Essays in Industrial Organization

by

Yeon Ju Baik

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2022

Date of final oral examination: 05/19/22

The dissertation is approved by the following members of the Final Oral Committee:
Alan Sorensen, Professor, Economics
Jean-Francois Houde, Professor, Economics
Kenneth Hendricks, Professor, Economics
Sheldon Du, Associate Professor, Agricultural, Applied Economics

To my family, Ye Gu and Hannah

ACKNOWLEDGMENTS

I would like to thank my advisors, Alan Sorensen, Jean Francois-Houde, and Kenneth Hendricks, for their support. Their advice and guidance helped me from the start, such as finding good insights from the data, and a nice model that can be used for analysis, writing, and finishing the job market. I have learned a lot from presenting and discussing with them. I also would like to thank Chris Sullivan, Lorenzo Magnolfi, and Xiaoxia Shi, who provided me with valuable comments. I'm also grateful to work with Darcy F. Kamal and Madhu Viswanathan, who allowed me to explore econ-related but new areas.

Also, a big thank you to fellow Ph.D. students. In particular, great support from the IO group, Natalia, Erika, Sarah, Cici, Renata, and Xinrong, assisted me in navigating the job market successfully. A special mention goes to Jeff Rudd, who supported writing the job market paper. I also thank Itzel, Gaurav, Jiarui, Min Sun, Young Jae, Annie, Joo Hyun, Soo Jeong, Hae Young, Hyun Kyeong, Ho Young, Jingnan, Deborah, Hye Jin, Yong Ki, and other fellow Ph.D. friends.

Last but foremost, I want to acknowledge the support from my family. Ye Gu, Hannah, and my parents are the source of my energy and motivation.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis contains two essays on industrial organization, particularly matching and strategic interaction between the players. The first chapter examined the matching in the online platform and focused on how the advertising changes the matching pattern. The last chapter studies matching in the electricity market.

In the first chapter, I examined how advertising affects equilibrium prices in an online marketplace by using unique data from a Korean online resale market. The posting frequency in the used market is used as a measure for advertising. Using the posted price, I infer how sellers compete in price using the empirical implications of the Armstrong and Vickers (2020) framework of oligopolistic sellers. The analysis shows that sellers with more frequent listing charge prices with first-order stochastically dominant distributions than sellers who advertise less. Sellers who post less face more elastic demand than the frequent posters, resulting in higher markups for the frequent posters. Repeated posting benefits frequent posters and platforms by increasing the market price and chances of sales but can harm consumer welfare.

The second chapter studies matching in the energy market using the case of the power purchase agreement, which is a way of selling and purchasing electricity generated from renewable sources at a fixed price over long periods. This paper investigates the link between wholesale market risk and the equilibrium prices of these contracts. We first present a stylized model of PPA equilibrium to fix intuition on the relationship between PPA prices, wholesale prices, and market volatility. We then test the model predictions using data on all utility-scale wind projects. Results suggest that the mean retail electricity prices and wholesale price volatility positively correlate with PPA prices, whereas the volatility measure does not strongly correlate with the equilibrium price. These findings highlight

how the participation of both buyers and sellers would affect equilibrium prices in the PPA market.

1 POSTING FREQUENCY AND PRICING IN AN ONLINE PLATFORM

1.1 Introduction

In the online marketplace, product visibility, such as search ranking, plays a critical role for the consumer choice. Therefore, sellers put considerable effort to make the product has a better search ranking, often called as "Prominent position" in the literature. In a secondhand market such as Facebook Marketplace or Craigslist, the sellers use "reposting", which is a way to renew their postings to make them visible at the top of the listings.

The sellers' diverse level of visibility can affect the market equilibrium by changing the matching probability of each product, and the probability of sales. Eventually, it will affect how sellers charge price of their products. The changes in price that varies by the level of advertising or the product visibility can be used as a tool to uncover how sellers view the relationship between the price competition and product visibility.

The direction of price movement induced by the increase in advertising is not yet clear. Various reasons support a negative direction: the sellers with higher visibility are likely to face a broader group of consumers often having less willingness to pay, leading to lower price(Rhodes (2011),Armstrong et al. (2009)). Also, sellers may enjoy economies of scale by selling the product in a large amount, and lower the price (Bagwell and Ramey (1994)). On the other hand, the search cost incorporated with the consumer choice can give more advantage to the visible product, which makes it easier for the seller to increase the price(Ursu (2018)).

This study uses sellers' frequency of postings in the secondhand online platform as an analogy to the advertising in the online market. This paper aims to understand the role of advertising and how it shapes the price competition by connecting the posting frequency and the sellers' pricing choices. By collecting hourly level listings and pricing decisions of individual sellers in an unique used cellphone trading platform, the choice of posting frequency and pricing are tracked.

To understand the relationship between the advertising and pricing, I borrowed the framework of competitive interaction from Armstrong and Vickers (2020) to analyze underlying sellers' pricing competition that is formed by effort for advertising. I identified the competition form by applying the prediction of model(Armstrong and Vickers (2020)) on pricing patterns of the data. Consequently, the suggested price competitive framework enables me to quantify online sellers' market power, which is obtained from the visibility effort. The market power can explain why a few power sellers often dominate online markets, and being prominent can increase the market price, decreasing consumer welfare.

The data is from an online platform,Cetizen, that specializes in the used cellphone trade. The platform provides an empirical setting for studying both the decision for posting to increase visibility (i.e., advertisement) and the pricing decision. I used posting frequency as a proxy to measure advertisement intensity, and the number of postings determines the size of potential consumers that the seller can approach. The platform does not intervene in the ranking process, and there is no sponsored option for attaining visibility or algorithmic pricing. Thus, the environment works as a clean setting to look only at the re-posting and pricing decisions of the sellers. The empirical analysis applies Chetverikov et al. (2020)'s recently developed stochastic monotonicity test, and Wilcox et al. (2014) tests and quantile regression to identify competitive interaction structures.

The analysis suggests a nested reach interaction structure based on the seller's pricing decisions, where seller with less posting face more elastic demand. The practice of re-posting can increase the platform's revenue by allowing the power sellers to raise prices and also by helping them to sell the products. Advertising, i.e., reposting, plays a role as a medium that can link or match the potential sellers and consumers. The empirical analysis suggests that the advertising affects not only the probability of choice that much of the literature focused on (Ursu (2018),Hunold et al. (2020),Mela et al. (1997),Tellis (1988)) but also on the price level. How the advertising matches the seller and buyer determines the sellers' competition. The price competition can directly affect how much market power the sellers can attain.

This analysis fills some gaps in the literature by providing evidence about the online resale market and seller behavior. Although the resale market is a rapidly growing industry, including Facebook Marketplace, Craigslist, and Offer-up, most of the empirical research focuses on certain types of products like used books (Hong and Shum (2006), Ellison and Ellison (2018)) or video games (Ishihara and Ching (2019)). Moreover, unlike the extensive studies of consumer search behavior, few publications describe individual seller behavior in the online platform (Einav et al. (2018)).

Section 2 places this study in the context of relevant literature. Section 3 introduces the data derived from Cetizen's online platform, and section 4 describes the theoretical model and predictions of Armstrong and Vickers (2020), linking competitive interaction structures and pricing choices. Finally, section 5 describes and applies the empirical testing strategy. Then I'll discuss the motive of the platform and the implications of the finding, with a conclusion.

1.2 Literature

This study complements and intersects with three areas of research: 1) advertisement and pricing strategy of a seller; 2) search and consideration process identification, 3) effect of search ranking. First, this paper links to the literature on seller's pricing, beginning with the theoretical studies on how the sellers compete with advertising, and the corresponding pricing equilibrium (McAfee (1994),Armstrong and Vickers (2020),Stahl II (1994)). A foundational paper by Butters (1977) used advertisement, limited by specific theoretical conditions, to develop an equilibrium analysis of the homogeneous goods market. Stahl II (1994) extended the Butters (1977) model and attached consumer search to a simultaneous decision of pricing and advertisement. McAfee (1994) added the effect of endogenous choice of an advertisement on the equilibrium pricing strategy. In contrast to the Stahl II (1994) simultaneous decision model, the McAfee (1994) model resembles a sequential decision: seller determines the level of advertisement which affects visibility, then decides the price. McAfee (1994) showed why the seller with the largest availability rate is likely to charge a monopoly price. The theoretical literature extended advertising by incorporating the consumer search model and applied the framework to understand the seller's pricing decisions.

Empirical papers studying sellers' pricing behavior documented heterogeneity across the sellers in the online marketplace. Huang (2021) looked at the friction in pricing behavior in Airbnb and quantify its impact on the equilibrium, focusing on the difference between the single listing seller and the multi-listing sellers. In addition, he examined the dynamic pricing aspect of Airbnb and heterogeneity in seller's pricing, which is prevalent in the online commerce market. Jolivet et al. (2016) looks at how reputation affects pricing. They describe the causal effect of reputation on the pricing and how it differs by the seller type and category of the items. Similarly, Hui et al. (2016) quantified the effects of reputation badges and buyer protection programs on the price. The reputation badge, which indicates top-rated sellers in the platform, is positively associated with the price. This empirical literature shows that seller heterogeneity in the online marketplace is one of the critical factors in the pricing decision, and it relates to the reputation and information asymmetry among the sellers.

Second, this paper is closely related to the literature on the search, consideration process identification. Empirical identification of a consideration set is important for accurate analysis of price elasticity and substitution behavior. Most research in this area relies upon additional data which looks at search process (De los Santos et al. (2012)), directly observed consumer search process (Honka (2014)) and survey (Honka et al. (2017)) to identify consideration set. The studies that focus on testing the search models are closely related to this paper. Hong and Shum (2006),Chen et al. (2007) used observed price data to match sequential search and fixed sample search models, finding qualitatively little difference between the two. De los Santos et al. (2012) used the data on web browsing and online purchase to test between sequential search and simultaneous search models. The key difference between the sequential and simultaneous search is whether the consumer revisits the store. Honka and Chintagunta (2017) used price variation inside the consumer's consideration set and matched with the price predictions from the two aforementioned search models. Even without the sequence of searches, they suggested a way to identify search methods from the data. They suggested a simultaneous search model for the consumer shopping behavior in the US auto insurance industry.

Marketing and economics scholars have extensively studied, from a consumer-centric perspective, consumer search cost and its relation to price (Bronnenberg and Vanhonacker (1996), Hauser and Wernerfelt (1989)). However, few studies locate the seller at the center of their investigation into price range and consideration set formation. Pancras (2010) studied the relationship between the pricing decision of a seller and consideration set formation in the discrete choice setting. Eliaz and Spiegler (2011) studied the seller's competitive marketing tool to influence the consideration sets using a theoretical model. Although this study focused on a seller's strategic response to consideration set formation, it did not consider seller's pricing strategy. Armstrong and Vickers (2020) classified competitive interaction patterns and studied how these patterns affect a seller's pricing behavior. Their work focused on the homogeneous goods market and integrated segmented literature with different settings of search models (Rosenthal (1980), Varian (1980), Burdett and Judd (1983), McAfee (1994)). This study's empirical perspective builds on the work emerging in the field of the seller- or seller-centric studies into interaction structures, considerations sets, advertisement intensity, and price range. It could provide insights into how advertising interacts with a seller's decisions about which competitors are significant.

Lastly, this paper closely relates to the literature on "prominence" and ranking in the online market. Prominence refers to the advantage of being sampled first by the consumers. The theoretical literature on prominence predicts the lower price of the more visible options. Rhodes (2011) argues consumers can have different values of the product, and their choice might depend on the implicit value. For example, the more common option with higher advertising intensity is likely to be chosen by the consumers who value less due to the search cost. Therefore, they predict prominent options with higher visibility would have lower prices than other options. Similarly, Armstrong et al. (2009) look at the difference in types of demand that are attracted to prominent option. The prominent option is likely to face fresh demand sensitive to price, while the non-prominent options face recurrent demand that is less sensitive. Thus, the prominent option would have lower prices. Bagwell and Ramey (1994) look at the seller's advantage of having a higher chance at the prominent position. Because of the higher chance of sales, the seller in a prominent position can acquire economies of scale by selling more, leading to lower prices. In this view, search cost created by advertising causes a price difference.

Empirical papers which studied search ranking found the prominent option has higher prices. Ursu (2018) look at the effect of search ranking in Expedia on consumer choices and pricing equilibrium. She showed rankings reduce the search cost, increases the chances of consideration using the experimental data. In other words, the option on the high rank can charge a higher price since the consumer they face might have higher search costs, less willingness to search for other options. Additionally, many papers look at the causal effect of ranking on consumer choices. De los Santos and Koulayev (2017) focused on the intermediary decision on the rank to increase the click-through rates. The biggest challenge is that the ranking has potential endogeneity concerns. For example, sponsored item in Amazon requires an additional fee for attaining visibility, which can be transferred to the price (Armstrong and Zhou (2011)). Also, the sellers who advertise intensely are likely to have a higher quality product with a higher price (Athey and Ellison (2011)).

This study relies upon an econometrics method to bridge the divide between theoretical models and the empirical data. The relationship between sellers' pricing behavior and advertisement choices may be expressed by stochastic dominance of price distribution. In this paper, I applied Chetverikov et al. (2020)'s method to statistically test whether the pricing distribution follows first order stochastic dominance pattern. The method of testing stochastic monotonicity is studied in Lee et al. (2009), Chetverikov et al. (2020), Delgado and Escanciano (2012). Lee et al. (2009) developed the hypothesis of stochastic monotonicity, and derived asymptotic distribution of test statistic. Delgado and Escanciano (2012) uses the Copula function approach to circumvent the smoothness problem while Chetverikov et al. (2020) uses adaptive testing, which can be adapted to the unknown smoothness level of the function, $F(Y|X)$. The test statistics are based on the differences of the conditional CDF for different values of the conditioning variable X. Chetverikov et al. (2020) enables testing the null hypothesis stochastic monotonicity of price distribution for reach.

1.3 Data

The data is from one of the biggest used cell phone trading platforms in Korea, Cetizen. About 20% of used cell phones in Korea are traded on the platform. I used a python crawler to collect hourly level listings of the two main cell phone manufacturers, Samsung and Apple, products from February 5th to August 29th, 2020. I automate the crawler to run every hour to collect all the postings that appear on the platform. Each posting includes information about the product, such as condition, warranty information, memory size, etc.

This platform provides a nice environment to study the role of advertising in online market place because this platform doesn't have any additional tool to boost up the visibility in the website. Unlike Ebay or Amazon, the platform do not require any additional fee for promoting a specific product. and there is no inside bargaining between the sellers and the platform for a product's position. Additionally, when the item is sold, the item is withdrawn from the list, which leaves a sign of being removed[1]. I collected the signs to track if the item finally sold.

Buyer's Choice

A buyer who enters the platform can search keywords. Also, the buyer can filter product categories to search for multiple options. The filter contains the model name, manufacturer, price range, and other product characteristics. After filtering the upper-level category, the buyer faces a list of postings organized in chronological order. For example, assume a buyer searches for a specific model such as Galaxy S9. The platform provides a list of products that have been posted recently by various sellers. For example, Figure 1.1 contains the list of Galaxy Quantum 2 sold by various sellers. The order of the listing is chronological - the newest listing

[1] Consellered by the platform manager.

goes on top of the page. Each posting consists of product characteristics such as product condition, whether they are under the warranty period, timestamp of the listing, model name, seller identification, and price.

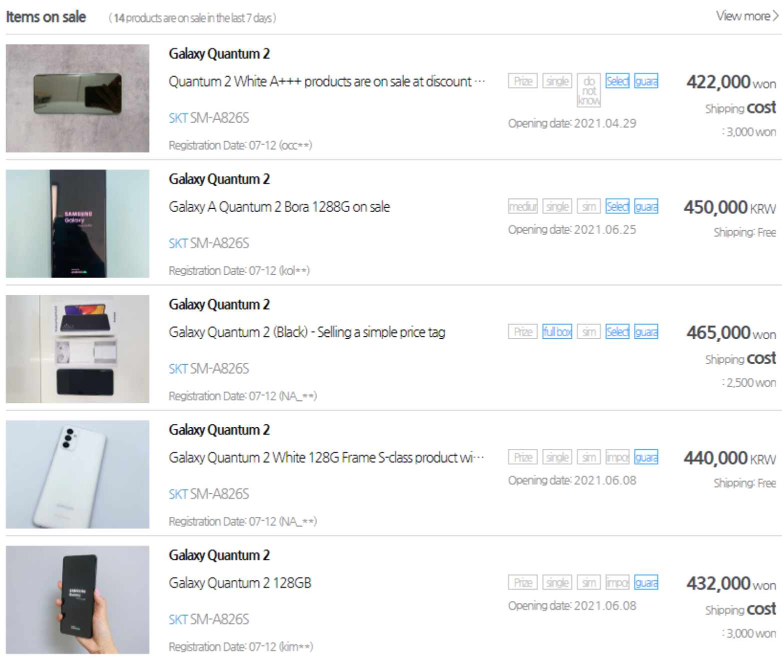


Figure 1.1: Model Listings: "Galaxy Quantum 2"

Seller's Choice

Since the listing is in chronological order as in Figure 1.3 with the old postings disappearing quickly, the sellers have an incentive to remain at the top of the listing to attract more consumers. Therefore, the sellers have an incentive to repeatedly post the product to stay at the top of the listing.

There could be two different strategies of the seller to increase the website's visibility. The first strategy is to make duplicated postings. This strategy is observed in Figure 1.3; at the top of the screenshot, there is "(Unused Refurbished) Note10 Plus Black Samsung Official New Product,"

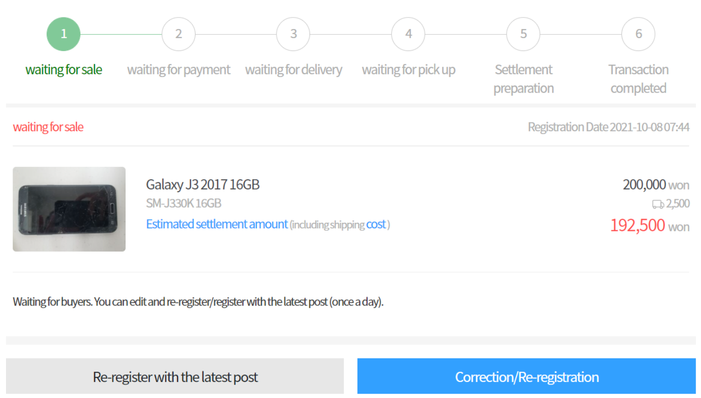


Figure 1.2: Seller panel

and it is repeated in the middle of the page. The sellers post the same listing multiple times to occupy more spaces.

The second strategy is to use reposting, which is a way of renewing the posting in the list as in Figure 1.2. After the point of initial listing, the sellers may see the option to re-register or renew the posting. By doing so, the listing can be refreshed and would bump to the top of the page.

I'm calling these two different strategies "reposting." The main reason for reposting is to continually occupy a position near the top and front pages of the listings, attracting more attention from consumers.

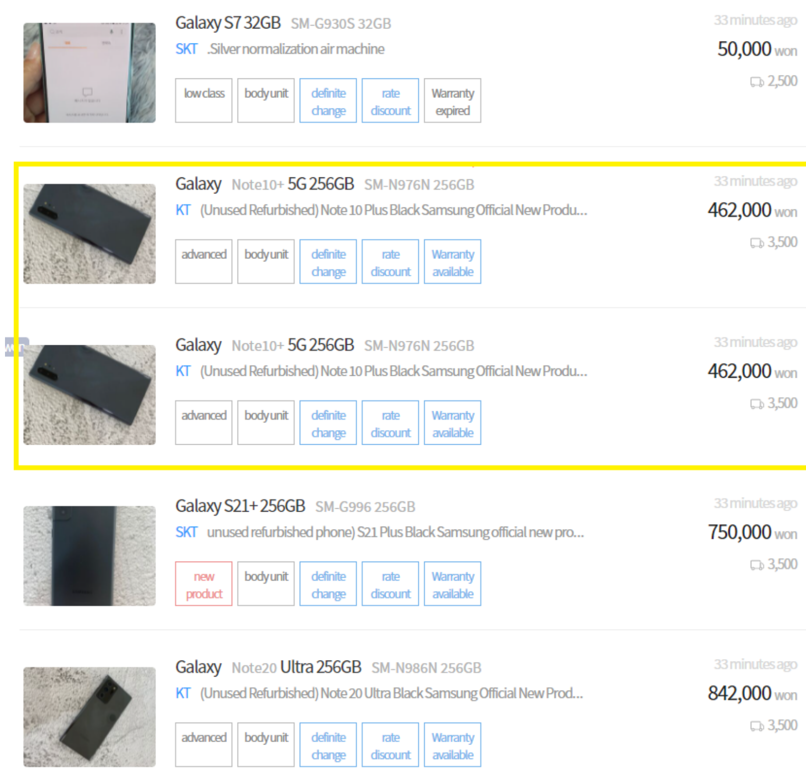


Figure 1.3: Default Website Listings

Among the total number of listings that I scraped (810,585), only 500,482 listings contain the whole characteristics of the product, including condition, information on the warranty, and memory size. These listings can contain multiple listings made for the same model due to the duplicated listing or re-posting. The number of unique listings is 104,173. Out of 500,482 listings, I'm using about 48% (248,497) of the listings with the new model price (such as Galaxy S9, S10, etc.). The number of models included in the filtered list is 15. I'm normalizing the posted price by dividing the value by the new product price. Also, by analyzing the price ratios, we can see the items sold are likely to be cheaper than the others. The ratio between the listed price and the average price in the platform of sold items is 0.964, which means the price of sold items is only about 96.4% of the market price. The summary statistics are in the appendix.

Table 1.1: Number of Observations

Data cleaning	Number of observation
Total number of postings	810,585
Unique postings	104,173
Sold items (including duplicated postings)	116,018
Sold items (only unique postings)	10,774
With original price	242,326
Number of Models	15
Price ratio of sold items	0.964 (0.143)
Price ratio of unsold items	0.98 (0.19)

* Price ratio means the ratio between the posted price to the average market price of a particular cellphone model. Price ratio of sold or unsold items are calculated in each unique posting level. Standard deviations are in the parentheses.

Empirical Foundations: Stylized Facts

First, the effort of reposting, or updating your posting again to remain in the top of the listing page, seems to help the sales of the product. In Table 1.2, I ran linear probability regression model to see if the number of repeated posting affects the selling probability. The unit of observation is a unique listing and the dependent variable is whether the item is withdrawn from the listings, which represents the status of being sold in the platform. Based on the statistically significant positive coefficient in the number of repeated posting daily rows which captures how frequently the posting is repeated, duplicated postings and reposting are helping the sellers sell their products on the platform. The negative coefficient of the price ratio shows that consumers primarily respond to the price. Within the same quality product, consumers would choose the cheaper option in general.

Table 1.2: Probability of Product Sales

LPM	Sold	Sold	Sold
Number of repeated posting/day	0.0470***	0.0468***	0.0478***
	(0.00531)	(0.00477)	(0.00473)
Price ratio ($)	-0.0593***	-0.0651***	-0.0786***
	(0.00802)	(0.00843)	(0.0102)
Model share	-0.380		
	(0.443)		
Average number of postings by a seller/day		-0.000568	-0.000643
		(0.000300)	(0.000330)
Day gap			0.00112***
			(0.000218)
Const	0.157***	0.0989***	0.106***
	(0.0291)	(0.0115)	(0.0137)
Controls	Yes	Yes	Yes
Seller FE	Yes	No	No
N	85376	104169	73115
R-sq	0.049	0.007	0.009

Note: Standard errors clustered in month and model level are in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, "Number of repeated posting (Day)" is number of repeated posting of a unique listing. "Price Ratio" is the product price ratio to the average market price. A "model share" represents the posting share of a specific model within the total postings on the website that are observed within a day. "An average number of postings by a seller/day" is the number of average daily postings that a seller uploads to the platform. "Day Gap" is the average number of days that elapse between the duplicated postings. Controls include conditions of the product, whether they are under warranty, and memory size. Column 1 uses only the individual postings made by sellers who sell more than 5 different cellphone models in the market. Column 2 uses observations made by any sellers, and Column 3 only uses the listings that are posted multiple times.

