Essays in FinTech and Macro-Finance

by

Chenyu Wang

Business Administration
Duke University

Defense Date: March 28, 2024

Approved:

S.Viswanathan, Supervisor

Adriano Rampini

Attila Ambrus

Simon Gervais

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Business Administration
in the Graduate School of Duke University
2024

ABSTRACT

Essays in FinTech and Macro-Finance

by

Chenyu Wang

Business Administration
Duke University

Defense Date: March 28, 2024

Approved:

S.Viswanathan, Supervisor

Adriano Rampini

Attila Ambrus

Simon Gervais

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Business Administration
in the Graduate School of Duke University
2024

Abstract

Data collection and analytics are the core of firms' development in digital economies and have an enormous impact on consumer welfare. We build a monopolistic competition model with heterogeneous firms to incorporate both data collection and analytics investment. The model studies how the complementary effect between data collection and analytics affects firms' pricing, profit and consumer welfare. Data is divided into two categories: raw data and effective data. Raw data is a byproduct of production and does not benefit firms on its own. Effective data is a signal on consumers' taste and must be produced with both analytics and raw data. We then find analytics can not only reduce firms' uncertainty but also lower user cost of capital and markup. Lower cost of data analytics can increase consumers' welfare by increasing competition. We allow firms to differ in the size of complementary effect. The model shows that cheaper analytics has asymmetric effects on heterogeneous firms' product quality and profit. Firms with strong complementary effects produce higher quality goods, charge lower price-per-utile and benefit from the cheaper analytics. The opposite is true for firms with weak complementary effects.

In the second paper, We build a model to incorporate the buy-now-pay-later (BNPL) platform and study its welfare implication. BNPL platforms lend money to consumers, provide private data to partner firms and charge fee from in-platform merchants. Data can lower production cost. Two types of data are available: public data and private data. Data size of both types increases in the number of firms. Private data is only available for in-platform merchants. We find BNPL platforms can hurt non-platform users. The reason is that the platform fee can decrease the number of firms in the market and reduce public data, which increases out-of-platform firms' product prices. We then study a duopoly model with two platforms competing with each other. The model predicts that competition between platforms benefits non-platform users but can hurt platform users. The intuition is that competition splits the in-platform merchants and reduces private data for both platforms.

Contents

List of Figures

Acknowledgements

I thank my dissertation chair S. "Vish" Viswanathan as well as my dissertation committee members Attila Ambrus, Simon Gervais, and Adriano Rampini for their invaluable advice and support. I also thank Paymon Khorrami, Melanie Wallskog for their suggestions and comments. All remaining errors are mine.

1. Introduction

Over the last two decades, firms have had access to increasingly rich datasets about their consumers. Together with the outburst of data is the rapid development of analytics. Many algorithms and techniques have been developed to help extract useful information from datasets, including, for example, machine learning and artificial intelligence. This suggests that analytics plays a significant role in shaping firms' decisions and in turn the entire economy. The current literature discusses how data and technology can affect our economy (e.g., Farboodi and Veldkamp (2021) and Jorgenson (2001)). In the first paper, we explicitly consider how firms must process the data that they gather in the course of their operations in order to benefit from it. Our analysis highlights how the interaction of data and analytics can affect firms' production and consumer welfare.

In particular, the paper builds a general equilibrium model with differentiated firms to show how analytics can affect the aggregate price index, firms' product quality and welfare. We decompose the impact of analytics on firms' investment decisions into two effects: direct effect and dynamic effect. The dynamic effect can be affected by analytics investment due to the complementary effect between capital and analytics. In many cases, the dynamic effect plays a bigger role. We show that with cheaper analytics, it helps lower the price index and then reduce the markup as well. Consumer welfare increases as a result. Firms in high predictive accuracy industries have strong incentives to offer discount to consumers and get data in return. They are even willing to sell higher quality products at a lower price under certain circumstances. The impact of analytics price on firms' profit is two-fold. Firms in high predictive accuracy industries benefit from cheaper analytics while firms in low predictive accuracy industry suffer. These results show that not only data but also analytics plays an important role in shaping the overall economy.

A leading feature of the financial market in the market is the rise of various FinTech firms and the usage of alternative datasets. In the second paper, I study a new type of FinTech firms: the buy-now-pay-later (BNPL) platforms. As a new FinTech business model, BNPL platforms have grown rapidly over the past 5 years. Berg, Fuster, and Puri (2021) estimate FinTech lending in the BNPL segment to be $25 billion in the US, with growth over 100%. Comparing to it, FinTech business lending has an annual growth rate 43% according to Ziegler, Shneor, Wenzlaff, Suresh, de Camargo Paes, Mammadova, Wanga, Kekre, Mutinda, Wang, Closs, Zhang, Forbes, Soki, Alam, and Knaup (2021). How do BNPL platforms affect consumers' welfare?

This paper explores the impact of BNPL platforms on consumer welfare. We find the introduction of BNPL platforms does not always increase the welfare of all consumers. In particular, rich consumers who do not use the platforms can get hurt. The intuition is that if the lending capability of platforms is not strong, the platform fee can prevent new entrants. Fewer firms competing with each other has a negative impact on the market because less public data is available. Firms out of the platform then have higher production cost. Product prices out of the platform go up. As a consequence, consumers who are not using the platform services get hurt.

The second question that this paper tries to answer is whether platform competition can benefit everyone. The model predicts that under certain circumstances, platform competition can decrease the welfare of those consumers who use the BNPL services. In particular, if the entry cost is relatively high and it is hard to get public data, platform competition will make private data provision less efficient. The intuition is that the platforms are now partnering with fewer firms and it is harder to provide better private data support. This leads to higher production cost for in-platform merchants and higher in-platform prices. As a consequence, consumers who are using the BNPL services get hurt.

2. Investing in Data Analytics

2.1 Introduction

Over the last two decades, firms have had access to increasingly rich datasets about their consumers. Together with the outburst of data is the rapid development of analytics. Many algorithms and techniques have been developed to help extract useful information from datasets, including, for example, machine learning and artificial intelligence. This suggests that analytics plays a significant role in shaping firms' decisions and in turn the entire economy. The current literature discusses how data and technology can affect our economy (e.g., Farboodi and Veldkamp (2021) and Jorgenson (2001)). In this paper, we explicitly consider how firms must process the data that they gather in the course of their operations in order to benefit from it. Our analysis highlights how the interaction of data and analytics can affect firms' production and consumer welfare.

Why is such interaction important? If we only consider data as a free byproduct that can help decrease marginal cost or lower demand uncertainty, then one cannot justify the fact that firms are investing large amount of money in analytics tools. Without these tools, it is hard for firms to learn from big datasets. It is also inappropriate to treat data as the product of an investment, like some patent or innovation emanating from R&D, as this ignores the fact that firms amass consumer for free in the course of their normal operations (for example, as they sell goods to their customers). In other words, firms could produce for the purpose of getting data. This motive can be even stronger if a firm possesses advanced algorithms that can efficiently convert data into useful information.

In the model, we show that data can have two effects on firms' investment decisions. One is the direct effect. The idea is that firms are uncertain over consumer' preference. Data then provides information that improves estimation. This effect has been considered in the previous literature (e.g., Eeckhout and Veldkamp (2021)).[1] The other effect, the dynamic effect, is the focus of this paper. Capital is now used not only for production

[1] Some papers treat data as direct boost to TFP, e.g., Jones and Tonetti (2020). It is also common in the literature on innovation to model innovation as a reduction of firms' marginal cost, e.g., Aghion and Howitt (1992).

today, but also for gathering data that will improve tomorrow's estimation. Farboodi and Veldkamp (2021), and Jones and Tonetti (2020) have studied this dynamic effect, but firms in their model are not able to control the magnitude of the dynamic effect. Different from theirs, we allow firms to invest in analytics in addition to capital. Powerful analytics will increase the estimation accuracy, so capital and analytics are complementary to each other. The magnitude of the dynamic effect is then determined by firms' investment in analytics. An equivalent way of understanding this is that advanced analytics lower the user cost of capital in the sense of Jorgenson (1963). Indeed, analytics can provide firms with incentives to produce more and collect more data.

While more data can increase a firm's market power as it learns more about the demand for the industry's products, we show that cheaper analytics actually increase market competition and decrease prices when there is free entry. So consumer welfare goes up. Interestingly, the two effects have opposite impact on each firm's pricing strategies. The direct effect always increases the price because it improves product quality. In contrast, the dynamic effect always decreases the price because it increases capital investment and total supply. When the firm is very productive, the second effect dominates the first one. In this way, some firms end up producing high-quality products while charging low prices.

Our paper also asks: can all firms benefit from a reduction in the cost of analytics tools? It is perhaps tempting to readily conclude that lower cost should lead to higher profit. However, this argument ignores the negative effect from the more intense competition prompted by cheaper analytics. In particular, prices go down, lowering firms' profits in the process. We distinguish two types of firms by assuming that some firms can convert each unit of data into more precise signals than other firms. In other words, one type of firms' dynamic effect is stronger. We show that firms with a strong dynamic effect benefit from lower cost of analytics but firms with a weak dynamic effect lose out. The intuition is that firms with a strong dynamic effect invest more heavily in algorithms, so their profit depends more on the analytics cost than on prices. In contrast, firms with a weak dynamic effect invest more heavily in traditional capital, and so their profit depends more on prices.

An important contributor to a firm's data generation process is its economic production. Data obtained in this way is almost free. In contrast, analytics can be expensive and often requires large investments. Data and analytics are complementary to each other. Without data, analytics is irrelevant. Without advanced analytics, we cannot extract useful information from datasets. We call the processed data effective data. Effective data is used by firms to learn consumers' preference and then to produce goods matched better with consumers' taste. In this way, our model microfounds the benefit of data to the economy.

Although effective data can benefit all the firms, demand for analytics varies a lot across industries. According to a 2020 survey conducted by Michigan State University and Insights Association on Top 50 Research & Data Analytics Industry in the U.S. (Bowers (2020)), there exists a large difference in the demand for data analytics across industries. Bowers (2020) provides a revenue decomposition of the top 50 research & data analytics firms. On average, 32% comes from entertainment, media, and advertising, 21.4% comes from healthcare products, services, and OTC medicines; however, less than 1% comes from travel/tourism or agriculture. As far as we know, the existing literature does not take this heterogeneity in data demand into account. This model addresses this.

We allow firms to differ in two dimensions: effective data productivity and consumers' preference for products. Industries consisting of firms with high (low) effective data productivity are called high (low) predictive accuracy industries. Industries consisting of firms with high (low) consumers' preference are called premium (non-premium) industries.[2] We show that firms in high predictive accuracy or premium industries produce relatively highly customized products (match consumers' taste better). However, these two industries take two different paths. High predictive accuracy industries rely heavily on analytics and are analytics intensive. Meanwhile, premium industries try to expand production and collect more free data. They are not analytics intensive.

In this paper, we build a monopolistic competition model in the spirit of Dixit and

[2] To understand this, consider the comparison between the auto industry and the sneaker industry. In general, consumers have higher preference for a car. So we refer to the auto industry as premium.

Stiglitz (1977), with heterogeneous firms, in which firms take both data and analytics into account. We distinguish two broad industries of firms. One is a competitive industry, in which firms have little incentive to acquire data or analytics. It produces numeraire goods. The other is a monopolistic competition industry; each firm in this industry produces a single differentiated good. These firms have strong incentives to acquire data and analytics because they face demand uncertainty. Demand uncertainty is modelled as follows: firms do not know consumers' taste over different features of the products. They need to pick features of the consumption goods they produce. If these features match consumers' tastes well, the valuation of that product will be high. Otherwise consumers will not value the product highly.

Similar to Farboodi and Veldkamp (2021), we model data as signals on unknown state variables (here, consumers' tastes). However, instead of directly treating data as signals as in Farboodi and Veldkamp (2021), we further divide data into two categories. One is the raw data which requires analysis (we will call it data for the rest of the paper) and the other is the effective data. Data is a byproduct of production and does not require any further effort or investment. This is similar to the treatment of data in Jones and Tonetti (2020). What's different is that data cannot bring direct benefits to the firms. Ultimately, only the data that firms process and analyze, namely effective data, matters. Each unit of effective data is a noisy signal about consumers' tastes. In order to produce effective data, firms need both data and investment in analytics. Figure 2.1 illustrates the production process of effective data. This setup captures the importance of both the size of datasets and the efficiency of analytical tools. The more advanced analytics a firm has, the more productive the production function for effective data is. In this framework, we are able to answer the following two questions that the existing literature does not address:

1. How do analytics affect the quality of goods produced by firms?
2. Does the cost of analytics affect different firms' profit asymmetrically?

We construct a price-per-utile (PPU) index that measures the cost of different goods on a per-utile basis. High match quality goods is defined by the feature chosen by the firm

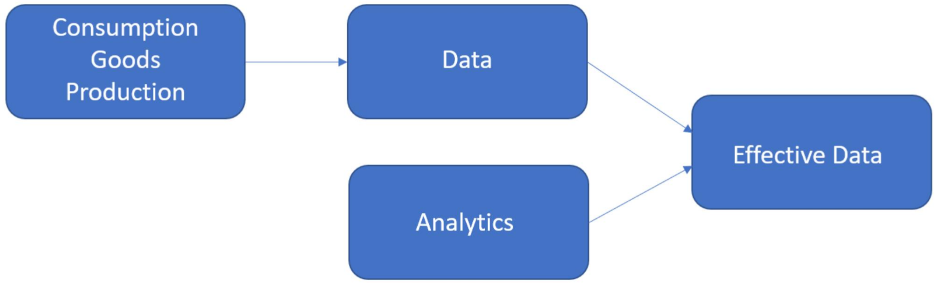

FIGURE 2.1: Production of Effective Data

that is matched with consumers' taste better. We find that surprisingly, the PPU of a high match quality good can be lower than a low match quality good even if its actual price is higher. This justifies why in practice consumers often filter products which better fit their taste instead of just picking the cheapest one. In fact, many popular products never mark down their price.

We also find the price of products in a premium industry is always higher than products in a non-premium industry. The reason is simply that consumers value premium products more. For the impact of data analytics on PPU and prices, we find that for firms in high predictive accuracy industries, cheaper data analytics always help them reduce PPU. The intuition is that firms in high predictive accuracy industries try to improve products' match quality as data analytics cost goes down.

For the second question, we find that a cost reduction of analytics benefits big firms at the expense of small firms. In the context of this model, big firms are those with stronger dynamic effects. The intuition is that as all the firms start to take advantage of cheaper analytics, the market price goes down. However, the magnitude of price cut is a weighted average of firms' dynamic effect. For big firms, their dynamic effect is stronger and can make better usage of data. The price cut is then smaller than that required to make big firms feel indifferent. In other words, they can lever up the benefits of cheaper analytics more efficiently than small firms. In contrast, for small firms, they are unable to recoup the loss from lower price index owing to their relatively inefficient usage of data. Therefore, small firms can struggle even more in a better-developed digital economy.

The literature on data, such as Eeckhout and Veldkamp (2021), often care about whether too much data can increase the market power of the firm and thereby hurt consumers. Our model implies that a lower cost of analytics can decrease the aggregate consumer's price index in the sense of Dixit and Stiglitz (1977) and lower markup. The markup in this model is defined by the ratio of the expected price to the marginal cost. This result illustrates one positive effect of cheap analytics to the society. With advanced analytics, firms will have more ways to take advantage of customers' data. Due to more intense competition, they have to offer a price discount. This mechanism drives the consumer's price index down. There are many such examples in practice. One is the safe driving auto insurance. Some insurance companies ask their customers to install the app in their smartphones to track their driving data. As an exchange, the insurance premium goes down. Similarly, online shopping platforms offer discounts to users to encourage them to try new services or products. For example, Amazon offers discounts to customers paying with credit card points. Similarly, credit card companies offer discount to cardholders if they try different payment methods such as ApplePay and GooglePay.

As for markup, we find that it only relates to the dynamic effect and cannot be affected by the direct effect. The intuition is that the direct effect can move markup in two opposite directions. On the one hand, it increases match quality and in turn product prices which push up the markup. This is similar to the first channel in Eeckhout and Veldkamp (2021). On the other hand, lower demand uncertainty increases production scale and then increases the marginal cost due to the concavity of the production function. This will lower the markup. This is similar to the risk premium channel in Eeckhout and Veldkamp (2021). Ultimately, these two effects cancel out perfectly. However, the dynamic effect lowers the user cost of capital, so firms produce more, prices go down, and marginal cost goes up. The stronger dynamic effect always lowers the markup.

The rest of the paper is organized as follows. Chapter 2.2 presents some results on firms' investment in analytics. Chapter 3.2 provides literature review. Chapter 3.3 lays out the basic setup of the model. Chapter 2.5 defines the stationary equilibrium and establishes its uniqueness. Chapter 2.6 shows various implications of the model. Chapter 2.7 shows the impact of data analytics in a general equilibrium model. Chapter 3.6 concludes.

