\xi_{jmt}, \omega_{jmt})$ are realized, and firms choose the retail prices for their products in market m.

Firms observe fixed costs for all potential products when making their product decisions. However, the demand and marginal cost shocks $(\xi_{jmt}, \omega_{jmt})$ are realized after firms have chosen which products to sell. This timing assumption is the same as that in Eizenberg (2014), Wollmann (2018), and Fan and Yang (2020), but different from two recent papers on entry or product repositioning in the airline industry (Ciliberto, Murry and Tamer, 2021 and Li, Mazur, Park, Roberts, Sweeting and Zhang, 2022), which assume that firms observe demand and marginal cost shocks as well as fixed cost shocks when making entry decisions. In other words, they account for selection based on all these shocks. By contrast, we allow for selection based only on unobserved fixed cost shocks and address selection based on unobserved demand and marginal cost shocks by including a large number of fixed effects in our demand and marginal cost functions. Specifically, we include product-, market-, and time-specific fixed effects. The remaining unobservables are month-to-month product/market-level transient shocks. We find it reasonable to assume that firms do not observe them when making

[16] As mentioned, federal and state laws prohibit practices that hinder the entry of craft breweries into the retailer market. These laws motivate our assumption that breweries make their own entry and product variety decisions. This simplification keeps our model tractable.

product choices. We also show later that the estimated shocks play a small role in explaining demand and marginal costs.

Stage 2. Pricing In month t, firm n chooses prices p_{jmt} for all $j \in \mathcal{J}_{nm\tau}$ to maximize its variable profit:

$$\max_{p_{jmt}, j \in \mathcal{J}_{nm\tau}} \sum_{j \in \mathcal{J}_{nm\tau}} \left(p_{jmt} - mc_{jmt}\right) D_{jmt}\left(p_{jmt}, p_{-jmt}\right). \tag{13}$$

The marginal cost mc_{jmt} is decomposed into a product fixed effect FE_j^{mc}, a market fixed effect FE_m^{mc}, a month fixed effect FE_t^{mc}, the effect of facility-market distance $\boldsymbol{d}_{jm}\gamma$ to account for any transportation cost, and a product-market-month specific shock ω_{jmt}:

$$mc_{jmt} = FE_j^{\text{mc}} + FE_m^{\text{mc}} + FE_t^{\text{mc}} + FE_m^{\text{mc, craft}} + FE_t^{\text{mc, craft}} + \boldsymbol{d}_{jm}\gamma + \omega_{jmt}. \tag{14}$$

We again allow both the market fixed effects and the month fixed effects to be different for craft and non-craft products.

This pricing model essentially assumes that the retail price of a product is the wholesale price plus a fixed markup for the distributor and the retailer. In fact, because we include product-, market-, and month-specific fixed effects in our specification of the brewery marginal cost, this markup can vary at the product, market, and month levels. We only need to assume that markups for distributors and retailers do not change in our counterfactual simulations.[17]

Stage 1. Entry and Product Decisions At the beginning of year τ, firm n is endowed with a set of potential products $\mathcal{J}_{n\tau}$ and decides on the set of products $\mathcal{J}_{nm\tau}$ to offer in market m. The profit-maximization problem at this stage is:

$$\max_{\mathcal{J}_{nm\tau} \subseteq \mathcal{J}_{n\tau}} \pi_{nm}\left(\mathcal{J}_{nm\tau}, \mathcal{J}_{-nm\tau}\right) - C_{nm}\left(\mathcal{J}_{nm\tau}\right), \tag{15}$$

where $\pi_{nm}\left(\mathcal{J}_{nm\tau}, \mathcal{J}_{-nm\tau}\right)$ is the expected variable profit and $C_{nm}\left(\mathcal{J}_{nm\tau}\right)$ is the fixed cost. We now derive the former and specify the latter.

To derive firm n's expected variable profit $\pi_{nm}\left(\mathcal{J}_{nm\tau}, \mathcal{J}_{-nm\tau}\right)$, we plug the second-stage equilibrium prices into its profit function, take the expectation over the transitory demand and marginal cost shocks, and sum over all months in a year. Formally, we use $\mathcal{J}_{-nm\tau}$ to denote the set of products that firm n's competitors sell in market m. Let $p_{jmt}\left(\mathcal{J}_{nm\tau}, \mathcal{J}_{-nm\tau}\right)$ and $Q_{jmt}\left(\mathcal{J}_{nm\tau}, \mathcal{J}_{-nm\tau}\right)$ denote the second-stage equilibrium price and

[17]Miller and Weinberg (2017) show that a double marginalization model where a brewery first sells to retailers does not significantly change their merger simulation results.

quantity, respectively, which depend on the observable covariates ($\boldsymbol{x}_j$, $\boldsymbol{d}_{jm}$), fixed effects (FE_j^{demand}, FE_m^{demand}, FE_t^{demand}, FE_j^{mc}, FE_m^{mc}, FE_t^{mc}) as well as the shocks (ξ_{jmt}, ω_{jmt}) for all products in market m. Let $\xi_{mt} = (\xi_{jmt}, j \in \mathcal{J}_{nm\tau} \cup \mathcal{J}_{-nm\tau})$ be the collection of demand shocks for all products in market m and define ω_{mt} for the marginal cost shocks analogously. Let $\mathcal{T}_\tau$ represent all months of year τ. Then, firm n's expected variable profit, $\pi_{nm}(\mathcal{J}_{nm\tau}, \mathcal{J}_{-nm\tau})$ in (15) is:

$$
\begin{aligned}
&\pi_{nm}(\mathcal{J}_{nm\tau}, \mathcal{J}_{-nm\tau}) \\
= &\sum_{t \in \mathcal{T}_\tau} E_{\xi_{mt}, \omega_{mt}} \left(\sum_{j \in \mathcal{J}_{nm\tau}} (p_{jmt}(\mathcal{J}_{nm\tau}, \mathcal{J}_{-nm\tau}) - mc_{jmt}) \cdot Q_{jmt}(\mathcal{J}_{nm\tau}, \mathcal{J}_{-nm\tau}) \right).
\end{aligned}
\tag{16}
$$

The fixed cost function in (15) is specified as

$$
C_{nm}(\mathcal{J}_{nm\tau}) = \sum_{j \in \mathcal{J}_{nm\tau}} (W_{jm}\theta + \sigma_\zeta \zeta_{jm\tau}), \tag{17}
$$

where W_{jm} is a vector of covariates including, for example, whether product j is produced by an independent craft brewery, whether its brewery is in CA, and the market size. We assume the fixed cost shock $\zeta_{jm\tau}$ is i.i.d. and follows a standard normal distribution in our baseline specification.[18]

5 Estimation

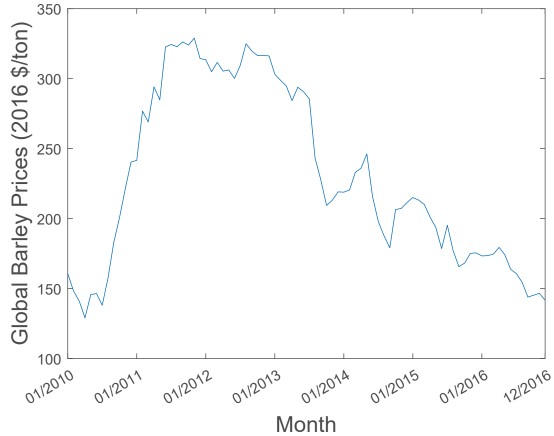
5.1 Estimation of Demand Parameters and Marginal Costs

5.1.1 Estimation Procedure

We combine the aggregate product/market/month-level data on prices, product characteristics, and market shares with the individual/month-level panel data on household purchases to estimate demand parameters. Specifically, we rely on the market share data to identify the mean taste coefficients (α, β) and fixed effects (FE_j^{demand}, FE_m^{demand}, FE_t^{demand}). We exploit the panel data and the correlations between household income and beer purchases to identify the standard deviations of the unobservable consumer heterogeneity (σ parameters) as well as the effect of household income on consumer tastes (κ parameters). We estimate

[18] We consider two robustness analyses in Supplemental Appendix SC. In the first extension, we add a market-specific shock $\lambda\zeta_{m\tau}$ to (17) in a robustness analysis, where $\zeta_{m\tau}$ is common to all products in a market, and follow the estimation procedure in Section 2.5. In the second robustness analysis, we extend our model and estimation method to allow for (dis)economies of scope in fixed costs. Our results are robust to both extensions.

Figure 3: Price of Barley



these parameters using the Generalized Method of Moments approach, where we combine a set of macro moments with two sets of micro moments.[19]

We construct macro moments based on instrumental variables consisting of the interactions of global barley prices with beer types to address potential price endogeneity. Barley is a common ingredient in almost all beers. Figure 3 plots the monthly price of barley in dollars per metric ton and shows fairly large monthly variations.[20]

We construct a new set of micro moments based on the persistence of a household's purchasing decisions to identify the standard deviation parameters. For example, a large value of σ^{craft} indicates that a household's preference for craft products is highly correlated across months. That means if a household ever purchases a craft product in a year, it is likely to purchase many craft products throughout the year. More generally, if we use $q_{i\tau}^f = \sum_{t \in \mathcal{T}_\tau} q_{it}^f$ to denote a household's purchase of beer type f ($f \in \{\text{ale}, \text{lager}, \ldots\}$) in year τ, then matching the conditional mean $E\left(q_{i\tau}^f \mid q_{i\tau}^f \geq 1\right)$ helps to identify the parameter σ^f.

Similar moments are also useful for identifying the correlation between taste shocks. For example, if a household that prefers type-f products tends to dislike type-f' products, then conditional on a household ever purchasing a type-f product, the household should buy few if any type-f' beers throughout the year.

Specifically, in constructing these micro moments, we match the model predictions of the following moments to their empirical counterparts:[21]

[19]Grieco, Murry, Pinkse and Sagl (2021) suggest combining the macro data with the micro data into one likelihood function for estimation. As pointed out in their paper, the efficiency gain from doing so is modest when the size of the micro data is small compared to the market size, which is the case in our paper. Moreover, we include a large set of fixed effects, which makes GMM estimation more computationally tractable than MLE estimation.

[20]Data source: https://fred.stlouisfed.org/series/PBARLUSDM

[21]See Supplemental Appendix SB for computational details.

- A household's expected annual purchase of a certain type of beer conditional on purchasing at least one unit of this type of beer in the year, i.e., $E\left(q_{i\tau}^{f} \,|\, q_{i\tau}^{f} \geq 1\right)$. Matching these moments helps to identify σ^{f}.

- A household's expected annual purchase of a certain type of beer conditional on purchasing at least one unit of craft beer in the year, i.e., $E\left(q_{i\tau}^{f} \,|\, q_{i\tau}^{\text{craft}} \geq 1\right)$. Matching these moments helps to identify the taste correlation between craft and type-f beer.

- A household's expected annual purchase of beer conditional on purchasing at least one unit of beer in the year, i.e., $E\left(q_{i\tau} \,|\, q_{i\tau} \geq 1\right)$, where $q_{i\tau}$ is a household's total beer purchase amount over a year. Matching this moment helps to identify σ_0.

We construct a second set of micro moments similar to those in Petrin (2002) to identify the effect of household income on consumer tastes:

- The ratio of average expenditure over average purchase quantity in a year among households whose income falls into a bin $\mathcal{I}$, i.e., $E\left(\text{expenditure}_{i\tau} \,|y_i \in \mathcal{I}\right) / E\left(q_{i\tau} \,|y_i \in \mathcal{I}\right)$, where the log-income bins $\mathcal{I}$ are $log\left(0, \$50K\right]$, $log\left(\$50K, \$100K\right]$, or $log\left(\$100K, +\infty\right)$.[22] Matching these moments helps to identify the income effect on price sensitivity, κ_α.

- $E(q_{i\tau}^{\text{craft}} \,|\, q_{i\tau}^{\text{craft}} \geq 1, y_i \in \mathcal{I})$, which helps to identify κ^{craft}.

- $E(q_{i\tau} \,|\, q_{i\tau} \geq 1, y_i \in \mathcal{I})$, which helps to identify κ_0.

Our estimation of marginal costs is standard and follows Berry, Levinsohn and Pakes (1995). We back out marginal costs based on the first-order conditions of the profit maximization problem in (13).

5.1.2 Results on Demand and Markups

Table 4 reports the demand estimation results. The estimated σ parameters indicate significant heterogeneity in preferences for craft products, imported products, and flavor types. For example, the estimated standard deviation of the unobservable heterogeneity in consumer taste for craft products $\hat{\sigma}^{\text{craft}}$ is 2.45. To understand the magnitude of this estimate, we compare it to the price coefficient of a household with an income of \$50,000, which is $-1.50 + 0.08 \cdot log\left(\$50,000\right) = -0.63$. Therefore, the estimated $\hat{\sigma}^{\text{craft}}$ is equivalent to a price change of 2.45/0.63, or 3.89 dollars.

[22] An alternative moment is the average price $E\left(\frac{\text{expenditure}_{i\tau}}{q_{i\tau}} \,|\, y_i \in \mathcal{I}\right)$. However, computing this moment is more cumbersome. It requires drawing both v_i^f and ε_{ij} to simulate this moment but only v_i^f to simulate the moment in the text.

Table 4: Demand Estimates

Unobserved	σ_0	0.001	Income Effect	κ_0	-1.34
Heterogeneity		(0.01)			(0.02)
	σ^{ale}	2.03		κ^{craft}	0.84
		(<0.01)			(0.02)
	σ^{lager}	0.86		κ_α	0.08
		(<0.01)			(<0.01)
	σ^{light}	2.56			
		(<0.01)	Price Coefficient	α	-1.50
	σ^{import}	1.82			(0.03)
		(<0.01)			
	σ^{craft}	2.45	Distance bin/Craft FE		Yes
		(<0.01)	Product FE		Yes
	$\rho^{\text{craft-light}}$	-0.75	Market/Craft FE		Yes
		(<0.01)	Time/Craft FE		Yes

Note: Standard errors are in parentheses.