Second, the intensity of repeated posting affects the probability of sales and the absolute size of prices. This result is consistent with the predictions from the literature(McAfee (1994), Armstrong and Vickers (2020)), that the product with more advertisement (larger reach) has a higher price. The regression result is in Table 1.3. The 2nd and 3rd rows that used the number of hourly postings made by each seller positively correlate with the price. The number of postings made by a seller per hour captures seller heterogeneity, whether it is frequent or less frequent. In the 3rd row, the number of postings of a cellphone model by a seller captures variation within a seller. Each seller's number of postings shows that the actively posting sellers are charging higher prices than the others. So the type of seller can affect the price. The variation within the seller also shows that frequently posted products are likely to have a higher price.

By making duplicated postings, the sellers may attain a favorable position in the platform by occupying the listings' top or front page. This increases sellers' chances of meeting consumers who have higher search costs, and greater willingness to pay. Eventually, these sellers have incentives to charge higher prices because consumers with higher search costs will not leave a particular seller if the marginal benefit of continuing the search is less than the cost of staying on the same page. This claim is supported by Figures 1.4 and 1.5 which show the posted price distribution for two different models. Here, the price is defined by the ratio between the posted price and the model's original price. The two groups are defined by the number of sellers' daily postings. Group 3 is the sellers posting on average more than 10 times per day, while Group 1 is the sellers who post less than 3 times a day. The groups show a distinct difference. Group 3's posted price of more than 10 times a day skews right more than Group 1 with less than 5 postings.

Table 1.3: Price Difference and Number of Postings

	Price ($)	Price ($)	Price ($)
Num posts for a cellphone/Hr	-0.263*		-0.408***
	(0.102)		(0.0989)
Num posts by a seller/Hr		0.0588***	0.0206
		(0.0120)	(0.0177)
Num posts by a seller for a cellphone/Hr			1.670***
			(0.417)
Controls	Yes	Yes	Yes
Model FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes
N	38961	38961	38961
R-sq	0.948	0.948	0.948

Note: The regression is conducted for the sellers who posts more than 20 postings within one hour. Unit of analysis is a posting with a unique description, Standard errors clustered in month, model level are in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, "Num posts for a cellphone/Hr" is the total number of postings that are posted within one hour for a specific model, "Num posts by a seller/Hr" is the number of postings that are made by a seller within one hour. "Num posts by a seller for a cellphone/Hr" is the number of postings made by a seller for a specific model within one hour. Controls include conditions of the product, whether they are under the warranty, and memory size.

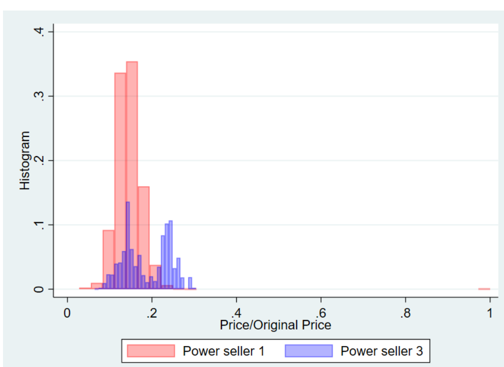


Figure 1.4: Price : Galaxy A8

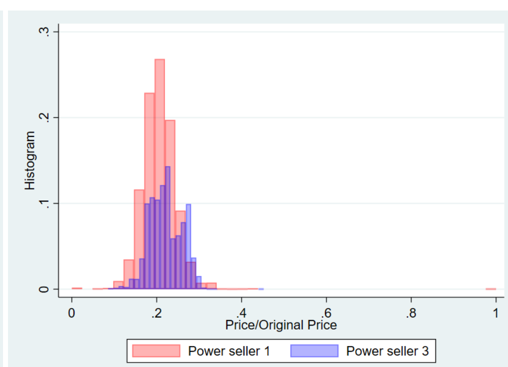


Figure 1.5: Price : Galaxy S9

[2]

[2] Group 3 are the sellers who post more than 10 times for a day on average, while

Potential Models of Seller Pricing Competition

The previous regression result shows the positive association between the frequency of the listings and higher prices. However, the analysis does not show how the frequency of the postings plays a role in the pricing decision. The posting frequency can affect seller's pricing decisions by forming how the sellers compete with each other. For example, assume one seller posts a lot more than other competitors in the market. This seller's consumers would be much broader than the others, and only the sellers with similar advertising levels would be the competitors that this seller might consider.

We can easily see the seller who make more postings would have broader, larger consumer groups. In other words, the size of consumer groups of each seller would be proportional to the frequency of postings. On the other hand, whether each seller's matching probabilities are independent is unclear. For example, if all the consumers of less frequent sellers face postings by frequent sellers most of the time, then the matching probability of less frequent sellers and more frequent sellers would not be independent.

Armstrong and Vickers (2020) explored the pricing pattern of sellers in oligopolies with symmetric and asymmetric interaction structures. They linked each seller's probability of being considered to sellers' pricing decisions. If competitors interact with each other by sharing the group of consumers who are sensitive to price, then the pro-competitive effect causes the price to be lower. If the competitors do not share consumer groups,i.e., have more "captive groups," the price range is likely higher. The interaction structure implies sellers' competition; it predicts how sellers price their products within the different interaction structures. The size of consumers that are shared with the other competitors forms the shape of interaction structure, and it is the function of matching probability. The

Group 1 are the sellers who post less than 3 times a day. The price is normalized between 0 and 1 by dividing the market price by the original price.

next section discusses how the formal model of interaction and theoretical predictions is applied to data.

1.4 Model of Interaction

This section explains the competitive interaction model of Armstrong and Vickers (2020) to assess conditions of price distributions.

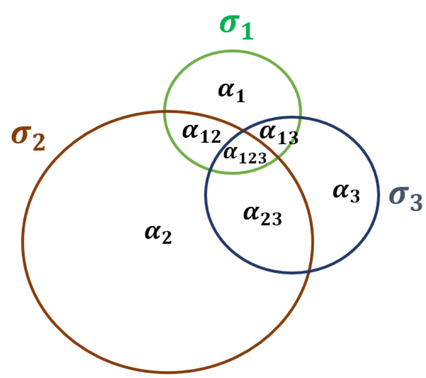


Figure 1.6: Interaction structure

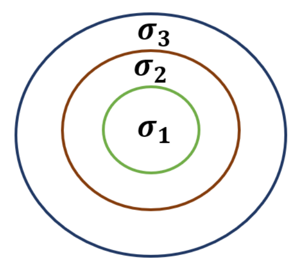


Figure 1.7: Nested

Armstrong and Vickers (2020) assumes each seller supplies a homogeneous product to the market, and consumers are willing to pay up to 1 for a unit of the product. The consideration set of each consumer can be diverse since it is the subset of all possible combinations of sellers in the market. The subset is defined as S, which is the subset of all sellers in the market $N = \{1, 2, \cdots, n\}$. The intensity of a seller's advertisement (i.e., posting activity) determines the probability of being included in the consumer's consideration set, and it is the size of a seller's circle in Figure 1.6 and 1.7. The model calls the matching probability "reach," or σ. For example, the likelihood of meeting seller i is σ_i, representing the fraction of consumers considering products from a particular seller i.

σ or the reach is the sum of various types of consumers with consideration sets. For example, the consumers who consider the products inside the set S are the sum of consumers who only consider each different product inside the subset of S'. $\sigma_S = \sum_{S'|S' \subset S} \alpha_{S'}$. Graphically, as in Figure 1.6, the reach of a seller is the sum of various consumer types who have different consideration sets. The type of consumers is α, and the number in subscript means the options that a consumer considers. For instance, seller 1, σ_1 is the sum of $\alpha_1 + \alpha_{12} + \alpha_{123} + \alpha_{13}$. So σ_1 is the measure of consumers considering seller 1's product and other sellers' products. The consumers only consider each different product inside the set S, which is the subset of all sellers in the market N, is represented as α_s. More intense advertisement causes a larger reach, increasing the likelihood more consumers will consider the product.

Figures 1.6 and 1.7 depict two competitive interaction structures among three sellers (S1, S2, S3) that correspond to consumers' different consideration sets. Figure 1.6 is an example case where three sellers are competing, and each seller's consumers are overlapped. Figure 1.7 is a nested reach case. One extreme case of Figure 1.6 is the independent case where the probability of meeting each seller is independent. To be specific, the independent structure can be defined as follows. For any seller $i, j \in N$ and a subset S, the size of the consumer type who only considers the product by sellers in the subset S is defined as α_S. The size of consumers who also consider the products by sellers in S is σ_S.

$$\alpha_S = (\prod_{i \in S} \sigma_i)(\prod_{j \notin S}(1 - \sigma_j)), \quad \sigma_S = \prod_{i \in S} \sigma_i \tag{1.1}$$

We can understand how posting frequency plays a role in price competition by interpreting it under the two polar structures. The first case is the independent structure, where each seller has an independent probability of matching. The second case is the nested structure—the key difference between the two lies in how consumer groups are shared with one another. Two cases are different views on how each seller's posting frequency plays a role in constructing the consumer's consideration set.

The first case, independent structure, is constructed on various types of consumers who are "captured" by a specific seller (e.g., $\alpha_1, \alpha_2, \alpha_3$). There are also consumers at the intersection of one or more circles who considers multiple options. Each seller's matching probability or the size of reach (σ) is independent so that we can derive the size of intersection set as in Equation 1.1.

The second case is the "nested" interaction structure as in Figure 1.7. The concentric circles in the nested structure represent how the consumers are shared within the sellers S1, S2, and S3. In this case, any consumers who consider seller 1's product also put other bigger sellers' products in its consideration set. Similarly, seller 2's consumers always consider buying from seller 3 in the nested interaction structure. In other words, a seller with less intense advertising always shares the consumers with the seller with more intense advertising. The relative size of a seller's captive consumers, or the size of reach σ, determines its market power.

The two structures can be mapped into two different ways of search. In one way, a consumer would not put that much effort into looking for the options. In that case, the consumer would only consider the product at the top of the page, which is the most visible option in the list. The consumer would face the products sold by different sellers, and the frequency of matching is proportional to the seller's number of posting. Here, each sellers arrive at the platform independently, and the probability of one seller's product being included in the consideration set is independent from the other competitors.

A second type of consumer choice happens when the consumer enters the platform at a certain point of time and decides among the options observed on the platform. The consumer would search multiple options and pick the best one among the list. It would look a lot like fixed sample search, or "extensive search",in which the consumer looks at a certain number of options and makes a choice from the list. In this case, even though the arrival process of each sellers is independent, the probability of the product included in the consideration set is not independent across the sellers. Specifically, let's assume seller 1 that arrives at the platform with the probability of p_1 and seller 2 with p_2, and p_1 is much higher than p_2. Consumers search among n different options. Then, the probability of the seller 1's product is in the consideration set would be $1-(1-p_1)^n$ and the seller 2 as $1-(1-p_2)^n$. The conditional probability of considering seller 1 given considering seller 2 is near 1, while it is just p_1 in the first case.

The former way can be referred to as "independent structure", where each seller's probability of included in the consideration set is independent. The latter is "nested structure" where the consumer's consideration set with the less frequent seller is included in the more frequent sellers' consideration set. Formally, if $p_1 > p_2$ S_1, S_2 are the event of seller 1 and 2 is included in the consideration set, then $P(S_1|S_2)$ is near 1.

The interaction structures illuminate sellers' pricing power and price dispersion. Sellers' ability to charge higher prices is proportional to the sizes of their captive consumers relative to all consumers considering their products. Seller 1 in the nested structure faces much more competition with sellers S2 and S3 than S1 in the independent structure, generally resulting in lower prices.

Armstrong and Vickers (2020) work generated several predictions about the relationships between the competitive interaction structures and price ranges. The five predictions are summarized below.

- **Prediction 1 (Sanity check)** : First order stochastic dominant price distribution with respect to the reach (σ) is likely to be observed in both independent and nested cases.

The intuition behind Prediction 1 is that the seller with larger reach(σ) has higher price distribution. It means both in independent and nested structure, the seller with higher reach (σ) would have the largest group of captive consumers, leading to higher price distribution. The average price is higher for the larger reach group due to the first order stochastic dominant structure.

- **Prediction 2 (Interaction identification)** : In independent structure cases, the minimum price of sellers with different size of reach (σ) should be the same; the nested case will show a monotonically increasing pattern.

  The demand for the smallest seller in the nested case is more elastic than the independent case, because the types of seller 1's consumer group are differ in the two structures. Seller 1's consumers in the nested option compare all possible choices available while that type in the nested is smaller as it is the multiplication of all other σ. Therefore, the minimum price of smaller sellers in nested cases should be less than the outside sellers, leading to a monotonically increasing pattern.

- **Prediction 3**: If the interaction follows nested pattern, the price distribution difference between a seller with larger reach (σ) and smaller reach sellers is likely to increase substantially when the size of the captive group is larger for the higher reach (σ) seller.

  The prediction uses the relationship between the size of the captive group and the price. If the size of reach (σ) of a larger seller increases, it would have a larger set of captive consumers. Eventually, that would increase the price, widening the difference in price distribution.

- **Prediction 4**: If the interaction follows independent structure, the entry of new small sellers can lower the price of other sellers, but such new sellers will not cause changes for a nested case.

This prediction uses the difference between independent and nested structure. The entry of new sellers into an independent structure such that they increase an independent structure's intersecting consideration sets and reduce other sellers' captive consumers, will cause α_i to decrease. In equilibrium, the minimum price p_0 will also decrease. Since all sellers share the minimum price p_0 , new sellers' entry reduces the average price as well. In contrast, the entry of a new, smaller seller into a nested structure, will not affect the price distribution of larger sellers. The aggregate price distribution across sellers is less likely affected by the entry of small sellers.

1.5 Empirical Strategy

In this section, predictions from Armstrong and Vickers (2020) are applied to the data from Cetizen to describe the pricing pattern. The Armstrong and Vickers (2020) model provides a framework to understand the price distribution and dispersion. I used the model to quantify the market power of the sellers that is acquired by the advertising.

In order to apply predictions from Armstrong and Vickers (2020), the assumptions of the model need to be tested against the data. If the assumptions do not apply, then the predictions are not reliable for the platform's data. After establishing the validity of the assumptions for the data set, two steps of statistical testing are applied using price distribution, particularly price support, to identify the interaction structure. Two main components that distinguish the interaction pattern are 1) the stochastic dominance pattern of sellers' pricing strategy and 2) the size of price support and minimum, maximum price that a seller charges. Two consecutive empirical testings are, first, a stochastic monotonicity testing procedure (Chetverikov et al. (2020)) which determines whether the seller with higher reach(σ) shows first-order stochastic dominance to other sellers with smaller σ; and second, whether the two groups with different sizes of reach (σ) have a distinct size of minimum price (Wilcox et al. (2014)). This two-step testing process distinguishes between the nested and independent structures.

The identification of the interaction structure is essential to understand the role of advertisement in the online marketplace. Advertisements influence how consumers consider available options, but the actual competition structure created from the advertising is not directly observed. Most of the time, the papers use clickstream data to identify consumer consideration procedure(Chen and Yao (2017),Koulayev (2014),Bronnenberg et al. (2016)). Without detailed data on consumer search procedure, it is not possible to measure the consideration process. The predictions from Armstrong and Vickers (2020) provide a nice tool for analyzing the pricing behavior, and thereby inferring the competition structure induced by advertising.

Assumptions of the Model

The primary assumption of the Armstrong and Vickers (2020) model is that the market consists of homogeneous goods. One concern is that the products in the data are not homogeneous due to the differences in quality or other characteristics. For example, the listings which sell Galaxy S9 do not have the same quality. Some listings might sell products with expired warranty, and some might be relatively new with a good product condition.

Under the assumption of homogeneous consumer preference, the common value component with observable product characteristics can be normalized. It is a way to approach the observable price to match with the model since imposing heterogeneous seller and consumer at the same time would make it impossible to identify both sides. Therefore, I'm imposing homogeneous consumer preferences to make the analysis simple.

Also, I assumed the observable characteristics are additively separable in the utility to focus on the price deviations from the mean. This is how the heterogeneity is dealt with in the differentiated goods market and auction literature (Wildenbeest (2011)) where observable characteristics can be additively separable in the function of price. The following two assumptions are based on Wildenbeest (2011), which uses an additively separable function to remove the common value component from the price function.

Assumption 1. *Each consumer has the same preferences towards the quality.*

The utility for purchasing a product from seller j with a quality q_j and homogeneous valuation of the product x would be as follows.

$$u_j = x + q_j - p_j$$

Assumption 2. *Sellers obtain quality of the product from the perfectly competitive markets and quality production function follows constant returns to scale.*

Under the assumptions, consumers valuation for a product with q_j would be $v(q_j) = x + q_j$ and seller's valuation cost markups would be $v(q_j) - r(q_j)$ where $r(q_j)$ is the marginal cost for quality q_j. Then due to the perfect competition, the $r(q_j) = q_j$ and the markups would be x, which is the same across the various qualities. In other words, sellers are symmetric for various qualities that they offer.

Based on the assumption, the price can be written as the function of observable characteristics. Assume p_{jt} is the product price by seller j at time t. Then, the price can be the function of observable characteristics(δ_{jt}) and the residuals(ϵ_{jt}). Under the assumption that shape of price distribution is the same across the different sellers, and observable characteristics is already captured in the price with the homogeneous consumer utility function, then the price at which sellers are competing would be the residuals ϵ_{jt}.

$$p_{jt} = x + q_{jt} - u_{jt} = \delta_{jt} + \epsilon_{jt}$$

Based on the assumption, I ran a price regression to remove other characteristics' effects on the price. Each observation contains quality (high/mid/low condition), warranty (under warranty/expired warranty) information, memory size. The price regression equation is constructed below. Regression result is in the appendix. Regression with price in dollar and log(price) are both summarized and log price regression is used in the rest of the analysis.

$$p_{ijt} = QT\beta_1 + GR\beta_2 + \text{Size}\beta_3 + \gamma_j + \text{Month FE} + \epsilon$$

By only using model fixed effect coefficient and residuals, the residualized price is reconstructed as:

$$\hat{p}_{ijt} = \gamma_j + \epsilon$$

The model is based on unit price, so the price is normalized between 0 and 1 by dividing the market price by the original price of a new product $(\frac{\text{Market price}}{\text{Original price}})$.

Another implicit assumption of the Armstrong and Vickers (2020) model is that each market is independent, which is defined in my analysis as a time window (1 month). The markets are defined by time for the following reasons: First, the observation window has to be long enough to measure the mixed price strategy as a distribution. Second, one month is long enough to separate the effect from the previous prices which are close. A month timeframe is long enough to establish market independence since what the seller posted in the previous month has little chance of affecting the next month's decision.

However, since markets are defined by time (window of a week or a month), there is some possibility of serial correlation or dependent pricing structure. Therefore, whether each seller's pricing strategy is independent is tested in this section.

To verify whether there is auto-correlation embedded in seller's pricing strategy, two statistical tests are conducted: Bartels' Rank test and Yule-Walker test are conducted to check potential auto-correlation structure. Bartels' rank test checks the randomness or non-autocorrelation structure by ranking the samples from the smallest to largest[3]. I tested the null hypothesis of being random (H_0). Yule-Walker tests potential AR structure that can be exist in the sample. The null hypothesis that is tested is there is no AR process within the variable. For this test, I calculated weekly average price of each sellers for different models, such as Galaxy S9 and used only the sellers of a particular model that posts more than 10 times. The average P-value of all the sellers is summarized in Table1.4, which is higher than the criterion value of $\frac{\alpha}{2} = 0.025$ (DiCiccio et al. (2020)). Therefore, it does not reject the null hypothesis, and the sellers' pricing distribution is random.

Based on both results, each observation that contains sellers' pricing has little serial correlation and also is close to random.

The last assumption of the pricing game in the previous sections is that the sellers use a mixed price strategy. It means no sellers are constantly selling a product at high or low prices. If the relative position of the sellers' price changes over time, it supports mixed price equilibrium. Following the literature (Chandra and Tappata (2011), Pennerstorfer et al. (2020), Lach and Moraga-González (2017)), I calculate rank reversal statistics, which can measure how many changes the sellers are making in the pricing game. Rank reversal statistics measure the frequency of the price rank

[3]The test statistics $U = \sum_{i=1}^{n-1} (R(X_i) - R(X_{i+1}))^2 / (n(n^2-1)/12)$ is used for the testing. By using the U statistics' distribution, which follow normal distribution with mean 2 and variance $\sigma_4^2 = 4(n-2)(5n^2-2n-9)/(5n(n+1)(n-1)^2)$

Table 1.4: Statistical Test Results

Model	Yule-Walker Average Pvalue	Bartel's Rank test Average Pvalue
SM-N950	0.376	0.291
A1901	0.301	0.242
A1905	0.298	0.213
A2097	0.365	0.267
A2105	0.350	0.245
A2215	0.403	0.316
A2221	0.367	0.308
SM-A530	0.265	0.175
SM-G960	0.288	0.182
SM-G973	0.377	0.284
SM-G975	0.354	0.274
SM-G977	0.351	0.266
SM-J330	0.243	0.144
SM-N960	0.369	0.293
SM-N976	0.395	0.300

changes among the sellers.

The rank reversal statistics for seller i and j are calculated as follows. If i and j participate in the market t, and if j's price is higher than i's price over 50% of the times when they meet each other in the market, then count the number of the days that the rank flipped. Thus, the rank reversal statistics calculates the number of times the rank flipped throughout a specific period.

$$r_{ij} = \frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} I(p_{jt} > p_{it}) \quad \text{when} \quad \frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} I(p_{it} > p_{jt}) > 0.5 \tag{1.2}$$

The average number for rank reversal statistics of Cetizen's data is around 0.1108, which means that a seller charging a higher price than the competitor for more than 50% of the time also posts a lower price around 11% of the time. These results are similar to the range found in Lach and

Moraga-González (2017) and Chandra and Tappata (2011). Based on the size of rank reversal statistics, the sellers seem to use a mixed pricing strategy in the market.

Table 1.5: Rank Reversal Statistics by Chandra and Tappata (2011)

Model	Rank Reversal
SM-A530	0.148
SM-G960	0.124
SM-G973	0.120
SM-G975	0.122
SM-G977	0.141
SM-J330	0.139
SM-N950	0.135
SM-N960	0.127
SM-N976	0.120
A1901	0.119
A1905	0.145
A2097	0.140
A2105	0.138
A2215	0.140
A2221	0.140

Note: The rank reversal statistics are calculated by (1.2). The market definition is a week from Feb. 5th, 2020 to Aug. 29th, 2020.

Construction of the Measure of reach (σ)

One key component in the model is the size of reach (σ), which is the matching probability of each sellers. This section explains how the size of reach (σ) is constructed. The size of reach (σ),i.e., the group of consumers considering a seller's product is determined by the amount of repeated posting. In other words, the frequency of posting is used as a proxy for advertisement effort or intensity. However, the definition of time window whether to calculate frequency in terms of an hour, week or a month can change the size of frequency measure.

First, I calculated the frequency measure using the average number of postings made by each seller within an hour. So if there is λ_m number of postings for one hour on average that sells cellphone model m, if a seller j posts about λ_{jm} of the model m, then the frequency measure is calculated as $\lambda_{jm}/\lambda_m = \sigma_{jm}$. Using the calculated σ_{jm}, I classified the sellers into 3 groups as in Table 1.6.

Table 1.6: Construction of σ Group

Seller quintile	σ (mean)	σ (median)	Std	Min	Max
1	0.134	0.128	0.048	0.066	0.363
2	0.224	0.209	0.074	0.070	0.488
3	0.350	0.320	0.155	0.084	0.934
Total	0.229	0.194	0.135	0.066	0.934

Another way to construct the size of reach is to calculate the fraction of the listing stock, or the share of listings made by a seller. Assume there is a seller i, product j, time t, number of postings $k \in \{1, \cdots, K_{ijt}\}$ and the random variable I, J, T, K, which represent the group of seller, product, time(market) and the postings. The sellers who are selling in the market are defined as $\tilde{I}$ and $1(\cdot)$ is the indicator function. Then the share of listings, σ, is constructed as (1.3). Essentially, it captures the fraction of listings made by seller i among all postings in the platform observed during a certain period of time (a month).

$$\sigma_{ijt} = \frac{\sum_{k=1}^{K_{ijt}} 1(I = i, J = j, T = t, K = k)}{\sum_{i' \in \tilde{I}} \sum_{k=1}^{K_{i'jt}} 1(I = i', J = j, T = t, K = k)} \tag{1.3}$$

The definition of σ determines how we view the seller type. For example, if many new listings are refreshed every hour, creating a list of products on the first page that changes every hour, then hourly frequency would be a better measure of σ. In the main document, I present analysis results based on the first definitions of the reach based on hourly level posting shares. As a robustness check, the estimation results with various definitions of σ are in the appendix and section 1.5.