2.2 Stylized Facts

In recent years, firms' demand in analytics and platforms providing such service has surged up. In particular, the revenue in big data and analytics (BDA) software market had doubled from 2011 to 2019. The revenue in 2011 is USD 32.14 billion. In 2019, the revenue is USD 67 billion.[3] Now, the BDA software market is migrating to the cloud. As of 2019, the cloud services portion takes up about a quarter of the total revenue of the BDA software market. Take Microsoft's Azure as an example. Its quarterly revenue in March have almost tripled in the past six years: from USD 6 billion in March 2015 to over USD 18 billion in March 2022. It not only provides hardware services e.g., cloud storage, but also offers lots of analytics tools. The largest three categories of products and services offered on the Microsoft Azure marketplace are: IT & management tools, analytics and compute. All these categories are important for data analysis.

If we look at firms' investment ratio of analytics to property plant and equipment (PPE), we can find a strong upward trend. Figure 2.2 plots the times series of investment ratio of software to PPE in the past decade. Data comes from Compustat. The ratio has gone up by roughly 38%. It shows that firms start to focus more on analytics since 2010.

Even though, all firms are investing proportionally more in software, firms' demand uncertainty exhibits different time series patterns. According to Gaur, Kesavan, Raman, and Fisher (2007), stock analysts' forecasting variance can serve as a proxy for demand uncertainty. We construct an index called scaled dispersion of forecasting. It is the standard deviation of forecasting of yearly earnings per share(EPS) scaled by market cap. The forecasting data is from The Institutional Brokers' Estimate System (IBES). We group all firms into two bins. One bin consists of firms with high software investment ratio. This bin represents the high predictive accuracy industries. The second bin contains firms with low

[3] Data is provided by Statistia.

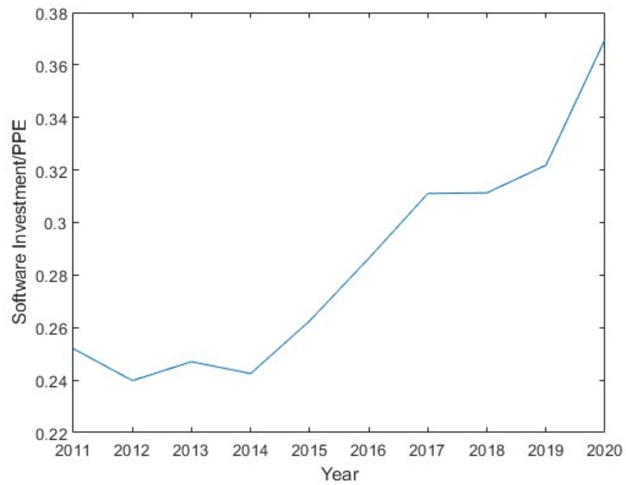

FIGURE 2.2: Ratio of Software Investment to PPE Investment

software investment ratio. This bin represents the low predictive accuracy industries.

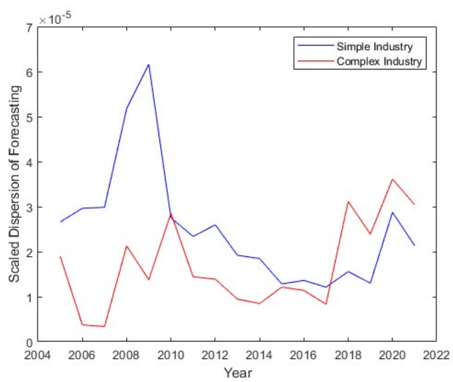

FIGURE 2.3: Forecasting Dispersion in Different Industry

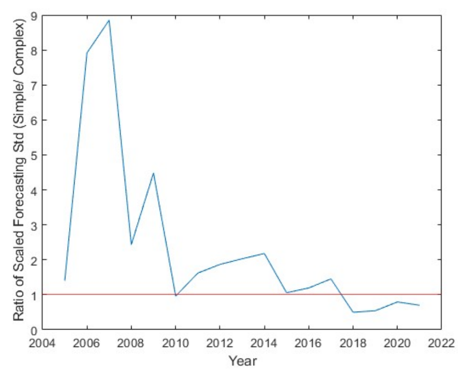

FIGURE 2.4: Forecasting Dispersion Ratio

Figure 2.3 plots the forecasting dispersion in both simple and low predictive accuracy industries. Figure 2.4 plots the dispersion ratio across these two industry. We can see from two figures that firms in high predictive accuracy industries tend to have lower forecasting dispersion or demand uncertainty. Surprisingly, firms in low predictive accuracy industries are facing higher demand uncertainty. It is consistent with our prediction that data analytics can have asymmetric impact on firms in different industry. It is those firm in high predictive accuracy industries that can take benefits from cheaper data analytics.

2.3 Related Literature

There is a growing literature on data in recent years. Only few of these papers touch on the analytics and differentiate raw and effective data. Begenau, Farboodi, and Veldkamp (2018) analyze the impact of hardware (data capacity) on firms' cost of capital and growth. Different from theirs, this paper is mainly interested in the software part of analytics, e.g., algorithms. Abis and Veldkamp (2021) estimate how structured data can change firms' production function. However, they only consider the analytics (data managers in their model) and ignore the process of data collection. Jovanovic and Rob (1987) consider an economy where firms can directly invest in information technology and show firm's size is positively autocorrelated across time. He considers information technology as a perfect substitute for data, but our paper treats data and analytics as complementary inputs. Moreover, in models of learning-by-doing, knowledge acquired by firms is not tradeable (e.g., Jovanovic and Nyarko (1996)).

Many papers have studied the impact of data alone. In this branch of literature, data is considered as the byproduct of the production. See Farboodi and Veldkamp (2022) for a survey on this topic. There are two popular ways of modelling data. One is modelling it as a direct boost to TFP, e.g., Jones and Tonetti (2020). Their paper answers the question of how different data ownership can affect welfare and equilibrium allocation based on non-rival property of data. The other is treating data as signals to help firms improve estimation precision, e.g., Farboodi and Veldkamp (2021). They derive a S-shaped curve of the forecast precision accumulation function and deal with the question whether data alone can sustain economy growth. Both Jones and Tonetti (2020) and Farboodi and Veldkamp (2021) focus on the feedback loop of data. The key difference between this paper and theirs is that they take the data feedback loop as given but we allow firms to choose the optimal data feedback loop. More specifically, firms can invest in analytics to improve the profitability of the data feedback loop. Analytics investment and capital investment are complementary to each other. The capability to invest in analytics pushes firms to invest more in capital.

This leads to different predictions in the product markup. Our model predicts that any difference in parameters that affects sales revenue but not effective data generation should have no impact on the markup. However, if fixing analytics, an increase in parameters that increase sales revenue but do not affect effective data generation will increase the markup. The reason is that firms want to expand production scale when profitability increases, but without investment in analytics, the user cost of capital also goes up. So comparing to our model, the production scale is smaller and the markup goes up. Farboodi, Mihet, Philippon, and Veldkamp (2019) show that data accumulation can increase the skewness of the firm size distribution in a partial equilibrium model.

Our paper answers the question that how heterogeneous firms optimally produce, set prices and acquire data to improve their competitive positions. We are able to construct the price-per-utile index to compare prices of products with different qualities. It also derives an interesting result that it is possible for firms in the high predictive industries to produce high quality products and charge low prices. Eeckhout and Veldkamp (2021) study the impact of data stock asymmetry on social welfare. In their paper, data stock is exogenously given. Our model endogenizes the data collecting process and shows that a universal cost reduction in analytics only benefits firms in high predictive accuracy industries.

There is a widespread concern over whether data can create market power. Among those papers on this question, markup is the most commonly studied measure to represent monopoly power. Empirically, Loecker, Eeckhout, and Unger (2020) show that the markup has increased from 21% to 61% since 1980s. Theoretically, Ichihashi (2020) shows how firms use consumer data to price discriminate. Eeckhout and Veldkamp (2021) build a model to illustrate data's impact on different measures of markup. They show given different parameter regions, data can either increase markup or decrease markup. Their paper is a static one and cannot measure the impact of the dynamic effect on the product markup. This paper directly shows that the dynamic effect has a significant impact on markup. Ridder (2021) points out intangibles reduce marginal costs and raise fixed costs, which in turn deterring other firms from entering. Complementing this literature, we predict cheaper analytics can help lower markup and analytics intensive firms have lower markup in comparison to capital intensive firms. Importantly, consumers' welfare decreases in the cost of analytics.

Our paper also relates to data market. See Bergemann, Bonatti, and Gan (2022) for a survey of the literature on the data markets. Bergemann, Brooks, and Morris (2015) show how a monopolistic firm can use consumer information to implement third degree price discrimination. Bergemann, Bonatti, and Smolin (2018) study how a data seller should design and sell statistical experiments. Bergemann and Bonatti (2019) study the excessive data sharing due to the fact that consumers do not internalize that their data can reveal information about other consumers. Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2022) also consider excessive data sharing and low data prices because of platforms.

Our model is also related to the literature on concentration of the economy. Kwon, Ma, and Zimmermann (2021) say that corporate concentration has increased persistently over the past century. They further show that R&D and IT investment contribute to such concentration. Crouzet and Eberly (2021) find that rising investment in intangibles is associated with industry leaders and they have higher markup. Begenau, Farboodi, and Veldkamp (2018) derive the asymmetric impact of data capacity on firms in a different framework relying on adjustment cost and risk averse investors. This paper has predictions consistent with the empirical findings. We show that in the digital economies, only firms in high predictive accuracy industry (they are relatively bigger) are able to benefit from cheaper analytics and they will grow even bigger. Meanwhile, firms in low predictive accuracy industry get hurt if analytics cost goes down.

2.4 Model

The model is in discrete time with infinite horizon, $t = 0, 1, ...$ The economy is populated with consumers and firms. There exists one representative consumer; however firms have two types. One type of firms produces consumption goods indexed by $i = 0$ in a competitive market. We normalize the measure of such firms to 1 and the price of the consumption goods they produce to 1. This is the numeraire. The other type of firms produces heterogeneous consumption goods. In particular, each firm i is a monopoly in industry i, and produces consumption good i. At time t, denote total measure of firms by n_t. So at time t, there is a total of $n_t - 1$ ($i \in [1, n_t]$) types of consumption goods produced by monopoly firms and 1 numeraire in the economy.[4]

2.4.1 Consumer

The consumer in the model lives for only one period and one new consumer is born at the beginning of each period. Each new consumer has endowment $W > 0$. He can allocate the endowment in two ways: a bundle of consumption goods $i \in [1, n_t]$ and numeraire $i = 0$. Utility from consuming these goods has an additive form,

$$U_t = b_0 c_{0t} + \left\{ \int_1^{n_t} [m_i - (\tilde{\theta}_{it} + \epsilon_{it} - \theta_{it})^2] c_{it}^{\rho} di \right\}^{\frac{\alpha}{\rho}} \tag{2.1}$$

where $\rho, \alpha \in (0, 1)$, and $\rho > \alpha$.

We can see from above that the utility from the numeraire is linear and the utility from the bundle is a variant of Dixit-Stiglitz utility. α measures the substitutability between the numeraire and the bundle. ρ measures the substitutability among different goods within the bundle. If α is close to 1, the consumption bundle and the numeraire are almost perfect substitutes. If ρ is close to 1, consumption goods within the bundle will be perfect substitutes. We are mainly interested in analyzing the behavior of these monopoly firms that produce similar but not the same goods. It is valid and without loss of generality to assume that ρ is bigger than α. Throughout the rest of the paper, we will make the following assumption that $\rho > \alpha$.

ASSUMPTION 1. *The substitutability ρ is bigger than α.*

[4] We differentiate two broad types of industries based on the observation that different industries have dramatically different demands for data and data analytics. In a survey conducted by Insight Association and Michigan State University in 2020, 32% revenue of top 50 data research companies (e.g., Nielsen) comes from entertainment and media industry, but less than 1% revenue comes from agricultural industry. These numbers show that it is important to take innate differences in data demand into account.

$\tilde{\theta}_{it} + \epsilon_{it}$ is the consumer's favorite feature of goods i at time t, and θ_{it} is the feature chosen by firm i at time t. One can interpret the feature as the color of shirts or the flavor of drinks. If a firm chooses the correct feature $\tilde{\theta}_{it} + \epsilon_{it}$, the consumer's utility from one unit of good i will be maximized at m_i. Otherwise, there will be a penalty measured by the quadratic term in the integrand. We assume that m_i is drawn independently from a continuous distribution F_m with support $[\underline{m}, \overline{m}]$ for each firm i. m_i is the consumer's preference for product i. Consumer's favorite feature $\tilde{\theta}_{it} + \epsilon_{it}$ consists of two parts. One is a persistent term $\tilde{\theta}_{it}$ and the other is a transient term $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$, which is drawn i.i.d. at the beginning of each period. $\tilde{\theta}_{it}$ is independent of $\tilde{\theta}_{jt}$. Neither of these two parts are observable by firms. The persistent term follows from a mean-reverting process[5]

$$\tilde{\theta}_{i,(t+1)} = \theta_i + \tau(\tilde{\theta}_{it} - \theta_i) + \eta_{i,(t+1)} \tag{2.2}$$

where $\eta_{it} \sim N(0, \sigma_\eta^2)$.

For simplicity, denote $\left[m_i - (\tilde{\theta}_{it} + \epsilon_{it} - \theta_{it})^2\right]^{\frac{1}{\rho}}$ and $\left[m_i - (\tilde{\theta}_{it} + \epsilon_{it} - \theta_{it})^2\right] c_{it}^{\rho}$ by b_{it} and $v_{it}(c_{it})$. We will omit the time subscript whenever no confusion arises. b_{it} represents the utility from consuming one unit of good i. $v_{it}(c_{it})$ represents the utility from consumption in goods i. The consumer's problem is to choose the optimal amount of goods i, c_i. In particular, his objective function is as follows,

$$\max_{c_i} \qquad b_0 c_0 + \left(\int_1^n (b_i c_i)^\rho di \right)^{\frac{\alpha}{\rho}}$$

$$s.t. \quad c_0 p_0 + \int_1^n c_i p_i di = W$$

where we drop the time subscript and $n-1$ is the measure of monopolistic firms and p_i is the price of goods i.

The problem can be solved in two steps. One, suppose the consumer puts W_1 in the consumption bundle from $i = 1$ to $i = n$. We solve the optimal consumption policy for

[5] Although, the representative consumer is short-lived, we assume the new-born consumer inherits the old consumer's taste to some extent.

goods within the bundle. Two, given the optimal consumption policy above, we solve for optimal W_1. We will assume that b_0 is big enough so that $W_1 < W$ in equilibrium. The consumer's problem at the first step is

$$\max_{c_i} \quad \int_1^n (b_i c_i)^\rho \, di$$
$$s.t. \quad \int_1^n c_i p_i di = W_1$$

The first step is similar to the standard optimal consumption problem with Dixit-Stiglitz utility. The following two lemmas characterize the optimal consumption policy for c_i, $i = 1, 2, ...$

Lemma 1. *Consumer's demand function for goods i is $c_i = (\frac{p_i}{P})^{-\sigma}\frac{W_1}{P}b_i^{\sigma-1}$, where $\sigma = \frac{1}{1-\rho}$ and $P = (\int_1^n (\frac{p_i}{b_i})^{1-\sigma} di)^{\frac{1}{1-\sigma}}$, and W_1 is the wealth allocated to the consumption bundle.*

Proof: see appendix.

P is called aggregate consumer price index for the bundle.[6] For simplicity, we will refer to it as the price index throughout the rest of the paper. This is the same as the one defined in Dixit and Stiglitz (1977). The next lemma shows the consumer's maximal utility from adopting the optimal policy.

Lemma 2. *Maximal utility from spending W_1 in goods $i \in [1, n_t]$ equals to $(\frac{W_1}{P})^\alpha$.*

Proof: see appendix.

Obviously, the consumer's welfare decreases in the price index and the price index is the only variable that affects the consumer's welfare. In later analysis, we will focus on how a sudden shock to the economy can affect the price index as well as the consumer's welfare.

[6] Even though b_i is a random variable, P is deterministic. Because $\tilde{\theta}_{it}$ are *i.i.d.*, we can apply the law of large numbers of a continuum of random variables. See Uhlig (1996) for the reference of law of large numbers for a continuum of random variables.

At the second step, the consumer's problem is

$$\max_{W_1, W_0} \quad \left(\frac{W_1}{P}\right)^{\alpha} + b_0 \frac{W_0}{p_0}$$

$$s.t. \quad W_1 + W_0 = W.$$

This problem is a simple wealth allocation problem with two possible choices. The results are shown in the following lemma and corollary.

Lemma 3. $W_1 = \alpha^{\frac{1}{1-\alpha}} \left(\frac{p_0}{b_0}\right)^{\frac{1}{1-\alpha}} P^{-\frac{\alpha}{1-\alpha}}$.

Proof: see appendix.

Lemma 3 shows two important factors which can push consumers to substitute the numeraire by the consumption bundle. One is the price index. The other is b_0, which represents the consumer's preference for the numeraire. High price index and larger b_0 can discourage the consumer from purchasing the consumption bundle. Combining the above three lemmas, we can derive the inverse demand for the goods in the bundle. This will be the key for later welfare analysis.