Table 5: Selected Micro Moments on Persistence in Purchasing Decisions

		Data	Model
(1)	$E\left(\sum_{t=1}^{12} q_{it} \mid \sum_{t=1}^{12} q_{it} \geq 1\right)$	7.50	7.72
(2)	$E\left(\sum_{t=1}^{12} q_{it}^{\text{ale}} \mid \sum_{t=1}^{12} q_{it}^{\text{ale}} \geq 1\right)$	3.10	4.10
(3)	$E\left(\sum_{t=1}^{12} q_{it}^{\text{lager}} \mid \sum_{t=1}^{12} q_{it}^{\text{lager}} \geq 1\right)$	5.56	4.07
(4)	$E\left(\sum_{t=1}^{12} q_{it}^{\text{light}} \mid \sum_{t=1}^{12} q_{it}^{\text{light}} \geq 1\right)$	8.03	8.02
(5)	$E\left(\sum_{t=1}^{12} q_{it}^{\text{import}} \mid \sum_{t=1}^{12} q_{it}^{\text{import}} \geq 1\right)$	2.86	2.57
(6)	$E\left(\sum_{t=1}^{12} q_{it}^{\text{craft}} \mid \sum_{t=1}^{12} q_{it}^{\text{craft}} \geq 1\right)$	3.93	4.12

The dispersion parameters $\sigma^{(\cdot)}$ are estimated by matching the micro moments that capture the persistence in a household's purchasing decisions. Table 5 shows the model fit for these micro moments. For example, from Row (6), we see that the average per-household annual craft purchase among households that purchase at least one unit of craft beers is 3.93 in the data and 4.12 according to our estimates. Compared with the unconditional average per-household annual craft purchase of 0.37 units in the data, this micro moment implies that craft beers are purchased by a set of dedicated craft consumers, leading to a significant estimate of σ^{craft}.

The estimation results also indicate a negative correlation between consumer taste for craft and light beers ($\hat{\rho}^{\text{craft-light}} = -0.75$). This finding is consistent with the summary statistics in Table 3, which show that light craft beers account for only 0.71% of the craft beer sales while light beers in general account for 41% of all beer sales. We find that allowing for a correlation between ν_i^{light} and ν_i^{craft} is helpful for matching the conditional purchases of light beers given at least one craft purchase. The moment $E\left(q_{i\tau}^{\text{light}} \mid q_{i\tau}^{\text{craft}} \geq 1\right)$ is 1.27 in both the data and according to our estimated model, and it would be 1.94 according to an estimation where such a correlation is not allowed in the model.

Moreover, we find heterogeneity in consumer tastes across income levels. Specifically, high-income households are less likely to purchase beer ($\hat{\kappa}_0 < 0$), have a stronger preference for craft products $\left(\hat{\kappa}^{\text{craft}} > 0\right)$, and are less price sensitive ($\hat{\kappa}_\alpha > 0$).

Overall, the estimated demand parameters imply that the substitution within craft products is larger than the substitution between craft and non-craft products. Table 6 reports the own and cross elasticities among the top-3 non-craft and top-3 craft products in 2016.[23] These elasticities suggest little substitution between craft and non-craft products. Similarly, Figure 4 presents the histogram of the diversion ratio for a craft product to non-craft products (Panel (A)) and that for a craft product to other craft products (Panel (B)). Panel (A) shows that for most craft products, almost no sales would be captured by non-craft products if the focal craft product's price is increased. By contrast, the distribution of the diversion ratio to other craft products in Panel (B) has a mode of around 20%.

Table 6: Elasticities of Top-3 Craft Products and Top-3 Main Products (%)

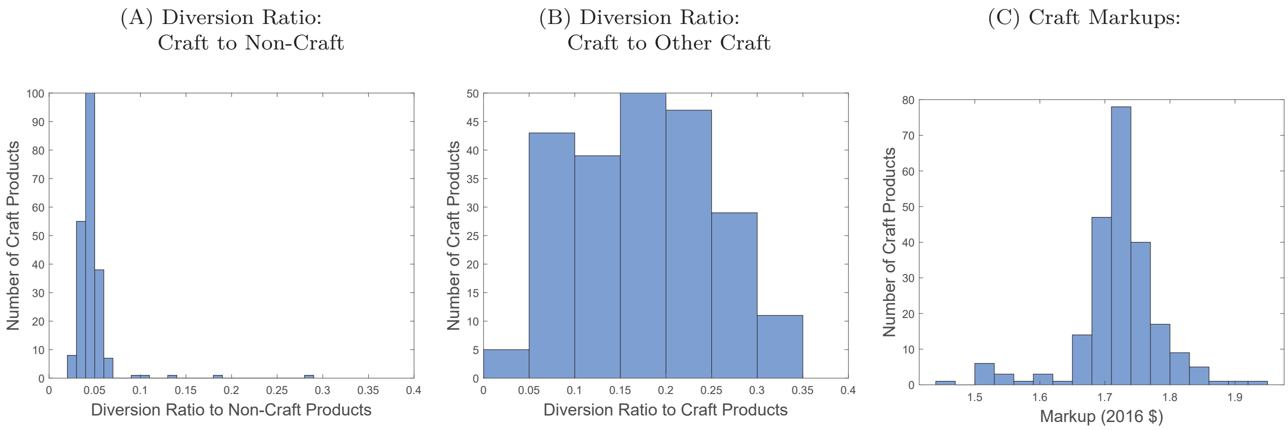
	Craft			Main		
	-10.05	0.13	0.02	0.01	0.01	0.03
Craft	0.21	-9.40	0.02	0.01	0.01	0.03
	0.04	0.02	-9.07	0.01	0.01	0.08
	<0.01	<0.01	<0.01	-5.25	0.56	0.04
Main	<0.01	<0.01	<0.01	0.57	-5.25	0.04
	<0.01	<0.01	0.01	0.04	0.04	-8.87

We back out the marginal costs using the first-order conditions at the pricing stage of the game and present the distribution of the quantity-weighted markup in Panel (C) of Figure 4. The median markup of craft beers is about \$1.7 in 2016 dollars. Some industry sources (e.g., Satran, 2014) put the brewer's margin at 8% of the retail price, or \$1.4 for an average price of \$17, in line with our estimates.

Finally, our observed explanatory variables account for the majority of the variations in demand and marginal costs. The R^2's from regressing the mean utility and marginal cost on observable covariates and fixed effects are both above 0.9, implying that after controlling for the product-, market-, and time-specific fixed effects, the month-to-month transient shocks

[23] Per our data agreement, we refrain from discussing the identities of beers or breweries in the data. Hidalgo (2023) estimates the elasticities of the top macro and craft beer brands to be -3.4 and -8.3, comparable to the elasticities reported in this table.

Figure 4: Histograms of Diversion Ratios and Markups



$(\xi_{jmt}, \omega_{jmt})$ play, at most, a small role.

5.2 Estimation of Fixed Cost Parameters

We estimate the fixed cost parameters year by year. In this section, we focus on craft products and present our fixed cost estimation results using the 2016 data for consistency with our later counterfactual analyses. In our estimation, one unit of observation is a potential product j and market m combination. In this part of the estimation, we exclude small craft products[24] and markets without craft products, resulting in a total of 95 potential products, 149 markets, and 14,155 potential product/market combinations in 2016.

In this section, we first explain a reformulation of our empirical model consistent with the notation in Section 2.2 to follow the estimation procedure described there. We also show data patterns that help with identification. Since we estimate the fixed cost parameters for each year separately, we suppress the year subscript τ in this section for exposition simplicity.

5.2.1 Reformulation of the Model

To be consistent with the model outlined in Section 2.2, we rewrite the profit function in (15), i.e., $\pi_{nm}\left(\mathcal{J}_{nm}, \mathcal{J}_{-nm}\right) - \sum_{j \in \mathcal{J}_{nm}} \left(W_{jm}\theta + \sigma_\zeta \zeta_{jm}\right)$ as

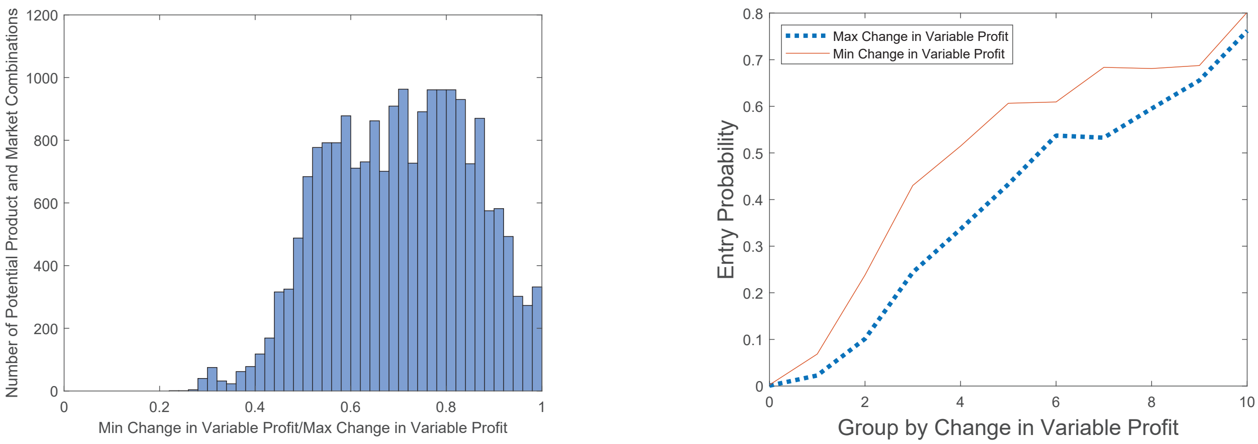
$$\pi_n\left(\boldsymbol{Y}_{nm}, \boldsymbol{Y}_{-nm}, X_{nm}\right) - \sum_{j \in \mathcal{J}_n} Y_{jm}\left(W_{jm}\theta + \sigma_\zeta \zeta_{jm}\right). \tag{18}$$

[24]We exclude craft products that appear in fewer than 36 market-month combinations, which account for 0.22% of the craft quantity in 2016.

Figure 5: Data Patterns Aiding Identification

(A) Histogram of $\underline{\Delta}_j\left(X_{nm}\right)/\overline{\Delta}_j\left(X_{nm}\right)$

(B) Variations in $\underline{\Delta}_j\left(X_{nm}\right), \overline{\Delta}_j\left(X_{nm}\right)$ and Entry Probability



Specifically, we now use a vector of indicators $\boldsymbol{Y}_{nm} \in \{0,1\}^{\#\mathcal{J}_n}$ to denote a firm's product portfolio $\mathcal{J}_{nm} \subseteq \mathcal{J}_n$, where $\mathcal{J}_n$ represents the potential products that firm n is endowed with. We use Y_{jm} to denote the element of $\boldsymbol{Y}_{nm}$ that corresponds to product $j \in \mathcal{J}_n$, where $Y_{jm} = 1$ if $j \in \mathcal{J}_{nm}$ and 0 otherwise. Therefore, the expected variable profit $\pi_{nm}\left(\mathcal{J}_{nm}, \mathcal{J}_{-nm}\right)$ can be written as $\pi_n\left(\boldsymbol{Y}_{nm}, \boldsymbol{Y}_{-nm}, X_{nm}\right)$, where the vector X_{nm} includes all demand and marginal cost covariates (including fixed effects). Similarly, the total fixed cost $\sum_{j\in\mathcal{J}_{nm}}\left(W_{jm}\theta + \sigma_\zeta\zeta_{jm}\right)$ can be written as the summation over all products with $Y_{jm} = 1$, i.e., $\sum_{j\in\mathcal{J}_n} Y_{jm}\left(W_{jm}\theta + \sigma_\zeta\zeta_{jm}\right)$.

5.2.2 Data Patterns That Help with Identification

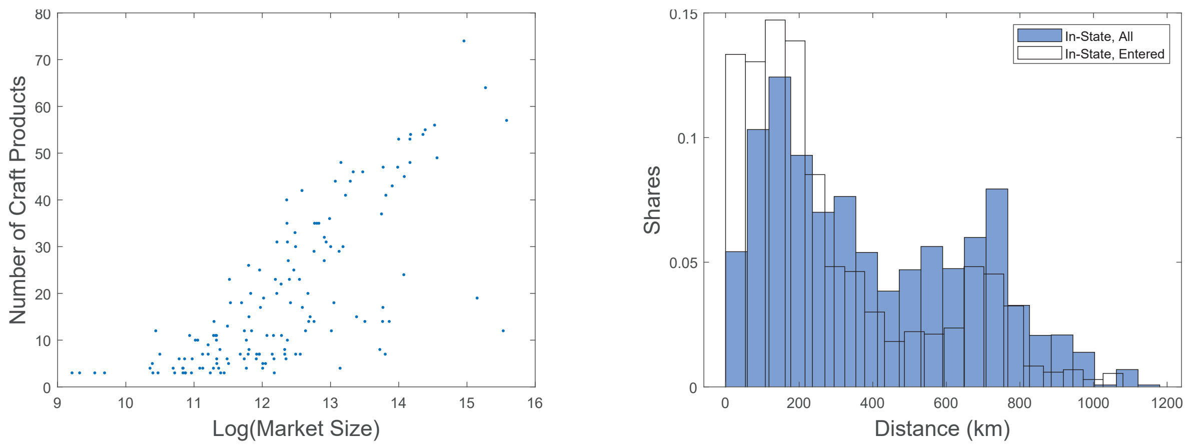
We have discussed identification in abstract for a general model in Section 2. Here, we present data patterns in our empirical setting that help with identification.

First, for a large proportion of the observations, the minimum and maximum changes in variable profit, i.e., $\underline{\Delta}_j\left(X_{nm}\right)$ and $\overline{\Delta}_j\left(X_{nm}\right)$, are relatively close, resulting in tight conditional choice probability bounds for these products. Panel (A) of Figure 5 plots the histogram of the ratio $\underline{\Delta}_j\left(X_{nm}\right)/\overline{\Delta}_j\left(X_{nm}\right)$ across all combinations of potential products and markets in 2016. The median of the ratio is around 0.7. In other words, for 50% of the observations, the ratio is larger than 0.7, where a larger ratio reflects a smaller difference between the minimum and maximum. The tight bounds reflect that the diversion ratios of many craft products are low (Figure 4).