Table 1.7 shows a slice of listing structure and the seller composition in the website. Within each week, three different sellers are classified by the frequency of postings as the first definition of σ. Group 1, whose sellers had the least number of postings, accounts for about 38% of the entire postings while the sellers in Group 3 composes about 31%. The proportion of postings are affected by both the number of sellers and the frequency of the postings. The number of sellers in Group 1 is about seven times more than Group 3(as shown in the second panel), while the average postings of a seller in Group 1 are only 1/3 of a seller in Group 3. Thus, the difference between the listing proportion of Group 1 sellers and that of Group 3 is not significant.

Table 1.7: Postings Composition (1 Week Window)

Posting Composition	Mean	SD	P25	P50	P75
Group1	0.38	0.07	0.33	0.38	0.42
Group2	0.30	0.06	0.27	0.31	0.34
Group3	0.31	0.07	0.28	0.32	0.36
Number of Sellers	Mean	SD	P25	P50	P75
Group1	88.46	33.53	66.00	87.00	109.00
Group2	24.08	8.94	18.00	24.00	30.00
Group3	14.59	6.89	9.00	13.00	20.00
Average # Postings (Each Group)	Mean	SD	P25	P50	P75
Group1	1.22	1.07	0.29	1.00	2.25
Group2	2.16	2.10	0.66	1.04	5.45
Group3	3.35	2.88	1.53	1.04	7.76

Note: The first panel shows the fraction of postings by each group of sellers that are defined by the frequency of postings (Group 3 is the largest). The second panel is the number of sellers of each group that are observed in 1 week window. The third panel is the average number of postings within one hour that are made by the sellers in each group within 1 week window.

To compare the price range, a popular cellphone model in this platform is picked: Galaxy S9. Figures 1.8 and 1.9 show the average price ranges of the sellers, classified by the intensity of reposting (σ) in April and July, 2020. Overall, the price support (region) is likely to be increased for the sellers with more frequent postings than the others with less σ. However, how the price support increases by the size of σ for each group is not monotone. For example, the Group2 in April has the highest maximum price than the other 2 groups. But overall, Group3 with the largest size of σ has the highest price support.

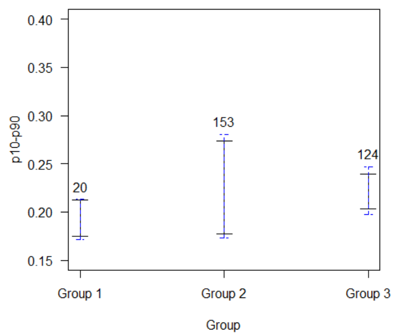


Figure 1.8: p10 − p90 in April

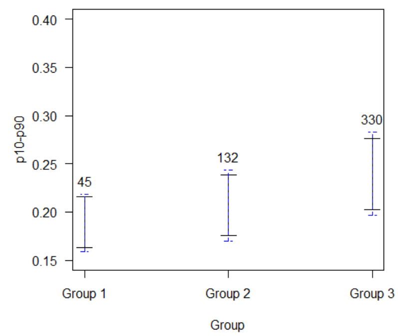


Figure 1.9: p10 − p90 in July

Note: The price support between 10% - 90% quantile of the Galaxy S9 price distribution in April and July

Prediction 1 and 2: Distinguishing Interaction Structure

This section tests predictions 1 and 2 of the competitive interaction model by comparing the price distribution and the size of reach σ. The model shows both the independent and nested structures generate stochastic monotonicity with respect to the frequency of posting (or reposting) which is measured by σ. The problem is to distinguish the two different interaction patterns.

The nested structure is a stronger form of independent structure: the less frequent posters cannot attain captive consumers with their advertisement effort in nested while independent structure allows less frequent sellers to get a certain group of captive consumers. In nested structure, the winner of competition (largest reach) will dominate the market and the loser (with much fewer postings) will face severe competition. Therefore, the different level of price competition affects the size of minimum price: In independent case, the minimum price of a seller is likely to be unchanged by the size of σ. In contrast, under nested structure, as the seller has a higher σ, they are likely to have a higher minimum price than the less frequent sellers.

The statistical testing procedure follows to identify competitive interaction patterns follows two steps. First, I test whether the price distribution of the power sellers is first-order stochastic dominant relative to the price distribution of the small sellers is tested. If the hypothesis of first order stochastic dominance is rejected, it suggests the interaction pattern does not follow an independent or nested pattern.

If the hypothesis is not rejected in the first step, the next step is to distinguish independent or nested. The main difference between the two will be the size of the minimum price. Under the independent case, the market power of small sellers is higher than the nested case due to the captive consumers. Therefore, the model predicts that sellers need to have a lower minimum price in nested structure while there is no difference in the independent structure's minimum price. Thus, the minimum price should be compared in the second step to distinguish between independent and nested interaction structure.

Below are the steps of the testing procedure. First, as a sanity check, first order stochastic dominance of price distributions is tested using stochastic monotonicity testing. Second, the two patterns are distinguished by the comparison of the minimum or the lower quantiles.

1. Step 1: under the assumption of independent and nested interaction , the price distribution follows stochastic monotonicity with respect to the size of reach (σ). The null hypothesis tested in the step is that the stochastic monotonicity holds (either independent or nested structure holds).

   - Test $H_0 : F_{p|\sigma}(p|\sigma') \geqslant F_{p|\sigma}(p|\sigma)$ for all p, $\sigma, \sigma' \in (0,1)$ with $\sigma \geqslant \sigma'$.

   If the hypothesis is rejected, it would suggest that neither independent nor nested pattern applies to the data.

2. Step 2: If the null hypothesis in the first stage is not rejected, it would suggest two cases - either independent interaction or nested cases. In Step 2, under the null of independent interaction, we could test the null hypothesis that the minimum price of the sellers with various sizes of reach (σ) do not differ. If the hypothesis is rejected, it means the minimum price changes, suggesting a nested structure.

   - Test $H_0 : F_p^{-1}(\tau|\sigma') = F_p^{-1}(\tau|\sigma)$ for $\tau < 0.05$, $\sigma, \sigma' \in (0, 1)$ with $\sigma \geqslant \sigma'$

Based on the two step procedure, the nested and independent interaction can be distinguished.

Step 1: Stochastic Monotonicity (Sanity Check)

The test statistics are from Chetverikov et al. (2020). The method, which is called "stochastic monotonicity", tests whether the distribution of a dependent variable increases with respect to a certain independent variable.

Let $\{(p_i, \sigma_i) : i = 1, \cdots n\}$ is the i.i.d. random sample from (p, Σ). p is the underlying distribution of price and Σ is the distribution of the reach, σ_i. If we assume $K(\dots)$ denote a one dimensional kernel function with bandwidth h, $K_h(x) = h^{-1}K(x/h), x \in R$. The null hypothesis of stochastic monotonicity can be written as (1.4).

$$H_0 : F(p|\sigma_i) \leqslant F(p|\sigma_j) \quad \text{if} \quad \sigma_i \geqslant \sigma_j \tag{1.4}$$

Basically, it tests first order stochasic dominance of price distribution with respect to a frequency of the postings (σ). The null hypothesis also can also be written as (1.5).

$$H_0 : E\left(1(p_i \leqslant p) - 1(p_j \leqslant p))sign(\sigma_i - \sigma_j)K_h(\sigma_i - \sigma)K_h(\sigma_j - \sigma)\right) \leqslant 0 \tag{1.5}$$

The null hypothesis in Step 1, which is $H_0 : F_{p|\sigma}(p|\sigma') \geqslant F_{p|\sigma}(p|\sigma)$ for all p, $\sigma, \sigma' \in (0,1)$ with $\sigma \geqslant \sigma'$, is tested.

I classified the data into the various windows of months and weeks. Also, to incorporate the heterogeneity of the sellers, I classified a group of sellers as "professional sellers";professional sellers sell more than 5 models a month. Additionally, sellers are grouped by the proportion of a certain brand of their products selling in the market. If a seller sells Samsung products more than 80% of their product lines, they are classified as a seller with brand > 0.8. The test result is summarized in Table 1.8.

The analysis shows first-order stochastic dominance of price distribution with respect to the size of reach σ on most of the sample period. The null hypothesis of monotonic stochasticity is not rejected in most of the sub-samples, which suggests that the pricing follows the prediction of the model.

To derive a more general conclusion from the statistical tests conducted on various subsamples, I used DiCiccio et al. (2020) to generalize the result. The main logic of the method is to use average p value of various sub-samples. Suppose there are M different samples that are selected by various criterion as in Table 1.8. Each samples would test hypothesis $H_0 : F^S_{p|\sigma}(p|\sigma') \geqslant F^S_{p|\sigma}(p|\sigma)$. Based on the conservative method that uses average p-value (2.2 of DiCiccio et al. (2020)), for M number of multiple splitted data, H_0 is rejected if $\frac{1}{M}\sum_{m=1}^{M} \hat{p}_{n,m} \leqslant \frac{\alpha}{2}$, where α is the type 1 error and $\hat{p}_{n,m}$ is the P-value of sub sample m. The average P-value over 10 sub-samples is 0.95, which is larger than 0.025, or $\frac{\alpha}{2}$. Therefore, the null hypothesis $H_0 : F_{p|\sigma}(p|\sigma') \geqslant F_{p|\sigma}(p|\sigma) \quad \text{if} \quad \sigma' > \sigma$ is not rejected for the analysis with entire samples. In other words, stochastic monotonicity generally holds.

Table 1.8: Step 1 Result: Galaxy S9

Samples	Galaxy S9
Monthly	Hypothesis Test
April	0.600
	(0.99)
April, prof	0.985
	(0.53)
July, prof	0.550
	(0.97)
April, prof, brand>0.8	1.449
	(0.06)
April, prof, brand<0.6	0.832
	(0.57)
Weekly	
12th wk	0.670
	(0.94)
12th wk, prof	1.073
	(0.49)
20th wk, prof	0.950
	(0.6)
12th wk, prof, brand>0.8	1.800
	(0.0)
12th wk, prof, brand<0.6	0.627
	(0.845)
Whole data	0.8
	(0.81)
P value mean	0.618
Criterion P value	0.025

Note: The estimated result is from Chetverikov et al. (2020). The null hypothesis (H_0) is whether stochastic monotonicity with respect to reach (σ) holds. P-values are in the parentheses. Professional sellers are the ones who sell more than 5 models within a month. "Brand" refers to the proportion of Samsung products that the sellers are selling in the market.

Step 2: Interaction Identification

The procedure in Step 2 is to compare lower quantiles of price distributions with various sizes of reach. Due to the irregularities of price distribution and the size of reach σ at an individual seller level analysis,I conducted a group comparison. The sellers are classified into 3 groups by the size of σ. I applied a distribution comparison methods, which is Harrell-Davis quantile estimator (Wilcox and Erceg-Hurn (2012), Wilcox et al. (2014)) using Wilcox's robust statistics (WRS) package in R. This method evaluates two distributions by comparing the quantile value of each distribution. Harrell-Davis estimate of q quantile is a weighted average of order statistics. [4]

The null hypothesis is that the q quantile of i group price distribution minus q quantile of j group price distribution is equal to zero.

$$H_0 : \hat{p}^*_{iq} - \hat{p}^*_{jq} = 0 \tag{1.6}$$

[4] I used method M in Wilcox and Erceg-Hurn (2012), which uses bootstrap of Harrell-Davis estimate.

Table 1.9: Step 2 Result: Galaxy S9

	Group1	Group3	Diff	P value
$q_{0.01}$	0.103	0.154	-0.051	0.0000
			(-0.057,-0.044)	
$q_{0.05}$	0.145	0.168	-0.023	0.0000
			(-0.032,-0.017)	
$q_{0.1}$	0.158	0.183	-0.025	0.0000
			(-0.034,-0.019)	
n	976	976		
	Group2	Group3	Diff	P value
$q_{0.01}$	0.104	0.154	-0.050	0.0000
			(-0.056,-0.033)	
$q_{0.05}$	0.129	0.168	-0.039	0.0000
			(-0.043,-0.034)	
$q_{0.1}$	0.139	0.184	-0.044	0.0000
			(-0.053,-0.035)	
n	958	958		

* Note: This table is the quantile difference test results which applies Wilcox and Erceg-Hurn (2012) method. The values in Group 1, Group 3 columns are the 1%, 5% , 10% quantiles of the price distribution. "Diff." column is the difference between the two, "P-value" is the P-value of the statistical test which compares the two quantiles. 95% confidence regions are in the parentheses.

The estimated result is summarized in Table 1.9. The sample used in the analysis is Galaxy S9 professional sellers that sell Samsung cellphone models in more than 80% of the postings in July. The sellers are grouped into 3 in terms of their size of σ. In both Group 1, 2, and Group 1, 3 comparison, Group 3 is likely to have a higher value in 0.01, 0.05, 0.1 quantile. It suggests the minimum price is likely to be higher for the sellers with larger σ. It supports the prediction from the nested interaction pattern.

Based on the previous results, we see that the sellers' pricing distribution follows first-order stochastic dominance in terms of the posting frequency (σ). Also, the comparison of price distribution in 0.01, 0.05, and 0.1 quantiles shows the power sellers have a higher minimum price (lower quantile) than the smaller sellers. Thus, it suggests interaction is a nested pattern. [5].

Similarly, two steps of testing are conducted for the various models of phones that are sold on the platform. The result is summarized in Table 1.10. The P-values of the Step 1 results are summarized in the 2nd column. They are the average P-values of the tests conducted on the 3 different sub-samples: April, April only professional sellers, June, only professional sellers. Column 3 is the result of the tests. Column 4 is the average P-value of the Step 2 tests conducted on the various subsamples. Based on the two test results, whether the data follows nested structure is summarized in column 6. About half of the models that are traded in the platform follow the nested pattern.

Table 1.10: Summary of Tests: Results on Various Cellphone Models

Summary	Step 1	Step 1 test	Step 2	Step 2 test	Result
Model	Mean P value	Reject	Mean P value(0.05)	Reject	Nested
iPhone X	0.233	0	0	1	1
iPhone 8	0.867	0	0	1	1
iPhone XS	0.090	0	0	1	1
iPhone XR	0.000	1	0.093	0	0
iPhone 11	0.003	0	0.133	0	0
Galaxy A8	0.047	0	0	1	1
Galaxy S9	0.347	0	0	1	1
Galaxy S10	0.723	0	0.147	0	0
Galaxy J3	0.000	1	0.8	0	0
Galaxy Note8	0.143	0	0	1	1
Galaxy Note9	0.583	0	0.24	0	0
Galaxy Note10	0.000	1	0.013	0	0

Note: Step 1 uses Chetverikov et al. (2020) to test stochastic monotonicity, Step 2 uses Wilcox and Erceg-Hurn (2012) to compare price quantiles. P-values are the average P-value across the different subsamples, In column 6, "1" indicates the data follow nested pattern while "0" is not.

[5] As a robustness check, I used the time-lapse between the two consecutive postings to measure advertising intensity. The analysis results which use the additional measure are in the appendix.

The market characteristics can affect the testing result, but it's hard to summarize the factors. Table 1.11 shows some hints of the difference between the markets that are tested as the nested and non-nested. [6]

First, the cellphone models in the nested group are the ones that have less frequent postings by the sellers in general, as in the "Ave # postings/hour" row. Second, the cellphone models classified as a nested group have fewer sellers in general. It shows the cellphone models with a nested structure are likely to have less frequent refresh by the sellers, less number of sellers. Third, the models in the nested group are sold more often. Because of the slower refresh rate in the list, the sellers enjoy the benefit of staying longer in the platform, getting more exposed to potential consumers, which gives more market power to the sellers. In other words, a thicker market is likely to show a nested structure.

Table 1.11: Nested and non-nested model

Variable	Nest		Non nest		Diff (Non nest-Nest)	
	Mean	SD	Mean	SD	β	t
G1 frequency	0.154	0.047	0.127	0.026	-0.027***	(-92.164)
G3 frequency	0.391	0.144	0.366	0.134	-0.025***	(-22.298)
Difference (G3-G1)	0.237	0.115	0.246	0.140	0.008***	(16.248)
Sold probability	0.228	0.161	0.119	0.121	-0.109***	(-189.675)
# G1 sellers	80.730	31.803	96.006	33.471	15.275***	(116.654)
# G2 sellers	20.482	6.134	27.581	9.794	7.099***	(217.299)
# G3 sellers	12.687	6.290	16.339	6.836	3.652***	(138.633)
Ave # postings / hour	8.572	2.751	11.738	2.901	3.166***	(279.205)
Observations	122171		126326		248497	

Quantile Regression In this subsection, a complementary approach to study the effect of advertisement intensity on price distribution is presented. The quantile regression estimates the conditional quantile of the dependent variable, so it can be used to see if the conditional 5% or 10% quantiles of the price distribution changes by the size of σ. If the coefficient

[6]The table is comparing the cellphone models in Table 1.D.4 which show the nested pattern in the analysis with the σ constructed based on the description of the listing.

of the advertisement intensity is positive and statistically significant, then it would mean the intensity measure (σ) would likely increase the certain quantile of the price. Formally, the coefficient solves the minimization problem of (1.7). $\rho_\tau(u) = (\tau - 1(u < 0))$ is the check function while τ is the quantile, i is the seller and T_i is the period when the seller i sells a product in the market.

$$\min_{\beta} \sum_{i} \sum_{t \in T_i} \rho_\tau(p_{it} - \sigma_{it}\delta) \tag{1.7}$$

I included the model (product) fixed effect and the month fixed effect to control the unobserved factors that might affect the price distribution. The parameter δ_τ is estimated for the various size of $\tau \in \{0.05, 0.1, 0.5, 0.9, 0.95\}$. The empirical specification of the quantile function would be written as (1.8). $\mu_{m(i)\tau}$ is the model dummy (Galaxy S9, iPhone) and $\gamma_{T(t)\tau}$ is the month dummy. The parameter of interest is δ_τ.

The quantile regression result is in Table 1.12. The result shows that coefficients for lower and upper quantiles are statistically significant and positive. They imply that both minimum and maximum price distribution are likely to be increased as the advertisement intensity is elevated. Also, as all coefficients in various quantiles are positive, stochastic dominance structure holds. To interpret this number, let's assume a cellphone like Galaxy S9 has 10 postings on average appear in the platform within 1-hour window. If seller A posts 4 listings on average on the 1-hour window, then the share of seller A is 40%. If there is a seller B with 5 listings per hour, its share is 50%, which is 10% higher than seller A. If seller B has a 10% higher share of listings than others, as in the previous example, then the price distribution of seller B is more elevated than seller A. The 5% in seller B's price distribution would be increased by 4.146$ and 95% by 3.549$. In sum, both two-step testing and quantile regression results show price distribution that implies nested structure.

$$p_{it} = \delta_\tau \sigma_{it} + \gamma_{T(t)\tau} + \mu_{m(i)\tau} + u_{\tau,it} \tag{1.8}$$

Table 1.12: Quantile Regression

τ	Estimate of δ		
	(1) Time invariant	(2) Time variant	(3) Time variant
0.05	56.28***	41.46***	44.72***
	(3.203)	(4.540)	(4.166)
0.1	53.04***	45.32***	46.10***
	(4.624)	(3.566)	(4.102)
0.5	59.84***	55.52***	56.55***
	(3.592)	(2.972)	(3.713)
0.9	41.15***	43.94***	42.15***
	(4.254)	(2.603)	(4.915)
0.95	29.90***	35.49***	33.67***
	(5.557)	(6.314)	(6.578)
No. Models	14	14	14
Month FE	O	X	O
No. Observations	23098	23098	23098

Note: Each observation contains the monthly average price and reach calculated in a hour window. The dependent variable is the average monthly price ($) of each seller that removes the effect of other factors(conditions, warranty, memory size). The first column (1) Time invariant uses σ that uses hourly average across the whole observation period, while (2), (3) use σ constructed on each month. Standard errors are in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Robustness check

In this section, I conducted several robustness checks to circumvent the potential endogeneity problem of the price.

First, frequent postings by each seller may represent the size of the seller's inventory. Therefore, the higher price that these sellers charge may come from seller's reputation or ability to have a large inventory. To control the effect from the inventory or seller's reputation, I constructed the size of reach (σ) based on each listing level. In principle, the website restricts the sellers to sell only one product for one listing. I assumed each listing with the same title is selling one product. Therefore, the reach size (σ) constructed on the individual listing level can be a better measure of visibility without potential concern from the seller fixed effect by comparing the listings within one particular seller. The unique listing is defined by the product's description, like "SKT Galaxy Golder G150 White". I calculated the share of listing out of the number of all listings that are tracked within one month. The result is in Table 1.13. Even though the coefficient size gets larger since the unit of σ is much smaller than the previous analysis, they are all statistically significant and positive in both lower and upper quantile.

Second, the price could be endogenous in the sense that it can contain the unobserved demand shock of a particular model. To handle potential endogeneity, I used two different instruments for the price. In the spirit of the Hausman instrument, the first instrument used is the price of other models posted within the same hour by the same seller. It still shows significant and positive coefficients in both lower and upper quantile price distributions. Also, there could be an issue of aggregate demand shocks such as increasing in used phones due to changes in regulation or new product releases. To control time-variant factors that can affect the demand, I used both time-variant and invariant measures of σ to see if they show a similar pattern. The results are in Table 1.14, and they also show a similar nested structure as in the previous analysis. The second instrument is the initial price of repeated listing with the same description since the later price, and posting decisions can be affected by other demand factors. The instrument shares the same spirit of uniform price instrument used in Huang (2021). He used the tendency for uniform pricing of each room for multiple nights as a price instrument in his paper. The analysis with the second instrument also shows significant and positive coefficients for both in lower and upper quantile of the price distribution, which support

Table 1.13: Quantile regression : Description level

τ	Estimate of δ	
	(1)	(2)
0.05	520.2***	517.0***
	(34.50)	(44.95)
0.1	484.4***	473.3***
	(42.66)	(51.30)
0.5	1161.7***	1160.8***
	(72.39)	(66.05)
0.9	1410.7***	1400.8***
	(113.4)	(126.5)
0.95	1405.5***	1215.2***
	(134.0)	(146.2)
No. Models	14	14
Month FE	X	O
No. Observations	51028	51028

Note: The reach size (σ) is constructed on each listing level using the unique description of the product like "SKT Galaxy Golder G150 White". Each observation contains the monthly average price and reach(σ) calculated in a month window. The dependent variable is the average monthly price ($) of each seller that removes the effect of other factors(conditions, warranty, memory size). Standard errors are in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the hypothesis of nested structure.

Table 1.14: Quantile regression: Other models

τ	Estimate of δ		
	(1) Time invariant	(2) Time variant	(3) Time variant
0.05	0.168***	0.119***	0.131***
	(0.0106)	(0.0115)	(0.0108)
0.1	0.160***	0.102***	0.128***
	(0.00868)	(0.00917)	(0.00756)
0.5	0.0627***	0.0492***	0.0507***
	(0.00380)	(0.00420)	(0.00461)
0.9	-0.0948***	-0.0605***	-0.0672***
	(0.0140)	(0.0107)	(0.00896)
0.95	-0.159***	-0.0996***	-0.103***
	(0.0250)	(0.0189)	(0.0188)
No. Models	14	14	14
Month FE	O	X	O
No. Observations	12422	12422	12422

Each observation contains the monthly average price and reach calculated using the frequency measure in an hour window. The dependent variable is the average monthly price ($) of the other product that are sold in the same hour by the seller after removing the effect of other factors(conditions, warranty, memory size). The first column (1) Time invariant uses σ that uses hourly average across the whole observation period, while (2), (3) use σ constructed on each month. Standard errors are in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Verification of Nested Structure

In this section, based on the hypothesis of being in the nested structure, whether the latter two predictions on the data pattern is verified.