Corollary 4. *Consumer's inverse demand for good i is:* $p_i = \left(\alpha \frac{p_0}{b_0}\right)^{\frac{1}{(1-\alpha)\sigma}} P^{-\frac{\alpha}{(1-\alpha)\sigma} + \frac{\sigma-1}{\sigma}} c_i^{-\frac{1}{\sigma}} b_i^{\rho}$.

Notice that the exponent of P in the inverse demand function can be rewritten as $\frac{1}{\sigma}(\frac{\rho}{1-\rho} - \frac{\alpha}{1-\alpha})$. For ease of notation, we denote $\frac{\rho}{1-\rho} - \frac{\alpha}{1-\alpha}$ by ϕ and denote $\left(\alpha \frac{p_0}{b_0}\right)^{\frac{1}{(1-\alpha)\sigma}}$ by φ. Because ρ is bigger than α, market price p_i increases in P. This result comes from two different substitution effects. One is the substitution effect within the bundle and the other is the substitution effect between the bundle and the numeraire $i = 0$. The first effect dominates the second if $\rho > \alpha$. So market price increases in price index P. The price index captures the impact of market competition.[7] There are two other factors affecting the inverse demand. One is firm's supply c_i. The other is consumers' utility from good i, b_i. Bigger b_i implies higher quality. Corollary 4 tells us that quality can come with some cost: higher price.

[7] This will be clearer when we discuss the impact of analytics on the price index later.

2.4.2 Firms

In this subsection, we study firms' problem and formally model data and analytics. Each firm i is a monopolist in industry i. Every monopolistic firm survives to the next period with probability $1 - \delta$. δ is called the death rate. At the beginning of each period, a set of new firms are born and they must decide whether or not to enter the industry at a fixed cost c. Any incumbent firms can remain in the industry for free. If a newborn firm does not enter, it will leave and never return. Firm i can produce consumption goods i, invest in analytics, and trade data. It also needs to choose the feature of goods i, θ_{it}. It wants to pick θ_{it} as close as consumer's favorite one $\tilde{\theta}_{it} + \epsilon_{it}$. The production function for consumption goods is Ak_{it}^d, $0 < d < 1$. Capital is rented from an outside sector at a fixed price $r_k > 0$.

Now we discuss our approach to modeling data. We model two types of data. One is raw data and the other is effective data. Raw data can be understood as raw sales data. Itself cannot bring any benefits but can be used as input to produce effective data. Raw data can be thought to be firms' spreadsheets that contain all transaction information from consumers. Without analysis, it is hard for firms to learn much about consumers' tastes. However, a histogram on the spreadsheet showing the number of transactions for each type of products will provide very useful information to firms about their consumers. Effective data can be treated as a histogram or a regression table such that everyone can learn something easily. When producing consumption goods, the capital will automatically produce the same amount of raw data at the same time as byproduct. For simplicity, we will omit the adjective raw for the rest of the paper. In particular, the production function for data is Ak^d, where we drop the time subscript. We assume firm A's data is a dataset that can contain two types of information. One is tailored information on firm A's own interested variables. The other is generic information on variables that other firms want to learn as well.

J units of the effective data is a noisy signal on $\tilde{\theta}_{i,(t+1)}$, $s_{it} = \tilde{\theta}_{i,(t+1)} + \xi_{it}$, where $\xi_{it} \sim$

$N(0, \frac{\sigma^2_{\xi i}}{j})$. It is produced by both data and analytics. The effective data produced at time t can only be used by firms at time $t+1$. The production function for effective data is $g(D,z) = Dz^{1-d} = Ak^d z^{1-d}$, where D is the size of data and z represents analytics. For simplicity, denote $\sigma^{-2}_{\xi i}$ by $\tau_{\xi i}$. For the rest of the paper, we will refer to $\tau_{\xi i}$ as the signal precision whenever no confusion exists. For each monopolistic firm i, $\tau_{\xi i}$ is drawn independently from a continuous distribution F_{ξ}. Denote $\sigma_{f\xi}$ by the standard deviation of F_{ξ}. Analytics is rented from an outside sector at a fixed price r_z. To understand the role of analytics, we can consider it as the superiority of platforms, such as powerful cloud computing or as advanced models for inference. In both cases, firms need to pay some cost to improve its analytics.

Firms cannot observe $\tilde{\theta}_i$, ϵ_i and η_i but know the law of motion of $\tilde{\theta}_i$ and distributions of ϵ_i and η_i. They can also observe market prices p_i, choices of product feature θ_i and numeraire's sale c_0. Denote firm i's information set at period t by I_{it}. Then $I_{it} = \{p_{il}, \theta_{il}, c_{0l}, s_{i(l-1)} | l \leqslant t\}$. Firms want to learn $\tilde{\theta}_{it}$ by acquiring more effective data. Denote $Var[\tilde{\theta}_{it}|I_{it}]^{-1}$ by Ω_{it}, which measures the forecast precision for firm i at the beginning of period t. For each new firm, its initial precision is Ω_0. New firms need to make entry decisions before observing consumers' preference for product i, m_i and signal precision $\tau_{\xi i}$.

Firms can trade raw data with each other at market price r_ω. Notice that even if firms' interested variables $\tilde{\theta}_{it}$ are independent from each other, one firm can still get information from analyzing other firms' datasets. In particular, if firm A purchases one unit of data from firm B, then this dataset provides precision 1 if its analytics level $z \geqslant 1$. Otherwise the firm cannot learn anything from the dataset. Here $z \geqslant 1$ imposes the minimum requirement of being able to analyze others' datasets. The cutoff $z = 1$ is not important. One can choose any cutoff such that every firm's investment in analytics in the equilibrium is above the cutoff.[8] This reflects the friction that other industries' data can provide only limited information. Or, another way to justify this is that to process other industries' data, one

[8] The reason for choosing 1 is that for any firm with $z = 1$, analyzing 1 unit of its own data, can gain precision 1. One can choose this cutoff freely as long as it is more productive to analyze firms' own datasets.

needs to implement a different algorithm. In order to model nonrival property of data (see Romer (1990a, 1990b)), we adopt the approach in Farboodi and Veldkamp (2021). If a firm sells ω units of data, it only loses $\kappa\omega$ in precision, $\kappa \in (0, 1]$.[9] $\omega > 0$ represents selling data and $\omega < 0$ represents buying data. In general, we can denote $\zeta = I[\omega < 0, z \geqslant 1] - \kappa I[\omega > 0]$. Then if a firm trades (buys or sells) ω units of data, its precision will change by $\zeta\omega$.

We next discuss the dynamics of firms' forecast precision Ω_{it}. Given realized market price p_{it}, firm's forecast precision over $\tilde{\theta}_{it} + \epsilon_{it}$ will increase by σ_ϵ^{-2}. The conditional variance will be $\frac{1}{\Omega + \sigma_\epsilon^{-2}}$. However, to predict $\tilde{\theta}_{i,(t+1)}$, the above conditional variance needs to be scaled by τ due to depreciation as in Equation 2.2. $\tilde{\theta}_{i,(t+1)}$ contains a new temporary shock η_{t+1} and it will increase the variance by σ_η^2. In sum, if the firm only uses yesterday's information to predict $\tilde{\theta}_{i,(t+1)}$, the conditional variance will be $\tau^2(\Omega + \sigma_\epsilon^{-2})^{-1} + \sigma_\eta^2$. Its inverse is the forecast precision provided by yesterday's information. Through production, firms can also get $Ak^d z^{1-d}$ units of effective signals, which provide an overall precision of $Ak^d z^{1-d} \sigma_{\xi i}^{-2}$. Moreover, if the firms trade ω units of data, then the precision will change by $\zeta\omega$. From standard Bayesian updating, we know given time t's forecast precision Ω_{it}, time $t + 1$'s forecast precision will be

$$\Omega_{i,(t+1)} = [\tau^2(\Omega_{it} + \sigma_\epsilon^{-2})^{-1} + \sigma_\eta^2]^{-1} + Ak_{it}^d z_{it}^{1-d} \sigma_{\xi i}^{-2} - \zeta\omega_{it}. \tag{2.3}$$

Firm i's problem is to choose optimal capital k_{it}, data analytics z_{it} to rent, data ω_{it} to trade, and feature of its goods θ_{it} in order to maximize expected profit. Its objective function is as follows:

$$\max_{\theta_{it}, k_{it}, z_{it}, \omega_{it}} \sum_{t=0}^{\infty} (1 - \delta)^t E\left[p_{it} A k_{it}^d - r_k k_{it} - r_z z_{it} + r_\omega \omega_{it} | I_{it}\right].$$

The revenue of a firm consists of four parts: sales revenue, cost of capital and data analytics, and data trading revenue. Even though capital and analytics are rented in each

[9] Because in the model, data sharing has no cost, κ cannot be zero. Otherwise, firms will share all of their data to others.

period, they can affect firms' forecast precision over all the future dates. This will affect a firm's expected continuation value. Firm's choice θ_{it} will affect its inverse demand function p_{it}. One useful observation is that optimal θ^*_{it} will be the conditional expectation of $\tilde{\theta}_{it}$, $\theta^*_{it} = E[\tilde{\theta}_{it}|I_{it}]$. The following lemma will be crucial to solve the firm's problem. It shows the conditional expected inverse demand function faced by each firm.

Lemma 5. *Firm i's conditional expected inverse demand function is:*

$$E\left[p_{it}|I_{it}\right] = \varphi^{\frac{1}{(1-\alpha)\sigma}} P^{-\frac{\alpha}{(1-\alpha)\sigma}+\frac{\sigma-1}{\sigma}} c_i^{-\frac{1}{\sigma}} [m_i - \Omega_{it}^{-1} - \sigma_\epsilon^2] \tag{2.4}$$

, where $\Omega_{it}^{-1} = Var[\tilde{\theta}_{it}|I_{it}]$

Proof: see appendix.

With the help of Lemma 5, we can transform firms' problem into a deterministic dynamic problem; the only state variable is firm i's precision Ω_{it}. We can rewrite firm i's problem as follows:

$$\max_{k,z,\omega} \quad E[p|I]Ak^d - r_k k - r_z z + r_\omega \omega + (1-\delta)V(\Omega'). \tag{2.5}$$

$$s.t. \quad \Omega' = [\tau^2(\Omega + \sigma_\epsilon^{-2})^{-1} + \sigma_\eta^2]^{-1} + Ak^d z^{1-d} \sigma_\xi^{-2} - \zeta\omega. \tag{2.6}$$

Equation 2.6 is the law of motion for forecast precision Ω_t. For simplicity, denote $[\tau^2(\Omega + \sigma_\epsilon^{-2})^{-1} + \sigma_\eta^2]^{-1}$ by $f(\Omega)$. $f(\cdot)$ represents the precision provided by yesterday's information.

The next theorem shows the value function is well defined and characterizes some properties of the value function.

Theorem 6. *The value function exists, increasing. It is also concave if $\underline{m}$ is sufficiently large.*

Proof: see appendix

The first-order conditions for this problem are as follows,

$$k: \quad \underbrace{\varphi P^{\frac{\phi}{\sigma}}(m-\Omega^{-1}-\sigma_\epsilon^2)A^\rho d\rho k^{d\rho-1}}_{\textbf{Direct effect}} + \underbrace{\frac{r_z d}{1-d}\frac{z}{k}}_{\substack{\Gamma \\ \textbf{Dynamic effect}}} = r_k, \tag{2.7}$$

$$z: \quad \frac{r_\omega}{\zeta}(1-d)A\left(\frac{k}{z}\right)^d \tau_\xi = r_z, \tag{2.8}$$

where $\Gamma \equiv \frac{r_z d}{1-d}\frac{z}{k} = r_z^{1-\frac{1}{d}}\left(\frac{A r_\omega \tau_\xi}{\zeta}\right)^{\frac{1}{d}}$.

Equation 2.7 shows the first-order condition of capital investment. It consists of two parts. The first term measures the marginal contribution of capital to current sales revenue. The magnitude of this term is affected by firms' forecast precision Ω. It measures the direct effect of data on firms' revenue. The second term measures capital's convenience yield to firms' continuation value. It has nothing to do with demand uncertainty which will be clear later. Denote this second term by Γ. This is the most important variable in this model. Γ measures the dynamic effect of data and analytics on capital investment. Larger investment in analytics can increase the dynamic effect Γ. The dynamic effect reflects the complementary effect between capital and analytics. $r_k - \Gamma$ is the user cost of capital in the sense of Jorgenson (1963). If Γ is big, the dynamic effect is very strong. As a consequence, the user cost of capital goes down. An immediate observation of Equation 2.7 is that both high precision Ω and large Γ will increase firms' capital investment k. Moreover, from the formula of Γ, we can see that Γ decreases in analytics cost r_z but increases in signal precision τ_ξ. It implies the necessity of cheap analytics and high signal precision in order to make firms expand production. In Farboodi and Veldkamp (2021), firms cannot optimally choose the magnitude of the second term Γ, because it only relates to capital investment in their model. One can consider Farboodi and Veldkamp (2021)'s data generating process as a special case of our model where analytics investment is fixed and free. However, we allow firms to directly change the size of Γ by letting them optimally choose analytics investment z. Intuitively, firms in our model are more willing to expand production because higher investment in analytics can lower the user cost of capital. However, in their model, more capital investment always lead to higher user cost of capital, which then discourages firms from expanding further. This will lead to different implications on markup, which will be discussed later in Chapter 2.6.

To solve firm's problem, it is easy to first determine the optimal data trading strategy. The result is shown in the following theorem.

Theorem 7. *Firms' optimal next period forecast precision increases in signal precision τ_ξ and consumers' preference m.*

Proof: see appendix

Theorem 7 is important and tells us what types of firms will pursue high forecast precision in the equilibrium. Because value function is increasing in forecast precision Ω, forecast precision can then represent the firm size. Theorem 7 then implies which firms will be big firms and which firms will become relatively small ones. Based on Theorem 7, we can define two criteria to separate different types of industries.

DEFINITION 1. *An industry i is called premium (non-premium) if consumers' preference for product i, m_i is big (small).*

DEFINITION 2. *An industry i is called high (low) predictive accuracy if per signal precision $\tau_{\xi i}$ is big (small).*

The first definition focuses on consumers' preference on different products. If m_i is big, obviously consumers value products in such industries more. One example of a premium industry can be luxury goods; people are willing to pay a lot for those high-end products. Another example of products in premium industry is goods sold by those famous or big brands. The second definition focuses on the difficulty in obtaining precise signals across different industries. Recall that $\tau_{\xi i}$ measures the precision provided by one unit of effective data. If it is big, then firms in these industries will get precise signals relatively easily. An example of a high predictive accuracy industry can be e-commerce and streaming market.

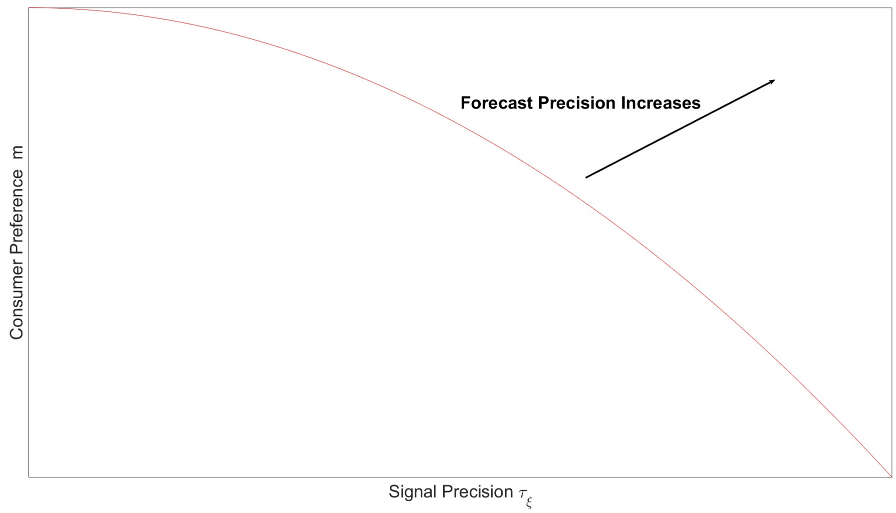

FIGURE 2.5: Who Have Higher Forecast Precision?

These firms (e.g., Amazon and Netflix) can use tools such as cookies to easily track consumers' browsing history and their shopping habits. Given these two definitions, Theorem 7 tells us that firms want to have high forecast precision for two motives: 1). consumers want to pay a high price for perfectly tailored products; 2). it is easy to obtain useful information on consumers' taste. The first motive relies on higher profit brought by having higher precision. The second motive relies on low cost from getting high precision.

Figure 2.5 illustrates firms' optimal choice Ω' given different signal precision and consumer preference m. Firms with attributes (m, τ_ξ) lying on the red curve all choose the same forecast precision. Firms at the bottom left region choose to attain low forecast precision. As indicated in Theorem 7, as consumer preference m or signal precision τ_ξ increases, firms start to accumulate higher forecast precision. An important implication of Theorem 7 is that the initial forecast precision with which a firm is endowed should not affect its optimal forecast precision choice in the future. All that matters is the pair of firms' attributes.