Second, there are considerable variations in $\underline{\Delta}_j\left(X_{nm}\right)$ and $\overline{\Delta}_j\left(X_{nm}\right)$, and these variations are informative about variations in entry probabilities. To see the variations in $\underline{\Delta}_j\left(X_{nm}\right)$

Figure 6: Exogenous Variations Aiding Identification

(A) Market Size and # Craft Products (B) Distribution of Distance



and $\overline{\Delta}_j(X_{nm})$, we note that their 10% and 90% percentiles are, respectively, (\$27, \$1,969) and (\$39, \$2,821), which differ by two orders of magnitude. To see the association between $\underline{\Delta}_j(X_{nm})$ and entry probabilities, we discretize $\underline{\Delta}_j(X_{nm})$ into 10 groups based on the deciles. For each group g, we compute the average entry probability for observations jm such that $\underline{\Delta}_j(X_{nm})$ is in this group as $\frac{\sum_{j,m} \mathbb{1}\left(\underline{\Delta}_j(X_{nm}) \in g\right) \cdot Y_{jm}}{\sum_{j,m} \mathbb{1}\left(\underline{\Delta}_j(X_{nm}) \in g\right)}$, where $Y_{jm} \in \{0,1\}$ represents the observed entry outcome. We repeat this exercise for $\overline{\Delta}_j(X_{nm})$. Panel (B) of Figure 5 displays the average entry probabilities associated with $\underline{\Delta}_j(X_{nm})$ (represented by the red solid line) and those associated with $\overline{\Delta}_j(X_{nm})$ (represented by the blue dotted line). From the figure, we can see that the average entry probabilities increase in both $\underline{\Delta}_j(X_{nm})$ and $\overline{\Delta}_j(X_{nm})$.

What exogenous variations in X_{nm} generate the variations in $\underline{\Delta}_j(X_{nm})$ and $\overline{\Delta}_j(X_{nm})$? In addition to variations in product characteristics as well as the fixed effects in the demand and marginal cost functions, variations in market size also play an important role, because everything else being equal, the returns to entry increase in the size of a market. This can be seen in Panel (A) of Figure 6, which depicts a strong positive correlation between the logarithm of market sizes and the number of craft products in a market.

Another source of exogenous variation is the distance between a production facility and a market. In Panel (B) of Figure 6, we plot the unconditional distribution of distances for all in-state craft potential product/market combinations and the conditional distribution for observed in-state craft product/market combinations (i.e., the product is in the market in the data).[25] Panel (B) shows that the conditional distribution has more probability mass at shorter distances than the unconditional one, suggesting a negative correlation between distance and entry. We account for these variations in variable profits by including controls for distances in both our demand and marginal cost functions.

5.2.3 Results on Fixed Costs

We follow the estimation procedure described in Section 2.2 to estimate the fixed cost parameters, which include the parameters of the mean fixed cost (θ) and the standard deviation of the fixed cost shock (σ_ζ).

Table 7 reports the 95% confidence set projected to each parameter. These estimates account for the statistical errors in the estimation of variable profits. We find a higher fixed cost for independent craft breweries and larger markets. The 95% confidence set projected to the coefficient of the craft indicator is [$255, $583], indicating that craft breweries incur higher fixed costs than non-craft breweries. This parameter is identified by the data pattern that products of craft breweries acquired by macro breweries are more likely to enter a market than those of independent craft breweries.[26] To study whether fixed costs vary with market size, we categorize markets into small, median, or large bins based on whether the market size is below 10^5, between 10^5 and $5{\times}10^5$, or above $5{\times}10^5$ units, and allow fixed costs to differ across bins.[27] We find that fixed costs are higher in larger markets and the standard deviation of the unobservable fixed cost shock also increases with market size.

6 Counterfactual Simulations

6.1 Counterfactual Designs

We consider a counterfactual merger where the largest firm in our sample, a so-called macro brewer, acquires the three largest craft firms in 2016, excluding Boston Beer Company and Sierra Nevada Brewing, which are unlikely merger targets given their size. In other words, we study a scenario where the trend of acquisitions in the craft beer industry continues to

[25]Out-of-state craft products tend to be widely distributed and thus less affected by the distance between their production facilities and markets. A number of such craft products in California are brewed on or near the East Coast.

[26]To clarify, for a craft product acquired by a macro brewery but still maintaining its craft status according to the Brewer Association's craft designation, we set its craft indicator to 0 in the fixed cost specification. This is because the product benefits from the distribution and marketing networks of the larger firms, which may affect its fixed costs.

[27]The 25% and 75% quantiles of the market sizes are $0.7{\times}10^5$ and $4.2{\times}10^5$ units.

Table 7: Estimates of Fixed Costs: Projected 95% Confidence Intervals, 2016

Craft (θ_1)	[255.32, 583.95]
In State× Craft (θ_2)	[-276.04, 37.57]
Market-size specific fixed cost (θ_3)	
Small market	[426.52, 721.90]
Medium market	[720.77, 1014.10]
Large market	[2028.00, 2453.69]
Market-size specific std. dev. (σ_ζ)	
Small market	[0.02, 105.80]
Medium market	[1.18, 207.45]
Large market	[767.59, 1044.93]

Note: Estimates in 2016 US dollars.

the point where the concentration of the craft market approaches the level in the overall beer market.[28]

In our simulation, we allow firms to adjust their craft products and hold the non-craft product choices fixed as observed in the data (but allow their prices to change). We do so for two reasons. First, solving a product choice game is computationally challenging because each firm can choose any subset from its set of potential products and there are $2^{\#\text{ of potential products}}$ such subsets. We compute the post-merger product equilibrium using the algorithm outlined in Fan and Yang (2020).[29] We further ease the computational burden by holding the non-craft product choices fixed. Second, this simplification is justified by the estimated small substitution between craft and non-craft products.

A potential entrant is a firm observed in any market in our sample. Each firm is endowed with a set of potential products comprised of the firm's craft products observed in any market in the 2016 data. In each market, a firm chooses a subset from its potential products, and an empty subset denotes no entry. The potential product set for the merged firm consists of the combined set of potential products. We assume that firms maximize profits at both the

[28]During our sample period, there are four observed acquisitions where a macro brewer acquired an independent craft brewer in our sample (and other mergers involving smaller craft brewers not in our sample). Of the four observed acquisitions, one brewer was not in our sample prior to the transaction. Among the other three, we observe an increase in entry and product variety post-merger, consistent with the simulated outcomes based on our estimated model, allowing for fixed-cost merger efficiency (see Section 6.2).

[29]Fan and Yang (2020) develop a heuristic algorithm to find a firm's optimal product portfolio given the portfolios of its competitors, and embed this optimization algorithm in a best-response iteration to solve for the post-merger product choice equilibrium. The algorithm starts with an initial vector of product decisions and evaluates whether it is profitable to add or drop a product. If a profitable deviation is found, the product vector is updated to the most profitable deviation among all such one-product deviations, and the process is repeated until no more profitable one-product deviations are possible. The algorithm is run with different initial vectors to check for multiple equilibria. We find identical results for two starting points, one based on the observed product decisions in the data and another where each firm chooses all potential products.

product choice and pricing stages.

We conduct three counterfactual simulations. We quantify the overall effects as well as the effects due to firm entry and those due to product changes. Specifically, in the counterfactual simulation described above (CF1), we allow for three adjustment margins — new entry, product adjustments, and price adjustments. In the second counterfactual (CF2), we allow for only incumbent product adjustment and price adjustment by removing the products added by new entrants in CF1 and recomputing the pricing equilibrium. In the third counterfactual (CF3), we allow for only the price effect of the merger by restoring pre-merger market products and recomputing the pricing equilibrium. The difference between the outcomes in CF1 and CF3 gives us the overall product variety effect of the merger, which can be further decomposed into the product variety effect due to new entry (CF1 - CF2) and that due to incumbent product adjustments (CF2 - CF3). For all simulations, we sample 10 vectors of parameter values from the 95% confidence set of the fixed cost parameters.

We draw three sets of shocks: demand, marginal cost, and fixed cost shocks. We draw demand and marginal cost shocks directly from their estimated distributions. For each sampled fixed cost parameter vector, we draw fixed cost shocks from the estimated distribution conditional on the observed pre-merger equilibrium to ensure that pre- and post-merger outcomes are comparable. Details on how we draw our fixed cost shocks can be found in Appendix B.

For each sampled parameter vector, we compute the simulated merger effects averaged across the simulation draws of the demand, marginal cost, and fixed cost shocks. We report the range of this average effect across the sampled parameter vectors as the 95% confidence interval of the average effects.

6.2 Counterfactual Results

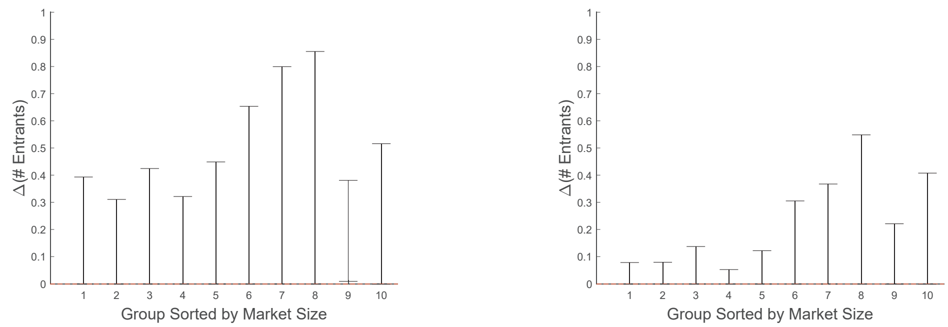
We present two sets of counterfactual simulation results, allowing for a merger efficiency in one and ignoring it in the other. A merger efficiency can arise because craft breweries could benefit from using the marketing networks of the acquirer and enjoy reduced fixed costs (Elzinga and McGlothlin, 2021).[30] According to our estimate of the parameter θ_1 in the fixed cost function, independent craft breweries face higher fixed costs. Recall that this parameter is identified by the difference in product decisions between independent craft breweries and craft products owned by the macro breweries, i.e., the latter products are more likely to enter

[30] Mergers involving craft breweries may not quickly realize efficiency gains in marginal costs as craft breweries often remain operationally independent and their beers continue to be brewed in the same facilities in the short run. This arrangement contrasts with mergers among macro breweries, where the merged firms relocate production and economize on transportation costs from production facilities to markets (Ashenfelter, Hosken and Weinberg, 2015).

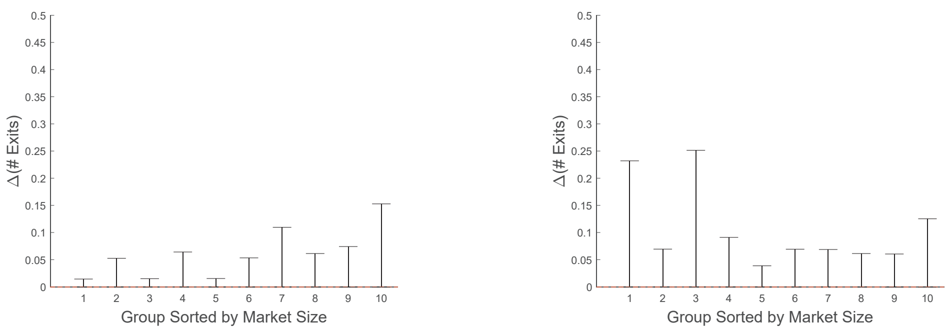
Figure 7: Merger Effects on Entry, Product Variety, and Prices

No Merger Efficiency With Merger Efficiency

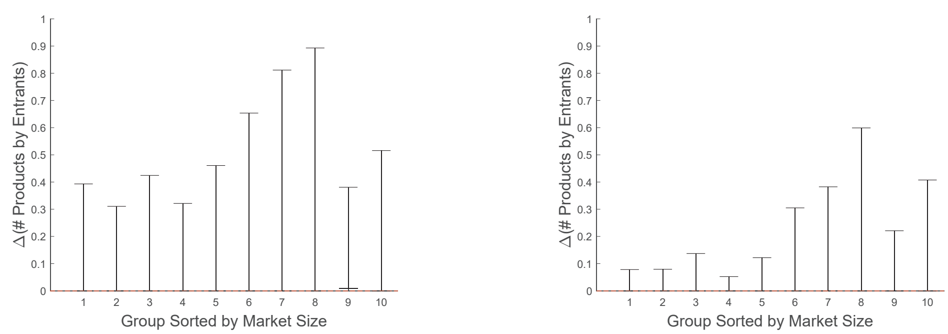
(A) Number of Firm Entries



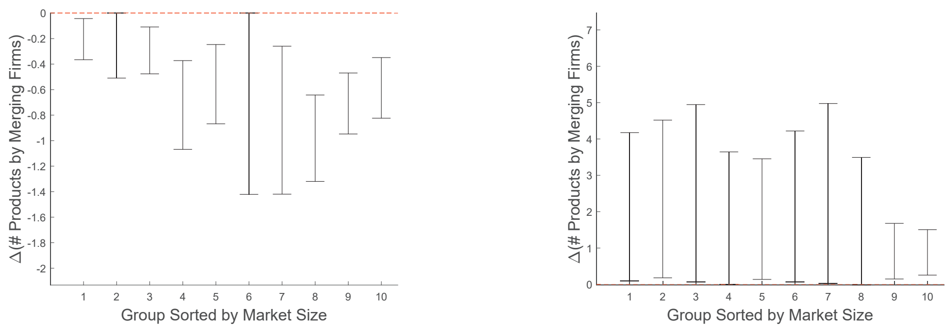
(B) Number of Firm Exits



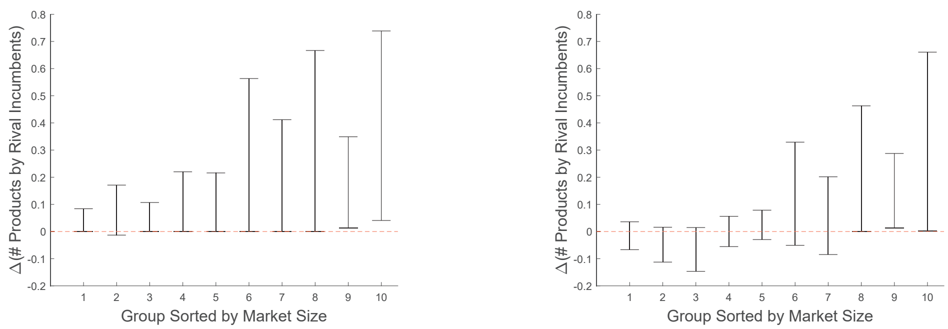
(C) Number of Products Added by Entrants