Prediction 3: Price Distribution Difference

According to prediction 3, the gap of price distribution between the power and non-power sellers is likely to be larger for the concentrated nested structure. This section analyzes the quantile gap of the two groups and its relationship with σ difference between the larger and smaller group.

Table 1.15: Quantile regression: Initial price

τ	Estimate of δ	
	(1)	(2)
0.05	458.7***	459.7***
	(42.68)	(48.61)
0.1	411.3***	404.4***
	(36.58)	(53.78)
0.5	1187.5***	1184.2***
	(76.24)	(73.46)
0.9	1500.0***	1416.6***
	(130.2)	(92.81)
0.95	1532.8***	1346.0***
	(183.4)	(143.4)
No. Models	14	14
Month FE	X	O
No. Observations	51028	51028

Note: The reach size (σ) is constructed on each listing level using the unique description of the product like "SKT Galaxy Golder G150 White". Each observation contains the initial price of the unique listings and reach calculated in a month window. The dependent variable is the initial price of repeated listing($) after removing the effect of other factors(conditions, warranty, memory size). Standard errors are in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. $p < 0.001$.

Table 1.16 summarizes analysis of the regression of the price distribution difference between Group 3 and 1 on the σ difference. σ difference is created as the average σ of group 3/average σ of group 1, which shows the relative ratio of the two event. However, it can be endogeneous since unobserved factors that can affect the entry decision, such as demand shock, can change both the σ difference and the price distribution.

To circumvent the issues of endogeneity, I used an instrumental variable approach similar to Dai et al. (2014), Borenstein and Rose (1994). I included the market level variables that can affect σ structure and are assumed to be unrelated to price dispersion. I used the event of new product release, brand, and the number of sellers in the previous week as instruments. The new product release event, such as the launch of Galaxy S20 and Note 20, can affect the supply of other used cellphones but is less likely to affect price dispersion. Also, the number of sellers of a particular product in the previous week can affect the concentration of the following week with little linkage to the price dispersion. In Table 1.16, both OLS, IV results show the positive effect of the listing concentration which is measured by σ ratio on the price distribution. It suggests there is some positive effect on price distribution difference.

Table 1.16: Concentration and Price Distribution Difference

	OLS		IV	
	p10(G3)-p10(G1)	p10(G3)-p10(G1)	p10(G3)-p10(G1)	p10(G3)-p10(G1)
$\sigma_{G_3}/\sigma_{G_1}$	2.733*	2.516*	8.241*	9.397**
	(1.184)	(1.174)	(3.508)	(3.492)
# sold		0.0271		0.0238
		(0.0146)		(0.0150)
Const	-0.418	-0.423	-16.66	-21.39*
	(4.830)	(4.825)	(9.183)	(9.210)
Model FE	Yes	Yes	Yes	Yes
N	465	465	450	450
R-sq	0.462	0.467	0.457	0.441
1stage F stat			19.37	20.11

Note: Standard errors clustered in month, model level are in the parentheses. The dependent variable is the difference of 10% price quantile of Groups 1 and 3 defined in Table 1.6. $\sigma_{G_1}/\sigma_{G_3}$ is constructed based on the average number of postings that is made by each sellers in the groups within one week window. "# Sold (Wk)" captures the number of a particular model that are sold within one week. Instruments that are used in columns 3 and 4 are the event of new product release, brand and the number of sellers in previous week.

Prediction 4: Entry of New Sellers

Prediction 4 states that more sellers with small size of σ who are classified as group 1 are in the market will decrease price distribution in the independent case and cause no changes in a nested structure. The prediction is assessed in Table 1.17.

Table 1.17 summarizes the regression of difference in Groups 1 and 3 prices on the changes in number of Group 1 sellers. "Δ # Sold" captures the changes in number of sold postings within one week to control the aggregated demand effect. In order to rule out aggregate competition effect, I included "Δ # Seller(wk)" to capture the changes in the number of sellers in the market. The regression result reveals whether the sellers are affected by the presence of the other sellers. If the price changes, sellers make strategic choices as they see more Group 1 sellers inside the platform. The result suggests a statistically insignificant relationship between the number of Group 1 sellers and Group 3 sellers' pricing after controlling the total number of sellers and the amount sold in a week or a month window. Therefore, how the sellers react to the number of competitors in the market also suggests a nested interaction pattern.

Table 1.17: Pricing after Group 1 Entrants

	Δ Group 1 price(wk)	Δ Group 2 price(wk)	Δ Group 3 price(wk)
Δ # seller(wk)	0.0383	0.143	0.206
	(0.101)	(0.0787)	(0.153)
Δ # sold item(wk)	-0.00952	-0.0105	-0.00183
	(0.00621)	(0.00682)	(0.00868)
Δ # Group1 seller(wk)	-0.0544	-0.149	-0.254
	(0.121)	(0.0816)	(0.178)
Const	0.0405	-0.386***	1.127***
	(0.0656)	(0.0548)	(0.101)
Model FE	O	O	O
N	450	450	450
R-sq	0.014	0.015	0.013

Note: Standard errors clustered in month, model level are in the parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The unit of the dependent variable is a dollar ($).Δ# Seller (Wk), Δ# Group 1 Seller (Wk),Δ Group 1,2,3 price (Wk) are the changes in the number of entire sellers, group 1 sellers, and the average price of each group within a week window. Δ # Sold is the changes in the number of sold postings within one week. Δ # Seller(wk) is the changes in the number of sellers in the market.

Market power of the sellers

This section quantifies the size of market power that can be acquired by the intensity of advertising, or the posting frequency. I use profit functions in different interaction structures to infer the optimal advertising decision.

The recover the cost parameter, I assumed the sellers' following decision process.

First, sellers form a belief of price and reach distribution based on the observable data. Based on the belief, the sellers privately observe their own cost parameter and decide the size of reach (σ). After deciding the size of reach, each seller chooses price distribution.

In the nested structure, the seller in the middle concentric circle or the seller 2, has to consider the surrounding(seller 3) and the inside sellers(seller 1)' pricing decisions. The profit function of the seller 2 incorporates the advertisement (σ) choice; then, it would be written as (1.9).

$$p((\sigma_2 - \sigma_1) + \sigma_1(1 - F_1(p)))(1 - F_3(p)) - c(\sigma_2) = \pi_2 \tag{1.9}$$

The optimal choice of the reach size (σ_i), or the advertisement intensity, is determined by the marginal revenue of increasing the size of reach and the marginal cost of advertising. The marginal revenue of increasing σ would be $p(1 - F_3(p))$ in (1.9). Intuitively, if a seller increases the reach size (σ) by increasing the repeated posting, the seller will have a higher chance of meeting captive consumers, who have higher search costs. Therefore, the sellers can charge the maximum price in their price range to those groups, but the price needs to be lower than the seller 3's product. In other words, the marginal benefit of increasing the size of reach (σ) is from the competition with the seller with a larger reach (σ).

As the sellers are conducting a mixed price strategy, the advertisement choice condition can be written with the minimum price(L_i). Also, under the assumption of cost function as a quadratic form($c(\sigma_i) = \frac{1}{2}\sigma_i^2 w_i$), the cost parameter can be written as a function of the price distribution as in Equation (1.10). The relationship can be generalized. The marginal revenue of increasing σ is affected by the maximum price of the seller and the price distribution of the sellers with a larger reach. $(1 - F_3(L_i))$ means the probability of the sellers with a larger reach charging a higher price. In practice, it can be estimated empirically by observing each sellers' pricing distribution.

$$H_2(1 - F_3(L_2)) = c'(\sigma_2) = w_2\sigma_2 \tag{1.10}$$

Based on the construction, the cost parameter can be written as (1.11). It can be generalized if L_2, σ_2 are plugged in for the minimum price and the size of reach of any seller, and $(1 - F_3(L_2) = 1)$ as the (empirical) price distribution of the sellers with larger reach(σ). The criterion defines G1 to G3 sellers in 1.5 by the size of reach, G3 is the sellers with the largest σ, G1 are the smallest.

I calculated markups, the ratio $\frac{p_i - w_i\sigma_i}{p_i}$, in Figure 1.10.

$$w_2 = \frac{L_2(1 - F_3(L_2))}{\sigma_2} = \frac{L_2}{\sigma_2} \tag{1.11}$$

If the interaction structure is independent, then the cost parameter for advertisement choices should be estimated from a different first-order condition. The profit function of a seller in independent structure is (1.22), $\pi_i = \sigma_i p_0$. As the seller increases the reach size (σ_i), the marginal revenue the seller gets from the increase is the minimum price p_0. We can write the profit function incorporating the cost and the optimal choice of advertisement as (1.13).

$$\pi_i = \sigma_i p_0 - c(\sigma_i) \tag{1.12}$$

$$p_0 - c'(\sigma_i) = p_0 - w_i\sigma_i = 0 \tag{1.13}$$

Then the advertisement cost parameter in independent case is

$$w_i = \frac{p_0}{\sigma_i}$$

The markups of the three groups of sellers in the independent case are in Figure 1.11. The calculated markups are from the listings for iPhone XR, whose test result suggests an independent structure. The comparison between the groups and the two different interaction structures is summarized in Table 1.18. The difference between group 1 and group 3 sellers is much more significant in the nested case. The group 3 seller is likely to have 2.21 times more markups than the group 1 seller on average in the nested structure. On the other hand, group 3 sellers show even lower markups than the group 1 seller in the independent case. The result shows that the nested structure helps the frequent seller who makes many posts to acquire higher markups.

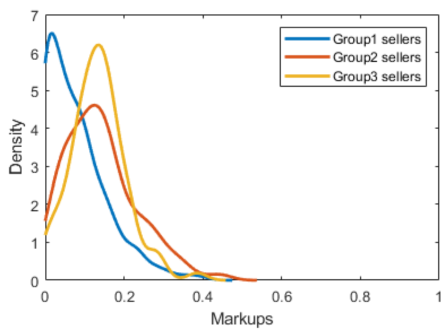


Figure 1.10: Markups of the sellers : Galaxy S9, Nested

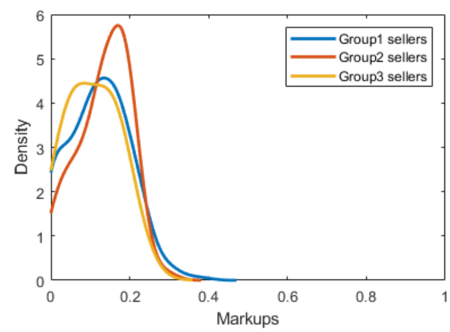


Figure 1.11: Markups of the sellers : iPhone XR, Independent

Table 1.18: Markups in interaction structure

Nested	Median	Mean	Std	Min	Max
G1 seller	0.061	0.079	0.078	0.000	0.410
G2 seller	0.128	0.140	0.090	0.000	0.446
G3 seller	0.135	0.134	0.069	0.000	0.385
Independent	Median	Mean	Std	Min	Max
G1 seller	0.124	0.122	0.078	0.000	0.388
G2 seller	0.141	0.134	0.066	0.000	0.301
G3 seller	0.084	0.104	0.061	0.007	0.215

1.6 Platform's Motive

The preceding analysis of power sellers and sellers' reach helps explain how interaction structure affects the sellers' general market price and pricing strategy. The natural question that arises is, how does the structure affect the platform's profit? As the platform receives transaction fees or commission rates, it has two complementary incentives: (1) To maximize the sale of higher-priced goods and (2) to increase total sales.

Under the nested case, the concentration of the postings can raise the chance of selling one good. If the expected return of Group 3 power sellers is higher than the less frequent group of sellers (Group 1), that will explain another reason why the platform prefers a nested structure. If the posting frequency increases the price and the sales probability, the platform would prefer the sellers with higher reach (σ) due to higher expected return. The formula of each listing's expected return is in (1.14), which is the sum of the probability of sales under the various size of σ multiplied by the price. The expected sales probability of each posting is imputed using the probit regression with the posting characteristics.

$$E(p\hat{q}_{sell}|\sigma, X) = \int p\hat{q}_{sell}(p, \sigma) f(p|\sigma, X) dp \tag{1.14}$$

The average returns of each listing by sellers aggregated in groups classified by posting frequency are in Table 1.19. Frequent posting can raise the price of the product and increase the chances of matching, neutralizing the negative effect of the price on selling probability. Eventually, this power seller-consumer dynamic within the nested structure increases platform revenue.

Table 1.19: One listing expected return, Galaxy S9

Seller Group	Mean
Group 1	5.718
	(1.549)
Group 2	6.082
	(2.128)
Group 3	7.32
	(2.572)

1.7 Discussion

This paper studies the sellers' advertising effort, measured by frequency of posting, and its relationship to the pricing strategy of third-party sellers in the online marketplace. A general framework of Armstrong and Vickers (2020) identifies and describes consumers and sellers' interaction patterns, explaining the seller's pricing strategy. The two polar cases of interaction patterns, independent and nested, are matched with the data to infer the closest structure.

The analysis shows the statistical significance of the price distribution difference between the sellers with various advertising levels, following the nested structure's predictions. Then, the nested structure is verified with the data using the model's predictions. First, if the postings made by power sellers with higher reach (σ) dominate the listing page, the price distribution gap between the sellers with smaller reach (σ) would get wider. Second, the entry of small sellers does not affect the pricing strategy of the sellers. That shows the listings made by frequent sellers are the key factor for the other seller's pricing strategy.

One potential concern would be the endogeneity of advertising. For example, one might argue the difference in quality is the source of listing frequency heterogeneity, leading to pricing patterns described by Armstrong and Vickers (2020) suggests. The sellers with higher quality products have more incentive to invest in higher rank(Athey and Ellison (2011)), and consequently, the price can be higher. This concern is bypassed by constructing the frequency measure throughout the whole observation period to reflect the time-invariant seller type.

Another potential reason for the higher price could be the difference in the sellers' ability for inventory management. For example, if the frequent sellers have a higher capacity of products than the less frequent sellers with various retail channels other than Cetizen, they can charge a higher price. Although this explanation cannot be analyzed well without better information about third-party sellers, the measure of reach σ constructed on the listing title level still shows stochastic monotonicity. That means there are other factors other than inventory that plays a role in pricing.

The estimated markups in various interaction patterns show that the market power of the power sellers in the platform is likely to be higher in the nested case than in the independent structure. In addition, it indicates that the sellers with higher visibility in the online platform may have higher market power than the traditional brick-and-mortar retail channels in terms of advertising. Also, how the platform is designed and what kind of rules are implemented for the listing can heavily affect the interaction structure, followed by the price distribution.

Finally, analyzing the platform's profit under the different interaction structures shows that the general market price can be affected by the advertising intensity of power sellers. The size of power sellers' captive group of consumers can raise the overall price in the market, which could be beneficial for the platform's profit. In addition, advertisement intensity (i.e., re-posting) can increase the chance of sales, which suggests it is beneficial for an online platform to allow category killers- or the power sellers to dominate the advertisement competition even without any additional fees for favorable search ranking.

1.8 Conclusions

This paper investigates online market sellers' pricing strategy and its relation to visibility using an interaction model developed by Armstrong and Vickers (2020). The visibility, measured by the number of repeated postings, increases the tendency to charge higher prices. Seller's price distribution identifies the interaction pattern between the sellers by using two statistical testing procedures; stochastic monotonicity testing and comparison of the quantile.

The result suggests a nested pattern among the sellers. The power sellers with a higher reach(σ) tend to charge higher price and market power. In the nested structure, the consumers are likely to consider the products of power sellers more than the less frequent sellers. As a result, the power sellers can enjoy the demand from the consumers with higher search costs.

The finding of the nested structure suggests the platforms can enjoy higher profit in the concentrated posting structure of power sellers. The study provides anecdotal evidence that concentration on a few power sellers can increase the market price in general and the chance of sales. It can also support why the online platform puts much effort in attracting power sellers or category killers through sponsored search.

This paper provides a general framework that can be applied to various pricing cases to identify interaction patterns. Future research could investigate extensions of the model. First, the model can incorporate the decision of repeated posting or the effort of getting visibility. Second, though the platform prefers active participation of the power sellers, whether consumer welfare enhanced or deteriorated from the domination of the power sellers is not clear. Finally, future research can incorporate an extensive margin of the market if both consumer and sellers' choices on the platform are available. By doing so, the optimal design of the platform can be discussed.

APPENDIX

1.A Derivation of the predictions from Armstrong and Vickers (2020)

This section expresses the relationship between sellers and consumers and interaction structures and price more formally. The availability size is called reach (σ_i), representing the fraction of consumers considering products from a particular seller i. The consumer type is defined by its consideration set and characterized as α_s with s consideration set.

First, we can prove how the reach size maps into the price distribution or the size of price support. Proposition 1 and 2 describes relationship between the size of σ and the price distribution.

Proposition 1. *If there are 5 group of sellers,*$\sigma_1 \leqslant \sigma_2 \leqslant \ldots \leqslant \sigma_5$*, each seller has price range with* $[\underline{p}_i, \bar{p}_i]$*. Then if the interaction is symmetric,* $\underline{p}_i = p_0$ *for all* i*. Under all circumstances,* $\bar{p}_1 \leqslant \bar{p}_2 \leqslant \ldots \leqslant \bar{p}_n$*.*

Proposition 2. *If the sellers are interacting independently, then for each p*$\in P$*,* $F_{p|\sigma}(p|\sigma) \leqslant F_{p|\sigma'}(p|\sigma')$ *whenever* $\sigma \geqslant \sigma' \quad$ *for* $\quad \sigma, \sigma' \in \Sigma$

The proof of proposition 1 and 2 follow Lemma 1 in Armstrong and Vickers (2020). Intuitively, every seller has an intersection set which includes consumers considering products from two or more sellers in independent interaction cases. Theoretically, the minimum price is the same across the sellers otherwise a seller with a higher minimum price will have an incentive to lower the minimum price to attract the consumers. Therefore, equilibrium with different lower bounds cannot exist. The expected profit in the independent case is $\sigma_i p_0$, which is is proportional to the size of reach (σ_i).

Given the fact that the minimum price is the same across all the sellers, there exists an equilibrium price that provides the common support of the seller's pricing distribution.

Proposition 3. *If n sellers have nested reach with* $\sigma_1 \leqslant \ldots \leqslant \sigma_n$*, seller i has minimum, maximum price range* L_i, H_i *with* $p_0 = L_1 = L_2 \leqslant L_3 \leqslant \ldots L_n, H_1 < H_2 < \ldots < H_{n-1} = H_n = 1$.

Armstrong and Vickers (2020) offered and proved Proposition 3 to describe the derivation of price regions in a nested interaction structure. In economic terms, the demand that small sellers encounter is likely to be elastic, eventually leading to lower prices. In the nested structure, consumers of small sellers' products also consider and compare product options from the larger consumers, and the perfect substitutability leads to lower prices.

Also, monotonically increasing price distribution can be observed in nested competition structures. First, the size of sellers' captive consumers determines conditions for the stochastic monotone relationship of price distribution. To see how the stochastic monotonic relationship works, assume p is included in the price region of seller i and i+1. In other words, $p \in [L_{i+1}, H_i]$ Then the price p needs to satisfy the profit condition of both seller i and i+1. Assume L_i is the lower bound of the price region of seller i. Then $F_i(L_i) = F_{i+1}(L_i) = 0$.

$$p(\beta_{i+1} + \beta_i(1 - F_i(p))) = \pi_{i+1} \tag{1.15}$$

$$p\beta_i(1 - F_{i+1}(p)) = \pi_i \tag{1.16}$$

Therefore,

$$\beta_{i+1} H_{i+1} = \pi_{i+1}$$

Using this condition, the price cdf can be written as follows.

$$\frac{1}{\beta_i}\left(\frac{\pi_{i+1}}{p}-\beta_{i+1}\right)=1-F_i(p)=\frac{\beta_{i+1}\left(H_{i+1}-p\right)}{\beta_i p} \tag{1.17}$$

$$F_i(p)=1-\frac{\beta_{i+1}}{\beta_i}\left(\frac{H_{i+1}-p}{p}\right) \tag{1.18}$$

The condition to make $F_i(p)$ decreases with i would be as follows.

$$F_i(p)=1-\frac{\beta_{i+1}}{\beta_i}\left(\frac{H_{i+1}-p}{p}\right)\leqslant F_{i-1}(p)=1-\frac{\beta_i}{\beta_{i-1}}\left(\frac{H_i-p}{p}\right) \tag{1.19}$$

$$\left(\frac{\beta_i}{\beta_{i-1}}\times\frac{\beta_i}{\beta_{i+1}}\right)\leqslant\frac{H_{i+1}-p}{H_i-p} \tag{1.20}$$

This suggests stochastic monotonicity of price distribution with respect to consideration probability, σ_i, and leads to the next proposition.

Proposition 4. *In nested reach case, stochastic monotonicity of* $F_{p|\sigma}$ *also would hold if for any* $i \geqslant 3$, $\frac{\beta_i}{\beta_{i-1}} \leqslant \frac{\beta_{i+1}}{\beta_i}$.

If the size of a seller's captive consumers (higher σ) is larger than non-power sellers, then the first-order stochastic dominant price distribution is expected. Additionally, the difference in consideration probability (σ_i) in the nested case suggests price distribution differences among and between groups. In other words, the price distribution gap would be derived as Equation (1.21). Additionally, the difference in reach (σ_i) in nested case can suggest price distribution difference of the groups. In other words, the price distribution gap would be derived as (1.21). If $\frac{\beta_{i+1}}{\beta_i} \geqslant \frac{\beta_i}{\beta_{i-1}}$, then the gap would get widened.

$$F_{i-1}(p)-F_i(p)=-\frac{\beta_i}{\beta_{i-1}}\left(\frac{H_i-p}{p}\right)+\frac{\beta_{i+1}}{\beta_i}\left(\frac{H_{i+1}-p}{p}\right) \tag{1.21}$$

Equation (1.21) shows the size of β_i and how its increase affects price distribution among the groups of sellers. If $\frac{\beta_{i+1}}{\beta_i} \geqslant \frac{\beta_i}{\beta_{i-1}}$, then the growth rate of the size of reach is higher in the power seller group, indicating a concentration of consideration probability. In the empirical context of Cetizen's platform, the consideration probability gap β describes the level of concentration of the platform's listings. If β is highly concentrated towards a handful of power sellers, then those sellers will dominate interactions with potential consumers.

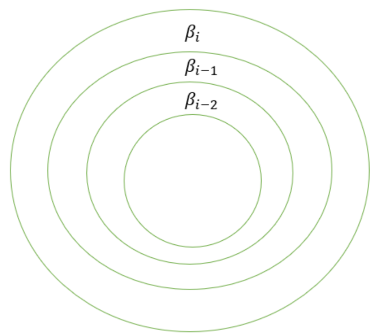


Figure 1.A.1: Captive Consumers: The Size of Captive Consumers (β_i) of Each Sellers

As the concentration of a seller's product listings increases, the power seller's reach will overwhelm the smaller set of a non-power seller. Importantly, power sellers' concentrations of listings causes two opposing forces to impact the price distribution, affecting price dispersion. Greater concentration of product postings implies that a power seller controls a large group of of captive consumers. Therefore, power sellers benefit from higher price dispersion. However, an increased concentration of product postings by a group of power sellers also increases the oligopolistic competition among the sellers; more specialized sellers lowers price dispersion. The propositions and proofs expanding upon Armstrong and Vickers (2020) (2020) predicts the concentration of consideration set structure will increase the price dispersion, but the degree of increase is not yet clear.

Proposition 5. *Concentration of reach (increasing $\frac{\beta_{i+1}}{\beta_i}$) would raise the price distribution ($F_i(p) < F_{i-1}(p)$). The direction of price dispersion is not clear.*

Additionally, the level of price distribution will fluctuate in relation to demand for the product. High product demand relative to the number of product listings leads to smaller demand elasticity, higher price, and higher dispersion. On the other hand, low demand relative to the number of listings reduces demand elasticity, resulting in a lower price and smaller dispersion.