Next, we discuss the entry problem that every new firm faces. A firm will enter the industry if and only if its expected profit exceeds the entry cost c. The expected profit for each new firm is

$$\iint V(\Omega_0, m, \tau_\xi) dF_m dF_\xi. \tag{2.9}$$

Given initial forecast precision Ω_0, consumer's preference m_i and signal precision $\tau_{\xi i}$, a new firm's profit is $V(\Omega_{i0}, m_i, \tau_\xi)$ upon entering. Because new firms cannot observe m_i and $\tau_{\xi i}$ before entering, taking expectation gives their expected profit. Denote a firm's entry decision by E. $E = 1$ represents entry and $E = 0$ otherwise. A new firm's objective function is

$$\max_E \left(\iint V(\Omega_0, m_i, \tau_{\xi i}) dF_m dF_\xi - c \right) I[E = 1]. \tag{2.10}$$

2.5 Equilibrium

In this section, we will derive the stationary equilibrium of the model. Also, properties of the equilibrium will be characterized. To begin with, we introduce some notation for simplicity. We will omit the time subscript whenever no confusion can arise. Denote the distribution of signal precision τ_ξ in the market by F_ξ^m. It equals to nF_ξ, where n is the number of firms. Denote the distribution of consumer's preference m in the market by F_m^m. It equals to nF_m. Notice that F_m and F_ξ remain constant over time but F_m^m and F_ξ^m do not because the number of firms in the market n changes over time. Denote the distribution of forecast precision Ω in the market by λ_t^m. Theorem 7 implies that given price index P_t, for a subset of $n_t^{'}(n_t^{'} \leqslant n_t)$ firms, the distribution of forecast precision will be $\frac{n_t'}{n_t}\lambda_t(\Omega; n_t, P_t)$. We can focus on the scaled distribution $\lambda_t = \frac{\lambda_t^m}{n_t}$(with a slight abuse of notation for simplicity), which has total measure 1. From Theorem 7, given price index P, λ_t can be determined by F_m^m and F_ξ^m. In particular, it needs to satisfy the following law of motion,

$$n_{t+1}\lambda_{t+1}(\Omega) = \int I\left[\Omega_{t+1}(\tau_\xi, m, P) \leqslant \Omega\right](1-\delta)n_t F_m(dm) F_\xi(d\tau_\xi) + e_t I[\Omega_0 < \Omega] \tag{2.11}$$

where e_t is the number of new entrants at period t and $I[\cdot]$ is the indicator function. The first term measures the number of incumbents with precision smaller than Ω at period

$t+1$. The second term measures the number of new entrants with precision smaller than Ω at period $t+1$.

DEFINITION 3. *Fix initial parameters $(\tilde{\theta}_{i0})_i$ and a pair of distributions (F_m, F_{ξ}). Then an equilibrium consists of a sequence of price indexes P_t, a sequence of feature choice, capital, analytics and data investment $(\theta_{it}, k_{it}, z_{it}, \omega_{it})$ for each incumbent firm, a sequence of the number of firms in the market, n_t, a sequence of the number of new entrants, e_t and a sequence of distribution triplets, $(F^m_{mt}, F^m_{\xi t}, \lambda_t)$ such that the following conditions hold.*

1. *For each firm i, the sequence of $(\theta_{it}, k_{it}, z_{it}, \omega_{it})$ maximizes firms' expected profit.*
2. *New entrants have zero expected profit.*
3. *Consumption goods market clears.*
4. *$F^m_{mt} = n_t F_m$ and $F^m_{\xi t} = n_t F_{\xi}$.*
5. *$n_t = (1-\delta)n_{t-1} + e_t$.*
6. *The date market clears.*
7. *λ_t satisfies Equation 2.11.*

Theorem 8. *The competitive equilibrium exists.*

Proof: see appendix.

We are mainly interested in the stationary equilibrium and the rest of the paper will focus on the stationary equilibrium. The stationary equilibrium is defined as follows.

DEFINITION 4. *Fix initial parameters $(\tilde{\theta}_{i0})_i$ and a pair of distributions (F_m, F_{ξ}). Then an equilibrium consists of a price indexes P, a set of feature choice, capital, analytics and data investment $(\theta_i, k_i, z_i, \omega_i)$ for each incumbent firm, the number of firms in the market, n, the number of new entrants, e and a triplet of the distribution, $(F^m_m, F^m_{\xi}, \lambda)$ such that the following conditions hold.*

1. *For each firm i, $(\theta_i, k_i, z_i, \omega_i)$ solves the optimization problem in Equation 2.5*
2. *The price index clears the market.*
3. *New entrants have zero expected profit.*
4. *$F^m_m = nF_m$ and $F^m_{\xi} = nF_{\xi}$.*

5. $e = \delta n$.

6. *The data market clears.*

7. λ *satisfies the following equation:*

$$n\lambda(\Omega) = \int I\left[\Omega'(\tau_\xi, m, P) \leqslant \Omega\right](1-\delta)nF_m(dm)F_\xi(d\tau_\xi) + eI[\Omega_0 < \Omega]. \quad (2.12)$$

Theorem 9. *There exists a unique stationary competitive equilibrium.*

Proof: see appendix.

2.6 Implication of the Stationary Equilibrium

In this section, we will fully characterize the stationary equilibrium and show a number of implications derived from our results. Because the model can offer answers to many different questions, we divide the section into several subsections. Each subsection contains results on one particular aspect of the stationary equilibrium. In particular, we are interested in the following key variables: price index P, individual firm's pricing strategy p_i, individual firm's forecast precision choice Ω_i', firms' profit and consumer welfare. We start by characterizing firms' dynamics and then discuss how the cost of analytics r_z can affect the key variables we are interested in.

2.6.1 Firm Dynamics

The next theorem characterizes firms' forecast precision dynamics in the stationary equilibrium.

Theorem 10. *There exists τ_0, η_0, v_m, v_ξ and r_0 such that if $\tau < \tau_0$, $\sigma_\eta > \eta_0$, $r_z < r_0$, $Var(F_m) < v_m$ and $Var(F_\xi) < v_\xi$, upon entering the market, new entrants will purchase data such that their forecast precision reaches $\Omega_1^*(m, \tau_\xi)$ at $t = 1$. Starting from $t = 1$, all incumbents will sell data and choose forecast precision $\Omega_2^*(m, \tau_\xi)$ such that $\Omega_1^*(m, \tau_\xi) > \Omega_2^*(m, \tau_\xi)$.*

Proof: see Appendix.

The parameter restrictions in Theorem 10 enable us to focus on this particular interesting equilibrium and get analytic results. In practice, we can see that new small business

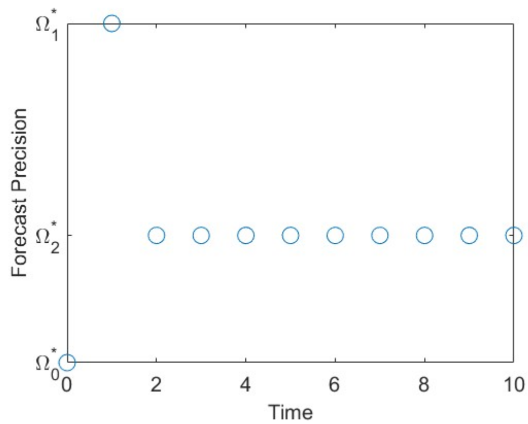

FIGURE 2.6: Dynamics of Firms' Forecast Precision

often needs data support from mature firms such as big platforms or banks.[10] For the rest of paper, refer to firms just entering as the new firms, to firms with age 1 as the young firms, and to firms with age over 1 as the mature firms. Theorem 10 tells us that new firms always purchase data but young and mature firms always sell data. Moreover, mature firms do not want to retain as much precision as they have when they are young. The intuition is that once a firm becomes mature and starts to sell data, the non-rival property of data enables a firm to sell 1 unit of precision for $r_\omega \frac{1}{\kappa}$ dollars which is bigger than r_ω. As a consequence, mature firms prefer to keep relatively lower precision and sell more to new firms. Figure 2.6 gives a illustration of firms' dynamics of forecast precision. Notice that a firm will reach its steady state starting from $t = 2$.

For the rest of this paper, we will maintain the assumption in Theorem 10.

ASSUMPTION 2. *There exists τ_0, η_0, v_m, v_ξ, and r_0 such that if $\tau < \tau_0$, $\sigma_\eta > \eta_0$, $r_z < r_0$, $Var(F_m) < v_m$ and $Var(F_\xi) < v_\xi$.*

2.6.2 Firm Size and Investment Decision

The impact from forecast precision exactly reflects the direct effect because the dynamic effect has nothing to do with firms' demand uncertainty. Firms with higher Ω enjoy higher

[10] Two leading examples are Facebook and American Express. Facebook offers data support for small business on its marketplace. American Express has a program called small business program to help those young firms.

profit. Hence we use precision level Ω as an index to measure firm size. We refer to firms with higher precision as big firms and to firms with lower precision as small firms. In this subsection, we ask how firm size affects firms' investment in capital and analytics. The main results are in the following two propositions.

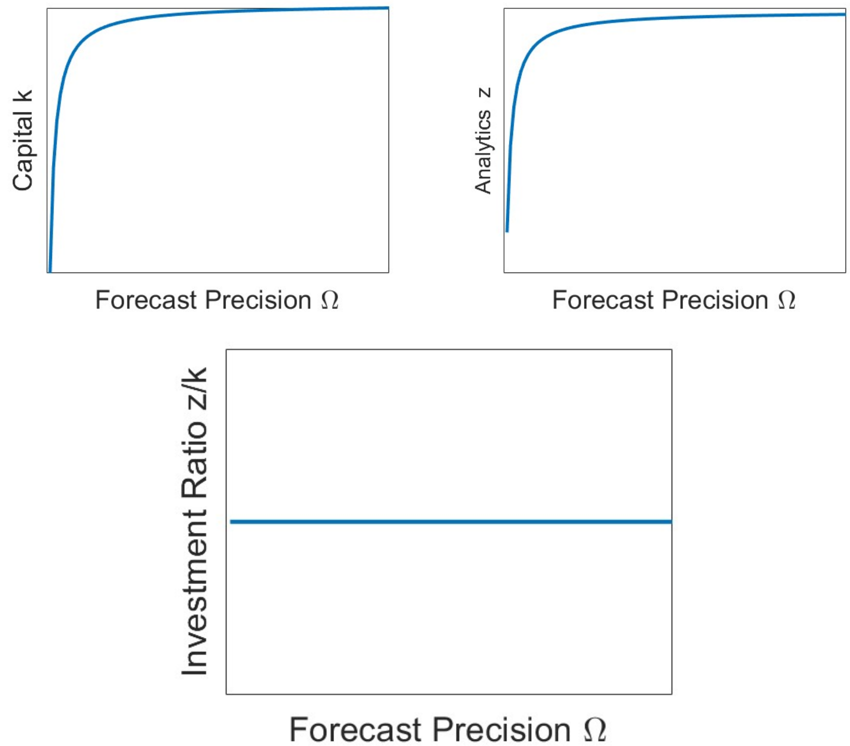

FIGURE 2.7: How Forecast Precision Affects Investment

Proposition 11. *Firms' analytics to capital ratio $\frac{z}{k}$ is constant regardless of precision level Ω.*

Proof: see appendix.

Proposition 12. *Firms' capital investment k and analytics investment z increase in precision Ω.*

Proof: see appendix.

Theses two propositions tell us that big firms invest more in both capital and data analytics. However, firms tend to increase investment in these two factors at the same speed as they grow (See Figure 2.7). The intuition is that $\left(\frac{k}{z}\right)^d$ measures the marginal payoff from investment in analytics because partial derivative of the effective data production function with respect to data analytics z equals to $\left(\frac{k}{z}\right)^d$. On the other hand, the marginal cost of investment in analytics z is fixed at r_z and the marginal cost (or benefit) from data trading is also a constant, $\frac{r_\omega}{\zeta}$. Forecast precision does not affect the production function of effective data directly. So a change in forecast precision will not change the investment ratio. This result again highlights the fact that only the dynamic effect can affect firms' decisions over investment ratio. However, an increase in forecast precision will increase the marginal profit from capital investment because of lower demand uncertainty. Therefore, big firms want to invest more in capital. Recall that there exists complementary effect between capital and analytics. Those big firms will increase investment in analytics at the same speed.

2.6.3 Cross-Section Investment Pattern

From Theorem 7, we know firm size is affected by two parameters: m_i and $\tau_{\xi i}$. As a sequel to the above subsection, we ask how these two parameters affect firms' investment decision. In particular, we compare firms' investment decision between those in premium (high predictive accuracy) industries and in non-premium (low predictive accuracy) industries.

Proposition 13. *Firms in premium industries tend to invest more in both capital and analytics. However, the investment ratio $\frac{z}{k}$ stays constant.*

Proof: see appendix.

This proposition implies that if consumers prefer one type of products, then firms in such industries have an incentive to invest more in both capital and analytics in order to increase production and reduce demand uncertainty. Since whether an industry is premium or not does not affect relative price of analytics, firms do not have incentives to change their investment ratio.

Figure 2.8 shows how consumer preference m affects firms' various investment decisions. As we can see from the top panel, m affects capital investment and analytics investment in

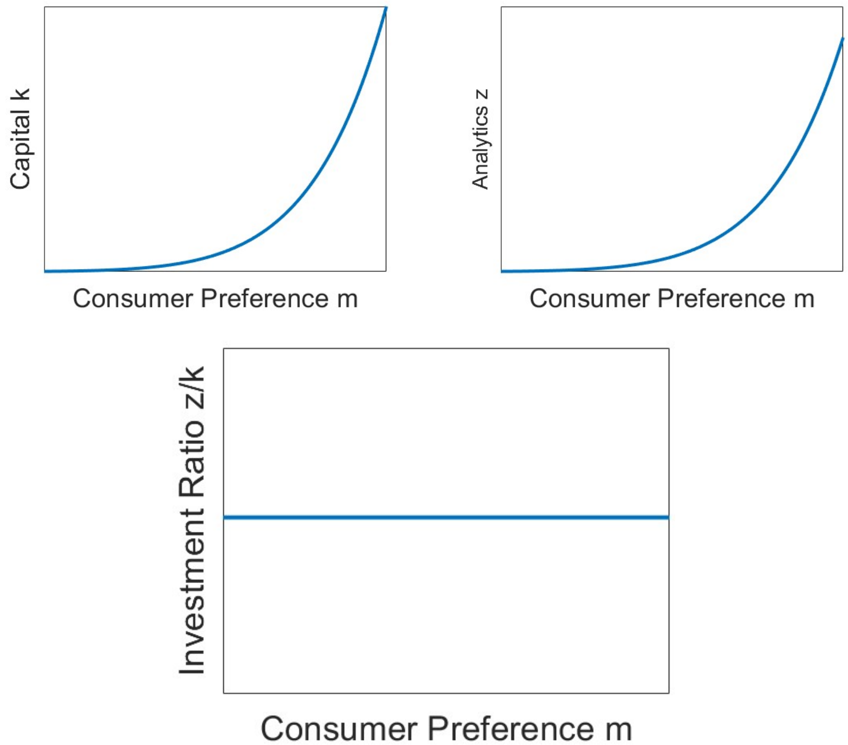

FIGURE 2.8: How Consumer's Preference m Affects Investment

the same way. The bottom graph shows that the investment ratio $\frac{z}{k}$ remains constant no matter how m changes.

Proposition 14. *Firms in high predictive accuracy industries tend to invest more in both capital and analytics. The investment ratio $\frac{z}{k}$ is also higher in these industries.*

Proof: see appendix.

The intuition for this proposition is similar to that in Proposition 13. The main difference here is that a shift in $\tau_{\xi i}$ changes the relative price of data analytics. High $\tau_{\xi i}$ means low relative price of analytics. In high predictive industries, it is easier to obtain precise signals, so firms will have incentives to invest proportionally more in analytics. Figure 2.9 shows how signal precision affects firms' investment decisions. The bottom panel confirms this pattern. The investment ratio $\frac{z}{k}$ strictly increases in signal precision. Looking at the expression for Γ, we can see that firms in the high predictive accuracy industry have higher Γ. So the dynamic effect of data for them are stronger. This proposition has an important

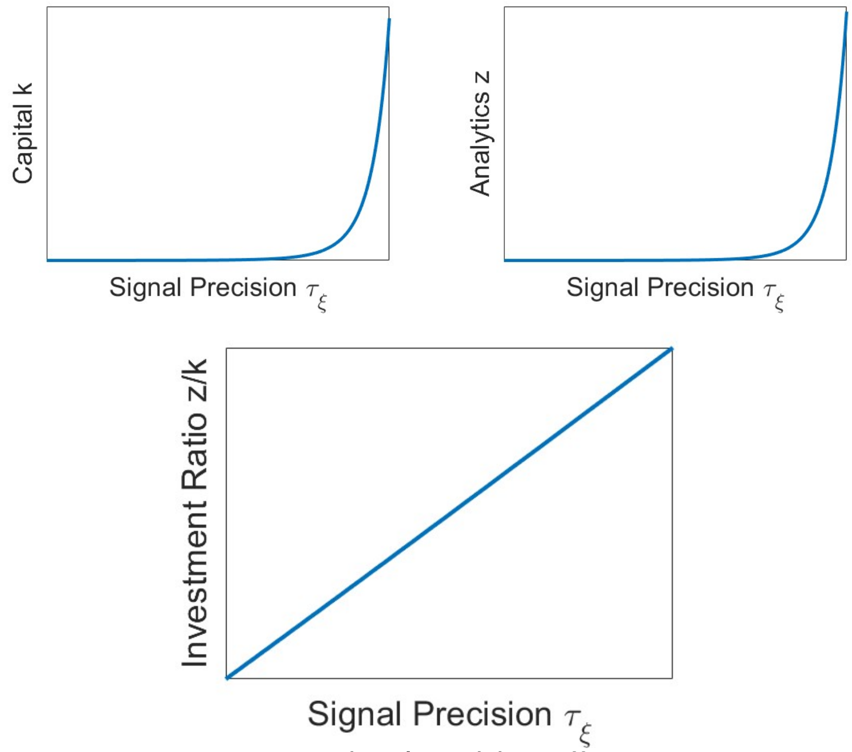

FIGURE 2.9: How Signal Precision Affects Investment

corollary: it tells us what types of firms want to become big data tech giants in equilibrium. The answer is those who can easily acquire precise information on consumers' preference. This is summarized as follows.