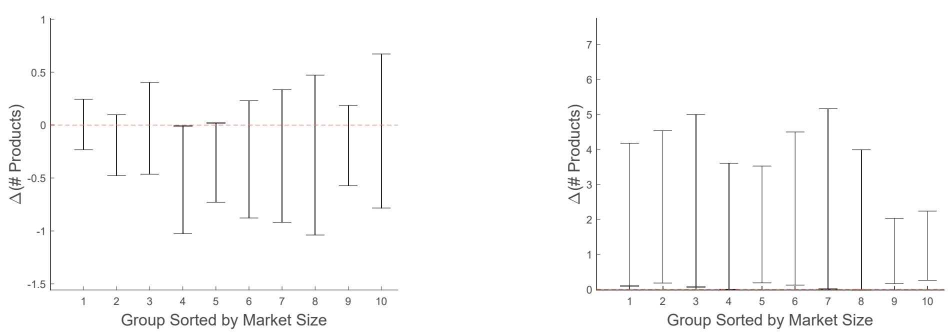
(D) Change in the Number of Products by Merging Firms



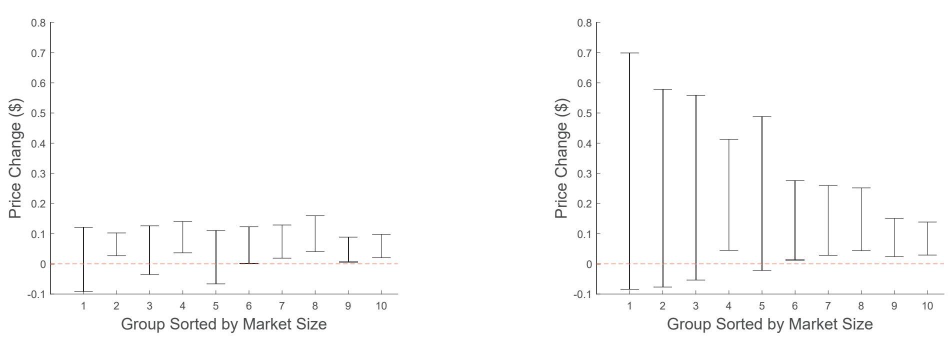
(E) Change in the Number of Products by Non-merging Incumbent Firms



(F) Change in the Number of Products



(G) Change in Craft Prices



a market. Therefore, to quantify the effects of a potential fixed cost reduction post-merger, we consider a scenario where the fixed costs of the acquired craft products decrease by the estimated θ_1, which is the extra fixed cost faced by independent craft breweries.

We report the merger effects on entry, product variety, and prices across markets in Figure 7 and the aggregate welfare effects in Table 8. For the purpose of presentation, in Figure 7, instead of presenting the merger effects in each of the 149 markets separately, we sort markets into 10 groups according to market size from group 1 (smallest) to group 10 (largest). There are 15 markets in groups 1 to 9 and 14 largest markets in group 10. Within each group, we average counterfactual outcomes across markets weighted by their market size.

The left panel of Figure 7 shows the merger's effects on firm entry, product variety, and prices without considering the potential merger efficiency. We see both firm entry and exit post-merger (Panels (A) and (B)), with more entries than exits. New entrants bring in new products (Panel (C)). The number of products added by new entrants is almost identical to the number of new entrants, implying that new entrants, on average, enter with one product. As for incumbents, merging incumbents drop products (Panel (D)) while non-merging incumbents add products (Panel (E)). As a result, the change in the overall number of products is ambiguous (Panel (F)). The increase in the quantity-weighted average craft beer price (Panel (G)) is centered around 5 cents, but could be as large as nearly 15 cents, which is about 10% of the average markup.

The left panel of Table 8 reports the welfare effects of the merger ignoring the merger efficiency. It shows that total surplus decreases (Row (1)): while there is a gain in producer surplus (Row (2)), consumers are worse off, with a decrease in consumer surplus of \$233,250 to \$1,075,040 (Row (3)). The effect of the product variety change on consumer surplus is ambiguous (Row (4)). A decomposition of the consumer surplus change due to the product changes indicates that while new entries recover the consumer welfare loss (Row (5)), incumbents' product adjustments reduce consumer welfare (Row (6)).

Overall, in the scenario without the merger efficiency gain, new entries occur and these new entrants bring new products to markets after the merger. Non-merging incumbents also add products. However, their positive effects are not enough to offset the negative welfare effects from the merged firm dropping products and the increased prices.

We now turn to simulations allowing for the merger efficiency in reducing fixed costs. The right panels of Figure 7 demonstrate that the merging firms now add products (Panel (D)). At the same time, there are fewer new entrants (Panel (A)) and more exits in some market groups (Panel (B)). The number of products added by new entrants is lower (Panel (C)). Similarly, the change in the number of products by non-merging firms also becomes

Table 8: Welfare Effects

		No Merger Efficiency	With Merger Efficiency
(1)	total surplus ($1000)	[-780.05, -200.23]	[-401.66, -31.24]
(2)	craft beer profits ($1000)	[33.02, 294.99]	[25.34, 285.68]
(3)	consumer surplus ($1000)	[-1,075.04, -233.25]	[-670.21, -152.21]
(4)	due to variety change	[-91.00, 111.55]	[6.08, 566.06]
(5)	due to entry	[1.21, 122.86]	[1.14, 69.95]
(6)	due to incumbent product adjustments	[-111.05, -9.11]	[4.94, 519.47]

Note: this table reports the aggregate welfare effects of the merger across markets. For each measure, we report the range across the vectors of parameters sampled from their 95% confidence set. The left panel shows the results without considering any merger efficiency, while the right panel reports the results that incorporate reductions in fixed costs when a craft brewery is acquired by a macro brewery.

smaller or even negative in very small markets (Panel (E)). However, the total number of products now increases (Panel (F)). At the same time, the upper bound of price changes becomes larger, indicating that prices could increase more (Panel (G)).

With the efficiency, the consumer welfare loss is smaller but not reversed. The range of the loss is now between $152,210 and $670,210, according to Row (3) in the right panel of Table 8. Both new entry and product adjustment by incumbents help to mitigate the negative welfare effect, as shown in Rows (5) and (6), leading to a positive consumer welfare effect due to changes in product variety (Row (4)). Nonetheless, both the overall consumer surplus change and the total surplus change are still negative.

We note two countervailing effects of the merger efficiency associated with fixed costs on the number of products in a market. On the one hand, the efficiency can induce new product entry by merging firms. On the other hand, the efficiency can depress new firm entry, increase rival firm exit, and discourage product entry by non-merging incumbents, thereby limiting the overall positive effect of efficiency gains on product variety and consumer welfare. The latter countervailing effect exists because while prices are often strategic complements, product offerings tend to be strategic substitutes.

In sum, the merger results in new firm entry in all markets as well as product entry by non-merging incumbents in larger markets. The effect on product variety depends on whether there is a merger efficiency in reducing fixed costs, and the merged firm may drop or add products accordingly. However, in both scenarios, the merger leads to a decrease in consumer surplus and total surplus.

7 Conclusion

We propose a new method for estimating discrete games and apply it to study merger effects on firm entry, product choice, and prices in the California retail craft beer market. The paper makes two contributions. Methodologically, we present a new method to estimate discrete games. This method is easy to compute and scalable to games with many firms or many firm decisions. Empirically, we study the effects of a merger on both firm entry and product variety. We simulate a merger that could significantly increase the concentration of the retail craft beer market. We find that, although new firm entry always occurs after such a merger, whether product variety decreases depends on a merger efficiency gain that reduces the fixed cost. However, even in our entry-favorable setting, the impact of new entry is insufficient to offset the overall negative effect of the merger on welfare. The merger efficiency can only reduce, but not reverse, the consumer surplus loss.

References

Anderson, Simon, Nisvan Erkal, and Daniel Piccinin (2020), "Aggregative games and oligopoly theory: Short-run and long-run analysis." *The RAND Journal of Economics*, 51, 470–495.

Andrews, Donald WK and Gustavo Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78, 119–157.

Aradillas-Lopez, Andres and Elie Tamer (2008), "The identification power of equilibrium in simple games." *Journal of Business & Economic Statistics*, 26, 261–283.

Ashenfelter, Orley, Daniel Hosken, and Matthew Weinberg (2015), "Efficiencies brewed: Pricing and consolidation in the US beer industry." *The RAND Journal of Economics*, 46, 328–361.

Asker, John (2016), "Diagnosing foreclosure due to exclusive dealing." *The Journal of Industrial Economics*, 64, 375–410.

Barkley, Aaron, Joachim R Groeger, and Robert A Miller (2021), "Bidding frictions in ascending auctions." *Journal of Econometrics*, 223, 376–400.

Beresteanu, Arie, Ilya Molchanov, and Francesca Molinari (2011), "Sharp identification regions in models with convex moment predictions." *Econometrica*, 79, 1785–1821.

Berry, Steven (1992), "Estimation of a model of entry in the airline industry." *Econometrica*, 60, 889–917.

Berry, Steven, Alon Eizenberg, and Joel Waldfogel (2016), "Optimal product variety in radio markets." *The RAND Journal of Economics*, 47, 463–497.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile prices in market equilibrium." *Econometrica*, 63, 841–890.

Berry, Steven and Joel Waldfogel (2001), "Do mergers increase product variety? Evidence from radio broadcasting." *The Quarterly Journal of Economics*, 116, 1009–1025.

Bresnahan, Timothy and Peter Reiss (1991), "Entry and competition in concentrated markets." *Journal of political economy*, 99, 977–1009.

Bronnenberg, Bart, Jean-Pierre Dubé, and Joonhwi Joo (2022), "Millennials and the takeoff of craft brands: Preference formation in the US beer industry." *Marketing Science*, 41, 710–732.

Cabral, Luıs (2003), "Horizontal mergers with free-entry: Why cost efficiencies may be a weak defense and asset sales a poor remedy." *International Journal of Industrial Organization*, 21, 607–623.

Caradonna, Peter, Nathan Miller, and Gloria Sheu (2021), "Mergers, entry, and consumer welfare." *Georgetown McDonough School of Business Research Paper*.

Chesher, Andrew and Adam M Rosen (2017), "Generalized instrumental variable models." *Econometrica*, 85, 959–989.

Ciliberto, Federico, Charles Murry, and Elie Tamer (2021), "Market structure and competition in airline markets." *Journal of Political Economy*, 129, 2995–3038.

Ciliberto, Federico and Elie Tamer (2009), "Market structure and multiple equilibria in airline markets." *Econometrica*, 77, 1791–1828.

Codog, Aric (2018), "The antitrust roadblock: Preventing consolidation of the craft beer market." *University of the Pacific Law Review*, 50, 403.

Croxall, Daniel (2019), "Helping craft beer maintain and grow market shares with private enforcement of tied-house and false advertising laws." *Gonzaga Law Review*, 55, 167.

Döpper, Hendrik, Alexander MacKay, Nathan Miller, and Joel Stiebale (2022), "Rising markups and the role of consumer preferences." *Harvard Business School Strategy Unit Working Paper*.

Draganska, Michaela, Michael Mazzeo, and Katja Seim (2009), "Beyond plain vanilla: Modeling joint product assortment and pricing decisions." *QME*, 7, 105–146.

Eizenberg, Alon (2014), "Upstream innovation and product variety in the US home PC market." *Review of Economic Studies*, 81, 1003–1045.

Elliott, Jonathan, Georges Vivien Houngbonon, Marc Ivaldi, and Paul Scott (2021), "Market structure, investment and technical efficiencies in mobile telecommunications."

Elzinga, Kenneth G and Alexander J McGlothlin (2021), "Has Anheuser-Busch let the steam out of craft beer? The economics of acquiring craft brewers." *Review of Industrial Organization*, 1–27.

Fan, Ying (2013), "Ownership consolidation and product characteristics: A study of the US daily newspaper market." *American Economic Review*, 103, 1598–1628.

Fan, Ying and Chenyu Yang (2020), "Competition, product proliferation, and welfare: A study of the US smartphone market." *American Economic Journal: Microeconomics*, 12, 99–134.

Galichon, Alfred and Marc Henry (2011), "Set identification in models with multiple equilibria." *The Review of Economic Studies*, 78, 1264–1298.

Gandhi, Amit, Luke Froeb, Steven Tschantz, and Gregory Werden (2008), "Post-merger product repositioning." *The Journal of Industrial Economics*, 56, 49–67.

Garrido, Francisco Andres (2020), "Mergers between multi-product firms with endogenous variety: Theory and an application to the ready-to-eat cereal industry." working paper, ITAM.

Grieco, Paul (2014), "Discrete games with flexible information structures: An application to local grocery markets." *The RAND Journal of Economics*, 45, 303–340.

Grieco, Paul, Charles Murry, Joris Pinkse, and Stephan Sagl (2021), "Efficient estimation of random coefficients demand models using product and consumer datasets." Technical report, Working paper, Pennsylvania State University.

Haile, Philip and Elie Tamer (2003), “Inference with an incomplete model of English auctions.” *Journal of Political Economy*, 111, 1–51.

Hidalgo, Julian (2023), “This craft’s for you! entry and market (ing) competition in the us beer industry.”

HIVERY (2023), “Retail reset challenges: A look inside typical grocery retailer.” URL `https://www.hivery.com/resources/category/blog/retail-reset-challenges-a-look-inside-typical-grocery-retailer/`.

Ho, Katherine (2009), “Insurer-provider networks in the medical care market.” *American Economic Review*, 99, 393–430.

Huang, Yufeng, Paul B Ellickson, and Mitchell J Lovett (2022), “Learning to set prices.” *Journal of Marketing Research*, 59, 411–434.

Illanes, Gastón (2017), “Switching costs in pension plan choice.” *Unpublished manuscript*.

Illanes, Gastón and Sarah Moshary (2020), “Market structure and product assortment: Evidence from a natural experiment in liquor licensure.” working paper 27016, National Bureau of Economic Research.

Jeziorski, Przemysław (2015), “Empirical model of dynamic merger enforcement–choosing ownership caps in US radio.” working paper, University of California, Berkeley.

Lewbel, Arthur (2019), “The identification zoo: Meanings of identification in econometrics.” *Journal of Economic Literature*, 57, 835–903.

Li, Sophia, Joe Mazur, Yongjoon Park, James Roberts, Andrew Sweeting, and Jun Zhang (2022), “Repositioning and market power after airline mergers.” *The RAND Journal of Economics*, 53, 166–199.

Mazzeo, Michael J (2002), “Product choice and oligopoly market structure.” *RAND Journal of Economics*, 221–242.

McFadden, Daniel (1989), “A method of simulated moments for estimation of discrete response models without numerical integration.” *Econometrica*, 57, 995–1026.