Corollary 1. *The amount of shift in price distribution due to the listing concentration (increasing β_i) is larger for the product with higher demand. As a result, price is more likely to be dispersed in high demanded product with greater concentration of reach.*

In independent structure, the seller can charge between $[p_0, 1]$. Since mixed strategy applies to pricing, the expected profit for p_0 and 1 should be the same. In other words, for seller i with the size of reach σ_i , minimum price p_0 and captive consumer α_i, the profit condition should satisfy the following.

$$\sigma_i p_0 = \alpha_i = \pi_i \tag{1.22}$$

Therefore, entry of new sellers into an independent structure such that they increase an independent structure's intersecting consideration sets and reduce other sellers' captive consumers, will cause α_i to decrease.. In equilibrium, the minimum price p_0 will also decrease. Since all sellers share the minimum price p_0 , new sellers' entry reduces the average price as well. In contrast, the entry of a new, smaller seller into a nested structure, will not affect the price distribution of larger sellers. The aggregate price distribution across sellers is less likely affected by the entry of small sellers.

Proposition 6. *In independent case, the entry of new sellers regardless of the size of reach (σ) can reduce the price of other sellers. In nested case, the effect on* *the price distribution is minimal.*

1.B Summary Statistics

Table 1.B.1: Summary Statistics of the Listings

Model name	Count	% of All postings
iPhone X	12730	5.12
iPhone 8	26319	10.59
iPhone XS	9590	3.86
iPhone XR	8806	3.54
iPhone 11 Pro	15301	6.16
iPhone 11	11592	4.66
Galaxy A8	9966	4.01
Galaxy S9	32084	12.91
Galaxy S10	13638	5.49
Galaxy S10+	6402	2.58
Galaxy S10 5G	23192	9.33
Galaxy J3	6171	2.48
Galaxy Note8	23800	9.58
Galaxy Note9	26430	10.64
Galaxy Note10	22476	9.04
Month	Count	% of All Postings
February	23009	9.26
March	46249	18.61
April	48036	19.33
May	23457	9.44
June	30471	12.26
July	40511	16.3
August	36764	14.79
Stat.	Price	Normalized Price
Mean	403.81	0.2623
Std.	(242.88)	(0.08)

Table 1.B.2: Summary Statistics of the Sellers

Variable	Mean	Median	SD	Min.	Max.
# Models(Month,Seller)	2.66	1.00	2.82	1.00	15.00
# Postings for a Model(Month)	6.96	2.00	19.97	1.00	617.04
% of Samsung	0.55	0.64	0.43	0.00	1.00
# Models(Week, Seller)	2.46	1.00	2.50	1.00	15.00
# Postings for a Model(Week)	3.77	2.00	6.95	1.00	167.13

Models(Month, Seller) refers to the number of models that a seller sells within a month window. # Postings for a model(Month) means the number of postings that sell a model within a month window. % of Samsung refers to the average ratio of postings with Samsung product within entire postings by a particular seller. # Models(Week,Seller) is the number of models that are sold by a particular seller within a week window. # Postings for a model(Week) is the average number of postings that are made by a seller to seller a particular model within a week window.

Table 1.B.4: Probability of Selling - Each Product

Model	Galaxy S9	iPhone XR	iPhone 11	Galaxy S10
Probit	Sold	Sold	Sold	Sold
# Repeated Posting (Day)	0.175**	0.341***	0.197***	0.119
	(0.0674)	(0.0775)	(0.0353)	(0.0747)
Price Ratio ($)	-0.652***	-1.763***	-1.359***	-1.563***
	(0.197)	(0.505)	(0.316)	(0.252)
Model Share	-10.85*	-13.95	-12.97**	-1.161
	(5.077)	(13.75)	(4.684)	(2.911)
Average Daily # Postings (Seller)	0.0106*	-0.0103	-0.00449	0.0185***
	(0.00438)	(0.00734)	(0.00657)	(0.00491)
Const.	-1.212***	1.251***	1.384***	-0.407
	(0.288)	(0.335)	(0.412)	(0.319)
Controls	Yes	Yes	Yes	Yes
N	6953	2090	6798	7727
Pseudo R-sq.	0.017	0.032	0.021	0.033
AIC	1707.6	2233.2	7040.6	2463.5
BIC	1782.9	2295.3	7115.7	2540.0

Note: Standard errors clustered in month, model level are in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, "# Repeated Posting (Day)" is number of repeated posting of a unique listing. "Price Ratio" is the ratio, product price/average market price. A model share represents the posting share of a specific model within the total postings on the website that are observed within a day. "Average Daily # Postings (Seller)" is the number of average daily postings that a seller upload in the platform. Controls include conditions of the product, whether they are under the warranty, and memory size.

Table 1.B.3: Power Sellers of Each Product

Model	Seller > 50 Postings/Mth
A1901	18
A1905	40
A2097	14
A2105	10
A2215	21
A2221	10
SM-A530	12
SM-G960	64
SM-G973	20
SM-G975	12
SM-G977	36
SM-J330	9
SM-N950	34
SM-N960	46
SM-N976	26

Note: The number of sellers with more than 50 postings for each model of smartphone (Galaxy S9, iPhone 10, etc.)

Table 1.B.5: Price Regression

Price regression	Price ($)	Log (Price)
Almost New	175.5***	0.773***
	(4.758)	(0.0160)
Good Condition	73.21***	0.473***
	(2.467)	(0.0117)
Acceptable Condition	44.87***	0.323***
	(1.962)	(0.0101)
Expired Warranty	0.626	-0.0337***
	(0.734)	(0.00513)
Under Warranty	6.628***	0.0118*
	(1.106)	(0.00468)
Memory Size	0.254***	0.000463***
	(0.0201)	(0.0000241)
Const.	-99.60***	3.256***
	(5.882)	(0.0778)
Model FE	Yes	Yes
Month FE	Yes	Yes
N	810578	810578
R-sq.	0.939	0.948

Note: Standard errors, clustered in month, model level are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

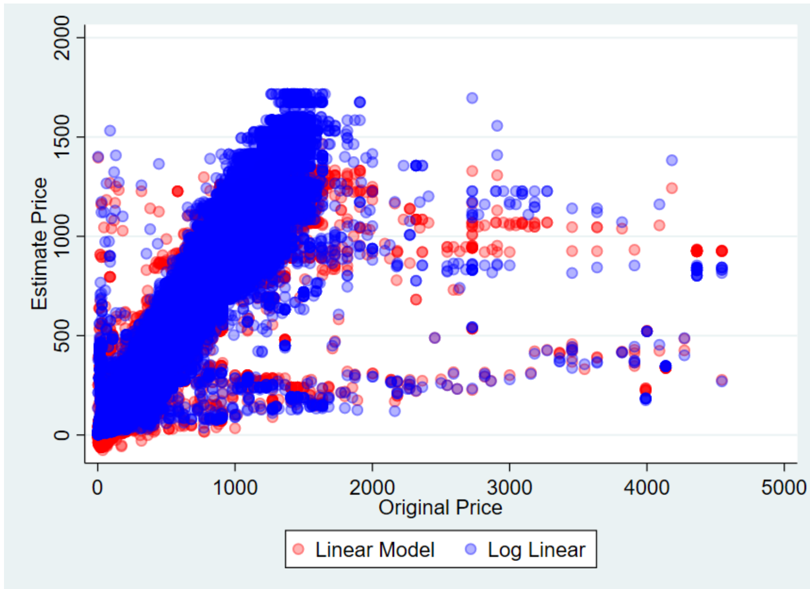


Figure 1.B.1: Estimated Price

Table 1.B.6: Difference between Regression Model and the Data

Stats	Linear	Log Linear
Mean	34.62	34.28
p25	11.08	8.42
p50	24.24	20.08
p75	45.02	42.49

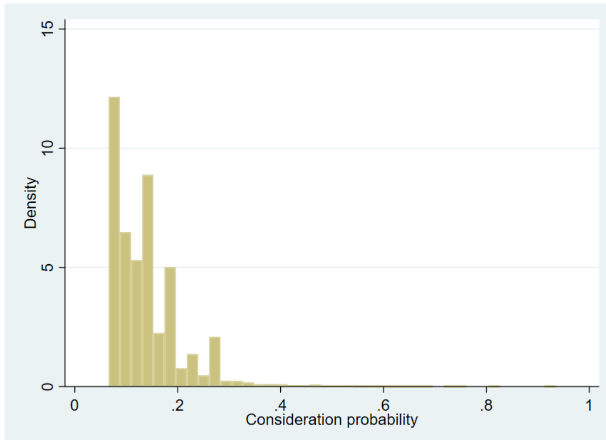


Figure 1.B.2: Histogram of σ_i

1.C Statistical Test

The stochastic monotonicity test Chetverikov et al. (2020) used in the empirical strategy section is in below.

First, the null hypothesis in Equation (1.5) can be simplified by using some notation. Assume

$$K_{ij,h}(\sigma) = sign(\sigma_i - \sigma_j)K_h(\sigma_i - \sigma)K_h(\sigma_j - \sigma)$$

Then

$$E\left[\sum_{i=1}^{n} 1(p_i \leqslant p) k_{i,h}(\sigma)\right] \leqslant 0 \tag{1.23}$$

Where

$$k_{i,h}(\sigma) = \sum_{j=1}^{n}(K_{ij,h}(\sigma) - K_{ji,h}(\sigma)) = 2\sum_{j=1}^{n} K_{ij,h}(\sigma)$$

Then the test statistics are chosen among the set of different values of bandwidth size. $h_{max} = 1, h_{min} = 1/n^{1-\delta}$ for some $\delta \in (0, 2/3]$.

$$B_n = \{h_{max} u^l | l = 0, 1, 2, \cdots, \left[\frac{\log(h_{max}/h_{min})}{\log(1/u)}\right]\}$$

for some $u \in (0, 1)$. Bandwidth is numerically selected within the values of B_n. Then the constructed test statistics are (1.24).

$$T = \max_{(\sigma,p,h) \in \Sigma_n \times p_n \times B_n} \frac{\sum_{i=1}^{n} k_{i,h}(\sigma) 1(p_i \leqslant p)}{\left(\sum_{i=1}^{n} k_{i,h}(\sigma)^2\right)^{1/2}} \tag{1.24}$$

1.D Two step results with various definition

- Various construction of σ

  A two-step analysis is also applied to the data with various reach(σ) construction. First, the σ is constructed by calculating the number of duplicated postings with the same title within 1 week window (same description). It is a granular way of defining the size of σ to circumvent potential seller heterogeneity problem. Second, σ is constructed based on the posting share of each seller within one month following (1.3). Third, the frequency (reposting) measure (λ_{jm}/λ_m) calculated using an hour window. Here, the type of the sellers are defined by each sellers' incidence rate that is measured within an hour window.

- Interpretation of the results

  The two step results are in Table 1.D.1, 1.D.2. Across the testing results with 3 different construction of σ shows rather consistent results in the first stage, volatile results in the second stage sensitive to the definition of groups. First order stochastic dominance structure is observed in most of the models, while the increasing minimum is only observed in some models. This suggests the amount of the benefits created from the frequent posting may differ across the cellphone models and the length of the observation window may matter due to the frequent entry and exit in the platform.

Table 1.D.4 shows the 2 step testing results with the two constructions of σ: (1)the frequency (reposting) measure (λ_{jm}/λ_m) calculated using an hour window , and (2) the percentage of total postings in a month of a seller within a cellphone model, and (3) the percentage calculated with the listing description, such as "Q92 128G White Unopened (normally closed)". The testing results with different construction of σ overlap. The testing result based on the description("Q92 128G White Unopened (normally closed)") is the most restrictive definition of the σ, creating the most conservative results. Even so, about half of the cellphone models support the nested structure.

Table 1.D.1: 2 Step testing for the Same Description

Summary	Step 1	Step 1 test	Step 2 (0.05)	Step 2 test	Result
Model	Mean P value	Reject	Mean P value	Reject	Nested
iPhone X	0.590	0	0.285	0	0
iPhone 8	0.777	0	0.000	1	1
iPhone XS	0.250	0	0.015	1	1
iPhone XR	0.000	1	0.000	1	0
iPhone 11	0.247	0	0.003	1	1
Galaxy A8	0.963	0	0.000	1	1
Galaxy S9	0.937	0	0.067	0	0
Galaxy S10	0.960	0	0.000	1	1
Galaxy J3	0.647	0	0.000	1	1
Galaxy Note8	0.850	0	0.052	0	0
Galaxy Note9	0.683	0	0.109	0	0
Galaxy Note10	0.000	1	0.011	1	0

Table 1.D.2: 2 Step testing with the Listing Share

Summary	Step 1	Step 1 Result	Step 2	Step 2 Result	Result
Model	Mean P value	Reject	Mean P value	Reject	Nested
iPhone X	0.347	0	0.000	1	1
iPhone 8	0.963	0	0.000	1	1
iPhone XS	0.440	0	0.000	1	1
iPhone XR	0.000	1	0.000	1	0
iPhone 11	0.030	0	0.020	1	1
Galaxy A8	0.673	0	0.433	0	0
Galaxy S9	0.990	0	0.000	1	1
Galaxy S10	0.937	0	0.327	0	0
Galaxy J3	0.247	0	0.273	0	0
Galaxy Note8	0.983	0	0.007	1	1
Galaxy Note9	0.973	0	0.447	0	0
Galaxy Note10	0.083	0	0.413	0	0

Table 1.D.4: Testing Results Based on Various Definition of σ

Model	**Frequency measure**	**Posting Share**	**Same description**
iPhone X	1	1	0
iPhone 8	1	1	1
iPhone XS	1	1	1
iPhone XR	0	0	0
iPhone 11	0	1	1
Galaxy A8	1	0	1
Galaxy S9	1	1	0
Galaxy S10	0	0	1
Galaxy J3	0	0	1
Galaxy Note8	1	1	0
Galaxy Note9	0	0	0
Galaxy Note10	0	0	0

Note: ’1’ refers to whether each cellphone models show nested structure. If it is not, the test result is written as ’0’. Column 1,"Frequency measure" uses σ constructed average number of postings within 1 hour window. Column 2,"Posting share" uses the percentage of listings made by a seller within 1 month window as σ. Column 3, "Same description" uses the percentage of postings made for each postings that has unique description, such as "Q92 128G White Unopened (normally closed)" within 1 week window.

Table 1.D.3: Quantile Regression: Only professional sellers

τ	Estimate of δ	
	(1)	(2)
0.05	56.57***	50.89**
	(10.21)	(18.05)
0.1	51.87**	49.47**
	(17.69)	(18.89)
0.5	261.8***	246.9***
	(22.69)	(27.66)
0.9	245.2***	215.1***
	(32.44)	(32.52)
0.95	193.5***	181.1***
	(29.29)	(38.39)
No. Models	14	14
Month FE	X	O
No. Observations	23098	23098

Professional sellers are defined as the sellers who sell more than 5 different models on average within 1 month.

2 AN EQUILIBRIUM ANALYSIS OF POWER PURCHASE AGREEMENT

2.1 Introduction

Compliance with emissions standards has incentivized firms to transform business operations (Lee, 2011; Brzeszczynski et al., 2019; Xia and Niu, 2020). In the U.S., the trend of emissions reduction can be credited to the advent of Renewable Portfolio Standards (RPS), which mandates the level of renewable energy consumption. Utilities and Independent Power Producers (IPPs) have to comply with the RPS criterion either by purchasing Renewable Energy Credit (REC) in the market or by directly purchasing renewable energy from the producer (EPA, 2021b). Since REC works as an instrument to circumvent the burden from the carbon footprints, more firms are signing physical or virtual PPAs with renewable energy producers to acquire RECs more directly (Level10Energy, 2021).

Increasing renewable energy penetration, requirements for carbon footprint, and decentralization of energy generation have made Power Purchase Agreements (PPAs) a common way of project financing power plants. PPAs are long-term contracts that enable the buyers of electricity to purchase electricity from renewable energy generators at a fixed price over 5 - 20 years (Bruck et al., 2018). More than 100 corporations across 23 countries used PPA to purchase renewable energy (Bloomberg Law, 2021). In addition, major corporations like Google, Amazon as well as U.S. state and local governments are using PPA to procure renewable energy (AWEA, 2019).

PPA incorporates the bilateral negotiation between a distributed energy resource and a buyer. The terms of the contract are mainly determined by the risks of renewable power generation: *market price risk* and *volumetric risk*. The *market price risk* stems from the fluctuation in the wholesale market whereas the *volumetric risk* is due to the volatility and the capacity in energy supply. These two risks mainly affect participants' entry decisions and equilibrium market price since it determines the overall excess benefit of using PPA compared to the trade in the wholesale market.

This paper analyzes the link between PPA prices and the risk factors. It is essential to study how PPA price is interrelated with these risk factors since it determines the financial return of an energy project. As distributed generation needs a bilateral contract at a price based on the wholesale market price (López-Lezama et al., 2010), PPA rates can represent the market equilibrium price. How the participants' characteristics and the other market factors such as wholesale market price affect the PPA price can provide hints on how the equilibrium rate of return is determined in a decentralized market. Additionally, revealing how the equilibrium price is formed and fundamental elements can help us answer what the regulator needs to focus on when they have a specific policy objective.

Even though PPAs have been a way of financing an energy project that has a long history (Thumann and Woodroof, 2021), an expansion of the use of PPA follows the same pattern as with the decentralizing trend in electricity supply. The concerns on distributed generation can be multiple (Rosen and Madlener, 2016), and finding economic reasoning is where PPA works. Cohen et al. (2020) considered PPA as a state incentive to attract more investment from the corporations. Distributed generation, which doesn't have that much capacity, would choose to make a bilateral contract with the buyer instead of selling energy in the wholesale market. Also, PPA provides a sales model with less risk by enabling third-party financing, and especially solar PV projects can avoid utility regulation (Cohen et al., 2020). Therefore, many cities are engaged in the initiative to use renewable electricity through PPA. In July 2017, local governments used 462 MW of renewable energy through PPA (Leung and Bailey, 2018).

We examine the equilibrium in the contract using the framework of a choice model of a buyer (off-taker) and the seller (generator) to sign a PPA for a wind project. We incorporate potential risks that a market participant might face, like the mean and volatility of wholesale market price and the contracted quantity of electricity. The model assumes a benchmark case where there is one representative buyer and seller. The utility function is borrowed from the literature on forward-contract (Bessembinder and Lemmon (2002), Brown and Sappington (2021)), where each agent has a certain level of risk preference. The conceptual choice model of generators (sellers) and power purchasers (buyers) yields several predictions of equilibrium PPA rates. First, wind generators are willing to accept lower PPA prices when the contracted quantity or the supply is higher. This is likely to be a form of quantity discount for reducing risk due to uncertainty from selling a large quantity of power in the wholesale market. Second, the presence of several buyers in the market will increase the equilibrium PPA prices due to competition. Third, if the buyer's return from selling the electricity in the wholesale market is high, they are willing to increase the PPA prices. Lastly, the volatility of the electricity price effect on the PPA price depends on the risk aversion parameter of the participants.

We verify the predictions from the theoretical model in the empirical analysis of this paper. We use the Heckman 2 Step procedure to account for the sample selection issues. We model the selection step using the project off-taker (buyer) type and size of the wind developer as the excluded instrument. The variation in risk preference amongst different project off-takers and wind developers causes variation in PPA adoption and, therefore, the PPA rates. The results show a positive association between the mean retail and wholesale electricity prices. As shown in the theoretical model, we find evidence of a significant negative association between contracted quantity (measured by the capacity of the project) and equilibrium PPA prices, indicative of a quantity discount.

In contrast to the other papers, this paper provides a unique view of the PPA equilibrium. First, unlike other papers which look at the return of the standalone project (Bruck et al., 2018), bilateral optimization problem (López-Lezama et al., 2010), this paper studies the equilibrium aspect of PPA with the choice of participation of both buyers and sellers. Levelized Cost of Energy (LCOE), which is a traditional method of calculating the profitability of energy projects (Bruck et al., 2018; Nissen and Harfst, 2019), cannot explain the general equilibrium effect stems from the participation choice. Since PPA is a bilateral contract that involves the interaction and competition among the buyers and sellers, LCOE is not a good measure to capture the strategic interaction between the buyer and sellers.

Second, this paper discovers how the wholesale market is related to the PPA market by looking at the risk component in the wholesale market's effect on the PPA rates. There has been a paper which studied the association between the electricity spot price and wind power production using a copula model and how the amount of wind power production will affect the PPA contract(Tranberg et al., 2020). Still, it does not fully incorporate the differences in buyer and seller characteristics and the market equilibrium effect. The inherent risk of PPA stems from the wholesale electricity price, and how it relates to PPA price shows burden-sharing between the buyers and sellers. The paper shares the same insights as in (Nissen and Harfst, 2019) by acknowledging the price effect.

Third, this paper shares a similar view of PPA as in Ghiassi-Farrokhfal et al. (2021) by focusing on both sides of PPA, which is quite rare in the literature. Some papers look at the incentive structure of sellers in PPA (Lei and Sandborn, 2018), but not many papers focused on both sides of PPA. Our paper focuses on the equilibrium rates in PPA, sharing the same insight of the papers which look at the contents in the contract (Mendicino et al., 2019). This paper provides a view on how to interpret the concept of PPA. : PPA is essentially a forward contract between buyer and seller.

Finally, the investigation of the PPA equilibrium provides several policy implications as government policy choices can affect not only the entry of market participants but also the contractual terms. The regulations can have implications of PPA prices in the long-run: Market openness, conditions on the participation, limitation on the contract length and buyer, seller screening criterion can affect equilibrium outcome in the PPA market. If a government aims to promote active participation in the PPA market, they need to consider the potential impacts of regulation on entry.

This article is structured as follows. The following section describes the general background of the PPA. Section 3 builds a model of PPA, which examines the choice of buyers and sellers of wind energy to participate in the PPA market. Section 4 describes the data used for the empirical analysis. Section 5 details the empirical analysis aimed at testing the predictions from the theoretical model. Section 6 provides a discussion of results and potential policy implications of the findings.

2.2 Background

Power Purchase Agreement

Power Purchase Agreement (PPA) is a long-term contract of buying power from a company that produces electricity such as a renewable energy generator like a utility or IPP (Thumann and Woodroof, 2021). The buyer, who is called an off-taker, agrees to purchase the electricity at a price agreed in the contract (EPA, 2021a). The buyer is typically a retailer or a private firm that sources the energy through a bilateral contract. The agreement includes conditions of the contract such as the amount of energy supply, length of the agreement, negotiated price, and other terms regarding the extra cost and additional burdens.

There are two types of power purchase agreements; Physical PPAs and Virtual PPAs. Both PPA types provide consistent pricing for the long-term by setting a fixed contracted price (Dormady, 2017). Physical PPA refers to the contract between a buyer and energy producer and physical delivery of the electricity (EPA, 2021a). The renewable energy generator and the buyer need to be located in the same market or grid to provide power physically. On the other hand, a virtual PPA is a financial contract with a fixed price guarantee that does not require physical electricity delivery. Therefore, there are no location constraints in a virtual PPA compared to a physical PPA. The buyers have to pay the difference between the contracted PPA price and the market price in a virtual PPA. By doing so, the buyer gets REC to offset the carbon footprint. Both types of PPAs enable the developer to hedge the market price risk of electricity by making a fixed price contract for a designated length of the period of electricity use (Morgan, 2018).

Because of the off-taker's environmental benefit from RECs or carbon offsets, PPA is utilized more frequently in renewable energy generation. In addition to that, PPA also provides financial stability to the energy projects by making the developers easy to borrow money from the bank (PEXA-PARK, 2021). PPA increases the revenue certainty, which is beneficial for both renewable energy generators and users.

The choice of signing a PPA can be analyzed on both sides: buyer and seller. PPA is a good way for sellers to secure future returns by selling the electricity at a pre-negotiated price, thereby reducing the uncertainty of project return. The seller would compare the PPA rates to the wholesale electricity price. Suppose the PPA price is lower than the wholesale price. In that case, the generator will sell the electricity in the wholesale market. Similarly, a buyer would perceive PPA price as the cost of energy procurement or the value of electricity from renewable sources. If it is higher than the market price, the buyers can choose to source the electricity from the wholesale market.