Corollary 15. *Analytics intensive firms are those in high predictive accuracy industry. Firms in low predictive accuracy industry are not analytics intensive.*

In practice, we can see that firms such as Tesla, Apple, Netflix and so on can get their users' information relatively easily through their sales. Meanwhile, they also invest a lot in getting and analyzing users' data. For example, in 2020, Netflix launched a free plan for mobile users in Kenya in order to get users' feedback data to see if it can help boost membership subscription.

2.6.4 Firm Attributes and Product Pricing

This subsection discusses how firm attributes affect the price of products that they produce. To better explain the idea of the main result, we first introduce the following definition.

DEFINITION 5. *A product is said to have high match quality if firm i's feature choice matches consumer's taste better: i.e., $\Omega_i^{-1} + \sigma_\epsilon^2$ is smaller.*

The utility penalty from mismatch is measured by $\Omega_i^{-1} + \sigma_\epsilon^2$. The higher precision a firm has, the product that it produces will have higher quality. So in general, big firms will produce high match quality products and small firms will produce low match quality products. Comparing across industries, Theorem 7 tells us that firms in premium or high predictive accuracy industries produce goods with higher quality. Before we compare actual prices of different products, we first define the concept of price-per-utile (PPU).

DEFINITION 6. *The price-per-utile of a product is the ratio of its expected price to the expected utility provided by one unit of consumption of it: $PPU_i = \frac{E[p_i]}{E[v_i(1)]}$.*

Price-per-utile is an index to measure the cost of getting one unit of utility. Because the actual price has taken its benefit to consumers into account, it is more useful to compare the cost of different products by comparing PPU instead of their actual prices.

Proposition 16. *Big firms (firms in premium or high predictive accuracy industries) produce higher match quality products but their price-per-utile is lower.*

Proof: see appendix.

The proposition is quite surprising. It tells us that even though high match quality products may cost a lot, they actually deserve the price as they are even more cost efficient to buyers. The intuition is that the expected price is a linear function of the expected utility provided by one unit of product. After scaling by utility, the only factor affects price-per-utile is the supply of the products. In particular, we can decompose the price-per-utile into the following two parts[11]:

$$PPU \propto \underbrace{P^{\frac{\phi}{\sigma}\left(1-\frac{d}{\sigma(1-d\rho)}\right)}}_{\text{Price Index}} \times \underbrace{\left(\frac{m-\Omega^{-1}-\sigma_\epsilon^2}{r_k-\Gamma}\right)^{-\frac{d}{\sigma(1-d\rho)}}}_{\text{Supply Effect}}. \tag{2.13}$$

Recall $m-\Omega^{-1}-\sigma_\epsilon^2$ measures the consumer's valuation for one unit of goods. It then captures part of the marginal benefit of capital today. The capital investment k is proportional to the ratio $\frac{m-\Omega^{-1}-\sigma_\epsilon^2}{r_k-\Gamma}$. Therefore, the second term in Equation 2.13 represents the supply effect. In equilibrium, big firms will produce more due to lower demand uncertainty or lower user cost of capital. That will help lower the price-per-utile. This proposition explains that why in practice, those best-selling products may not be the cheapest ones. For example, the leader in market share of smart phones in US is Apple (53% in $Q2$ 2021[12]), but their average price is the highest. Similarly, when we filter goods in online shopping platform, the most important criterion often is whether the product matches our preference. Those cheap items may cost us less, but they will not benefit consumers much either.

Now we compare the expected actual prices of different products. From Theorem 7, we know firm size is affected by two parameters: m_i and $\tau_{\tilde{\varsigma}i}$. We compare prices in two dimensions: 1). how expected prices differ between premium industry and non-premium industry; 2). how prices differ between high predictive accuracy industry and low predictive accuracy industry. The main results are contained in the following two propositions.

Proposition 17. *Firms with the same age in premium industry charge higher prices: the expected price increases in the consumer's preference m. The expected price can be decomposed in the fol-*

[11] Derivation of Equation 2.13 is in the proof of Proposition 16 at Appendix.

[12] The data is from Statistia. Sumsung has the second largest market share, 26%.

lowing way:

$$\mathsf{E}[p] \propto \underbrace{P^{\frac{\phi}{\sigma}(1-\frac{d}{\sigma(1-d\rho)})}}_{\text{Price Index}} \times \underbrace{\left(\frac{m-\Omega^{-1}-\sigma_\epsilon^2}{r_k-\Gamma}\right)^{-\frac{d}{\sigma(1-d\rho)}}}_{\text{Supply Effect}} \times \underbrace{(m-\Omega^{-1}-\sigma_\epsilon^2)}_{\text{Match Quality}} \tag{2.14}$$

$$= \underbrace{P^{\frac{\phi}{\sigma}(1-\frac{d}{\sigma(1-d\rho)})}}_{\text{Price Index}} \times \underbrace{(r_k-\Gamma)^{\frac{d}{\sigma(1-d\rho)}}}_{\substack{\text{User Cost of Capital}\\ \text{Dynamic Effect}}} \times \underbrace{(m-\Omega^{-1}-\sigma_\epsilon^2)^{1-\frac{d}{\sigma(1-d\rho)}}}_{\substack{\text{Match Quality}\\ \text{Direct Effect}}} \tag{2.15}$$

Proof: see appendix.

This proposition says that the premium industry charges a higher expected price. Notice that there are two factors influencing the expected price. One is the supply of products. The other is match quality (See Equation 2.14). Because supply is determined by demand uncertainty and user cost of capital, we can rearrange terms and get Equation 2.15. Now, the expected price is determined by the two effects mentioned at the beginning. Proposition 13 tells us that m cannot affect firms' investment ratio and Γ. The dynamic effect remains the same for firms with different consumers' preference. The only effect that makes a difference is the direct effect. For firms with higher m, they face lower demand uncertainty. They have higher match quality and stronger direct effect. As a result, the expected price is higher. This result matches the practice well. Bronnenberg, Dubé, Gentzkow, and Shapiro (2015) show that uninformative consumers often pay extra to buy national brands of headache remedies.

Proposition 18. *Suppose firm* 1 *and firm* 2 *with the same consumers' preference m and age, have signal precision* τ_ξ *and* τ_ξ' *respectively and* $\tau_\xi < \tau_\xi'$*. Then*

1. *If both firms are new, the expected price charged by firm* 1 *is higher:* $\mathsf{E}[p_1] > \mathsf{E}[p_2]$
2. *If both firms are mature and d is large, the expected price charged by firm* 1 *is higher:* $\mathsf{E}[p_1] > \mathsf{E}[p_2]$

This proposition answers the second part of the question: how expected prices differ between high predictive accuracy industries and low predictive accuracy industries? The result is relatively more complicated. Ignore the price index for now because it cannot be affected by any individual firm. Taking a close look at Equation 2.15, we see that two effects are affecting the expected prices in opposite directions. The direct effect increases the expected price but the dynamic effect will decrease the expected price. Which one is dominant depends on the weight of these two effects. Equation 2.15 tells us that the exponents of the latter two factors serve as the weights of them.[13] For example, if $\frac{d}{\sigma(1-d\rho)}$ is big, then dynamic effect tends to be dominant.

Now let us explain the results stated in the proposition. The first one is on new firms. For all the new firms, they are not able to adjust match qualities. The only factor here to be considered is the user cost of capital. We know it decreases in τ_ζ because for firms in the high predictive accuracy industry, capital's marginal contribution to continuation value is larger. Therefore, new firms in high predictive accuracy industry charge lower prices. This result captures firms' incentive to learn consumer's taste even at the cost of lower prices. In practice, when firms promote new products, it usually comes with large discount in order to attract consumers. With better data analytics, this discount can be even larger. One example is the auto insurance quote. If the driver has a smartphone and is willing to download the safe driving apps which track the driver's driving status, a decent discount can usually be applied to the bill.

The second result is on mature firms. The key difference between mature firms and new firms is that mature firms are able to adjust their match quality. So mature firms in different industry can design product with different quality, which will affect the final prices. To see which factor is dominant, notice that $\frac{d}{\sigma(1-d\rho)}$ increases in d. As d increases, capital is more productive and firms invest more in capital. The user cost of capital will be considered more in determining the final price, so firm 1's expected price is higher. This result has two interesting implications. One is that product's price is not necessarily monotone in quality. The other is that capabilities to obtain precise signals can benefit both consumers and firms. Firms in high predictive accuracy industry are willing to sell high

[13] The two exponents add up to 1 exactly.

quality products to consumers at low prices in exchange for consumer data. However, firms in low predictive accuracy industry cannot imitate such pricing strategies due to limitations on getting precise consumer data.

2.6.5 Markup

In this subsection, we study markup. Because of our simple dynamic setting and the incorporation of firm heterogeneity, we are able to answer two questions: 1.) how firms' markup evolves over time. 2.) how firm attributes affect their products' markup.

DEFINITION 7. *We define markup $\mathcal{M}$ by the ratio of expected price to marginal cost*

Theorem 19. *Any individual firm's markup takes the following form:*

$$\mathcal{M} = \frac{r_k - \Gamma}{\rho r_k} \tag{2.16}$$

Proof: see appendix.

Theorem 19 shows that the markup only depends on the user cost of capital $r_k - \Gamma$. Remember that Γ measures the dynamic effect of data. In other words, it is the combination of data and analytics that determines the mark-up of firms:

$$MC \propto \underbrace{(m - \Omega^{-1} - \sigma_\epsilon^2)^{\frac{1-d}{(1-d\rho)}}}_{\text{Direct effect}} \underbrace{(r_k - \Gamma)^{-\frac{1-d}{(1-d\rho)}}}_{\text{Dynamic effect}} P^{\frac{\phi(1-d)}{\sigma(1-d\rho)}}. \tag{2.17}$$

From Lemma 5 and Equation A.12, we have

$$\mathrm{E}(p) \propto \underbrace{(m - \Omega^{-1} - \sigma_\epsilon^2)^{\frac{1-d}{(1-d\rho)}}}_{\text{Direct effect}} \underbrace{(r_k - \Gamma)^{\frac{d}{\sigma(1-d\rho)}}}_{\text{Dynamic effect}} P^{\frac{\phi}{\sigma}\frac{1-d}{(1-d\rho)}}. \tag{2.18}$$

To see why only the dynamic effect matters, look at the expression for marginal cost and expected price. First of all, both effects will increase the marginal cost because firms now are expanding their production scale. The impact of the direct effect on the marginal cost is similar to the risk premium channel in Eeckhout and Veldkamp (2021). Because their model is static, they cannot measure the impact of dynamic effect on the markup. However, as mentioned in Proposition 18, the direct effect increases the expected price but the dynamic effect decreases the expected price. The impact of the direct effect on the expected price is similar to the first channel (more data raises product markup) in Eeckhout and Veldkamp (2021). Moreover, the direct effect on marginal cost and the expected price have the same magnitude and direction. We compute the markup as the ratio of marginal cost to the expected price. The direct effect is neutral on the markup. Notice that a lower user cost of capital (higher Γ) decreases the markup through both higher marginal cost and lower expected price. Therefore, only the dynamic effect can affect the markup.

Based on Theorem 19, we can immediately get answers to the two questions at the beginning of this subsection. These are summarized in the following proposition.

Proposition 20.

1. *Once a firm enters the market, its markup decreases for one period and then keeps constant.*
2. *Markup in high predictive accuracy industry is lower than markup in low predictive accuracy industry.*
3. *Markup in premium and non-premium industry are the same.*

Proof: see appendix.

Proposition 20 says a firm imposes a high markup when new but decreases the markup when mature. The reason lies in the nonrival property of data. When firms are new, they are purchasing data. So they have a low investment ratio $\frac{z}{k}$, and a weak dynamic effect Γ. This leads to a higher markup.[14] However, when firms become young, they start to sell data. Recall $\kappa < 1$. They want to keep a higher investment ratio $\frac{z}{k}$ and stronger dynamic effect Γ. This leads to a lower markup. The second part of the result implies that firms with strong abilities to get precise consumer information are actually imposing lower markup. There is a widespread concern over whether a large amount of information over consumers

[14] Some may think that new firms are usually doing promotions in order to penetrate into the market. So they may have lower markup. Our result on new firms is built for those who have already settled for a while but still have not quite understood their customers' taste.

can grant firms too much market power. Our answer to this question is no. Although it is true that lower demand uncertainty brought by data can increase firms' prices, the marginal cost also increases due to larger production scale. In fact, the increase in prices and the increase in the marginal cost cancel out perfectly. Moreover, complementarity between capital and analytics will further reduce the markup. This result then indicates that better infrastructure to collect consumer data can be beneficial to both firms and consumers. The last result predicts that whether a consumer's preference is high or low should not affect the markup. The reason is that consumers' preference does not affect firms' investment ratio $\frac{z}{k}$. This implication will be different if we do not allow firms to optimally choose investment in analytics as in Farboodi and Veldkamp (2021). Their setup will predict that markup in the premium industries should be higher than the markup in non-premium industries. The reason is that higher preference will lead to more capital investment and large capital investment results in weaker dynamic effects (Γ decreases) and higher user cost of capital. This reflects firms' lack of interests in expanding scale without advanced analytics. The third result can be extended to a more general result: any difference in parameters only affecting sales revenue instead of affecting effective data generation should have no impact on the markup. Similarly, if we omit analytics investment, any positive shocks to parameters which increase sales revenue and do not affect effective data generation always increase the product markup.

2.7 Impact of Data Analytics Price on General Equilibrium

In this section, we study whether a cost reduction for analytics can affect price index P and price-per-utile. If the cost of analytics decreases, it definitely benefits firms. However, will firms decrease their prices in turn? These two questions shape the impact of analytics on the whole economy.

2.7.1 Data Analytics Price and Aggregate Price Index

The first key variable we are going to study is the price index because it directly determines the consumer welfare.

Proposition 21. *Aggregate price index P increases in analytics cost r_z.*

Proof: see appendix.

This is one of the main result of this paper. It tells us if the cost of obtaining advanced analytics goes down, the price index will go down as a result. To explain this result, notice that r_z can affect the price index in three possible ways. One is the direct effect: reduction in cost is reflected as a decrease in monopoly price. The other one is the general equilibrium feedback effect: due to the decrease in monopoly price from the direct effect, firms are not willing to produce as many products as before, which pushes up the price. The third effect comes from data trading: lower data analytics cost changes data price and then affects firms' pricing strategies. However, the third effect is actually zero because ex ante, new firms' expected total data trading over lifetime is 0. As a consequence, the net effect of data price on new firms' expected profit is zero as well. Although the first two effects have opposite directions, the proposition tells us that it is the direct effect that dominates in equilibrium. The reason is that if price index increases as r_z decreases, then it will attract more new firms to enter the market. As a consequence, the supply will go up and the price index has to decrease until the new firms feel indifferent between entering or leaving. Proposition 21 indicates that if it is easier to get advanced analytics, firms are willing to offer a bigger discount. An example in practice is that now, many stores offer discount to consumers if they use digital payment method such as Google Pay and Apple Pay.

Proposition 22. *The elasticity of price index with respect to cost of analytics $\frac{\partial \ln(P)}{\partial \ln(r_z)}$ satisfies the following equation and inequality:*

$$\frac{\partial \ln(P)}{\partial \ln(r_z)} = \frac{1-d}{1-\frac{\alpha}{1-\alpha}/\frac{\rho}{1-\rho}} \frac{E\left[X_i \frac{\Gamma_i}{r_k-\Gamma_i} + X_i' \frac{\Gamma_i'}{r_k-\Gamma_i'}(1-\delta) + X_i'' \frac{\Gamma_i''}{r_k-\Gamma_i''}\frac{(1-\delta)^2}{\delta}\right]}{E[X_i + X_i'(1-\delta) + X_i''\frac{(1-\delta)^2}{\delta}]}, \tag{2.19}$$

$$\frac{\partial \ln(P)}{\partial \ln(r_z)} \geqslant \frac{1-d}{1-\frac{\alpha}{1-\alpha}/\frac{\rho}{1-\rho}} E\left[\frac{\Gamma_i}{r_k-\Gamma_i}\right], \tag{2.20}$$

where X_i is a function of Ω_i and m_i and we use $'$ to denote $t+1$.