Miller, Nathan and Matthew Weinberg (2017), “Understanding the price effects of the MillerCoors joint venture.” *Econometrica*, 85, 1763–1791.

Miller, Nathan H, Gloria Sheu, and Matthew C Weinberg (2021), "Oligopolistic price leadership and mergers: The United States beer industry." *American Economic Review*, 111, 3123–3159.

Miravete, Eugenio J, Katja Seim, and Jeff Thurk (2018), "Market power and the laffer curve." *Econometrica*, 86, 1651–1687.

Miravete, Eugenio J, Katja Seim, and Jeff Thurk (2020), "One markup to rule them all: Taxation by liquor pricing regulation." *American Economic Journal: Microeconomics*, 12, 1–41.

Miravete, Eugenio J, Katja Seim, and Jeff Thurk (2023), "Inflation, taxation & market power."

Pakes, Ariel, Jack Porter, Kate Ho, and Joy Ishii (2015), "Moment inequalities and their application." *Econometrica*, 83, 315–334.

Petrin, Amil (2002), "Quantifying the benefits of new products: The case of the minivan." *Journal of Political Economy*, 110, 705–729.

Reiss, Peter and Pablo Spiller (1989), "Competition and entry in small airline markets." *The Journal of Law and Economics*, 32, S179–S202.

Ryan, Tom (2016), "Kroger ignites category captain brewhaha." RetailWire, URL `https://retailwire.com/discussion/kroger-ignites-category-captain-brewhaha/`.

Satran, Joe (2014), "Here's how a six-pack of craft beer ends up costing $12." *https://www.huffpost.com/entry/craft-beer-expensive-cost_n_5670015*.

Schennach, Susanne (2014), "Entropic latent variable integration via simulation." *Econometrica*, 82, 345–385.

Seim, Katja (2006), "An empirical model of firm entry with endogenous product-type choices." *The RAND Journal of Economics*, 37, 619–640.

Spector, David (2003), "Horizontal mergers, entry, and efficiency defences." *International Journal of Industrial Organization*, 21, 1591–1600.

Sweeting, Andrew (2010), "The effects of mergers on product positioning: Evidence from the music radio industry." *The RAND Journal of Economics*, 41, 372–397.

Sweeting, Andrew (2013), "Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry." *Econometrica*, 81, 1763–1803.

Tamayo, Andrew (2009), "What's brewing in the old north state: An analysis of the beer distribution laws regulating North Carolina's craft breweries." *NCL Rev.*, 88, 2198.

Tamer, Elie (2003), "Incomplete simultaneous discrete response model with multiple equilibria." *The Review of Economic Studies*, 70, 147–165.

Tremblay, Victor J, Natsuko Iwasaki, and Carol Horton Tremblay (2005), "The dynamics of industry concentration for US micro and macro brewers." *Review of Industrial Organization*, 26, 307–324.

Wang, Shuang (2020), "Price competition with endogenous entry: The effects of Marriott & Starwood merger in Texas." working paper.

Watson, Bart (2016), "On-premise beer data and craft." *https://www.brewersassociation.org/insights/importance-on-premise-craft-brewers/*.

Werden, Gregory and Luke Froeb (1998), "The entry-inducing effects of horizontal mergers: An exploratory analysis." *The Journal of Industrial Economics*, 46, 525–543.

Wollmann, Thomas (2018), "Trucks without bailouts: Equilibrium product characteristics for commercial vehicles." *American Economic Review*, 108, 1364–1406.

A Details on Inference and Step-by-Step Calculation of the Confidence Set

Confidence Set

We construct our confidence set based on the inequalities in (8) by inverting the test in Andrews and Soares (2010). With an abuse of notation, we now use θ to denote the fixed cost parameters including both the coefficients of covariates (originally denoted by θ) and the parameters in the distribution of the unobservable fixed cost shock (originally denoted by σ_ζ). We also denote the moment functions in (8) by $Z_{m,\tilde{k}}(\theta)$, $\tilde{k} = 1, \ldots, 2K$. For the first K moment functions, $Z_{m,\tilde{k}}(\theta)$ is $\frac{1}{\#\mathcal{J}} \sum_{j \in \mathcal{J}} L(Y_{jm}, X_{nm}, W_{jm}, \theta) \cdot g^{(\tilde{k})}(X_{nm}, W_{jm})$ in (8). For $\tilde{k} = K+1, \ldots, 2K$, $Z_{m,\tilde{k}}(\theta)$ is $\frac{1}{\#\mathcal{J}} \sum_{j \in \mathcal{J}} H(Y_{jm}, X_{nm}, W_{jm}, \theta) \cdot g^{(\tilde{k}-K)}(X_{nm}, W_{jm})$ in (8).

Let $Z_m(\theta) = (Z_{m,1}(\theta), ..., Z_{m,2K}(\theta))'$. Then, the sample moment functions are

$$\bar{Z}_M(\theta) = \frac{1}{M} \sum_{m=1}^{M} Z_m(\theta). \tag{A.1}$$

Let

$$\hat{\Sigma}_M(\theta) = \frac{1}{M} \sum_{m=1}^{M} \left(Z_m(\theta) - \bar{Z}_M(\theta)\right)\left(Z_m(\theta) - \bar{Z}_M(\theta)\right)' \tag{A.2}$$

be the estimator of the covariance matrix of $\sqrt{M}\bar{Z}_M(\theta)$.

The test statistic is given by

$$T_M(\theta) = S(\sqrt{M}\bar{Z}_M(\theta), \hat{\Sigma}_M(\theta)), \tag{A.3}$$

where $S(Z, \Sigma) = \sum_{\tilde{k}=1}^{2K} [Z_{\tilde{k}}/\sigma_{\tilde{k}}]_+^2$ is the modified method of moments test function, $\sigma_{\tilde{k}}^2$ is the $\tilde{k}$th diagonal element of Σ, and the function $[\cdot]_+$ takes the value of the argument if it is positive and 0 otherwise.

The critical value for the null hypothesis $H_0 : \theta = \theta_0$, denoted by $\hat{c}_M(\theta_0, 1-\alpha)$, is the $(1-\alpha)$ quantile of $S(\hat{\Omega}_M^{1/2}(\theta_0)R + [\eta_M(\theta_0)]_-, \hat{\Omega}_M(\theta_0))$, where $R \sim N(0_{2K}, I_{2K})$, $\hat{\Omega}_M(\theta) = \hat{D}_M^{-1/2}(\theta)\hat{\Sigma}_M(\theta)\hat{D}_M^{-1/2}(\theta)$, $\hat{D}_M(\theta) = Diag(\hat{\Sigma}_M(\theta))$, $\eta_M(\theta) = (lnM)^{-1/2}M^{1/2}\hat{D}_M^{-1/2}(\theta)\bar{Z}_M(\theta)$, and $[\eta_M(\theta)]_- = ([\eta_{M,1}(\theta)]_-, ..., [\eta_{M,2K}(\theta)]_-)$.

We invert this test to construct our confidence set, which is $\{\theta_0 : T_M(\theta_0) \leq \hat{c}_M(\theta_0, 1-\alpha)\}$.

Step-by-Step Calculation

We now describe the steps in constructing the confidence set.

(1) We first compute $\underline{\Delta}_j(X_{nm})$, $\overline{\Delta}_j(X_{nm})$, and $g^{(k)}(X_{nm}, W_{jm})$ for each j, n, m and k.

(2) For each candidate parameter value θ_0, we compute the test statistic $T_M(\theta_0)$ as follows:

   (i) compute $Z_m(\theta_0)$ by first plugging $\underline{\Delta}_j(X_{nm})$ into $L(Y_{jm}, X_{nm}, W_{jm}, \theta_0)$ and $\overline{\Delta}_j(X_{nm})$ into $H(Y_{jm}, X_{nm}, W_{jm}, \theta_0)$ and then combining them with each $g^{(k)}(X_{nm}, W_{jm})$ to compute $Z_{m,\tilde{k}}(\theta_0)$,

   (ii) compute the sample moments $\bar{Z}_M(\theta_0)$ according to (A.1),

   (iii) compute the sample variance estimator $\hat{\Sigma}_M(\theta_0)$ according to (A.2),

   (iv) compute the test statistic $T_M(\theta_0)$ according to (A.3).

(3) For each candidate parameter value θ_0, we compute the critical value $\hat{c}_M(\theta_0, 1-\alpha)$ as follows:

   (v) compute $\hat{D}_M(\theta_0) = Diag(\hat{\Sigma}_M(\theta_0))$, $\hat{\Omega}_M(\theta_0) = \hat{D}_M^{-1/2}(\theta_0)\hat{\Sigma}_M(\theta_0)\hat{D}_M^{-1/2}(\theta_0)$, and $\eta_M(\theta_0) = (lnM)^{-1/2}M^{1/2}\hat{D}_M^{-1/2}(\theta)\bar{Z}_M(\theta_0)$,simulate NS draws of the $2K$-dimensional random vector R from the standard normal distribution: $\{R_r : r = 1, ..., NS\}$,

   (viii) find the $(1-\alpha)$ quantile of $\{S(\hat{\Omega}_M^{1/2}(\theta_0)R_r + [\eta_M(\theta_0)]_-, \hat{\Omega}_M(\theta_0)) : r = 1, ..., NS\}$.

(4) To construct the confidence set, we consider a large set of parameter values and include a point θ_0 in the confidence set if $T_M(\theta_0) \leq \hat{c}_M(\theta_0, 1-\alpha)$. In practice, we use a three-step stochastic search process.

   (4.1) We find $\theta_0^\star$ that minimizes the test statistic $T_M(\theta)$.

   (4.2) We repeat the following forward perturbation process 50 times. In each forward perturbation process,

   - We add to $\theta_0^\star$ a perturbation and check whether the test statistic at the new point is below the corresponding critical value.[31]
   - If it is, we save the new point as $\theta_1^\star$ and perturb it. Otherwise, we consider another perturbation to $\theta_0^\star$.
   - We continue to add perturbations in this fashion 50 times.

   Each forward perturbation process yields a set of points satisfying $T_M(\theta_0) \leq \hat{c}_M(\theta_0, 1-\alpha)$. We collect all such points across the 50 repetitions of the process.

[31]The size of the perturbation depends on the empirical application. We draw from a normal distribution with mean 0 and a standard deviation of $\frac{1}{5}$ independently for each parameter.

(4.3) We select 100 points obtained in (4.2) and for each of them, we perform a forward perturbation similar to the one in (4.2), but with larger perturbations.[32]

The 100 selected points consist of two groups. The first 50 points are sampled randomly from the set obtained in (4.2). The second 50 points are sampled randomly from the points at the boundary of the confidence set obtained in (4.2), i.e., at least one dimension of these points is the maximum or minimum along that dimension among all points in the set found in (4.2). We do this to ensure that we consider points beyond the bounds of the set obtained in (4.2).

In some cases, $\underline{\Delta}_j$ and $\overline{\Delta}_j$ may depend on estimated parameters (which are estimated before estimating θ). For example, in the empirical part of the paper, the change in variable profit is computed based on the estimated demand and marginal cost parameters. When the moment functions depend on the estimated parameters, the estimated covariance of the moments $\hat{\Sigma}_M(\theta)$ needs to be adjusted to account for the estimation errors in these parameters. In this appendix, β denotes the collection of demand and marginal cost parameters and $\hat{\beta}$ denotes their estimates. The following steps are adjusted:

- In Step (1), we compute $\underline{\Delta}_j(X_{nm}, \hat{\beta})$, $\overline{\Delta}_j(X_{nm}, \hat{\beta})$, and $g^{(k)}(X_{nm}, W_{jm}, \hat{\beta})$.
- In (i) and (ii), we plug in $\underline{\Delta}_j(X_{nm}, \hat{\beta})$, $\overline{\Delta}_j(X_{nm}, \hat{\beta})$, and $g^{(k)}(X_{nm}, W_{jm}, \hat{\beta})$ to compute $Z_m(\theta_0, \hat{\beta})$ and then $\bar{Z}_M(\theta_0, \hat{\beta})$. In (iii), we simulate the sample variance from the asymptotic distribution of the demand and marginal cost parameters as follows:
  - simulate $\widetilde{NS}$ draws of the demand and marginal cost parameter values denoted by $\{\beta_r : r = 1, ..., \widetilde{NS}\}$,
  - compute $\underline{\Delta}_j(X_{nm}, \beta_r), \overline{\Delta}_j(X_{nm}, \beta_r), g^{(k)}(X_{nm}, W_{jm}, \beta_r)$ for $r = 1, ..., \widetilde{NS}$,
  - compute $Z_m(\theta_0, \beta_r)$ and then $\bar{Z}_M(\theta_0, \beta_r)$ for $r = 1, ..., \widetilde{NS}$,
  - compute the adjusted sample variance estimator as

$$\hat{\Sigma}_M(\theta_0) = \frac{1}{M \times \widetilde{NS}} \sum_{m=1}^{M} \sum_{r=1}^{\widetilde{NS}} \left(Z_m(\theta_0, \beta_r) - \bar{Z}_M(\theta_0, \beta_r) \right) \left(Z_m(\theta_0, \beta_r) - \bar{Z}_M(\theta_0, \beta_r) \right)'.$$

- In (iv), we compute the test statistic $T_M(\theta_0, \hat{\beta}) = S(\sqrt{M} \bar{Z}_M(\theta_0, \hat{\beta}), \hat{\Sigma}_M(\theta_0))$.
- In computing the critical value in Step (3), we replace $\bar{Z}_M(\theta_0)$ by $\bar{Z}_M(\theta_0, \hat{\beta})$ and plug in the adjusted sample variance to obtain the critical value $\hat{c}_M(\theta_0, \hat{\beta}, 1 - \alpha)$.

[32] In practice, we draw from a standard normal distribution.

- In constructing the confidence set in Step (4), we follow the same procedure to find θ_0 s.t. $T_M(\theta_0, \hat{\beta}) \leq \hat{c}_M(\theta_0, \hat{\beta}, 1-\alpha)$.