Figure 2.2.1 shows a decreasing trend of PPA price for wind projects over the last decade. The decreasing trend is due to several factors. First, technological advancements in both solar and wind generation lowered the cost of renewable energy. Second, a rise in competition amongst energy sellers also lowered the PPA price (Level10 Energy, 2020). Finally, federal subsidies like Production Tax Credits and state level policies like Renewable Portfolio Standard have lowered PPA prices across the US. In Figure 2.2.1, the PPA price moves in the same direction as the off-peak Locational Marginal Price in Texas[1].

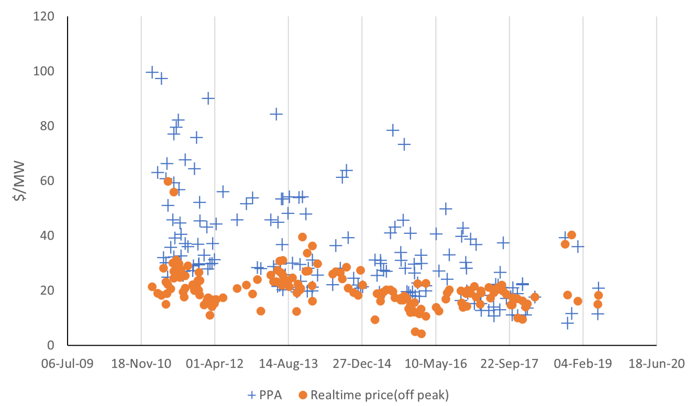


Figure 2.2.1: Wind PPA prices and Wholesale electricity price
Notes: Figure provides decreasing trend of PPA price which is from Berkeley Lab, Utility-Scale solar 2020 and the mean electricity price in the off peak hours in Texas (HB_{BUSAVG}). The electricity price is from S&P Global Market Intelligence platform.

While we expect the PPA price to move in the same direction as the wholesale market price, the gap between the two prices is determined by several sources. One of the sources is the buyer's and seller's bargaining process, and the other is their risk preferences.

We modeled the PPA equilibrium to analyze the relationship between the abovementioned sources and the PPA rates. There are several reasons why it is necessary to model a PPA market. First, predicting PPA prices would be complicated without a formal model of participants' choices. By modeling the participation choice of buyer and seller, we can simulate

[1]We compare PPA prices to the off-peak spot prices in Texas because most of the wind generation happens during the off-peak hours. This provides a reliable comparison between PPA rates and prices in the wholesale electricity market.

how the exogenous factors affect the market equilibrium, thereby making the prediction of PPA prices feasible. Second, a direct comparison between wholesale and PPA prices is misleading because it doesn't account for players' decisions on participation. To the best of our knowledge, this is the first paper with a formal equilibrium analysis of the PPA market incorporating the choice of participation.

2.3 Equilibrium Model of PPA

This section examines the choice of a buyer (off-takers) and a seller to sign a PPA for a wind project and the following equilibrium using an analytical model. We model this choice under potential risk factors: the size of market price w, market price volatility measured by the variance σ_w, the quantity demanded $\bar{Q}$, and the quantity generated x. The effect from the wholesale market price (w) is referred to as the *price risk*, while the impact from quantity fluctuations (x) is the *volumetric risk*.

Price risk is due to the variation of the wholesale market price (w). The PPA price would move in the same direction as the mean market price, and the market price volatility can affect the relative attractiveness of a PPA contract. The *volumetric risk* is due to the variation of the generated energy x and the contracted quantity M. Suppose there is excess energy generation over the contracted amount of electricity. In that case, the generator can sell the extra energy in the wholesale market. A rational generator would likely sign a contract for a fixed amount of electricity supply less than the expected amount of energy generation$(x \geqslant M)$. The following section presents the formal choice model of buyers and sellers based on the risk mentioned above factors.

The buyer and seller decide to participate in the contract after comparing the wholesale electricity price w and PPA price $\underline{p}$, which determines the future return and associated expected utility. If both are risk-averse, future return volatility will induce the agents to participate in the contract more. As a benchmark case, I borrowed the mean-variance utility function from the forward contract literature(Hirshleifer and Subrahmanyam (1993),Anderson and Danthine (1980),Brown and Sappington (2021)). There is one buyer and one generator in the market, who have different levels of risk preference which are represented by $A_i, i = \{b, g\}$. The argument inside the utility function is net revenue, which is written in π. The utility of each agent depends on the expected revenue and the variance of it $(E(\pi), var(\pi))$.

$$U_i(\pi_i) = E(\pi_i) - A_i var(\pi_i), \quad i = \{b, g\} \tag{2.1}$$

The formulation of demand and supply are designed to reflect the contract process of PPA. Usually, the agreement goes through several steps. In the first stage, a renewable energy project that needs financing or in the building process determines a rough structure of the contract and reach out to potential buyers. The buyers compare multiple offers and choose the best one. Then they negotiate the terms and sign a PPA contract. We can interpret the equilibrium as the negotiation process where details and the price of the electricity are determined. If there is only one buyer and supplier, it would be similar to one-to-one matching. If there are multiple buyers, it would be similar to the oligopolistic structure where buyers or suppliers compete.

First, I assumed electricity generation(x) and the wholesale market price (w) is independent. This assumption is realistic in many wind projects since the wind price in consideration(w) is the future price which is independent from the wind capacity of the generators.

Assumption 2.1. *The electricity capacity of each wind project (x) and wholesale market price (w) is independent*$(x \perp w)$.

Also, I assumed both buyer and seller are price takers in the wholesale and retail market. This assumption could be restrictive since the buyer and seller can trade large amounts of electricity, affecting the equilibrium in the wholesale and retail market. However, I'm departing from general equilibrium where the wholesale, retail and PPA market are related in this paper for the simplicity of the analysis, but it can be relaxed in future work.

Assumption 2.2. *Both buyer and seller are price taker in the wholesale and retail market.*

The market price volatility does not increase the utility of both buyer and seller. In other words, we assumed both buyer and seller are risk-averse. The reason why we don't include risk-neutral cases is as follows. In risk-neutral case, either buyer or seller would choose a corner solution. Therefore, the risk factors such as price variation or quantity fluctuation do not affect both buyer and seller's choice, making it harder to analyze the effect of the risk factors on the equilibrium.

Assumption 2.3. $A_i > 0, \quad i = \{b, g\}$

To make the problem simple, I assumed generators are homogeneous in terms of their risk preferences with only one A_g in the market. This assumption may oversimplify the reality because the wind projects are likely to be vastly different across developers and owners. Yet, it is a reasonable starting point since the generators offer the contracts to the buyers, and buyers have the option to accept or reject the offer. Furthermore, the buyers already observe the intention of participation from the seller's side when they agree. In other words, the buyers usually have bargaining power. So the model simplifies the matching as the buyers have heterogeneous intentions to participate in the contract. At the same time, the generators always want to participate in PPA with the same risk preference (A_g).

Moreover, we assumed a retailer in the electricity market as a buyer. Though the data show diverse off-taker types, including utility companies and many different kinds of organizations, the utility company case can generalize the participation choice of off-takers. Because other types of organizations get a fixed level of utility that depends on the actual usage of electricity, it can be a special case of the model.

The timing of the model is as follows. First, a buyer decide whether to participate in the contract after observing retail, wholesale market price , generator's risk preference parameter (A_g), expected generation level(x). Next, a buyer choose the amount of purchase(F) that depends on the contracted price, a generator decides the amount of sales (M) and the contracted price is determined.

Benchmark Model

First, I present a benchmark model where there is only one buyer and one generator in the market. It would be similar to the case where 1 buyer and 1 seller make an exclusive match, and each side makes an optimal decision on the amount of electricity traded. The chosen amount of electricity traded is essentially the quantity demanded and supply in a contract. Based on the demand and supply function, the equilibrium price is chosen. For the notation, r is the retail price, $\bar{Q}$ is the amount of electricity sold in the retail market, F is the amount of electricity purchased in PPA, M is the amount of electricity sold in PPA. x is the total electricity production by the generator.

The return of buyer and generator after making a PPA contract is in (2.2),(2.3). The buyer gets revenue from selling the electricity in the retail market $(r\bar{Q})$, and the cost of sourcing the electricity is divided into the wholesale market$(w(\bar{Q}-F))$ and PPA(F). Some part of electricity is from the wholesale market$((\bar{Q}-F))$, and the rest is from the PPA (F). The stochastic components in this formulation are two: the wholesale market price (w) and the total generation (x).

$$\pi_b = r\bar{Q} - w(\bar{Q} - F) - \underline{p}F \tag{2.2}$$
$$\pi_g = M\underline{p} + (x - M)w \tag{2.3}$$

As the utility depends on the expected and variance of the return, the values of both party are calculated as (2.4)-(2.9).

$$\begin{aligned}
E(\pi_b) &= E(r)\bar{Q} - E(w)(\bar{Q} - F) - \underline{p}F && (2.4)\\
var(\pi_b) &= E\left[(r - E(r))\bar{Q} - (w - E(w))(\bar{Q} - F)\right]^2 && (2.5)\\
&= var(r)\bar{Q}^2 + var(w)(\bar{Q} - F)^2 - 2cov(r, w)(\bar{Q} - F)\bar{Q} && (2.6)\\
E(\pi_g) &= M\underline{p} + (E(x) - M)E(w) && (2.7)\\
var(\pi_g) &= -2Mcov(xw, w) + var(xw) + M^2 var(w) && (2.8)\\
&= var(xw) + M(M - 2E(x))var(w) && (2.9)
\end{aligned}$$

As a subgame perfect equilibrium, the participating buyer and seller's demand and supply determine equilibrium price of the contract. The buyer maximize the utility by choosing the size of F (future demand) which maximizes the expected utility of the return.

$$\begin{aligned}
argmax_F E(u_b) &= argmax_F E(\pi_b) - A_b var(\pi_b) && (2.10)\\
&= E(r)\bar{Q} - E(w)(\bar{Q} - F) - \underline{p}F && (2.11)\\
&- A_b\left[var(r)\bar{Q}^2 + var(w)(\bar{Q} - F)^2 - 2cov(r, w)(\bar{Q} - F)\bar{Q}\right] && (2.12)
\end{aligned}$$

By solving the first order condition of the expected return, the optimal choice of future (F) is determined as (2.13)

$$E(w) - \underline{p} - A_b(2var(w)(F - \bar{Q}) + 2\bar{Q}cov(rw)) = 0$$

$$F = \bar{Q} - \frac{\bar{Q}\mathrm{cov}(r,w)}{\mathrm{var}(w)} - \frac{\underline{p} - E(w)}{2A_b \mathrm{var}(w)} \tag{2.13}$$

Similarly, we can solve the problem of the generator as in (2.14).

$$\mathrm{argmax}_M E(u_g) = \mathrm{argmax}_M E(\pi_g) - A_g \mathrm{var}(\pi_g) \tag{2.14}$$
$$= M\underline{p} + (E(x) - M)E(w) \tag{2.15}$$
$$- A_g [\mathrm{var}(xw) + M(M - 2E(x))\mathrm{var}(w)] \tag{2.16}$$

The optimal choice of future supply (M) is in (2.17).

$$M = x + \frac{\underline{p} - E(w)}{2A_g \mathrm{var}(w)} \tag{2.17}$$

In equilibrium, the quantity demanded and supply should be the same $(F = M)$.

$$\bar{Q} - x = \frac{1}{2\mathrm{var}(w)} \left[(\underline{p} - E(w))(\frac{1}{A_b} + \frac{1}{A_g}) + 2\bar{Q}\mathrm{cov}(r,w) \right]$$

The equilibrium PPA rates $(\underline{p}^*)$ is a function of risk aversion parameter (A_b, A_g) and quantity demanded, supply $(\bar{Q}, x)$.

$$\underline{p}^* = \frac{A_g A_b}{A_b + A_g}(2\mathrm{var}(w)(\bar{Q} - x) - 2\bar{Q}\mathrm{cov}(r,w)) + Ew$$

We analyzed the direction of $\underline{p}$'s movement using the comparative statics. First, the PPA rates would be higher if either buyer or generator is risk-averse. The more the player is risk-averse, the higher the PPA price. Also, an increase in wholesale market electricity prices would increase the PPA rates. If the quantity demanded is higher than the total electricity generation $(\bar{Q} \geqslant x)$, an increase in wholesale market price volatility would raise the PPA rates.

On the other hand, wholesale market price volatility would decrease the PPA rates if there is more energy generation ($x \geqslant \bar{Q}$). In other words, the relative size of demand and supply can be a crucial factor in the relationship between the wholesale market price volatility and the PPA rates. Because the wholesale market price represents the opportunity cost of both buyer and seller, PPA rates would be positively related to the wholesale market.

The one thing that's not clear is the covariance between the retail and wholesale market prices. If the wholesale market and retail price are aligned (higher $\text{cov}(r, w)$), then an increase in wholesale price would lead to a rise in the retail price, maintaining the same margins for the retailer. Therefore, the buyer would not have much incentive to make a PPA contract. In contrast, lower cov(r,w) means a lower retail price when the wholesale price is higher. Thus, it leads to lower margins. Therefore, the buyer would have a higher incentive to make a contract. In other words, PPA rates and cov(r,w) have a negative association.

Table 2.3.1: Comparative statics

Variable	$\underline{p}$
A_b, A_g	$+$
$\text{var}(w)$	$+(\bar{Q} > x), -(\bar{Q} < x)$
$E(w)$	$+$
$\text{cov}(r, w)$	-
$\bar{Q}$	$+$
x	-

In the first stage where the buyer decide whether to purchase electricity from the PPA or not, the buyer compares the option of purchasing from the PPA and from the wholesale market. If we call the utility of the buyer when it purchase F^* amount of the electricity from the PPA as $E(u_b, F^* > 0)$ and the utility of not purchasing $E(u_b, F^* = 0)$, then $E(u_b, F^* > 0) - E(u_b, F^* = 0) > 0$. As F^* is a function of $\underline{p}$, and $\underline{p}$ is the function of parameters in equilibrium, we can derive the condition of buyer's risk aversion parameter that enables the market equilibrium.

$$E(u_b, F^* > 0) - E(u_b, F^* = 0) \tag{2.18}$$

$$= (E(w) - \underline{p})F^* - 2\bar{Q}F^*\text{var}(w) + \text{var}(w)(F^*)^2 + 2\text{cov}(r,w)F^*\bar{Q} > 0 \tag{2.19}$$

$$F^* > \frac{2(\text{var}(w) - \text{cov}(r,w))\bar{Q} + \underline{p} - E(w)}{\text{var}(w)} \tag{2.20}$$

$$(1 - \frac{\text{cov}(r,w)}{\text{var}(w)})\bar{Q} - \frac{\underline{p} - E(w)}{2A_b\text{var}(w)} > \frac{2(\text{var}(w) - \text{cov}(r,w))\bar{Q} + \underline{p} - E(w)}{\text{var}(w)} \tag{2.21}$$

Plugging in $\underline{p}^* = E(\underline{p}^*) = \frac{A_g A_b}{A_b + A_g}(2\text{var}(w)(\bar{Q} - E(x)) - 2\bar{Q}\text{cov}(r,w)) + Ew$ provides a condition for the risk preference parameter A_b.

Proposition 7. *1. If $\bar{Q}(var(w) - cov(r,w)) + 2A_g[var(w)(\bar{Q} - x) - \bar{Q}cov(r,w)] < 0$, or $\underline{p}^* - E(w) < 0$ then the risk aversion parameter should be higher than a threshold. Risk averse buyers would participate in the PPA.*

$$A_b > \frac{-A_g[2\bar{Q}(var(w) - cov(r,w)) - xvar(w)]}{\bar{Q}(var(w) - cov(r,w)) + 2A_g[var(w)(\bar{Q} - x) - \bar{Q}cov(r,w)]} \tag{2.22}$$

And energy generator's risk preference parameter should be higher than a threshold (2.23).

$$A_g > \frac{\bar{Q}(var(w) - cov(r,w))}{-2\left(\bar{Q}(var(w) - cov(r,w)) - xvar(w)\right)} \tag{2.23}$$

2. If $\bar{Q}(var(w) - cov(r,w)) + 2A_g[var(w)(\bar{Q} - x) - \bar{Q}cov(r,w)] > 0$ and $2\bar{Q}(var(w) - cov(r,w)) - xvar(w) < 0$, risk loving buyers and sellers who satisfy (2.24,2.25) would participate in the PPA.

$$A_b < \frac{-A_g\left[2\bar{Q}(var(w)-cov(r,w))-xvar(w)\right]}{\bar{Q}(var(w)-cov(r,w))+2A_g[var(w)(\bar{Q}-x)-\bar{Q}cov(r,w)]} \tag{2.24}$$

$$A_g < \frac{\bar{Q}(var(w)-cov(r,w))}{-2\left(\bar{Q}(var(w)-cov(r,w))-xvar(w)\right)} \tag{2.25}$$

To interpret this findings, we need to show what the condition of $\bar{Q}(\text{var}(w)-\text{cov}(r,w))+2A_g[\text{var}(w)(\bar{Q}-x)-\bar{Q}\text{cov}(r,w)]$ means. By calculation, we can see the sign of $\bar{Q}(\text{var}(w)-\text{cov}(r,w))+2A_g[\text{var}(w)(\bar{Q}-x)-\bar{Q}\text{cov}(r,w)]$ represents the sign of $\underline{p}^*-E(w)$. Changing the notation as a function of $\underline{p}-E(w)$ and risk parameters, the condition can be written as follows.

$$\bar{Q}(\text{var}(w)-\text{cov}(r,w))+2A_g[\text{var}(w)(\bar{Q}-x)-\bar{Q}\text{cov}(r,w)] \tag{2.26}$$

$$=\text{var}(w)F+\frac{\underline{p}-E(w)}{2A_b}+\frac{A_g+A_b}{A_b}(\underline{p}-E(w)) \tag{2.27}$$

$$\bar{Q}(\text{var}(w)-\text{cov}(r,w))+2A_g[\text{var}(w)(\bar{Q}-x)-\bar{Q}\text{cov}(r,w)]<0 \tag{2.28}$$

$$\leftrightarrow \underline{p}-E(w)<0 \tag{2.29}$$

$$\leftrightarrow \text{var}(w)(\bar{Q}-E(x))-\bar{Q}\text{cov}(r,w)<0 \tag{2.30}$$

So in other words, if $\underline{p}^*-E(w)$ is sufficiently small, then risk averse buyers and sellers would participate in the contract. The level of risk aversion of the generators would increase if the gap between the PPA contract price and the wholesale price gets wider.

If $\underline{p}^*-E(w)$ is not sufficiently small, or even when the contracted price is higher than the average spot price, then only the risk loving participants choose the contract. PPA contracts wouldn't be attractive as much to the buyers and generators.

Also, this condition for risk aversion parameter shows that the buyer type who choose PPA would depends on the market environment including wholesale market, retail market volatility, quantity demanded, and the amount of electricity generation. Eventually, the participation decision of the buyers affects the equilibrium price $\underline{p}$. This supports potential selection problem incorporated in the PPA rates analysis.

We can also derive an equilibrium when there are multiple generators with the same risk preference A_g. If we assume the number of generators as K, then the equilibrium can be derived as follows.

First, each generator would generate x_i and contract M_i. The sum of total generation across various generators would be written as M.

$$M = \sum_{i=1}^{K} M_i = \sum_{i=1}^{K} (x_i + \frac{\underline{p} - E(w)}{2A_g \text{var}(w)}) = \sum_{i=1}^{K} x_i + \frac{(\underline{p} - E(w))K}{2A_g \text{var}(w)}$$

In equilibrium where the quantity demanded and supply is the same($M = F$), the equilibrium PPA rates would be determined as (2.31).

$$\underline{p} = \frac{2\text{var}(w)}{KA_b + A_g} A_b A_g \left\{ (1 - \frac{\text{cov}(r,w)}{\text{var}(w)})\bar{Q} - x \right\} + Ew \tag{2.31}$$

The effect of K on $\underline{p}$ depends on the sign of $(1 - \frac{\text{cov}(r,w)}{\text{var}(w)})\bar{Q} - x$, which can be interpreted as the sign of $\underline{p} - E(w)$. If there are more generators (with a larger K), the equilibrium PPA rate would get lower. This finding suggests the PPA rates would get lower if there are more generators in the market.

$$\begin{cases} (1 - \frac{\text{cov}(r,w)}{\text{var}(w)})\bar{Q} - x > 0 (\leftrightarrow \underline{p} - Ew > 0) & \frac{\partial \underline{p}}{\partial K} < 0 \\ (1 - \frac{\text{cov}(r,w)}{\text{var}(w)})\bar{Q} - x < 0 (\leftrightarrow \underline{p} - Ew < 0) & \frac{\partial \underline{p}}{\partial K} > 0 \end{cases} \tag{2.32}$$

2.4 Data

For the empirical analysis in this paper, we use data on all the wind projects within the US from 1981 to 2020 from the American Wind Energy Association (AWEA). The dataset would include information like the project off-takers, PPA rate if the project off-take type was a PPA, nameplate capacity, location, and other project characteristics. We merge the data from AWEA with the annual mean power sector data for each state from EIA, and the electricity price of different ISOs from S&P Global Market Intelligence. The integrated data includes the annual mean electricity price of each ISOs and states, the amount of electricity generated by various sources, including wind, petroleum, coal, natural gas, and the number of retail electricity consumers of each state published by EIA.

The aggregate information on the power sector across the states and years will serve as valuable explanatory variables for the empirical analysis of equilibrium PPA rates. For example, the variation in retail electricity prices across various states could be indicative of the association between market prices and PPA rates. The number of retail consumers can show how the electricity market size affects equilibrium PPA rates. The amount of electricity generated by various sources is converted to capture the ratio of wind power generation in a state over time.

The geographical distribution of wind project and PPA is in Figures 2.4.1 and 2.4.2. The wind projects are mostly populated in California, Texas, Oregon, and some Midwest states. The frequency of PPA goes proportional to the number of wind projects as in Figure 2.4.2. The participation of PPA is likely to be higher for the deregulated market such as Texas and California.

Table 2.4.1 summarizes the proportion of various types of PPA off-takers. Among 1,686 wind projects in our data, about 47 % of projects (795) use PPA. This table disaggregates projects using PPAs by the off-taker type. Among the various types of off-takers, less than half of the

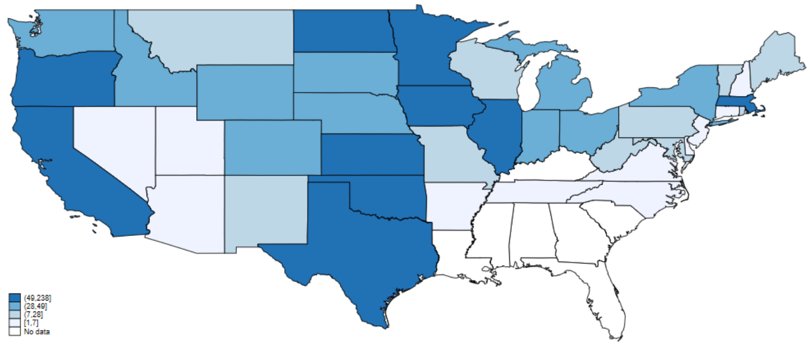


Figure 2.4.1: Wind project distribution

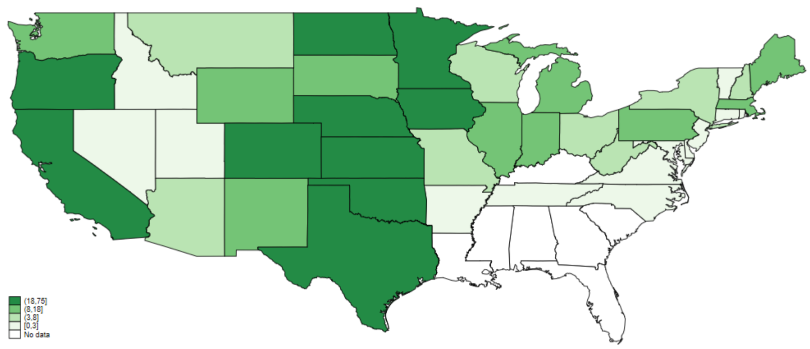


Figure 2.4.2: PPA distribution

PPAs (43%) are contracted with investor owned utilities, 10% with private firms (commercial & industrial), and 10% with regional governments (municipal). Generally, the proportion of off-taker type both in PPA and wind projects looks similar.