Proof: see appendix.

Proposition 22 consists of two parts. The first is a characterization of the elasticity of price index. The second offers a useful lower bound to estimate the elasticity of the price index. Equation 2.19 is the key to the following results. When $\kappa = 1$, Equation 2.19 simplifies to the following intuitive expression:

$$\frac{\partial \ln(P)}{\partial \ln(r_z)} = \frac{1-d}{1-\frac{\alpha}{1-\alpha}/\frac{\rho}{1-\rho}} \frac{E\left[X\frac{\Gamma_i}{r_k-\Gamma_i}\right]}{E[X]}. \tag{2.21}$$

Equation 2.21 shows that the elasticity of the price index is the weighted average of the ratio $\frac{\Gamma_i}{r_k-\Gamma_i}$ and $\frac{X_i}{E[X_i]}$ is the weight. What does $\frac{\Gamma_i}{r_k-\Gamma_i}$ mean? It measures the size of dynamic effect relative to user cost of capital. If firms invest a lot in analytics such that Γ_i are all very large, then the elasticity of price index will be large. However, if firms mainly invest in capital instead of analytics, then r_z only has small impact on the price index. The inequality in Equation 2.20 simplifies Equation 2.19 to some extent. It provides a lower bound on the price elasticity depending only on Γ. It implies that if in a digital economy where the dynamic effect of data is very strong (Γ is large), then the price index will be very sensitive to the cost of analytics. It implies that even a small cost reduction in those algorithms can significantly benefit consumers through much lower price index.

2.7.2 Data Price and Dynamic Effect of Data

In this subsection, we study how the cost of analytics can affect the price r_ω of data and the dynamic effect of data Γ. Particularly, its impact on Γ will be the core of later analysis. We will start with the data price.

Proposition 23. *Suppose $\kappa > \kappa_0$. The elasticity of data price with respect to the cost of data analytics $\epsilon_{\omega,z}$ is positive: $\epsilon_{\omega,z} = \frac{d\ln(r_\omega)}{d\ln(r_z)} > 0$.*

Proof: see appendix.

Proposition 23 implies that as the cost of data analytics goes down, it is cheaper for firms to get data from other partners. This result coincides with what we see in practice. As more and more advanced algorithms such as machine learning are developed, the demand for data increases quickly. Over the past decade, more and more data are collected for people to analyze. Datasets which were previously unavailable can even be accessible for free nowadays.

Proposition 24. *Suppose $d > d_0$ and $\rho > \rho_0$. The dynamic effect Γ and investment ratio $\frac{z}{k}$ decrease in the cost of data analytics r_z.*

Proof: see appendix.

From Proposition 24, we know that a lower cost of analytics r_z encourages firms to invest more in analytics. Intuitively, the lower r_z is, the more advanced analytics a firm will possess. Therefore, the firm are now more efficient in converting data to signals. That's why capital is now more valuable in increasing firms' continuation value. So Γ becomes bigger. In the next subsection, we will use this proposition to derive several implications on price-per-utile and markup. Proposition 24 also shows that all firms will gradually transform into more analytics intensive ones.

2.7.3 Product Quality, Forecast Precision and Data Analytics Price

In this subsection, we study whether a lower cost of data analytics can help all the firms improve their product quality. Surprisingly, we find only a subset of firms will want to improve their product quality, even if it becomes cheaper to invest in some data analytic tools. In particular, only firms in high predictive accuracy industry want to produce better goods. The key intuition behind this is cheaper analytics increase competition. As suggested in Proposition 21, the price index P has to decrease. This negative change discourages firms to learn more about the consumers. For those firms in low predictive accuracy industry, this negative effect is dominant.

Proposition 25. *Suppose $d > d_0$ and $\rho > \rho_0$. There exists ξ_0 and ξ_1, such that for all mature firms,*

1. *If $\tau_{\xi i} < \xi_0$, forecast precision increases in r_z.*

2. If $\tau_{\xi i} > \xi_1$, forecast precision decreases in r_z

Proof: see appendix.

Proposition 25 tells us the change of data analytics cost has asymmetric impact on firms' product design. As cost goes down, the better products will be made by firms in high predictive accuracy industry. However, cheaper analytics hurt firms in low predictive accuracy industry. Why can such an asymmetry arise? The key lies in the following equation,

$$(r_k - \Gamma)\frac{1}{r_z}\underbrace{\left[\epsilon_{r_\omega,z} - \frac{\phi}{\sigma}\frac{1}{1-d\rho}\epsilon_{p,z}\right]}_{\text{Net effect from capital}(-)} + \frac{\rho}{1-d\rho}\Gamma\frac{1}{r_z}\underbrace{\left[1 - d - \epsilon_{r_\omega,z}\right]}_{\text{Net effect from data analytics}(+)} = \beta_{oz}\frac{\partial\Omega}{\partial r_z} \tag{2.22}$$

where $\beta_{oz} < 0$.

Equation 2.22 shows the factors that can affect firms' decisions over forecast precision. What we are interested is the sign of $\frac{\partial\Omega}{\partial r_z}$ on the right-hand side. Notice that both capital investment and analytics investment have an impact here. First of all, suppose analytics becomes cheaper. Then firms are saving money, which is reflected in the term $1-d$ in the second bracket. Second, cheaper analytics will make firms invest more in analytics instead of purchasing data. Data price is affected and this effect is summarized by the data price elasticity in the second bracket. Under the assumption in this proposition, the data price elasticity is relatively small. Hence, the second bracket is positive.

What about the impact from capital investment? First, more capital investment will lower price index because supply goes up. This effect is reflected by the price elasticity in the first bracket. Moreover, a change in capital investment shifts firms' data stock and then affects firms' data sale. This effect is captured by the data price elasticity in the first bracket. Again, because data price elasticity is small, the first bracket is always negative. We then can see two effects move firms' forecast precision in opposite directions. Which one is the dominant depends on the magnitude. From Equation 2.22, it is obvious that the magnitude of the capital investment effect positively relates to the user cost of capital. The magnitude of analytics investment effect positively relates to the dynamic effect Γ. For a firm in high predictive accuracy industry, it has bigger Γ. So the second effect dominates. Their forecast precision decreases in the analytics cost. For firms in low predictive accuracy industry, they heavily rely on capital investment. So the first effect dominates. Their forecast precision increases in the analytics cost.

This result then tells us that as analytics improve or become more powerful, not all the firms will seek to improve product quality. There always exist some fraction of firms find it optimal to sell low quality products in order to control costs and maximize profit.

2.7.4 Price-Per-Utile, Markup and Data Analytics Cost

In this subsection and the next subsection, we analyze how the cost of analytics affects the price-per-utile and markup. Assume that the cost of analytics decreases permanently at $t = 0$: r_z goes down. This sudden change has an immediate effect on firms' current pricing strategy. However, remember that firms at $t = 0$ cannot adjust their forecast precision because the forecast precision Ω is a pre-determined variable. We refer to this effect at $t = 0$ as the short run effect of analytics cost. Starting from $t = 1$, firms' forecast precision will adjust in response to the sudden change in the cost of analytics until a new stationary equilibrium is reached. We refer to the effect on the new stationary equilibrium as the long run effect of analytics cost.

Proposition 26. *Markup $\mathcal{M}$ increases in the cost of analytics r_z.*

Proof: see appendix

This proposition is the direct consequence of Theorem 19 and Proposition 24. We already know that the markup is only determined by the user cost of capital $r_k - \Gamma$. As r_z decreases, the user cost of capital decreases as well. So the markup will decrease. This implies the advantage of analytics in lowering the markup.

Proposition 27. *Price-per-utile of firms with different ages responds to a change in the cost of analytics in the following directions:*

1. *New firms' price-per-utile increases in the cost of analytics.*
2. *In the short run, mature firms' price-per-utile also increases in the cost of analytics.*

Proof: see appendix.

Because new firms and mature firms in the short run cannot adjust their forecast precision or products' quality, the cost of analytics should have the same effect on these two types of firms. By Proposition 21, we know price index increases in the cost r_z. Moreover, from Proposition 24, we know that the user cost of capital $r_k - \Gamma$ increases in r_z. The expression for price-per-utile in Equation 2.13 shows that price-per-utile increases in price index and user cost of capital. Therefore, price-per-utile must increase in the cost of analytics as well. This result again shows that it benefits consumers whenever analytics becomes cheaper. The dynamic effect Γ is the dominant force and it becomes even more dominant when the cost r_z decreases.

Proposition 28. *In the long run, PPU of firms in high predictive accuracy industry increases in the cost of analytics.*

Proof: see appendix.

In the long run, firms will adjust their forecast precision as r_z changes. Proposition 25 shows that firms in high predictive accuracy industry always increase their forecast precision as the cost r_z decreases. This implies that their demand uncertainty decreases and they will further increase production scale. The impact from this channel is similar to the impact from the channels in Proposition 27. Therefore, the price-per-utile will decrease even more in the long run as the cost r_z goes down.

2.7.5 Data Analytics Price and Welfare

In this subsection, we analyze the impact of analytics on welfare. We start with a proposition on analytics' influence on consumers. Then we show an important result on how each individual monopolistic firm is affected by data anlytics. This result highlights the asymmetric impact of analytics cost on heterogeneous firms' profit.

We first look at the consumer. From Lemma 2, we know that consumers' welfare decreases in price index P. Proposition 21 tells us that the price index P increases in the cost of analytics r_z. Therefore, we can conclude that the consumer's welfare decreases in the cost of data analytics. This result is summarized in the following proposition.

Proposition 29. *Consumer's welfare decreases in r_z.*

This proposition shows the trade-off between product quality and prices. In most cases, consumers need to pay a higher price for better quality. Therefore, the key variable that determines welfare is the price index. This actually captures the competitiveness of the economy. If the price index is low, no firm is able to charge high monopolistic prices. Proposition 21 tells us that the lower cost of analytics r_z leads to a reduction in this price index by encouraging more firms to enter and compete. The reason is that complementary effect between capital and analytics makes firms' profit margin increase.

Proposition 30. *For mature firms, there exists a cutoff ξ_2 and ξ_3, such that*

1. *If $\tau_{\xi i} < \xi_2$ then firm i's profit increases in r_z.*
2. *If $\tau_{\xi i} > \xi_3$, then firms i's profit decreases in r_z.*

Proof: see appendix.

Proposition 30 indicates that a change in cost r_z affects firms with different signal precision $\tau_{\xi i}$ in different ways. The cost r_z can affect profit in three main channels. First of all, a higher cost of data analytics r_z can increase price index P as implied by Proposition 21. This effect tends to increase firms' profit. Second, a change in r_z can affect firms' profit through data trading. Third, bigger r_z directly increases the cost of investment in analytics, which hurts firms, as

$$\frac{dV(\Omega, P, r_\omega, r_z)}{dr_z} = \underbrace{\frac{\partial V}{\partial r_z} + \frac{\partial V}{\partial P}\frac{\partial P}{\partial r_z}}_{+} + \underbrace{\frac{\partial V}{\partial r_\omega}\frac{\partial r_\omega}{\partial r_z}}_{+} + \underbrace{\frac{\partial V}{\partial \Omega}\frac{\partial \Omega}{\partial r_z}}_{+}. \tag{2.23}$$

Equation 2.23 shows how these three effects can affect the firm value. Focus on the low predictive accuracy industry first. From Proposition 23 and Proposition 25, we know the last two terms are both positive. It then suffices to figure out the sign of the first two terms. Recalling Equation 2.21, if r_z increases, the increase of the price index P is a weighted average of the ratio $\frac{\Gamma_i}{r_k - \Gamma_i}$. However, firms in the low predictive accuracy industry have small Γ_i and lower $\frac{\Gamma_i}{r_k - \Gamma_i}$. Notice that firms' investment cost in analytics is positively related to $\frac{\Gamma_i}{r_k - \Gamma_i}$ because bigger Γ_i implies higher data analytics investment ratio. This implies that the increase in price index is enough to offset the loss from data analytics investment for firms in the low predictive accuracy industry. So the sum of the first two terms should be positive for those firms.

What about firms in high predictive accuracy industry? We can first derive a upper bound for the impact of analytics cost on their firm value:

$$\frac{dV(\Omega, P, r_\omega, r_z)}{dr_z} < \underbrace{\frac{z}{\delta}\left[-1 + \frac{\epsilon_{r_\omega, z}}{1-d}\right]}_{\text{Net effect from data analytics}} + \underbrace{\frac{\partial P}{\partial r_z}\left[\frac{\partial V}{\partial P} - \frac{\omega \mathrm{E}(\frac{\partial \hat{\omega}}{\partial P})}{\delta \mathrm{E}(\frac{\partial \hat{\omega}}{\partial r_\omega})}\right]}_{\text{Net effect from capital}} + \underbrace{\frac{\partial V}{\partial \Omega}\frac{\partial \Omega}{\partial r_z}}_{-}. \tag{2.24}$$

Similar to the decomposition in Equation 2.22, we can decompose the total impact into three different effects. The last one is from product quality. This is negative from Proposition 25. The other two comes from analytics and capital investment respectively. For the analytics investment effect, it consists of two parts. One is the direct impact which is -1 in the first bracket. The other is the indirect effect from data price, which is the second term in the bracket. The net effect from analytics is negative because of the small data price elasticity. The first impact is from the product prices. Similarly, for the capital investment, it has a direct impact from price index because capital investment affects supply and then average market price. This is measured by the first term in the second bracket. The second impact is from data trading due to a change in data stock. This is reflected by the second term in the second bracket. The first two effects decrease in τ_ξ. In particular, the investment in data analytics z is directly controlled by τ_ξ or Γ. Therefore, firms in the high predictive accuracy industry have large τ_ξ and Γ. They rely more on analytics investment. The sum of the first two effects should be negative for them. As a consequence, their profit decreases in the cost of analytics.

Proposition 30 sheds light on the concentration of industry over the past several decades. As discussed in Kwon, Ma, and Zimmermann (2021), bigger firms tend to grow even bigger. However, small firms are still struggling. Our results are consistent with this empirical finding. As the analytics becomes cheaper, only firms in the high predictive accuracy industry (they are indeed big firms) are able to take advantage of this. On contrast, firms in low predictive accuracy industry can even shrink.

2.8 Conclusion

This paper builds a general equilibrium model with differentiated firms to show how analytics can affect the aggregate price index, firms' product quality and welfare. We decompose the impact of analytics on firms' investment decisions into two effects: direct effect and dynamic effect. The dynamic effect can be affected by analytics investment due to the complementary effect between capital and analytics. In many cases, the dynamic effect plays a bigger role. We show that with cheaper analytics, it helps lower the price index and then reduce the markup as well. Consumer welfare increases as a result. Firms in high predictive accuracy industries have strong incentives to offer discount to consumers and get data in return. They are even willing to sell higher quality products at a lower price under certain circumstances. The impact of analytics price on firms' profit is two-fold. Firms in high predictive accuracy industries benefit from cheaper analytics while firms in low predictive accuracy industry suffer. These results show that not only data but also analytics plays an important role in shaping the overall economy.

3. Buy-Now-Pay-Later (BNPL) Platform and Consumer Welfare

3.1 Introduction

As a new FinTech business model, the buy-now-pay-later (BNPL) platforms have grown rapidly over the past 5 years. Berg, Fuster, and Puri (2021) estimate FinTech lending in the BNPL segment to be $25 billion in the US, with growth over 100%. Comparing to it, FinTech business lending has an annual growth rate 43% according to Ziegler, Shneor, Wenzlaff, Suresh, de Camargo Paes, Mammadova, Wanga, Kekre, Mutinda, Wang, Closs, Zhang, Forbes, Soki, Alam, and Knaup (2021). How do BNPL platforms affect consumers' welfare?

BNPL platforms mainly provide two types of service. One is lending money to the customers who cannot afford their desired products. The other is providing private data support for partner firms and charging fee on firms that want to join the platform.[1] Data can help firms to reduce marginal cost. One of the most important ways to get public consumers' preference data is to look at other firms' product design. It means more firms result in more public data.[2]

One may conjecture that BNPL platforms should increase all consumers' welfare by providing loans to financially constrained consumers. This paper argues that BNPL platforms can hurt unconstrained people by increasing out-of-platform merchants' product prices. The introduction of BNPL platforms can have two opposite effects on the number of firms in the market. On one hand, it enables consumers to purchase goods with higher quality by providing loans and lowering partner firms' prices due to private data support. This tends to increase the number of firms in the market because it is more costly for a firm to produce goods with higher quality. On the other hand, platform fee and private data support can

[1] According to the $10-K$ filling report of Affirm, it provides merchants with valuable data they can use to inform more tailored promotions and offers to consumers.