B Details on Fixed Cost Simulation Draws in Counterfactual Simulations

We draw fixed costs consistent with both the estimated distribution of fixed cost and the observed pre-merger outcome as an equilibrium. As explained in Section 6, it is important to maintain this consistency to properly compare the pre- and post-merger outcomes. To obtain one such set of draws in a market m, we proceed with the following steps:

1. For each potential product j of firm n, we calculate the change in firm n's expected variable profit when product j enters the market as defined in equation (4),

$$\Delta_j\left(\boldsymbol{Y}_{-jm}, X_{nm}\right) = \pi_n\left(Y_{jm}=1, \boldsymbol{Y}_{-jm}, X_{nm}\right) - \pi_n\left(Y_{jm}=0, \boldsymbol{Y}_{-jm}, X_{nm}\right),$$

   where we plug in the observed $\boldsymbol{Y}_{-jm}$. If product j is observed in market m, we define a range $(-\infty, \Delta_j\left(\boldsymbol{Y}_{-jm}, X_{nm}\right))$. Otherwise, we define a range $(\Delta_j\left(\boldsymbol{Y}_{-jm}, X_{nm}\right), \infty)$.

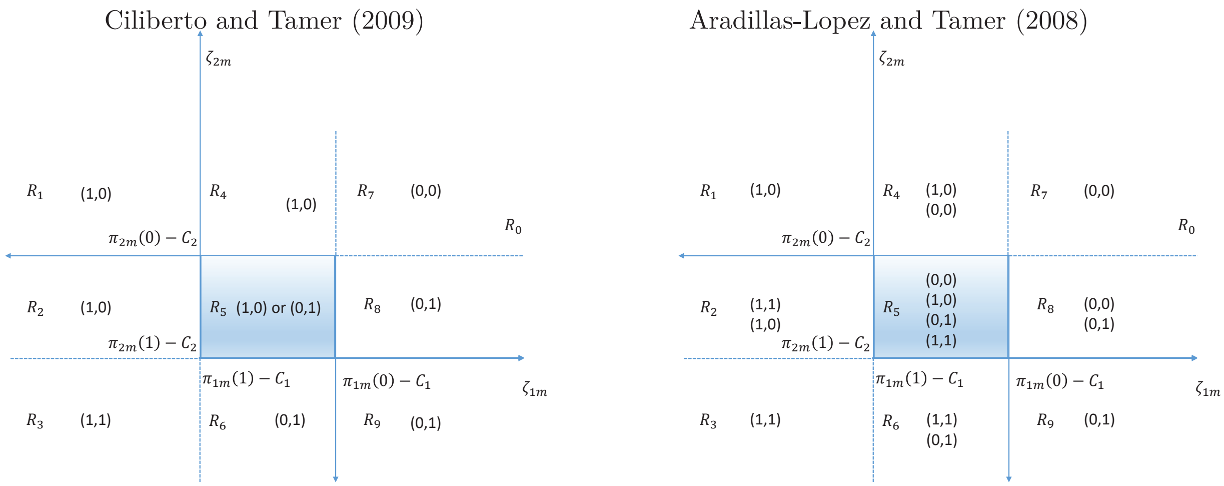
2. We simulate draws of the fixed costs for firm n from a truncated normal distribution with the underlying normal distribution parameterized by mean $W_{jm}\hat{\theta}$ and variance $\hat{\sigma}_{\zeta}^2$. The support of the truncated distribution is defined by the ranges obtained in step 1. These draws satisfy the necessary conditions for the observed outcome to be an equilibrium.

3. For each draw from step 2, we check whether firm n's best response to $\boldsymbol{Y}_{-nm}$ is indeed $\boldsymbol{Y}_{nm}$, where $\boldsymbol{Y}_{nm}$ and $\boldsymbol{Y}_{-nm}$, respectively, represent firm n's and its opponents' product decisions in market m in the data. We find each firm's best response by employing the algorithm in Fan and Yang (2020) using two starting points, i.e., $\boldsymbol{Y}_{nm}^0 = (0, ..., 0)$ and $\boldsymbol{Y}_{nm}^0 = (1, ..., 1)$. If the algorithm converges to $\boldsymbol{Y}_{nm}$ from both starting points, we keep the set of draws for firm n. If at least one of the starting points does not lead to $\boldsymbol{Y}_{nm}$, we go back to step 2 and re-draw the fixed costs.

4. We repeat this process for every firm n.

SA A Graphic Illustration of the Comparison of Our Bounds to the CT and the AT Bounds

Figure SA.1: Model Implications Under Different Assumptions



Comparison to CT

The left graph of Figure SA.1 lists all possible pure-strategy Nash equilibria in each region of (ζ_{1m}, ζ_{2m}) in the 2×2 entry game. We use $\Pr(R)$ to represent the probability that (ζ_{1m}, ζ_{2m}) is in region R. The CT bounds for $(Y_{1m} = 1, Y_{2m} = 0)$ are:

$$\sum_{\ell=1,2,4} \Pr(R_\ell) \leq \Pr(Y_{1m} = 1, Y_{2m} = 0) \leq \sum_{\ell=1,2,4,5} \Pr(R_\ell). \tag{SA.1}$$

The bounds for $(Y_{1m} = 1, Y_{2m} = 1)$ degenerate into an equation:

$$\Pr(Y_{1m} = 1, Y_{2m} = 1) = \Pr(R_3). \tag{SA.2}$$

Therefore, the CT bounds imply the following bounds for $\Pr(Y_{1m} = 1)$:

$$\sum_{\ell=1}^{4} \Pr(R_\ell) \leq \Pr(Y_{1m} = 1) \leq \sum_{\ell=1}^{5} \Pr(R_\ell). \tag{SA.3}$$

By contrast, our bounds are

$$\sum_{\ell=1}^{3} \Pr\left(R_\ell\right) \leq \Pr\left(Y_{1m}=1\right) \leq \sum_{\ell=1}^{6} \Pr\left(R_\ell\right). \tag{SA.4}$$

These bounds are wider than those in (SA.3). Intuitively, the CT bounds rely on stronger assumptions (Nash vs. level-1 rationality), exploit more model implications (probability of an equilibrium vs. an action), and are therefore tighter.

Comparison to AT

The right panel of Figure SA.1 shows all outcomes consistent with the level-1 rationality assumption. The AT bounds for $(Y_{1m}=1, Y_{2m}=1)$ and $(Y_{1m}=1, Y_{2m}=0)$ are, respectively:

$$\begin{aligned} \Pr\left(R_3\right) \leq \Pr\left(Y_{1m}=1, Y_{2m}=1\right) &\leq \sum_{\ell=2,3,5,6} \Pr\left(R_\ell\right), \\ \Pr\left(R_1\right) \leq \Pr\left(Y_{1m}=1, Y_{2m}=0\right) &\leq \sum_{\ell=1,2,4,5} \Pr\left(R_\ell\right). \end{aligned} \tag{SA.5}$$

In the region R_2, although the model implication for (Y_{1m}, Y_{2m}) is not unique, firm 1 always chooses the dominant strategy $Y_{1m}=1$. The AT bounds do not exploit such uniqueness of a firm's action while our bounds do. Therefore, the AT bounds are not necessarily sharper than ours.

Another way to see this point is as follows. In regions R_2, R_4, R_5, R_6, and R_8, $\Pr\left(Y_{1m}=1, Y_{2m}=1\right)$ and $\Pr\left(Y_{1m}=1, Y_{2m}=0\right)$ depend on the equilibrium selection rules. The possible values for $\Pr\left(Y_{1m}=1, Y_{2m}=1\right)$ and $\Pr\left(Y_{1m}=1, Y_{2m}=0\right)$ are

$$\begin{aligned} \Pr\left(Y_{1m}=1, Y_{2m}=1\right) &= \Pr(R_3) + \sum_{l=2,5,6} \alpha_l \Pr(R_l), \\ \Pr\left(Y_{1m}=1, Y_{2m}=0\right) &= \Pr(R_1) + \sum_{l=2,4,5} \beta_l \Pr(R_l), \end{aligned}$$

where $\alpha_l \in [0,1]$ is the probability that $(Y_{1m}=1, Y_{2m}=1)$ is the selected in region l and $\beta_l \in [0,1]$ is the probability that $(Y_{1m}=1, Y_{2m}=0)$ is selected in region l. Since $(Y_{1m}=1, Y_{2m}=1)$ and $(Y_{1m}=1, Y_{2m}=0)$ are the only two possible outcomes in R_2, we have $\alpha_2+\beta_2=1$. Similarly, we have $\alpha_5+\beta_5 \leq 1$.

The AT lower and upper bounds for $\Pr\left(Y_{1m}=1, Y_{2m}=1\right)$ in (SA.5) are essentially

$$\min_{\{\alpha_2,\alpha_5,\alpha_6\}\in[0,1]^3} \Pr(R_3) + \sum_{l=2,5,6} \alpha_l \Pr(R_l) \quad \text{and} \quad \max_{\{\alpha_2,\alpha_5,\alpha_6\}\in[0,1]^3} \Pr(R_3) + \sum_{l=2,5,6} \alpha_l \Pr(R_l),$$

and those for $\Pr(Y_{1m} = 1, Y_{2m} = 0)$ are

$$\min_{\{\beta_2,\beta_4,\beta_5\}\in[0,1]^3} \Pr(R_1) + \sum_{l=2,4,5} \beta_l \Pr(R_l) \quad \text{and} \quad \max_{\{\beta_2,\beta_4,\beta_5\}\in[0,1]^3} \Pr(R_1) + \sum_{l=2,4,5} \beta_l \Pr(R_l).$$

In other words, the AT bounds only require α_l and β_l to be between 0 and 1, but ignore the requirement that $\alpha_2 + \beta_2 = 1$. Below we show that our bounds use this condition.

The possible value for $\Pr(Y_{1m} = 1)$ is

$$\Pr(R_1) + \Pr(R_3) + \sum_{l=2,5,6} \alpha_l \Pr(R_l) + \sum_{l=2,4,5} \beta_l \Pr(R_l).$$

Since $\alpha_2 + \beta_2 = 1$, it can rewritten as

$$\Pr(R_1) + \Pr(R_2) + \Pr(R_3) + \sum_{l=5,6} \alpha_l \Pr(R_l) + \sum_{l=4,5} \beta_l \Pr(R_l).$$

Therefore, the minimum possible value for $\Pr(Y_{1m} = 1)$ is $\Pr(R_1) + \Pr(R_2) + \Pr(R_3)$, which is exactly our lower bound for $\Pr(Y_{1m} = 1)$.

In sum, our bounds as well as the CT and AT bounds exploit only a subset of conditions implied by a model. Different from the comparison between our bounds and the CT bounds, the AT bounds and our bounds rely on the same model restrictions but use different and non-nesting model implications.

SB Details on Micro Moments

In this section, we explain how we compute the model prediction for the micro moment $E\left(q_{i\tau}^f \mid q_{i\tau}^f \geq 1\right)$. The calculation for other micro-moments in Section 5.1 is similar.

Let $s_{jmt}(\boldsymbol{\nu}, y)$ denote the Logit choice probability of product j in month t when the vector of unobserved tastes is $\boldsymbol{\nu}$ and log-income is y. Let $G_m(\boldsymbol{\nu}, y)$ denote the distribution of $(\boldsymbol{\nu}, y)$, which can vary across markets and is thus indexed by m. We assume that each consumer has 8 opportunities to buy beer per month, which is the average number of household trips per month in the Nielsen Consumer Panel data. Then, the probability that a household with values $(\boldsymbol{\nu}, y)$ buys type-f products in market m in year τ is

$$\psi_{m\tau}^f(\boldsymbol{\nu}, y) = 1 - \prod_{t\in\mathcal{T}_\tau} \left(1 - \sum_{j\in\mathcal{J}_{m\tau}^f} s_{jmt}(\boldsymbol{\nu}, y)\right)^8,$$

where $\mathcal{T}_\tau$ denotes the months of the year and $\mathcal{J}_{m\tau}^f$ is the collection of all type-f products in market m in year τ. The conditional expectation of the annual purchase of type-f products for households in market m and year τ is, therefore,

$$E_{m\tau}\left(q_{i\tau}^{f} \mid q_{i\tau}^{f} \geq 1\right) = \int_{\boldsymbol{\nu},y} \frac{\sum_{t\in\mathcal{T}_{\tau}} \sum_{j\in\mathcal{J}_{m\tau}^{f}} 8 \cdot s_{jmt}\left(\boldsymbol{\nu}, y\right)}{\psi_{m\tau}^{f}\left(\nu, y\right)} dG_{m}\left(\boldsymbol{\nu}, y\right),$$

where $E_{m\tau}$ is the expectation specific to a market m and a year τ. To obtain the average across market/year combinations, we weigh these conditional means in each market/year combination by the expected number of households who purchase type-f products, which is the product of the market size and the unconditional probability of purchasing type-f products in a market/year, i.e.,

$$weight_{m\tau} = MktSize_{m\tau} \cdot \int \psi_{m\tau}^{f}\left(\boldsymbol{\nu}, y\right) dG_{m}\left(\boldsymbol{\nu}, y\right).$$

Therefore, the expected purchase of type-f products conditional on having at least one purchase is

$$E\left(q_{i\tau}^{f} \mid q_{i\tau}^{f} \geq 1\right) = \frac{\sum_{m\tau} E_{m\tau}\left(q_{i\tau}^{f} \mid q_{i\tau}^{f} \geq 1\right) \cdot weight_{m\tau}}{\sum_{m\tau} weight_{m\tau}}.$$

SC Robustness Analyses

In this section, we consider two extensions. First, we extend the fixed cost specification to allow for a market-level unobserved cost shock ζ_m:

$$W_{jm}\theta + \sigma_{\zeta}\zeta_{jm} + \lambda\zeta_{m}.$$

We assume ζ_m is i.i.d. across markets and has a standard normal distribution. We follow the estimation procedure in Section 2.5 to estimate $(\theta, \sigma_{\zeta}, \lambda)$ and report the projected 95% confidence interval in Table SC.1. For comparison, we also copy the baseline results in Table SC.1. We find that the standard deviation of the market-level shock ζ_m is comparable to the standard deviation of the product/market-level shock ζ_{jm} in medium-sized markets. We also find that the merger simulation results are robust (Figure SC.1 and Table SC.2).

For the second robustness analysis, we extend our model to allow for (dis)economies of scope in fixed costs and derive a new set of inequalities bounding the entry probability of a firm in addition to that of a product in order to estimate the additional parameter.