Figure 2.4.3 plots the total number of PPA contracts and the mean wind capacity (MW) over the years. We can notice that both project size and the number of PPA contracts have been increasing. We drop projects with less than 10 MW of nameplate capacity to restrict the empirical analysis on utility-scale projects. Table 2.4.2 shows the mean contract length and PPA rates for projects disaggregated into quintiles of nameplate capacity. We can notice that the mean capacity of the project in the first quintile is about 19 MW, whereas that in the fifth quintile is about 240 MW. There is an

Table 2.4.1: Type of Wind Project Off-takers

Type of Purchaser	PPA	PPA/wind	Wind project
CCA	3	0.75	4
Commercial & Industrial	94	0.45	211
Cooperative	104	0.65	161
Federal Power Authority	12	0.75	16
Government Agency	13	0.26	50
Investor Owned	341	0.46	738
Military	1	0.07	14
Municipal	83	0.48	174
Political Subdivision	32	0.57	56
Power Marketer	32	0.48	66
School	6	0.10	58
State	9	0.45	20
Transmission	1	0.50	2
Tribe		0.00	4
Unknown	64	0.57	112
Total	795	0.47	1,686

Note: Besides PPA, electricity generated from other wind projects is sold to off-takers. This table summarizes the types of off-takers of all wind projects. CCA refers to Community Choice Aggregator.

inverse relationship between project size and PPA rate, whereas the mean contract length is almost the same across the five quintiles. The decrease in PPA rates with higher project capacity is most pronounced for projects in the second quintile and higher.

This paper only focuses on the PPA contracts for the wind industry. The reasons why we didn't include solar are the following. First, solar PPAs often include rooftop solar for many homeowners and small businesses with different financial incentives than utility-scale solar projects. Second, due to the little bargaining power of individual sellers of rooftop solar, the PPA rates cannot fully reflect the strategic decision between a buyer and a

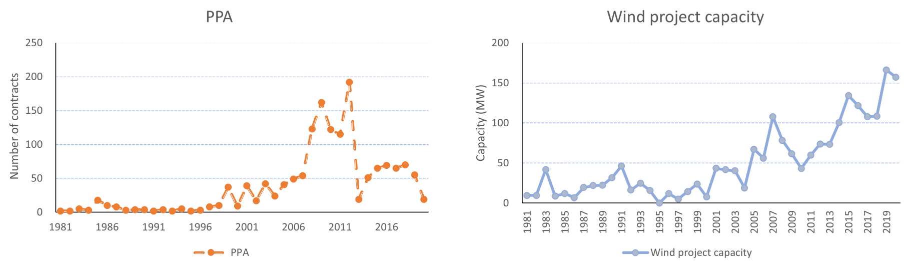


Figure 2.4.3: Growth of wind capacity and PPA

Table 2.4.2: Average Project Characteristics by Capacity Quintile

Capacity Quintile	Mean Capacity (MW)	Mean Contract Length (years)	Mean PPA Rate ($/MWh)
1	18.79	16.72	66.54
	(6.92)	(6.17)	(41.98)
2	52.71	18.52	49.45
	(11.98)	(6.75)	(20.85)
3	95.82	19.02	51.63
	(9.85)	(6.51)	(23.60)
4	150.43	18.10	57.57
	(20.98)	(6.06)	(32.07)
5	241.42	19.28	43.73
	(57.83)	(5.47)	(23.29)
Total	111.12	18.40	52.24
	(83.24)	(6.29)	(27.38)

seller. Finally, rapid price fall in the solar PV market goes along with the increase in PPA adoption rates. Therefore, it is hard to distinguish the pure market effect (electricity price effect) on PPA and solar system costs. In contrast, wind power shows a relatively smaller decrease in construction cost. From 2013 to 2018, costs for solar fell about 50% while wind power fell by 27% (EIA, 2020).

2.5 Empirical Analysis

Determinants of PPA rates

The conceptual choice model of generators (sellers) and power purchasers (buyers) yields several predictions of equilibrium PPA rates. The predictions from the model can be summarized as follows: First, wind generators are willing to accept lower PPA prices when the contracted quantity is higher. It is a quantity discount for reducing risk due to uncertainty from selling a large amount of power in the wholesale market. Second, the presence of several sellers in the market will decrease the equilibrium PPA prices if the PPA price is higher than the average spot price. Third, the PPA rates would move in the same direction as the wholesale market price. Lastly, the volatility of electricity prices would raise the PPA rates if there is more demand in the market ($\bar{Q} > x$). The equilibrium rates will be lower if there is more supply ($\bar{Q} < x$).

The empirical analysis in this section attempts to test the above predictions by using the data of wind projects in the US that signed the PPA. The estimating equation can be written as follows:

$$y_{it} = \alpha + \mathbf{X}_{it}'\Pi + \epsilon_{it} \tag{2.33}$$

where, y_{it} is the PPA rate (\$/MWh) of project i that began operation in period t. The vector $\mathbf{X}$ includes the relevant explanatory variables and fixed effects. We use contract length (years), mean and standard deviation of retail electricity price (\$/MWh), nameplate capacity (MW), total retail customers (in 1,000,000) as a proxy for market size, and proportion of electricity generated by wind in the state as the variables of interest. We use Independent System Operators (ISO) fixed effect and year fixed effects to control for variation in PPA rates that could be correlated to system-level regulations or specific characteristics. To control for time-varying shocks common to all projects, we use the year of operation fixed effects. The standard errors are clustered at the state-year level to account for correlation amongst the states.

Table 2.5.1 shows the results of OLS estimation of Equation 2.33. The baseline specification is Column (4). From Column (4) in Table 2.5.1 we see a negative association between the capacity quintiles and proportion of electricity from wind in the state suggesting an inverse relationship between the supply (M) and equilibrium PPA price.

To account for non-linearities in the association between nameplate capacity and PPA rates, we separate the capacity variable into five quintiles. The base group is the projects with a capacity of 10 - 30 MW. We see a decline in PPA prices for projects with capacity in [30, 78) MW and [78, 116] MW intervals compared to projects in 10 - 30 MW. However, from Column (4) we do not observe a monotonic pattern of price decline. The reason why PPA rates are lower for the group with higher capacity could be a sign of quantity discount, as generators can avoid uncertainty by making a contract with a large capacity. Then the generators would be willing to take a lower PPA price. Another possible explanation is that the sellers with larger capacity are involved in PPA more frequently, resulting in an increase in PPA supply and thus reduce the PPA rates.

The positive coefficients of the annual mean of residential and retail electricity price suggest a positive relationship between PPA price and market prices as predicted by the conceptual model. The residential and retail electricity price (total electricity price) reflects electricity price variation across the states, and they are used as proxy for wholesale electricity price in various locations. This implies that higher returns in the electricity market are likely to push the PPA price higher, thereby increasing the opportunity cost of sellers to engage in the PPA market.

On the other hand, the market price volatility does not substantially impact PPA; it still shows positive signs, which supports the case of $\bar{Q} > x$

where there is more demand than supply in the market.

Table 2.5.1: PPA Rates : OLS results

	PPA rates (1)	PPA rates (2)	PPA rates (3)	PPA rates (4)
Contract length	0.486	0.264	0.322	0.417
	(0.328)	(0.331)	(0.374)	(0.330)
Residential Electricity price(mean,Cents/Kwh)	4.109			
	(2.461)			
Residential Electricity std(monthly var,Cents/Kwh)	-10.09			
	(8.768)			
Total Electricity price(mean,Cents/Kwh)		6.160**	3.629*	3.179*
		(2.144)	(1.476)	(1.407)
Total Electricity std(monthly var,Cents/Kwh)		0.192	2.718	3.855
		(6.526)	(6.739)	(6.754)
Total # retail customer(1,000,000)	1.432	0.613	-0.285	-0.311
	(1.149)	(0.821)	(0.583)	(0.572)
Wind Proportion	-155.4***	-110.4**	-129.8***	-112.6***
	(34.45)	(34.19)	(34.98)	(31.31)
Capacity(MW) $\in$ [30,78)				-21.40*
				(9.889)
Capacity (MW) $\in$ [78,116)				-16.34
				(8.407)
Capacity (MW) $\in$ [119.3,198.5)				-17.32
				(9.854)
Capacity (MW) $\in$ [198,496)				-21.01*
				(8.962)
Total Capacity(MW)	-0.0331	-0.00336	-0.0121	
	(0.0253)	(0.0234)	(0.0161)	
Const	15.52	-11.77	35.73*	55.61**
	(27.18)	(17.49)	(16.25)	(18.32)
Year FE	No	Yes	Yes	Yes
ISO FE	No	No	Yes	Yes
N	213	213	213	213
R-sq	0.442	0.59	0.705	0.723

Notes:This table reports the results of the OLS regression of Equation 2.33. The sample includes wind projects with nameplate capacity larger than 10 MW. The residential electricity price is the retail electricity price sold to households. The total electricity price is the aggregated price across residential, commercial, industrial, and transportation electricity prices measured at the state level. Each price is calculated as annual mean (average across monthly mean prices) and standard deviation. Wind proportion means the ratio of wind power generation to total generation in each state. Capacity is divided into five quintiles to account for non-linearities in the association between capacity and PPA rates. Total capacity represents the wind power generation capacity of each different wind project. Robust standard errors clustered at the state-year level reported in parenthesis. Significance: * p<0.05, ** p<0.01, *** p<0.001

One concern in Equation 2.33 is the sample selection issue of signing PPAs. In other words, the selected samples that have PPA rates records are likely to be the ones that have a high incentive to participate in the PPA contracts. If risk preference parameters are the primitive of the PPA participation decision, then the equilibrium PPA rates are likely to be biased by the size of the parameters. Thus, coefficient estimates for the retail electricity market price from OLS estimation can also be biased.

To incorporate sample selection issues, We explicitly model the selection of a project into signing PPA by estimating the Heckman 2 Step Model. The first step involves estimating the selection equation, which in this case, is the binary choice of whether to sign a PPA or not. Assuming d_{it}^* as the latent variable that indicates whether the project i signed a PPA at year t, then the latent regression specification of the selection equation can be written as:

$$d^*_{it} = \mathbf{Z}'_{it}\Gamma + \mathbf{X}'_{it}\Psi + \eta_{it}; \tag{2.34}$$

$$d_{it} = \begin{cases} 1 & \text{if } d^*_{it} \geqslant 0 \\ 0 & \text{if } d^*_{it} < 0 \end{cases} \tag{2.35}$$

The decision to sign a PPA is again modeled as a function of covariates included in Equation 2.33 along with the project off-take type and whether the project was developed by one of the top 5 developers in the US. The off-take types can partially control the risk type of the buyers and top 5 developer also can refelect the risk type of the sellers. The project off-taker types include whether the project off-taker is Community Choice Aggregation (CCA), Co-operative, Commercial and Industrial firm, Federal Power Authority, Government Agency, Investor Owner Utility, Municipality, or a Power Marketer. As seen from Table 2.4.1, there is a significant difference across various off-takers for PPA adoption, which is likely due to different risk preferences for PPA adoption. Also, the size and types of developers affect the PPA adoption rate. We included the dummy variable, which indicates whether the project is developed by the top 5 developers - Avangrid, NextEra, EDP Renewables North America, EDF Renewables, and Invenergy. As shown in Table 2.5.2, big developers are more likely to sign PPAs than smaller developers. Therefore, we use project off-take type and developer size as the instruments denoted by $\mathbf{Z}_{it}$.

The next step is the estimation of the intensity equation. To correct for the selection bias in Equation 2.33, we use the Inverse Mills Ratio (IMR) obtained from the selection step and include it in the intensity equation.

Assuming $\eta_{it} \sim N(0,1)$, the estimate of IMR $\hat{\lambda}_{it}$ is:

$$\hat{\lambda}_{it} = \frac{\phi(\mathbf{Z}_{it}'\Gamma)}{\Phi(\mathbf{Z}_{it}'\Gamma)} \tag{2.36}$$

The identifying variation for the IMR is the variation in the risk preferences of different project off-takers and preference by developers size for participation in the PPA market. The intensity equation can be written as:

$$y_{it} = \alpha + \delta\hat{\lambda}_{it} + \mathbf{X}_{it}'\Pi + \epsilon_{it} \tag{2.37}$$

where, y_{it} is the PPA rate of project i. The intensity equation includes all the covariates and fixed effects as in Equation 2.34 and the estimate of IMR $\hat{\lambda}_{it}$ which corrects for the selection bias. As discussed above, the type of project off-taker and developer size summarized by the vector $\mathbf{Z}_{it}$ are the excluded instruments in the selection step.

Table 2.5.2: PPA Ratio by Developer Size

Big Developer	# Total PPA contracts	Ratio PPA/Wind
0	530	0.31
1	275	0.49

Big developers include top 5 developers - Avangrid, NextEra, EDP Renewables North America, EDF Renewables, and Invenergy. The observation period is the whole sample period (1981-2020). Ratio PPA/Wind is the proportion of wind projects that are financed by PPA.

Table 2.B.1 in appendix shows the results of the selection equation while the results of the second step (intensity equation) is summarized in Table 2.5.3. Similar to the OLS results, we see a negative association between project capacity and PPA prices in Table 2.5.3. The coefficient estimates imply that PPA prices are expected to be lower when contracted quantity is higher. This effect is robust across all the five specifications in Table 2.5.3

Additionally, I used the annual mean and standard deviation across the monthly mean price of electricity under the ISO(Independent system operator). In contrast to the retail price, the average ISO price negatively affects PPA rates, while the ISO price variation has a positive effect on PPA rates. This relationship implies that other factors may intervene in the linkage between the ISO electricity price and the PPA rates. One potential reason might be that the ISO with a high proportion of renewable energy generation is likely to have relatively higher electricity prices and high volatility, which is reflected in both PPA rates and electricity prices. Also, different levels of buyers' market power across the ISOs could create higher electricity prices and lower PPA rates. In addition, a diverse participation rate of both buyers and sellers in PPA across various ISOs due to the variation in regulations and market environment also affects the contrasting result of ISO price.

A simple scatter plot shows a very weak positive association between the PPA rates and ISO electricity prices as in Figure 2.5.1. This figure plots the relationship between the aggregated yearly mean electricity price measured in ISOs and PPA rates of wind projects. One potential reason why it's hard to see a clear association between the two is that PPA and its utilization of renewable energy are rather new. The project-level data is linked to electricity prices measured every year, and the number of data points is rather limited. Additionally, ISO electricity price is more limited since there are only 9 ISOs in the US, and not all of them use wind power generation.

2.6 Discussion and Policy Implications

This paper examines the underlying risk of making a long-term PPA contract between a power generator and a purchaser. We derive several predictions on the equilibrium PPA prices based on a stylized choice model of

Table 2.5.3: Results of the intensity equation from the Heckman 2 Step Procedure

Heckman 2 step	PPA rates	PPA rates	PPA rates	PPA rates	PPA rates
Contract Length	0.517*	0.444	0.468*	0.509	0.248
	(0.229)	(0.234)	(0.231)	(0.284)	(0.322)
Retail Electricity price(mean,Cents/Kwh)	2.992***	3.271***			
	(0.815)	(0.829)			
Retail Electricity std(monthly var,Cents/Kwh)	1.765	1.383			
	(5.114)	(5.191)			
Wind Proportion	-109.6***	-128.5***	-112.2***	-105.0**	-159.0***
	(29.10)	(29.46)	(29.52)	(34.85)	(38.20)
Capacity(MW)$\in$ [30,78)	-20.71***		-20.49***	-47.73***	
	(5.881)		(5.943)	(8.585)	
Capacity (MW)$\in$ [78,116)	-15.58**		-15.03*	-38.77***	
	(5.890)		(5.941)	(8.876)	
Capacity (MW)$\in$ [119.3,198.5)	-17.04**		-16.65**	-43.19***	
	(6.067)		(6.155)	(8.907)	
Capacity (MW)$\in$ [198,496)	-20.89**		-20.12**	-49.18***	
	(6.408)		(6.469)	(9.274)	
Total Capacity(MW)		-0.0148			-0.0362
		(0.0180)			(0.0268)
Total Electricity price(mean,Cents/Kwh)			2.83*		
			(1.053)		
Total Electricity std(monthly var,Cents/Kwh)			2.749		
			(5.838)		
ISO price (mean,$/MWh)				-2.268***	-2.388***
				(0.482)	(0.531)
ISO price (monthly var,$/MWh)				6.102***	5.598***
				(0.888)	(1.014)
IMR	2.915	4.024	3.18	-1.910	-1.103
	(3.871)	(3.957)	(3.888)	(3.995)	(4.366)
Const	47.87***	28.32*	51.99***	150.2***	130.6***
	(14.04)	(13.36)	(14.15)	(29.93)	(31.37)
ρ					
σ					
ISO FE	Yes	Yes	Yes	No	No
Year FE	Yes	Yes	Yes	Yes	Yes
N	593	593	593	490	490
Wald chi2	535.22	488.95	527.12	220.38	144.19
prob>chi2	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: This Table is the result of the Heckman 2 Step procedure Equation 2.37. The sample includes wind projects with nameplate capacity larger than 10 MW. The residential electricity price is the retail electricity price sold to households, while the total electricity price is the aggregated price across residential, commercial, industrial, and transportation electricity prices measured at the state level. I calculated the annual mean by averaging the monthly mean electricity price of the ISOs. Also, the standard deviation is calculated as a variation across the monthly average prices. The total number(#) of retail customers is the number of electricity buyers in various states. Wind proportion is the ratio between the energy generated by wind power and the total generated electricity. ISO price is measured by the annual mean of monthly price (mean of monthly price), the standard deviation of monthly price (variation of the monthly price). Capacity is divided into five quintiles to account for non-linearity in the association between capacity and PPA rates. Standard errors in parentheses. Significance: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

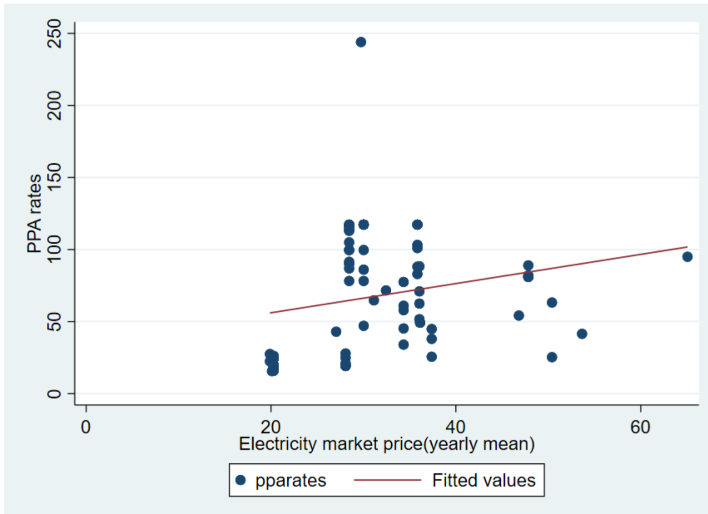


Figure 2.5.1: PPA rates and Wholesale electricity price (yearly mean)

PPA participation. First, the model predicts a negative association between the quantity supply and the PPA prices. Also, the equilibrium suggests a positive association between the wholesale and retail electricity prices with the PPA prices. On the other hand, how price volatility affects the equilibrium PPA rates depends on the market's relative size of demand and supply. We can see a positive association though it's not substantial.

We test the predictions using data on all utility-scale wind projects in the US. We use the Heckman 2 Step procedure, wherein we model the selection process as a function of risk preferences of the project off-take type and developer size. The results from the empirical analysis are in line with the predictions of the theoretical model and suggest several policy implications.

1. As shown in Table 2.B.1, buyer characteristics or project off-take types can affect the PPA price because of the difference in the risk preference. The theoretical model shows how the risk preference parameter size of both buyer and generator can create a difference in willingness to pay and equilibrium PPA rates. Policies targeting

entry of specific buyer types can therefore affect equilibrium PPA rates as the policy may affect the risk preference of the participants.

2. Wind generators with bigger capacities are likely to have lower PPA prices. Bigger wind farms have more incentive to participate in PPA to avoid uncertainty, lowering the equilibrium PPA price. Therefore, if the policy is favorable to large-sized energy generators because of some capacity restriction, it would decrease the PPA rates.

3. The regression results show the positive effect of mean retail electricity price and price variance on the PPA rates, while the impact from the price variance is not statistically significant. Various regression specifications show similar patterns in the association between retail electricity price and PPA rates. While the retail electricity price is not a perfect measure of the wholesale market price, we can easily see the return of both buyer and seller directly affects the equilibrium PPA rates. Thus, distortion of buyer and seller's returns due to the government's intervention can create changes in PPA equilibrium, leading to the entry and exit of the participants.

As of June 2019, 28 states along with Washington D.C. and Puerto Rico allow PPA. Texas has a system size limitation, which needs to be no more than the mean annual electricity consumption. Also, some states have restrictions on the term length; Arkansas has a limit of 5 years unless there is a relevant reason to extend, Connecticut with 20 years of limit, and quantity limit, which is 5% of the load. Michigan has a regulation on the buyer's condition; it needs to be public utilities with more than 500,000 customers, and it is allowed for less than 6 years unless they need to submit an additional application to show the need. Likewise, many states have some barriers for the minimum contract requirements and certain constraints on the term length (NCSL, 2015).

The regulations regarding the minimum requirements and term length can affect the entry decision of both buyers and sellers in the PPA market. The entry barrier could distort the participation decision of both buyers and sellers, changing the risk aversion parameter of the participants. Consequently, the participation decision would change the equilibrium price in the PPA market. For example, suppose the term length is too long. Then, only entities that can remain in the market over a long period with lower risk aversion parameters will enter the market, affecting the equilibrium outcome in the PPA market in the longer run.

Thus, how other environmental factors such as regulation on the contract design affect the financial burdens of both the supply and demand side and their incentives to participate. Barriers to entry due to contract constraints could affect a participant's incentives, resulting in less participation. Distorted participation may induce inefficiency in clean energy generation project financing by reducing or increasing PPA prices.

2.7 Conclusions

In this article, we discussed the equilibrium in the PPA market by analyzing a theoretical model of the two side participation decision and following an empirical analysis of wind power generation data in the US. We have found the relationship between risk factors such as the wholesale market price risk, generation level, and PPA rates. The model shows that the participation decision, which incorporates players' risk preference, could be critical in determining the equilibrium. The empirical analysis supports the predictions of the model. We could observe a positive association between the wholesale market price mean and volatility and the PPA rates and a negative association between the capacity and the PPA rates. These findings and framework can be used as a benchmark for future policy analysis.

APPENDIX

2.A Summary Statistics

Wind projects are dispersed across many ISOs. About 25 % of all wind projects are in the MISO region, and about 11% of wind projects are located in CAISO and ERCOT regions. The wind projects are not concentrated in a specific region and are distributed across the US. We matched the electricity price averaged across buses in the grid.

Table 2.A.1: Frequency of wind projects by Independent System Operators (ISOs) across the US

ISO	Frequency	Percent (%)
CAISO	184	11.23
ERCOT	191	11.65
ISO-NE	89	5.43
MISO	398	24.28
N/A	375	22.88
NYISO	43	2.62
PJM	131	7.99
SPP	228	13.91
Total	1,639	100

The summary stats of the electricity price is in table 2.A.2. "Avg residential price" is the average monthly mean residential, retail price of a state within a year. "Avg total price" is the average monthly mean retail price of states in a year. "Std residential price" and "Std total price" is the standard deviation across the monthly prices of each state. Avg ISO price is the yearly average price across monthly mean wholesale electricity prices of each state. Std ISO price is the standard deviation of the monthly average wholesale electricity price. The standard deviation of each measure across the states and years is in the parentheses.

Table 2.A.2: Summary Stats on the electricity price

Variable	Stats
Avg residential price(Cents/kWh)	11.28
	(3.07)
Avg total price(Cents/kWh)	9.1
	(2.81)
Std residential price(Cents/kWh)	0.58
	(0.32)
Std total price(Cents/kWh)	0.47
	(0.31)
Avg ISO price ($/MWh)	29.52
	(8.48)
Std ISO price ($/MWh)	9.63
	(7.92)

In figure 2.A.1, the distribution of yearly average residential prices in various states and PPA rates are plotted. There is some variation in the electricity prices, and PPA rates are generally more expensive than residential electricity prices.