[2] In practice, other firms' design can reflect the current trend to a large extent. If the design is a hit, then it directly reveals consumers' preference. If the design is a miss, it still says something about what consumers dislike. An example in video game industry is that leading works of firms such as Nintendo and Rockstar Games tell many other game developers what consumers really like in new games.

also lead to fewer firms competing with each other. One reason is that platform fee increases prices and lowers consumers' choice of quality. The number of firms then decreases because it is less costly to produce low quality products. The other reason is private data provision reduces the marginal cost of partner firms and enables them to expand scale. This then decreases the number of firms in the market as well. Which effect is the dominant one depends on platforms' lending capability. If the lending capability is strong (able to lend more to a consumer), the first effect dominates and BNPL platforms benefit everyone. Otherwise, only severely constrained consumers can benefit from BNPL services.

BNPL platforms are different from credit cards and e-commerce platforms. Credit cards usually have a very high annual APR if the card holder cannot pay the statement balance on time. However, these BNPL platforms can offer interest rate free loans with duration much longer than one month.[3] Sometimes, Chase or American Express will give members an offer called **Plan It**. It allows members to split a large purchase into several small installments and pay back within half a year without any interest rate. The difference is that you can never predict when these offers are available to you. In terms of data provision, most of credit cards issuing banks do not provide data support to firms. For e-commerce platforms, they do not provide loans in general. Actually, many large e-commerce companies have partnered with BNPL platforms to provide finance support for constrained consumers.[4]

We build a model to incorporate BNPL platforms and compare it to a benchmark model without such platforms. There are three types of agents. A continuum of consumers, firms and BNPL platforms. We first consider a model with one monopolistic BNPL platform and then present a model with two platforms competing with each other. There are a continuum of products differentiated by the quality. Each firm produces one type of products. The market for trading goods only opens at the beginning of the first period, $t = 0$. Each consumer receives an endowment at each period but she can only purchase at $t = 0$ and consume at $t = 1$. Rich consumers have larger endowment but poor consumers have smaller

[3] According to Affirm (2021), the APR for a 3 month loan is 0.

[4] For example, Amazon partnered with Affirm to delivery pay-over-time services at August 2021.

endowment. We fix a consumer's demand to be 1. So poor consumers have incentives to borrow if there is a BNPL platform.

Firms can enter the market by paying a fixed cost. Each firm optimally chooses what type of products to produce and competes with each other. So there are a continuum of perfectly competitive markets differentiated by the quality of products supplied. Firms' cost function is assumed to increase in the quality of products and decrease in the size of consumer data. Public data can be obtained freely and it increases in the number of firms in the market. The BNPL platform can lend money to consumers with an upper limit. Firms can choose to join the platform by paying a performance fee. This performance fee is some fraction of the revenue from consumers who are using the BNPL services.[5] In equilibrium, firms earn zero profit but the monopolistic platform can get positive profit by optimally choosing the performance fee.

Who are using the BNPL services? Our results show only financially constrained consumers use such services. This is consistent with what we see in practice. Several surveys and media coverage have shown that many BNPL users are college students. Alcazar and Bradford (2021) point out that for the BNPL product offered directly to consumers by FinTechs before a purchase is made, most users are Generation Z and financially underserved consumers.

Our model shows that the BNPL platform always increases very poor consumers' welfare. The benefit from loans is enough to compensate the potential loss from higher market prices. However, for consumers who are not very poor, the impact of BNPL platforms is not necessarily positive. We can first look at rich consumers who do not use the BNPL services. What these consumers really care about is the market price. The market price is determined by the number of firms in the market. If more firms are competing with each other, then more data will be available. Firms' cost goes down and the market price goes down as well. The platform can affect the total number of firms through three factors: 1.) lending; 2.) performance fee; and 3.) private data provision. Lending always

[5] This fee is similar to the one charged by credit cards or digital payment methods.

increases the number of firms because it lets consumers to buy goods with higher quality. However, it is more costly to produce higher quality goods, so more firms have to enter. Performance fee always decreases the number of firms because the fee will increase firms' prices. Higher prices will force consumers to buy lower quality products. Then fewer firms will be needed to fulfill the demand. The data provision can affect the number of firms in two opposite directions. On one hand, it lowers marginal cost and lowers prices. So quality of products purchased will go up. More firms have to enter the market. On the other hand, lower marginal cost can make in-platform merchants expand more than out-of-platform merchants. The increasing scale of in-platform merchants forbids new firms to enter. The total number of firms decreases. Our paper shows the net effect of private data provision on the number of firms is neutral. Therefore, what really matters are lending capability and the performance fee. If the platform is able to lend more to consumers but charge low performance fee, the number of firms will increase. Even out-of-platform merchants' cost will go down and out-of-platform prices go down as well. Those rich consumers' welfare then increases. Otherwise, the BNPL platform will hurt rich consumers.

We then look at consumers who are using the service but are not very poor. For these people, they are facing a trade-off between more slack budget constraints and possibly higher prices. In general, the benefit from loosening consumers' budget constraint gradually decreases as their endowment increase. As a consequence, in order to increase the welfare of BNPL users who are relatively rich, the performance fee charged by the platform cannot be high.

The theory also predicts that if the performance fee is very small, no out-of-platform merchants can survive. However, it still does not necessarily mean that the BNPL platform will benefit rich consumers. The reason that the platform dominates the market is that out-of-platform merchants cannot beat in-platform merchants by having cheaper prices. Now, rich consumers are forced to shop through platforms. If the lending capability of the platform is weak, the number of firms in the market can be much smaller than that in the benchmark model without the platform. It then implies that rich consumers may get hurt because they are now forced to purchase goods with higher prices.

Will the monopoly platform pick a low performance fee to benefit everyone? The answer again depends on the lending capability. If the lending capability is strong, then the equilibrium performance fee indeed can improve every consumer's welfare. Otherwise, if the lending capability is weak, then the platform always hurts rich people. The intuition is that no matter what performance fee a platform charges, it always has a negative impact on all consumers. In order to compensate this, the lending capability needs to be strong enough so that consumers are now able to purchase much better products. As a result, firms' cost goes down due to the fact that more firms will be in the market.

Then we consider a model with two platforms competing with each other in a Betrand way. Firms are allowed to join one of the platforms and consumers are free to decide which platform to use. If firms and consumers are indifferent between two platforms, they just roll a dice and join one of them with probability one half. In equilibrium, both platforms earn zero profit and their revenue should equal to entry cost. The model shows that as long as the entry cost is small, the consumer welfare in the duopoly competition model is always higher than that in the benchmark model. Competition forces both platforms to cut performance fee and it then can benefit everyone. Next, we compare this duopoly model to the monopoly model to see if such competition can improve consumers' welfare.

It is tempting to conjecture that consumers' welfare should be higher in the duopoly case. However, our model predicts that this conjecture can be wrong under some circumstances. One benefit from platform competition is that the performance fee will be lower. However, such competition also leads to some cost. Private data provision may not be that efficient as in the monopoly case. The reason is that now the in-platform merchants have to be split into two platforms. The private data support from one platform can be weaker than before. It implies that the BNPL users may get hurt due to platform competition. Which effect is dominant depends on the entry cost, private data provision ability and the lending capability. Here, stronger private data provision ability means that a platform can generate more data with the same number of partner merchants. If both platforms have very strong data provision ability and lending capability, then such split of in-platform merchants does not matter too much. Market prices in the platform in the duopoly model will be lower. It then has two important effects. One is that BNPL users (consumers) can purchase goods with higher quality. This effect will increase the number of firms in the market. The other effect is that more consumers will use the platform. Because the platforms have strong lending capability, it can further increase the number of firms in the market. In particular, if more consumers are using these platforms, more firms will enter the market. As a result, more firms are competing in the duopoly model due to both effects. All consumers' welfare will be higher in the duopoly model.

However, if the entry cost is high, public data provision ability and the lending capability is weak, then competition between platforms does not help too much in reducing the performance fee. In this case, the split of in-platform merchants will play a significant role. In particular, market prices in the platform will be lower in the monopoly model. It has two effects. One is that consumers who are using the BNPL platforms will get hurt in the duopoly model. The other is that more consumers are using the platform in the monopoly model. However, because the lending capability is weak, more users of BNPL platforms can only decrease the total number of firms in the market. So the number of firms in the duopoly model is still higher than the number of firms in the monopoly model. Rich consumers who are not using the BNPL services will have higher utility in the duopoly model.

The rest of this paper is organized as follows. Section 3.2 presents a summary of related literature. Section 3.3 solves the benchmark model without BNPL platforms. Section 3.4 presents the model where only a monopolistic platform exists and compares its allocation to the benchmark model. Section 3.5 solves a model with two platforms competing with each other. It compares the welfare of different types of consumers to the monopoly model in Section 3.4. Section 3.6 concludes.

3.2 Related Literature

Our paper is mostly related the literature on FinTech financing. There is a growing literature in this area. Papers such as Goldstein, Jiang, and Karolyi (2019) and Allen, Gu, and Jagtiani (2021) survey the important questions and results in FinTech. Many papers have studied how the FinTech firms can monitor and screen consumers better than a traditional bank. For example, Berg, Burg, Gombovic, and Puri (2019) show digital footprints left by consumers can have strong predictive power on consumers' credit risk. Fuster, Goldsmoth-Pinkham, Ramadorai, and Walther (2021) show that the introduction of statistical algorithm such as machine learning can deliver better prediction accuracy. Other papers focus on the improved relationship between consumers and lenders. Fuster, Plosser, Schnabl, and Vickery (2019) also show that FinTech lenders can process mortgage application 20% faster than other lenders. Buchak, Matvos, Piskorski, and Seru (2018) show improving lending technology among new shadow banks can lead to cheaper lending and better products.

As FinTech firms, the screening technology and shorter processing time are also two features of the BNPL platforms. However, this paper focuses on another important feature of BNPL platforms: its private data support for partner firms. In particular, how the combination of data support and lending can affect consumers' welfare? Now, the existing works on BNPL platforms are scarce. Little research has been done on the impact of BNPL lending on consumers' welfare. Berg, Burg, Gombovic, and Puri (2019) analyze BNPL lending with a focus on screening technology. Our paper complements the literature in this aspect. Because FinTech lending relies a lot on soft information and consumers can get loans instantly, the impulse buying problem can occur. Lots of previous papers have researched this self-control problem. Heidhues and Kőszegi (2010) show with credit cards, unsophisticated consumers can overborrow and suffer welfare loss in the end. Benton, Meier, and Sprenger (2007) point out the self-control problem leads to large size of uncollateralized debt held by U.S. households. Maggio and Yao (2021) show that FinTech lenders tend to lend first to risky borrowers and FinTech borrowers are significantly more likely to default than neighbors with similar characteristics.

Different from theirs, we show the impact of BNPL platforms on consumers' welfare is ambiguous without overborrowing. In particular, this paper finds any BNPL platform can have a positive impact on out-of-platform merchants and rich consumers by increasing the number of firms in the market. Their loans enable consumers to purchase better products and let more firms compete with each other. So the introduction of BNPL platform can create data spillover for all firms and help them to reduce cost. However, we also point out if the platform has limited lending capability, BNPL platforms can hurt rich people due to the performance fee.

Our paper is also related to literature on big data. We treat data as free public signals provided by each firm. There are a number of papers discussing the role of data in digital economy. One of the key defining feature of data is that they are free. Papers such as Jones and Tonetti (2020) and Farboodi and Veldkamp (2021) all take this view. Jones and Tonetti (2020) analyze the impact of ownership of data on social welfare. Farboodi and Veldkamp (2021) show data alone cannot sustain the economic growth. In Eeckhout and Veldkamp (2021), data is generated by firms freely and can be used to estimate demand. Also, see Bergemann and Bonatti (2019) for a survey of the literature on data market. This paper analyzes the impact of data in the framework of FinTech lending. We show whether the BNPL platforms can help generate more public data is the key to determine whether their existence can be a Pareto improvement for all consumers. This paper also finds that a monopolistic platform can benefit BNPL users more than duopolistic platforms under some circumstances because of data aggregation.

Finally, this paper is related to platform pricing. Several papers studied the data sharing and market power of platforms. Kirpalani and Philippon (2021) show that data sharing can improve gain from trade by improving match quality. However, such data sharing can lead to externalities that increase the market power of platforms. Caillaud and Jullien (2003) model platforms as matchmakers and analyze the competition with price discrimination.

Rochet and Tirole (2003) focus on network externalities and analyze competition among two-sided platforms.

3.3 Benchmark Model: No BNPL Platforms

The benchmark model is static. There are two types of agents. One is a continuum of consumers with total measure $n_c = 1$. The other is a continuum of firms whose measure will be endogenously determined in the equilibrium. There are a continuum of consumption goods. These goods are differentiated by their quality ν. For the rest of the paper, we will consider quality ν as the type of goods. Every firm can choose only one type of consumption goods to produce. Type ν consumption goods is supplied by a competitive market with price $p(\nu)$. The market only opens at $t = 0$. No discounting is considered in this paper. In the next two subsections, we will discuss consumers and firms in detail.

3.3.1 Consumer

Each consumer has a fixed demand 1 for consumption goods. In other words, she only needs to pick one type of consumption goods and purchase one unit to maximize her utility. The utility function is

$$u_b = \nu - p_b(\nu) \tag{3.1}$$

where the subscript b represents benchmark. Every consumer will receive endowment W at dates $t = 0, 1, 2, ...$ W is drawn i.i.d from a distribution with CDF $F(W)$ and density $f(W)$. The support of F is $[\underline{W}, \overline{W}]$. We will call W the type of consumers. Obviously, endowment W measures the wealth level of different consumers. In the benchmark model, there are no BNPL platforms, so consumers cannot borrow. They can only buy at $t = 0$ and consume at $t = 1$. Their objective function is as follows

$$\max_{\nu} \quad u_b = \nu - p_b(\nu) \tag{3.2}$$

$$s.t. \quad p_b(\nu) \leqslant W \tag{3.3}$$

The first order condition is

$$1 = (\lambda_b + 1)p_b'(\nu) \tag{3.4}$$

where λ_b is the multiplier for the budget constraint. Suppose the consumers are deep pocket. Then given the price schedule $p(\nu)$, consumers' favorite type ν_b^u is determined by the following equation

$$1 = p_b'(\nu_b^u) \tag{3.5}$$

where the superscript u for ν_b^u stands for unconstrained.

3.3.2 Firms

Any firm can enter the market at a fixed entry cost c_e. Upon entering, each firm has to choose one type of goods to produce. Because the market is competitive, firms take the price schedule $p_b(\nu)$ as given. We will also call ν, the type of firms. Denote $q_b(\nu)$ by the production scale of firm with type ν.

Firms' production cost relates to three key elements: 1.) quality of the product ν; 2.) production scale $q_b(\nu)$; 3). measure of public accessible data D_b. Cost increases in the first two elements. The third element is the key of this paper. In practice, in order to design the best products, firms need to acquire information on consumers' preference. Otherwise, firms' investment can be made in the wrong direction and lead to huge loss. Therefore, in our model, we model data as a way of reducing production cost. Then a natural question arises: where does data come from? One of the easiest and cheapest way comes into mind is looking at the other firms' strategies. For example, if during a period when most of firms have excellent sales in blue v-neck shirts, then one may know that consumers' favorite color of v-neck shirts is blue. We call this type of data public data. Moreover, the measure of data obtained in this way should increase in the number of firms in the market. The reason is that more firms mean more exploration. If there exists only a few firms, it is hard to know the complete preference of consumers.

Therefore, our choice of the cost function is

$$c(\nu, q, D_b) = \frac{1}{2}\nu^4 q_b^2 \phi(D_b(n_b)) \tag{3.6}$$

where n_b is the total measure of firms in the market. ϕ is strictly decreasing and D_b is strictly increasing. ϕ is concave. The function ϕ measures the contribution of data to reducing production cost. The more data a firm has, the smaller $\phi(D_b(n_b))$ will be. One natural restriction on ϕ is that $\lim_{D_b\to\infty}\phi(D_b)$ is strictly positive. The reason is that even if a firm knows perfectly about its customers, it still needs to make positive investment in order to produce. Since we are interested in the composite of $\phi \circ D_b(n_b)$, we can simply assume $D_b(n_b) = \alpha n_b$. For simplicity, we pick $\phi(D) = 1 + \frac{a}{1+D}$.[6]

Type ν firm's objective function is

$$\pi(\nu) = \max_{q_b} p_b(\nu)q_b - \frac{1}{2}\nu^4 q_b^2 \phi(\alpha n_b) - c_e \tag{3.7}$$

Incumbent firms' profit consists of three parts. The first is sales revenue. The second is production cost and the third is the entry cost. The first order condition for production scale q_b is

$$q_b(\nu) = \frac{p_b(\nu)}{\nu^4 \phi(\alpha n_b)} \tag{3.8}$$

Notice from Equation 3.8, firms' production scale decreases in the quality of the product but increases in market prices and the size of accessible data.

3.3.3 Equilibrium

In this subsection, we will characterize the competitive equilibrium and compute consumers' welfare.

DEFINITION 8. *The competitive equilibrium consists of a price schedule $p_b(\nu)$, firms' production scale $q_b(\nu)$ and a distribution of firms' type $G_b(\nu)$ such that*

1. *Given price schedule $p_b(\nu)$, $q_b(\nu)$ solves firms' problems.*
2. *Market of every type of consumption goods has to clear.*
3. *Firms earn zero profit.*

In the next several propositions, we will characterize consumers' optimal consumption choice and firm distribution in detail. We start with a result on the price schedule in the

[6] The form of ϕ does not matter in general. One can choose other ϕ and get the same results.

equilibrium.