Our baseline fixed cost specification is additively separable across products and thus does not allow for (dis)economies of scope. We consider the following extension of the fixed cost

Table SC.1: Fixed Cost Estimates: Projected 95% Confidence Interval, Allowing for a Market-Level Shock

	Baseline Model	With a Market-Level Shock
Craft (θ_1)	[255.32, 583.95]	[13.15, 447.83]
In State$\times$ Craft (θ_2)	[-276.04, 37.57]	[-215.23, 239.22]
Market-size specific fixed cost (θ_3)		
Small market	[426.52, 721.90]	[470.68, 874.90]
Medium market	[720.77, 1014.10]	[435.05, 875.24]
Large market	[2028.00, 2453.69]	[2201.91, 2819.98]
Market-size specific std. dev. (σ_ζ)		
Small market	[0.02, 105.80]	[0.01, 128.23]
Medium market	[1.18, 207.45]	[0.07, 177.45]
Large market	[767.59, 1044.93]	[590.89, 1087.97]
Market-level unobserved cost std. dev. (λ)		[0.00, 175.51]

Note: Estimates in 2016 US dollars.

Table SC.2: Welfare Effects, Allowing for a Market-Level Shock

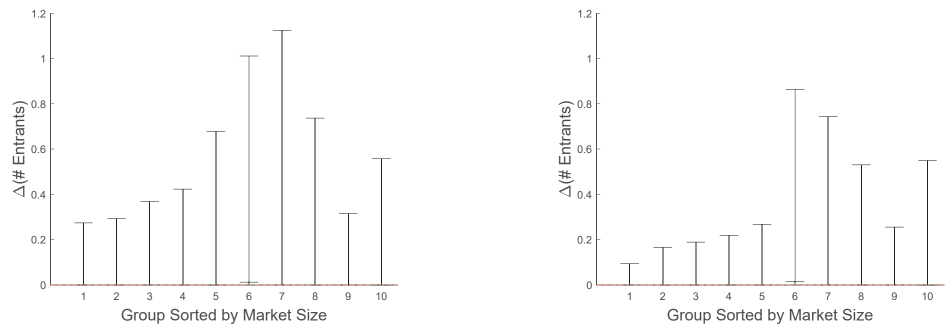
		No Merger Efficiency	With Merger Efficiency
(1)	total surplus ($1000)	[-781.57, -190.74]	[-583.33, -132.23]
(2)	craft beer profits ($1000)	[31.37, 295.32]	[24.41, 287.68]
(3)	consumer surplus ($1000)	[-1076.89, -222.11]	[-841.97, -158.60]
(4)	due to variety change	[-91.28, 162.04]	[-10.65, 386.12]
(5)	due to entry	[0.16, 133.87]	[0.13, 101.77]
(6)	due to incumbent product adjustments	[-108.36, 28.17]	[-48.47, 325.46]

Note: this table reports the aggregate welfare effects of the merger across markets. For each measure, we report the range across the vectors of parameters sampled from their 95% confidence set. The left panel shows the results without considering any merger efficiency, while the right panel reports the results that incorporate reductions in fixed costs when a craft brewery is acquired by a macro brewery.

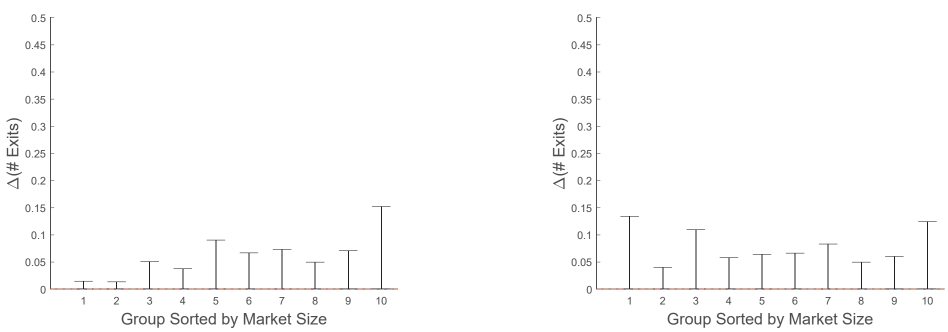
Figure SC.1: Merger Effects on Entry, Product Variety, and Prices, Allowing for a Market-Level Shock

No Merger Efficiency With Merger Efficiency

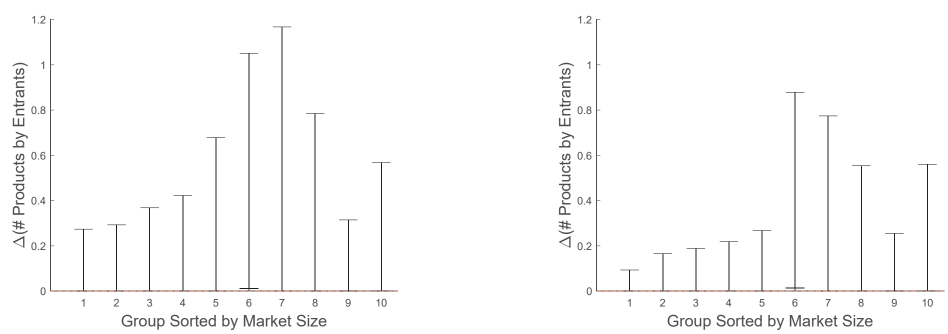
(A) Number of Firm Entries



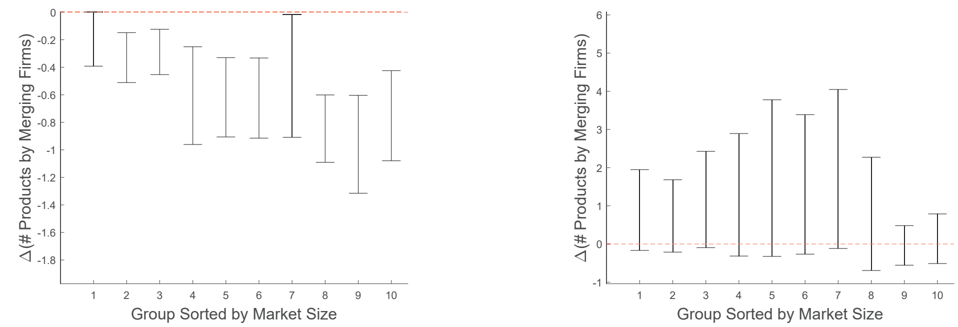
(B) Number of Firm Exits



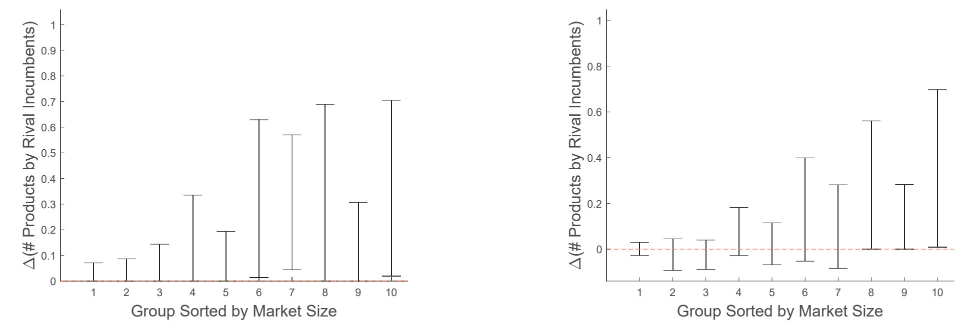
(C) Number of Products Added by Entrants



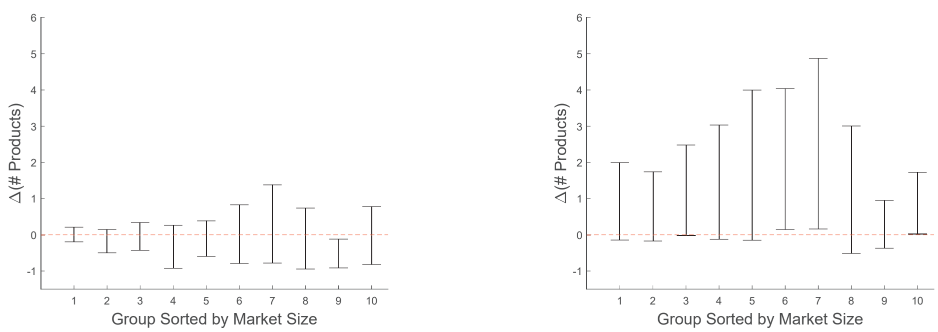
(D) Change in the Number of Products by Merging Firms



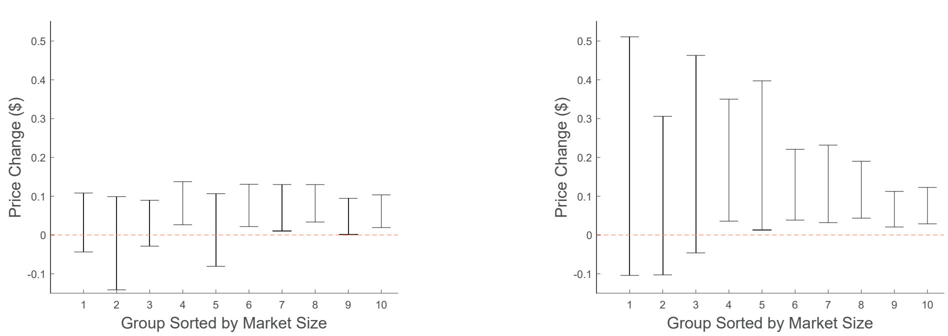
(E) Change in the Number of Products by Non-merging Incumbent Firms



(F) Change in the Number of Products



(G) Change in Craft Prices



function:

$$\theta_0 \mathbb{1}\left(\sum_{j\in\mathcal{J}_n} Y_{jm} > 0\right) + \sum_{j\in\mathcal{J}_n} Y_{jm}\left(W_{jm}\theta + \sigma_\zeta \zeta_{jm}\right),$$

which is no longer additive in the fixed cost of each product. The fixed cost function exhibits economies (or diseconomies) of scope if $\theta_0 > 0$ (or $\theta_0 < 0$).

To estimate θ_0, we additionally consider bounds for the conditional probability that a firm has at least one product in a market, denoted by $\Pr(\sum_{j\in\mathcal{J}_n} Y_{jm} > 0 \,|\, X_{nm}, W_{nm})$, where $W_{nm} = (W_{jm}, j \in \mathcal{J}_n)$ is the collection of the fixed cost covariates for firm n's potential products. We define $\zeta_{nm} = (\zeta_{jm}, j \in \mathcal{J}_n)$ analogously. Also, let $\boldsymbol{Y}_{nm} = (Y_{jm}, j \in \mathcal{J}_n)$ be firm n's product decision in market m and $\boldsymbol{Y}_{-nm}$ be the opponents' decisions. Finally, we denote firm n's maximum profit from entering the market for given $(\boldsymbol{Y}_{-nm}, X_{nm}, W_{nm})$ by

$$\begin{aligned}
&\Gamma_{nm}(\boldsymbol{Y}_{-nm}, X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta) \\
&= \max_{\{\boldsymbol{Y}_{nm} \text{ s.t. } \sum_{j\in\mathcal{J}_n} Y_{jm} > 0\}} \pi_n(\boldsymbol{Y}_{nm}, \boldsymbol{Y}_{-nm}, X_{nm}) - \sum_{j\in\mathcal{J}_n} Y_{jm}\left(W_{jm}\theta + \sigma_\zeta \zeta_{jm}\right) - \theta_0.
\end{aligned}$$

We define the minimum

$$\underline{\Gamma}_{nm}(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta) = \min_{\boldsymbol{Y}_{-nm}} \Gamma_{nm}(\boldsymbol{Y}_{-nm}, X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta)$$

and the maximum

$$\overline{\Gamma}_{nm}(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta) = \max_{\boldsymbol{Y}_{-nm}} \Gamma_{nm}(\boldsymbol{Y}_{-nm}, X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta).$$

Under the assumption that the observed brewery entry decisions are not dominated, the bounds for $\Pr(\sum_{j\in\mathcal{J}_n} Y_{jm} > 0 \,|\, X_{nm}, W_{nm})$ are

$$\begin{aligned}
&\Pr\left(\underline{\Gamma}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right) > 0\right) \\
&\leq \Pr(\sum_{j\in\mathcal{J}_n} Y_{jm} > 0 \,|\, X_{nm}, W_{nm}) \\
&\leq \Pr\left(\overline{\Gamma}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right) > 0\right).
\end{aligned}$$

Unfortunately, computing $\Gamma_{nm}(\boldsymbol{Y}_{-nm}, X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta)$ can be costly because there are many possible values for $\boldsymbol{Y}_{nm}$ in $\{\boldsymbol{Y}_{nm} \text{ s.t. } \sum_{j\in\mathcal{J}_n} Y_{jm} > 0\}$. Consequently, computing its lower and upper bounds, $\underline{\Gamma}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right)$ and $\overline{\Gamma}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right)$, is also costly. In what follows, we define a function $\underline{\underline{\Gamma}}_{nm}(\cdot)$ that is always smaller than or equal to $\underline{\Gamma}_{nm}(\cdot)$ and a function $\overline{\overline{\Gamma}}_{nm}(\cdot)$ that is always larger than or equal to $\overline{\Gamma}_{nm}(\cdot)$, and use these functions to construct the bounds for $\Pr(\sum_{j\in\mathcal{J}_n} Y_{jm} > 0 \,|\, X_{nm}, W_{nm})$.

Consider an example where firm n has three potential products, $j = 1, 2, 3$. The firm profit given the product entry decision $(1, 1, 0)$ is $\pi_n((1, 1, 0), \boldsymbol{Y}_{-nm}, X_{nm})$, which can be written as the sum of two differences: $\pi_n((1, 1, 0), \boldsymbol{Y}_{-nm}, X_{nm}) - \pi_n((1, 0, 0), \boldsymbol{Y}_{-nm}, X_{nm})$ and $\pi_n((1, 0, 0), \boldsymbol{Y}_{-nm}, X_{nm}) - \pi_n((0, 0, 0), \boldsymbol{Y}_{-nm}, X_{nm})$. Given the definition of $\underline{\Delta}_j(X_{nm})$ and $\overline{\Delta}_j\left(X_{nm}\right)$ in equation (4), we have

$$\underline{\Delta}_1 + \underline{\Delta}_2 \leq \pi_n((1, 1, 0), \boldsymbol{Y}_{-nm}, X_{nm}) \leq \overline{\Delta}_1 + \overline{\Delta}_2.$$

More generally,

$$\sum_{j \in \mathcal{J}_n} Y_{jm} \underline{\Delta}_j\left(X_{nm}\right) \leq \pi_n\left(\boldsymbol{Y}_{nm}, \boldsymbol{Y}_{-nm}, X_m\right) \leq \sum_{j \in \mathcal{J}_n} Y_{jm} \overline{\Delta}_j\left(X_m\right).$$

Define

$$\begin{aligned}
\underline{\underline{\Gamma}}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right) &= \max_{\{\boldsymbol{Y}_{nm} \text{ s.t. } \sum_{j \in \mathcal{J}_n} Y_{jm} > 0\}} \sum_{j \in \mathcal{J}_n} Y_{jm}\left(\underline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta \zeta_{jm}\right) - \theta_0, \\
\overline{\overline{\Gamma}}_{nm}(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta) &= \max_{\{\boldsymbol{Y}_{nm} \text{ s.t. } \sum_{j \in \mathcal{J}_n} Y_{jm} > 0\}} \sum_{j \in \mathcal{J}_n} Y_{jm}\left(\overline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta \zeta_{jm}\right) - \theta_0.
\end{aligned} \tag{SC.1}$$

We have $\underline{\underline{\Gamma}}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right) \leq \underline{\Gamma}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right)$ by the max-min inequality and $\overline{\Gamma}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right) \leq \overline{\overline{\Gamma}}_{nm}(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta)$ by the definition of $\overline{\overline{\Gamma}}_{nm}(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta)$.