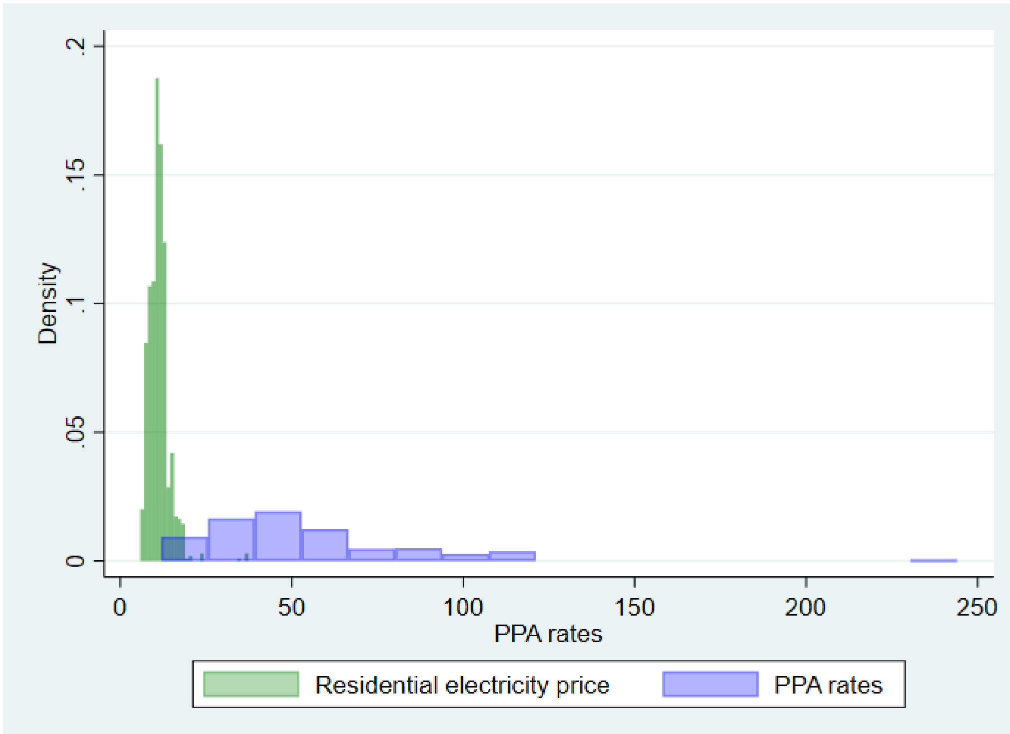


Figure 2.A.1: Residential electricity price and PPA rates variation

In figure 2.A.2, the relationship between average residential, total electricity price, and the average spot price in various ISOs. Total electricity price refers to the average price across industrial, residential, commercial, industrial, and transportation electricity prices measured at the state level. There is a weekly positive relationship between the average price and ISO electricity price.

The variation of residential retail electricity price standard deviation is in Figure 2.A.3a. The relationship between retail price variation and ISO price variation is in Figure 2.A.3b. The observation number of ISO price variation is limited, creating a weakly inverse relationship between the standard deviation of residential retail electricity price and ISO price standard deviation.

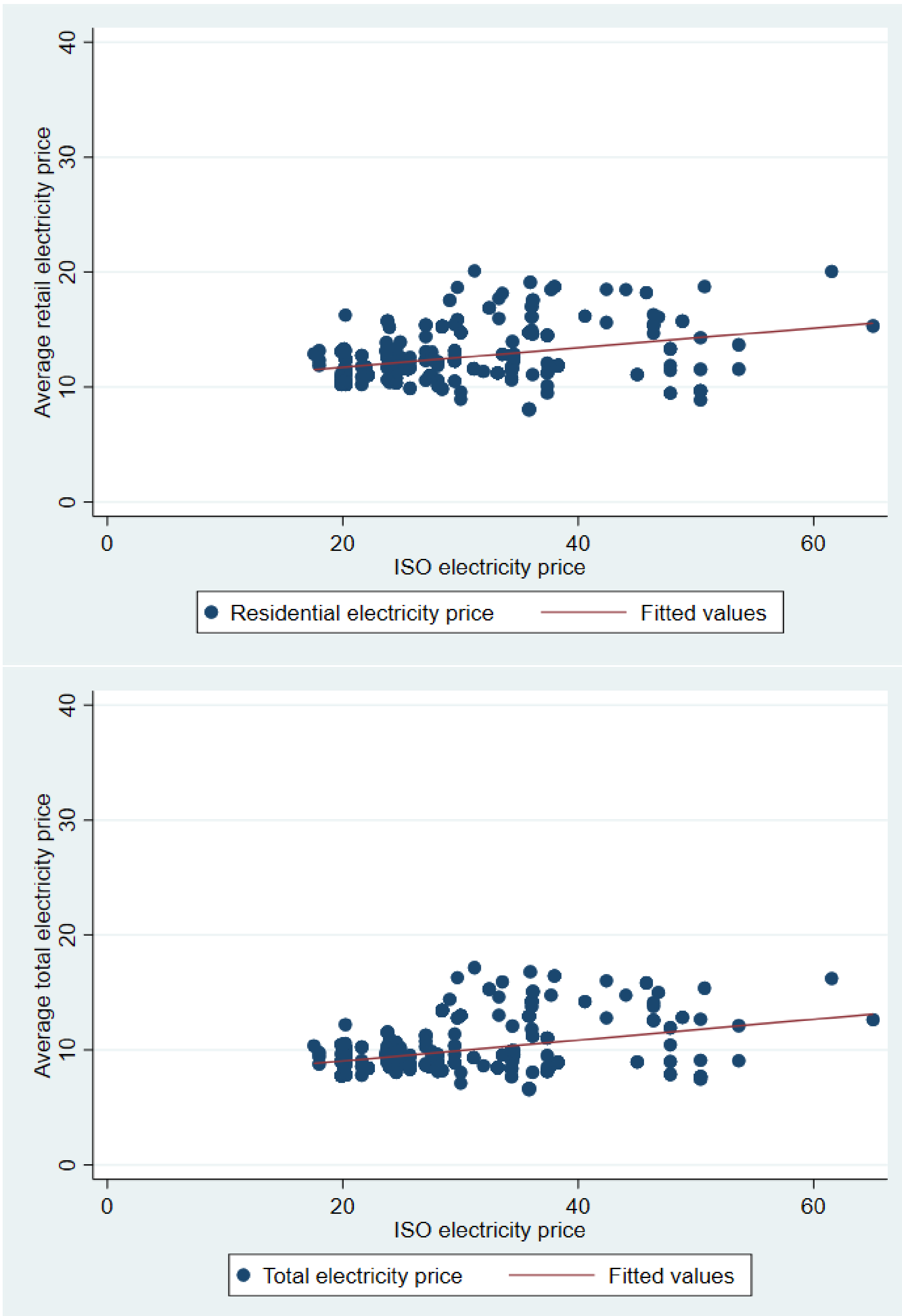
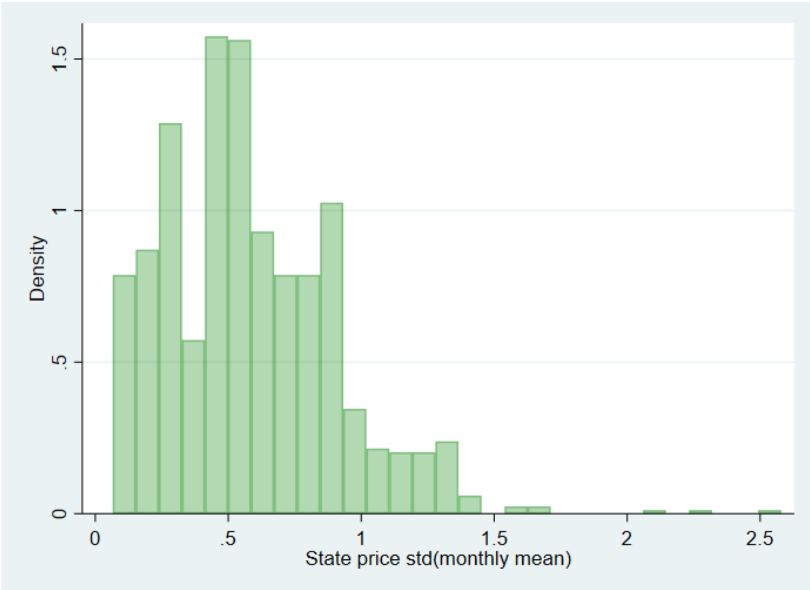
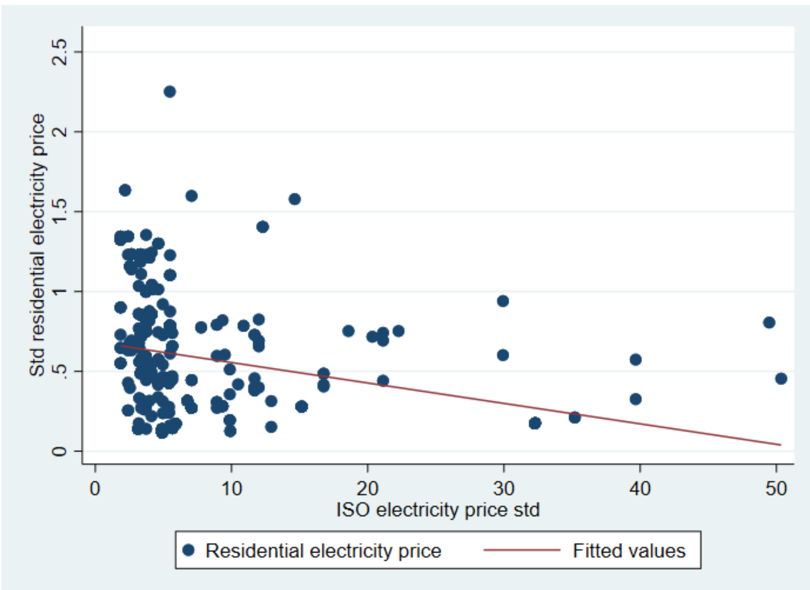


Figure 2.A.2: Retail electricity price and ISO price



(a) Residential retail electricity price standard deviation



(b) Retail electricity price standard deviation and ISO price variation

2.B Heckman first stage

The Heckman first stage result shows the variation of the PPA adoption rates across various off-taker types. As risk preference can affect the equilibrium price in the trade, the off-taker type is included in the first stage of Heckman analysis. The projects bought by federal power authorities and political subdivisions are likely to use PPAs more than other projects.

Table 2.B.1: Results of the Heckman first stage equation

First stage	PPA
Total number of commerical customer	0.0331*
	(2.44)
Big Developer	0.616***
	(5.24)
Commercial & Industrial	4.787***
	(6.25)
Cooperative	6.326***
	(8.57)
Federal Power Authority	7.135***
	(7.72)
Investor Owned	5.903***
	(8.12)
Municipal	5.813***
	(7.80)
Political Subdivision	6.388***
	(8.29)
Power Marketer	5.196***
	(6.70)
State	5.772***
	(5.54)
Const	-6.622***
	(-9.11)
N	508
Pseudo R^2	0.1468
Prob > chi2	0.0000

Notes: This table shows the results of the selection equation in Equation 2.34. For the sake of brevity we only the estimates of the excluded instruments. Big developers include top 5 wind developers in the US - Avangrid, NextEra, EDP Renewables North America, EDF Renewables, and Invenergy. Baseline PPA buyer type is CCA(Community Choice Aggregation). Standard errors in parentheses. Significance: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

REFERENCES

Anderson, Ronald W, and Jean-Pierre Danthine. 1980. Hedging and joint production: Theory and illustrations. *The Journal of Finance* 35(2):487–498.

Armstrong, Mark, and John Vickers. 2020. Patterns of price competition and the structure of consumer choice.

Armstrong, Mark, John Vickers, and Jidong Zhou. 2009. Prominence and consumer search. *The RAND Journal of Economics* 40(2):209–233.

Armstrong, Mark, and Jidong Zhou. 2011. Paying for prominence. *The Economic Journal* 121(556):F368–F395.

Athey, Susan, and Glenn Ellison. 2011. Position auctions with consumer search. *The Quarterly Journal of Economics* 126(3):1213–1270.

AWEA. 2019. US wind industry annual market report. Tech. Rep., American Wind Energy Association, Washington DC (United States).

Bagwell, Kyle, and Garey Ramey. 1994. Coordination economies, advertising, and search behavior in retail markets. *The American Economic Review* 498–517.

Bessembinder, Hendrik, and Michael L Lemmon. 2002. Equilibrium pricing and optimal hedging in electricity forward markets. *the Journal of Finance* 57(3):1347–1382.

Bloomberg Law. 2021. What's next for renewable power purchase agreements? Data retrieved from the website, `https://news.bloomberglaw.com/environment-and-energy/whats-next-for-renewable-power-purchase-agreements`.

Borenstein, Severin, and Nancy L Rose. 1994. Competition and price dispersion in the us airline industry. *Journal of Political Economy* 102(4): 653–683.

Bronnenberg, Bart J, Jun B Kim, and Carl F Mela. 2016. Zooming in on choice: How do consumers search for cameras online? *Marketing science* 35(5):693–712.

Bronnenberg, Bart J, and Wilfried R Vanhonacker. 1996. Limited choice sets, local price response, and implied measures of price competition. *Journal of Marketing Research* 33(2):163–173.

Brown, David, and David Sappington. 2021. Market structure, risk preferences, and forward contracting incentives. Working Papers 2021-12, University of Alberta, Department of Economics.

Bruck, Maira, Peter Sandborn, and Navid Goudarzi. 2018. A levelized cost of energy (lcoe) model for wind farms that include power purchase agreements (ppas). *Renewable Energy* 122:131–139.

Brzeszczynski, Janusz, Binam Ghimire, Tooraj Jamasb, and Graham McIntosh. 2019. Socially responsible investment and market performance: The case of energy and resource companies. *The Energy Journal* 40(5):17–73.

Burdett, Kenneth, and Kenneth L Judd. 1983. Equilibrium price dispersion. *Econometrica: Journal of the Econometric Society* 955–969.

Butters, Gerard R. 1977. Equilibrium distributions of sales and advertising prices. *The Review of Economic Studies* 44(3):465–491.

Chandra, Ambarish, and Mariano Tappata. 2011. Consumer search and dynamic price dispersion: an application to gasoline markets. *The RAND Journal of Economics* 42(4):681–704.

Chen, Xiaohong, Han Hong, and Matthew Shum. 2007. Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. *Journal of Econometrics* 141(1):109–140.

Chen, Yuxin, and Song Yao. 2017. Sequential search with refinement: Model and application with click-stream data. *Management Science* 63(12): 4345–4365.

Chetverikov, Denis, Daniel Wilhelm, and Dongwoo Kim. 2020. An adaptive test of stochastic monotonicity. *Econometric Theory*.

Cohen, Jed J, Levan Elbakidze, and Randall Jackson. 2020. Solar bait: How us states attract solar investments from large corporations. *The Energy Journal* 41(2).

Dai, Mian, Qihong Liu, and Konstantinos Serfes. 2014. Is the effect of competition on price dispersion nonmonotonic? evidence from the us airline industry. *Review of Economics and Statistics* 96(1):161–170.

Delgado, Miguel A, and Juan Carlos Escanciano. 2012. Distribution-free tests of stochastic monotonicity. *Journal of Econometrics* 170(1):68–75.

DiCiccio, Cyrus J, Thomas J DiCiccio, and Joseph P Romano. 2020. Exact tests via multiple data splitting. *Statistics & Probability Letters* 166:108865.

Dormady, Noah. 2017. Having your cake and eating it, too: Ohio's proposals to re-regulate through power purchase agreements. *The Electricity Journal* 30(5):42–46.

EIA. 2020. Average u.s. construction costs for solar and wind generation continue to fall. Tech. Rep., EIA.

Einav, Liran, Chiara Farronato, Jonathan Levin, and Neel Sundaresan. 2018. Auctions versus posted prices in online markets. *Journal of Political Economy* 126(1):178–215.

Eliaz, Kfir, and Ran Spiegler. 2011. Consideration sets and competitive marketing. *The Review of Economic Studies* 78(1):235–262.

Ellison, Glenn, and Sara Fisher Ellison. 2018. Match quality, search, and the internet market for used books. Tech. Rep., National Bureau of Economic Research.

EPA. 2021a. Physical power purchase agreements (physical ppas). Data retrieved from the website, `https://www.epa.gov/greenpower/physical-power-purchase-agreements-physical-ppas`.

———. 2021b. Renewable energy certificates (recs). Data retrieved from the website, `https://www.epa.gov/greenpower/renewable-energy-certificates-recs`.

Ghiassi-Farrokhfal, Yashar, Wolfgang Ketter, and John Collins. 2021. Making green power purchase agreements more predictable and reliable for companies. *Decision Support Systems* 144:113514.

Hauser, John R, and Birger Wernerfelt. 1989. The competitive implications of relevant-set/response analysis. *Journal of Marketing Research* 26(4):391–405.

Hirshleifer, David, and Avanidhar Subrahmanyam. 1993. Futures versus share contracting as means of diversifying output risk. *The Economic Journal* 103(418):620–638.

Hong, Han, and Matthew Shum. 2006. Using price distributions to estimate search costs. *The RAND Journal of Economics* 37(2):257–275.

Honka, Elisabeth. 2014. Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics* 45(4):847–884.

Honka, Elisabeth, and Pradeep Chintagunta. 2017. Simultaneous or sequential? search strategies in the us auto insurance industry. *Marketing Science* 36(1):21–42.

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino. 2017. Advertising, consumer awareness, and choice: Evidence from the us banking industry. *The RAND Journal of Economics* 48(3):611–646.

Huang, Yufeng. 2021. Seller-pricing frictions and platform remedies.

Hui, Xiang, Maryam Saeedi, Zeqian Shen, and Neel Sundaresan. 2016. Reputation and regulations: Evidence from ebay. *Management Science* 62(12):3604–3616.

Hunold, Matthias, Reinhold Kesler, and Ulrich Laitenberger. 2020. Rankings of online travel agents, channel pricing, and consumer protection. *Marketing Science* 39(1):92–116.

Ishihara, Masakazu, and Andrew T Ching. 2019. Dynamic demand for new and used durable goods without physical depreciation: The case of japanese video games. *Marketing Science* 38(3):392–416.

Jolivet, Grégory, Bruno Jullien, and Fabien Postel-Vinay. 2016. Reputation and prices on the e-market: Evidence from a major french platform. *International Journal of Industrial Organization* 45:59–75.

Koulayev, Sergei. 2014. Search for differentiated products: identification and estimation. *The RAND Journal of Economics* 45(3):553–575.

Lach, Saul, and José L Moraga-González. 2017. Asymmetric price effects of competition. *The Journal of Industrial Economics* 65(4):767–803.

Lee, Ki-Hoon. 2011. Integrating carbon footprint into supply chain management: the case of hyundai motor company (hmc) in the automobile industry. *Journal of cleaner production* 19(11):1216–1223.

Lee, Sokbae, Oliver Linton, and Yoon-Jae Whang. 2009. Testing for stochastic monotonicity. *Econometrica* 77(2):585–602.

Lei, Xin, and Peter A Sandborn. 2018. Maintenance scheduling based on remaining useful life predictions for wind farms managed using power purchase agreements. *Renewable Energy* 116:188–198.

Leung, Jessica, and Amy Bailey. 2018. How cities benefit from power purchase agreements. *Center for Climate and Energy Solutions. Retrieved from https://www. c2es. org/site/assets/uploads/2018/09/how-cities-benefit-from-ppas. pdf*.

Level10 Energy. 2020. Ppa price index, q1 2020. Tech. Rep., Level10 Energy.

Level10Energy. 2021. 4 ways to get renewable energy certificates: Pros,cons of each. Data retrieved from the website, `https://www.leveltenenergy.com/post/ways-to-get-renewable-energy-certificates`.

López-Lezama, Jesús María, Antonio Padilha-Feltrin, Javier Contreras, and José Ignacio Muñoz. 2010. Optimal contract pricing of distributed generation in distribution networks. *IEEE transactions on power systems* 26(1):128–136.

McAfee, R Preston. 1994. Endogenous availability, cartels, and merger in an equilibrium price dispersion. *Journal of Economic Theory* 62(1):24–47.

Mela, Carl F, Sunil Gupta, and Donald R Lehmann. 1997. The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing research* 34(2):248–261.

Mendicino, Luca, Daniele Menniti, Anna Pinnarelli, and Nicola Sorrentino. 2019. Corporate power purchase agreement: Formulation of the related levelized cost of energy and its application to a real life case study. *Applied Energy* 253:113577.

Morgan, J. P. 2018. Lifting the veil on hidden risk in renewable power purchase agreements.

NCSL. 2015. State policies for power purchase agreements. Data retrieved from the website, `https://www.ncsl.org/research/energy/state-policies-for-purchase-agreements.aspx`.

Nissen, Ulrich, and Nathanael Harfst. 2019. Shortcomings of the traditional "levelized cost of energy"[lcoe] for the determination of grid parity. *Energy* 171:1009–1016.

Pancras, Joseph. 2010. A framework to determine the value of consumer consideration set information for firm pricing strategies. *Computational Economics* 35(3):269–300.

Pennerstorfer, Dieter, Philipp Schmidt-Dengler, Nicolas Schutz, Christoph Weiss, and Biliana Yontcheva. 2020. Information and price dispersion: Theory and evidence. *International Economic Review* 61(2): 871–899.

PEXAPARK. 2021. What is a ppa? your definitive guide to power purchase agreement. Data retrieved from the website, `https://pexapark.com/solar-power-purchase-agreement-ppa/`.

Rhodes, Andrew. 2011. Can prominence matter even in an almost frictionless market? *The Economic Journal* 121(556):F297–F308.

Rosen, Christiane, and Reinhard Madlener. 2016. Regulatory options for local reserve energy markets: Implications for prosumers, utilities, and other stakeholders. *The Energy Journal* 37(Bollino-Madlener Special Issue).

Rosenthal, Robert W. 1980. A model in which an increase in the number of sellers leads to a higher price. *Econometrica: Journal of the Econometric Society* 1575–1579.

De los Santos, Babur, Ali Hortaçsu, and Matthijs R Wildenbeest. 2012. Testing models of consumer search using data on web browsing and purchasing behavior. *American economic review* 102(6):2955–80.

De los Santos, Babur, and Sergei Koulayev. 2017. Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science* 36(4):542–564.

Stahl II, Dale O. 1994. Oligopolistic pricing and advertising. *Journal of economic theory* 64(1):162–177.

Tellis, Gerard J. 1988. Advertising exposure, loyalty, and brand purchase: A two-stage model of choice. *Journal of marketing research* 25(2):134–144.

Thumann, Albert, and Eric Woodroof. 2021. *Energy project financing: resources and strategies for success*. CRC Press.

Tranberg, Bo, Rasmus Thrane Hansen, and Leopoldo Catania. 2020. Managing volumetric risk of long-term power purchase agreements. *Energy Economics* 85:104567.

Ursu, Raluca M. 2018. The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science* 37(4):530–552.

Varian, Hal R. 1980. A model of sales. *The American economic review* 70(4): 651–659.

Wilcox, Rand R, and David M Erceg-Hurn. 2012. Comparing two dependent groups via quantiles. *Journal of Applied Statistics* 39(12):2655–2664.

Wilcox, Rand R, David M Erceg-Hurn, Florence Clark, and Michael Carlson. 2014. Comparing two independent groups via the lower and upper quantiles. *Journal of Statistical Computation and Simulation* 84(7):1543–1551.

Wildenbeest, Matthijs R. 2011. An empirical model of search with vertically differentiated products. *The RAND Journal of Economics* 42(4): 729–757.

Xia, Jing, and Wenju Niu. 2020. Pushing carbon footprint reduction along environment with carbon-reducing information asymmetry. *Journal of Cleaner Production* 249:119376.