The integer programming problem in (SC.1) can be solved quickly given the additive structure. Specifically,

$$\begin{aligned}
&\underline{\underline{\Gamma}}_{nm}(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta) \\
= &-\theta_0 + \begin{cases} \sum_{j \in \mathcal{J}_n}\left[\underline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta \zeta_{jm}\right]_+ & \text{if } \exists j \text{ s.t. } \underline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta \zeta_{jm} > 0, \\ \max_{j \in \mathcal{J}_n}\left\{\underline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta \zeta_{jm}\right\} & \text{otherwise.} \end{cases}
\end{aligned}$$

In other words, we simply need to calculate the value of each $\underline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta \zeta_{jm}$, sum up all the positive terms, and subtract θ_0. If $\underline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta \zeta_{jm} < 0$ for all $j \in \mathcal{J}_n$, we calculate the bound as the maximum of $\underline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta \zeta_{jm} - \theta_0$. Similarly,

$\overline{\overline{\Gamma}}_{nm}(X_{nm}, W_{nm}, \zeta_{nm}, \theta)$ is given by

$$\begin{aligned}
&\overline{\overline{\Gamma}}_{nm}(X_{nm}, W_{nm}, \zeta_{nm}, \theta) \\
=& -\theta_0 + \begin{cases} \sum_{j\in\mathcal{J}_n} \left[\overline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta\zeta_{jm}\right]_+ & \text{if } \exists j \text{ s.t. } \overline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta\zeta_{jm} > 0, \\ \max_{j\in\mathcal{J}_n}\left\{\overline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - \sigma_\zeta\zeta_{jm}\right\} & \text{otherwise.} \end{cases}
\end{aligned}$$

In the end, we use the following lower and upper bounds for estimation:

$$\begin{aligned}
&\Pr\left(\underline{\underline{\Gamma}}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right) > 0\right) \\
&\leq \Pr\Big(\sum_{j\in\mathcal{J}_n} Y_{jm} > 0 \,|X_{nm}, W_{nm}\Big) \\
&\leq \Pr\left(\overline{\overline{\Gamma}}_{nm}\left(X_{nm}, W_{nm}, \zeta_{nm}, \theta, \sigma_\zeta\right) > 0\right).
\end{aligned}$$

We combine moments associated with firm entry and those associated with product entry for estimation. For moments associated with product entry, we modify the bounds of the conditional choice probability of an individual product's outcome to take into account θ_0:

$$\begin{aligned}
&F_\zeta\left(\zeta_{jm} < (\underline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - [\theta_0]_+)/\sigma_\zeta\right) \\
\leq& \Pr\left(Y_{jm} = 1\,|X_{nm}, W_{jm}\right) \\
\leq& F_\zeta\left(\zeta_{jm} < (\overline{\Delta}_j\left(X_{nm}\right) - W_{jm}\theta - [\theta_0]_-)/\sigma_\zeta\right).
\end{aligned}$$

The same set of non-negative g functions as those in the baseline estimation are used to construct the moments associated with individual products. For moments associated with firm entry, we use $\left(\underline{\Delta}_j(X_{nm}), \overline{\Delta}_j(X_{nm})\right)$ of the top three most profitable products of firm n to define the non-negative g functions. If a firm has only one (or two) potential products, we set the g functions corresponding with the second and third products (or the third product) to be 0.

We report the 95% projected confidence interval in Table SC.3. For comparison, we also copy the results from the baseline specification in the first column. We find evidence for diseconomies of scope in the small and large markets. The diseconomies of scope in the small markets may be explained by the limited shelf space of the retailers in these markets. In large markets, on the other hand, retailers typically require brewery sales representatives to ensure products are well stocked and expired products are removed quickly and the logistic challenges could increase with the number of products, again leading to diseconomies of scope.

We report the counterfactual simulation results in Figure SC.2 (merger effects on entry,

Table SC.3: Fixed Cost Estimates: Projected 95% Confidence Interval, Allowing for (Dis)Economies of Scope

	Baseline Model	(Dis)Economies of Scope
Craft (θ_1)	[255.32, 583.95]	[213.49, 635.58]
In State× Craft (θ_2)	[-276.04, 37.57]	[-384.14, 104.57]
Market-size specific fixed cost (θ_3)		
Small market	[426.52, 721.90]	[864.66, 1264.90]
Medium market	[720.77, 1014.10]	[697.71, 1174.91]
Large market	[2028.00, 2453.69]	[3251.24, 3722.16]
Market-size specific std. dev. (σ_ζ)		
Small market	[0.02, 105.80]	[0.08, 211.94]
Medium market	[1.18, 207.45]	[0.09, 251.42]
Large market	[767.59, 1044.93]	[20.85, 539.83]
Market-size specific firm entry cost (θ_0)		
Small market		[-692.29, -233.74]
Medium market		[-443.84, 21.76]
Large market		[-2566.82, -1986.40]

Note: Estimates in 2016 US dollars.

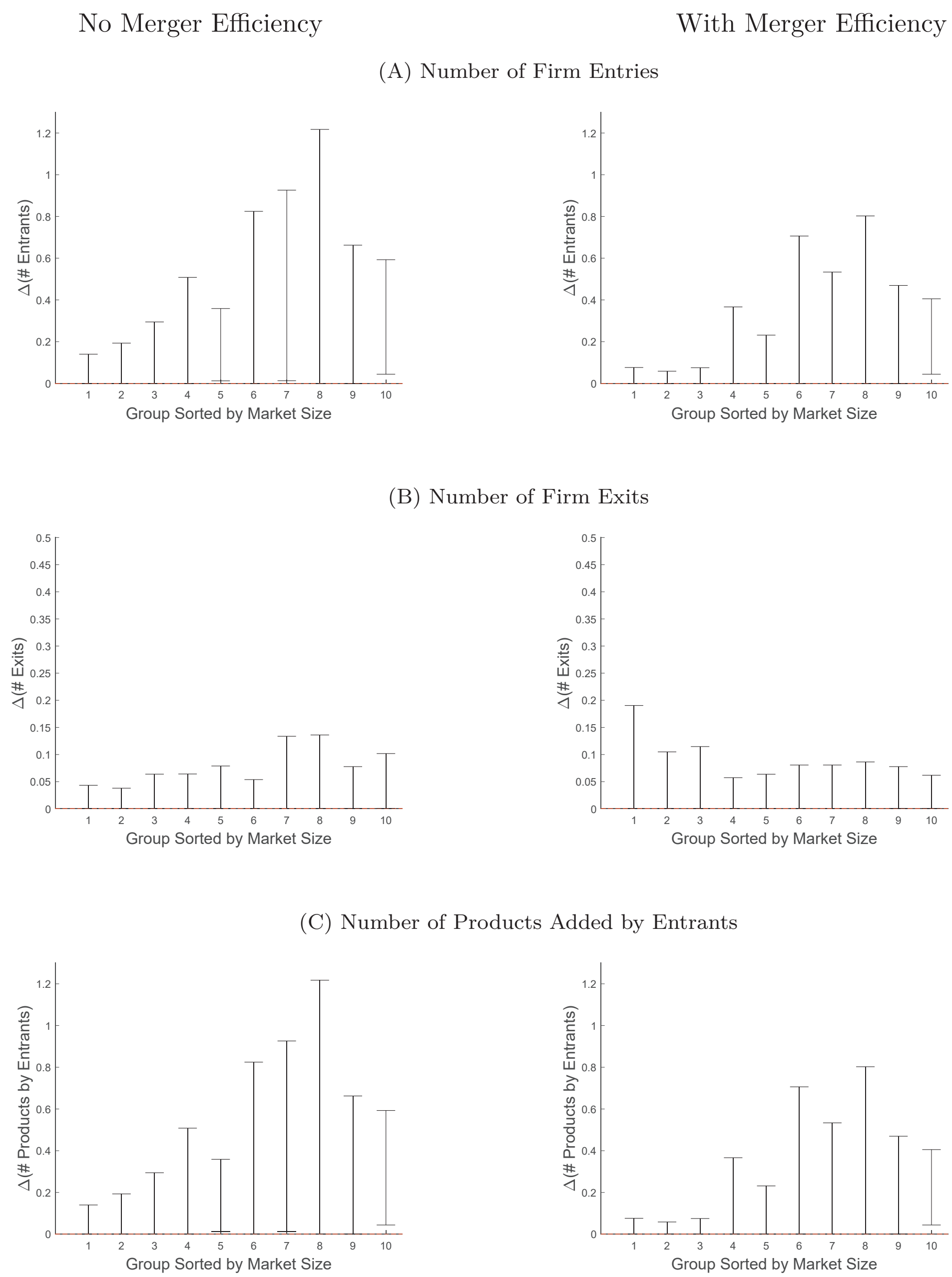
product variety, and prices) and Table SC.4 (welfare effects). From the comparison of Figure SC.2 to Figure 7 and Table SC.4 to Table 8, we can see that while the estimates are somewhat different, the merger effects are very similar to the baseline results.

Table SC.4: Welfare Effects, Allowing for (Dis)Economies of Scope

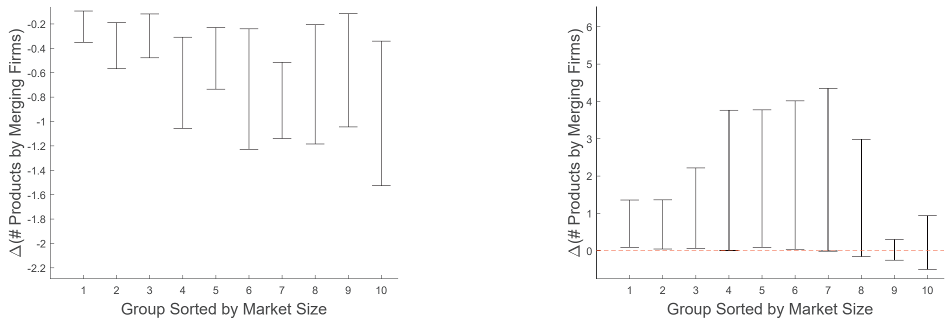
		No Merger Efficiency	With Merger Efficiency
(1)	total surplus ($1000)	[-839.00, -240.21]	[-478.45, -122.95]
(2)	craft beer profits ($1000)	[24.94, 323.19]	[2.20, 307.26]
(3)	consumer surplus ($1000)	[-1159.80, -265.15]	[-750.98, -165.63]
(4)	due to variety change	[-148.74, 48.36]	[-7.34, 486.75]
(5)	due to entry	[1.71, 128.49]	[1.10, 74.58]
(6)	due to incumbent product adjustments	[-183.70, -40.89]	[-8.45, 420.80]

Note: this table reports the aggregate welfare effects of the merger across markets. For each measure, we report the range across the vectors of parameters sampled from their 95% confidence set. The left panel shows the results without considering any merger efficiency, while the right panel reports the results that incorporate reductions in fixed costs when a craft brewery is acquired by a macro brewery.

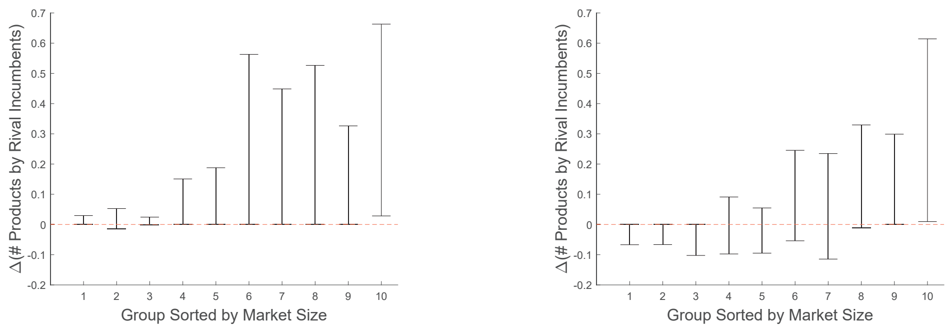
Figure SC.2: Merger Effects on Entry, Product Variety, and Prices, Allowing for (Dis)Economies of Scope



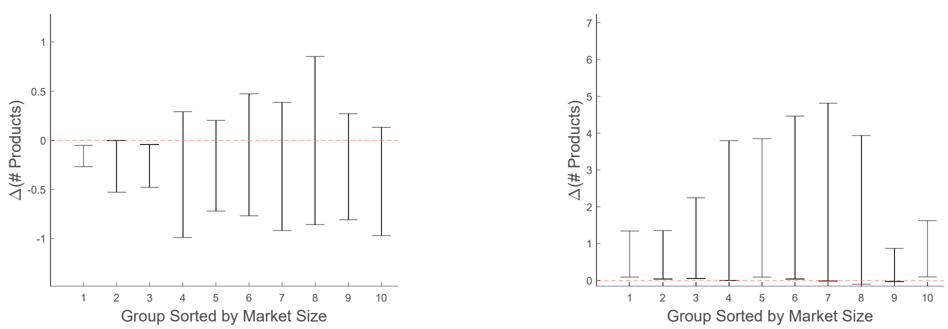
(D) Change in the Number of Products by Merging Firms



(E) Change in the Number of Products by Non-merging Incumbent Firms



(F) Change in the Number of Products



(G) Change in Craft Prices