Proposition 31. *The price schedule in the benchmark model is:* $p_b(\nu) = c_b\nu^2$*, where* $c_b = \sqrt{2c_e\phi(\alpha n_b)}$*.*

Proof see appendix.

Proposition 31 shows us that the price schedule is a quadratic function of the quality. Goods with higher quality will ask for higher price. There are two factors that can affect the shape of $p_b(\nu)$: entry cost c_e and data size αn_b. Higher entry cost will lead to higher price. Meanwhile, more data can help lower price. Recall that data size is determined by the total number of firms in the market n_b. It shows us that it is important to increase the number of firms in the market in order to lower prices. The next proposition will characterize consumers' optimal choice given the equilibrium price schedule.

Proposition 32. *There exists a cutoff* W_b^c *and* $W_b^c = \frac{1}{4c_b}$ *such that*

1. *If* $W < W_b^c$*, type* W *consumer chooses quality* $\nu_b^c = \sqrt{\frac{W}{c_b}}$*.*
2. *If* $W \geqslant W_b^c$*, type* W *consumer chooses quality* $\nu_b^u = \frac{1}{2c_b}$*.*

Proof see appendix.

Proposition 32 shows there exists a cutoff W_b^c. If consumers' endowment at $t = 0$ is bigger than W_b^c, they will purchase the unconstrained optimal quality ν_b^u. Otherwise, consumers' budget constraint will be binding. They are only able to buy the most expensive goods that they can afford. Notice that for all constrained consumers, different types will purchase products with different qualities. In other words, in the constrained region, endowment and product quality have a one-to-one correspondence. We will use this observation to derive firm distribution in the next proposition.

What is the impact of the number of firms n_b on consumers' choice? First of all, more firms means flatter price schedule because $\frac{1}{2}\nu^4q^2\phi(\alpha n_b)$ decreases in n_b. So all the consumers are now able to pick products with higher quality. However, interestingly, more consumers are constrained now: W_b^c increases in n_b. The intuition is that consumers' unconstrained optimal ν_b^u increases due to flatter price schedule. However, the price of the unconstrained optimal $p(\nu_b^u)$ increases as well due to the convexity of the price schedule. Therefore, in order to get the most desired product, consumers now need to pay even more now.

Before we state the result on firm distribution, it is useful to highlight an observation. Except for firms that produce type ν_b^u goods, the other firms are trading with exactly only one type of consumers. The reason is that constrained consumers with different endowment choose different qualities. Therefore, sorting firm according to product quality is equivalent to sort firms by the type of consumers who they are trading with. Denote the type of consumers who purchase type ν products by $W(\nu)$.

Proposition 33. *Firm distribution $G_b(\nu)$ has an atom at ν_b^u. The mass at ν_b^u is $\frac{W_b^c(1-F(W_b^c))}{2c_e}$. For any quality $\nu < \nu_b^u$, G_b has density g_b such that $g_b(\nu) = \frac{W(\nu)f(W(\nu))}{2c_e}$*

Proof see appendix.

Proposition 33 indicates the firm distribution G_b is a mixed distribution with only one atom. We then can easily characterize the total number of firms in the market n_b:

$$n_b = \int_{\underline{W}}^{W_b^c} \frac{Wf(W)}{2c_e} dW + \frac{W_b^c(1-F(W_b^c))}{2c_e} \tag{3.9}$$

It consists of two parts. The first part is the measure of firms serving constrained consumers. The second part is the measure of firms serving unconstrained consumers. Observe that the right hand side of Equatin 3.9 increases in the cutoff endowment W_b^c. Recall that W_b^c also increases in n_b. The next theorem shows that n_b is uniquely determined by Equation 3.9. Based on these observations, we can deduce that in order to have larger number of firms in the market, the cutoff endowment W_b^c must be bigger. One way to achieve this is that firms' data production function becomes more productive: α increases. The other way is to increase the data processing ability: same amount of data reduces more production cost ($\phi'(D_b)$ decreases). All of these show the positive feedback loop brought by data: more data results in more firms which then leads to even more accessible data.

Theorem 34. *There exists a unique competitive equilibrium.*

Proof see appendix.

Proposition 35. *Consumers' welfare increases in data processing ability a and data production productivity* α*.*

Proof see appendix.

We can observe from Proposition 35 that there are two ways to help increase consumers' welfare. One is to make firms more productive in generating data (α increases). The other way is to enable firms to take better advantage of datasets (α increases). The intuition are the same for both results. More data or more efficient use of data reduces firms' production cost. It induces more firms to enter the market. Larger number of firms further increases accessible data for all firms. Production cost will become even lower and then the market price will decrease as well. As a consequence, all consumers are able to purchase products with higher quality at a lower price.

3.4 Model with a Single BNPL Platform

In this section, we formally introduce the BNPL platform. To begin with, we solve a case where there exists only a monopolistic platform. In practice, these platforms play two roles. One is partnering with banks to lend money to consumers. The other is to provide service to their partner firms to help increase firms' revenue. One of the most important feature of BNPL platforms is that they are able to provide data support for firms that join them. This feature will also be the key in our model.

3.4.1 BNPL Platform

There exists one monopolistic platform in the market. It has two roles as mentioned above. One is that it provides loans to consumers at a zero interest rate. For a consumer of type W, she can borrow at most $(k-1)W$ from the platform at $t=0$, where $k \in \bar{k}$.[7] Why the interest rate is zero? Our model basically considers the short-term loan provided

[7] The choice of the upper limit $\bar{k}$ will be specified later.

by these platforms. One of the key feature of BNPL firms is that they provide the so-called service: pay in 4 installments with zero interest. The duration of such loans is usually 1 to 3 months.[8] k is an exogenous variable. It captures the enforcement ability of platforms or debt collecting agencies. The short-term loans provided by BNPL firms are unsecured. If the consumers default, the platform can either report to a credit bureau or rely on a debt collecting agency. Because the focus of this paper is the data support from platforms to partner firms, we will ignore the credit risk here and make k exogenous. If the consumers feel indifferent between purchasing via the platform and shopping out of a platform, we assume they always use the platform.

The second role of BNPL platforms is their relationship with partner firms. If firms want to join the platform, they have to pay a performance fee. In particular, supposing that a firm in the platform sells a product at a price to a consumer using the BNPL service, this firm has to pay θp to the platform. In practice, we can see firms may have to pay another type of entry fee upfront. Because this flat entry fee is determined by the negotiation between firms and the platform, it may depend on firm size. We then do not differentiate these two types of fees. θ should be considered as a composite rate charged to all partner firms.

To support firms, the platform provides data to them. We call the data provided by the platform private data and the data accessible to all firms as in the benchmark model public data. The production function of private data is $D_{pv}(n_p) = \varphi n_p$. n_p is the number of firms in the platform because platforms can only aggregate information from their partners. If φ is very big, then private data generation is far more productive.

The platform's problem is choosing θ in order to maximize its profit. Denote the price schedule for firms in the platform, firms' production scale and firm distribution by p_{mi}, q_m

[8] It the consumer wants to get loans with longer duration, positive interest rate is required. However, this is not the focus of this paper.

and G_m respectively.[9] Platform's objective function is

$$\max_{\theta} \int \theta p_{mi}(\nu) q_m(\nu) I\left[W(\nu) \in \Xi_m\right] dG_m(\nu) - c_p \tag{3.10}$$

where Ξ_m represents the type of consumers who will use the platform services. $W(\nu)$ is the type of consumer who purchase products with quality ν. c_p is platform's entry cost.

3.4.2 Consumer

In this model with the BNPL platform, consumers are now allowed to borrow. Their new budget constraint becomes $p_{mi} \leqslant kW$ if purchasing via the platform. Their objective function if using the platform is then

$$\max_{\nu} \qquad u_{mi} = \nu - p_{mi}(\nu) \tag{3.11}$$

$$s.t. \quad p_{mi}(\nu) \leqslant kW \tag{3.12}$$

where u_{mi} is the utility from using the platform. The first order condition is

$$1 = (\lambda_{mi} + 1) p'_{mi}(\nu) \tag{3.13}$$

where λ_{mi} is the multiplier for the budget constraint.

3.4.3 Firms

If a firm decides to join the platform, then it needs to pay θ fraction of the revenue to the platform. Meanwhile, the partner firm is able to get private data $D_{pv}(n_{mi})$ from BNPL firms, where n_{mi} is the number of firms in the platform in the monopoly case. Denote the total number of firms by n_m. Then firms' objective function in the platform is

$$\pi_{mi}(\nu) = \max_{q_{mi}} (1-\theta) p_{mi} q_{mi} - \frac{1}{2} \nu^4 q_{mi}^2 \phi(\alpha n_m + \varphi n_{mi}) - c_e \tag{3.14}$$

Notice that the total data stock of partner firms consists of two parts: public data αn_m and private data φn_{mi}. Denote $D_{mp} = \alpha n_m + \varphi n_{mi}$ by the total data stock of firms in the platform.

[9] All the subscript m represents monopoly case

In other words, firms in the platform trade part of its revenue for lower production cost. The first order condition for production scale q_{mi} is

$$q_{mi} = \frac{(1-\theta)p_{mi}}{\nu^4 \phi(D_{mi})} \tag{3.15}$$

Equation 3.15 tells us the platform can affect firms' scale in two possible ways. Performance fee θ can decrease firms' production scale. However, due to the extra private data, the production cost decreases. This channel helps increase firms' scale. Similar to the benchmark case, firms producing higher quality goods also have smaller scale.

3.4.4 Partial Equilibrium

Before we show the results for the general equilibrium, this subsection starts to analyze the result in a partial equilibrium where θ is taken as fixed. The intuition for the partial equilibrium can be easily generalized to the general equilibrium (θ is endogenous). Using the free entry condition, we can derive the price schedule in the market.

Proposition 36. *The price schedule in the platform is:* $p_{mi} = c_{mi}\nu^2$, *where* $c_{mi} = \frac{\sqrt{2c_e\phi(D_{mi})}}{1-\theta}$. *The price schedule* p_{mo} *out of the platform is:* $p_{mo} = c_{mo}\nu^2$, *where* $c_{mo} = \sqrt{2c_e\phi(D_{mo})}$, *and* $D_{mo} = \alpha n_m$.

Proof see appendix.

From Proposition 36, we know the only difference between the two price schedules is the coefficient in front of the quality. So there are two possible cases. One is $c_{mi} > c_{mo}$. In this case, price schedule in the platform is steeper. This is the only case where in-platform firms and out-of-platforms can coexist. The other is that $c_{mi} \leqslant c_{mo}$. In this case the price schedule in the platform is flatter than the price schedule out of the platform. This then immediately implies that there cannot exist any out-of-platform firms. We study the former case first and then present the latter case.

3.4.4.1 Coexistence of In-Platform Firms and Out-of-Platform Firms

We first derive consumers' optimal choices. A quick observation is that because price schedule is flatter out of platform, unconstrained consumers will always prefer to purchase from an out-of-platform merchant. It implies that the unconstrained optimal product must be sold by an out-of-platform merchant. The details are stated in the following proposition.

Proposition 37. *There exist two cutoffs W_m^c, and W_m^p such that*

1. *If $W \leqslant W_m^p$, type W consumer uses BNPL services and chooses quality $\nu_{mi}^c = \sqrt{\frac{kW}{c_{mi}}}$.*
2. *If $W \in (W_m^p, W_m^c)$, type W consumer chooses quality $\nu_m^c = \sqrt{\frac{W}{c_{mo}}}$.*
3. *If $W \geqslant W_m^c$, type W consumer chooses quality $\nu_m^u = \frac{1}{2c_{mo}}$.*

where $W_m^c = \frac{1}{4c_{mo}}$, and $W_m^p = \frac{1}{(k-1)^2}\left[\sqrt{\frac{k}{c_{mi}}} - \sqrt{\frac{1}{c_{mo}}}\right]^2$.

Proof see appendix.

Proposition 37 shows us that only poor people use BNPL services. This is consistent with what we see in practice. Moreover, we can divide consumers into three groups. For those who are extremely rich, they will directly purchase from an out-of-platform merchant and pick the unconstrained optimal type of products. For people who are relatively rich, they also purchase from out-of-platform merchants. However, they cannot afford the unconstrained optimal type and will buy the goods with the highest quality that they can afford. The last group consists of poor consumers. They prefer to borrow money from the platform. The difference is that even with loans, they are still constrained and cannot afford their favorite type of products offered in the platform. In general, there can be another type of consumers. They are borrowing but their budget constraint is slack. It requires k to be very large. We pick the upper limit $\bar{k}$ to be relatively small so that this type of consumers cannot exist. Based on Proposition 37, we can then derive firms' production scale and firm distribution.

Proposition 38. *Firm distribution $G_m(\nu)$ has one atoms at ν_m^u. The mass at ν_m^u is $\frac{W_m^c(1-F(W_m^c))}{2c_e}$.*

For any quality $\nu < \nu^c_{mi}$, G_m has density g_m such that $g_m(\nu) = \frac{k(1-\theta)W(\nu)f(W(\nu))}{2c_e}$. For quality $\nu \in (\nu^c_{mi}, \nu^u_m)$, $g_m(\nu) = \frac{W(\nu)f(W(\nu))}{2c_e}$.

Proof see appendix.

The difference between firm distributions in two models lies in the density of firms in the platform. With the BNPL platform, there exists an extra coefficient $k(1-\theta)$ in the density. If $k(1-\theta) > 1$, then the density will be bigger after the introduction of the BNPL platform. As we will see later, $k(1-\theta)$ plays a key role in determining the impact of BNPL platforms on consumer welfare. Using Proposition 38, we can easily deduce the number of firms in the platform n_{mi} and total measure of firms in the market n_m:

$$n_{mi} = \int_{\underline{W}}^{W^p_m} \frac{k(1-\theta)Wf(W)}{2c_e} dW \tag{3.16}$$

$$n_m = \int_{\underline{W}}^{W^p_m} \frac{k(1-\theta)Wf(W)}{2c_e} dW + \int_{W^p_m}^{W^c_m} \frac{Wf(W)}{2c_e} dW + \frac{W^c_m(1-F(W^c_m))}{2c_e} \tag{3.17}$$

Equation 3.17 tells us that $n_m < n_b$ if and only if $k(1-\theta) < 1$. In fact, if $k(1-\theta) < 1$, we have

$$n_m < \int_{\underline{W}}^{W^c_m} \frac{Wf(W)}{2c_e} + \frac{W^c_m(1-F(W^c_m))}{2c_e} \tag{3.18}$$

If we want to find a firm measure n such that n equals to the right hand side of the inequality 3.18, the only solution is n_b. The reason is that W^c_b and W^c_m only relate to the total number of firms in the market n. This result implies that if $k(1-\theta) < 1$, out-of-platform merchants charge a higher price than that in the benchmark case. It will hurt those rich people. The following proposition shows us in detail how the introduction of a BNPL platform can affect consumer welfare.

Proposition 39.

1. *Suppose $k(1-\theta) > 1$. All consumers' welfare increases.*

2. *Suppose $k(1-\theta) < 1$. There exists a cutoff W_m^b, such that type $W < W_m^b$ consumers benefits from BNPL services and other consumers get hurt.*

Proof see appendix.

Proposition 39 tells us that the BNPL platform may not benefit everyone. In particular, if the platform cannot provide services at low performance fee, it can hurt those rich consumers. The existence of BNPL platforms decreases the total number of firms in the market and increases out-of-platform merchants' production cost. As a consequence, out-of-platform price schedule has to go up. Now let's see how the platform may increase everyone's welfare. Suppose $k(1-\theta) > 1$. Now poor consumers choose quality $\nu_{mi}^c = \sqrt{\frac{k(1-\theta)W}{\sqrt{2c_e\phi(D_{mi})}}}$. Higher $k(1-\theta)$ means that poor consumers are now able to purchase products with higher quality. Then firms' production scale will be

$$q_{mi} = \frac{(1-\theta)kW}{\phi(D_{mi})} \times \frac{2c_e\phi(D_{mi})}{k^2(1-\theta)^2W^2} = \frac{2c_e}{k(1-\theta)W} \tag{3.19}$$

Equation 3.19 then shows us firms cut production scale because it is too costly to produce goods with higher quality at previous scale. Notice that there are two effects here. One is that products' higher quality forces firms to shrink scale. The other is that higher prices encourage firms to increase scale. However, the dominant force is the former one because the increase of price cannot compensate the increase of the cost. Because consumers' demand is fixed, that means more firms have to enter the market and fill the demand gap. That's why the total number of firms goes up if $k(1-\theta) > 1$. We call this effect **Quality Shifting Effect**. In particular, we call an effect improving constrained consumers' choice of quality **Quality Improvement Effect** and call an effect lowering constrained consumers' choice of quality **Quality Deterioration Effect**. According to Proposition 39, the quality improvement effect dominates when $k(1-\theta) > 1$ and the quality deterioration effect dominates when $k(1-\theta) < 1$.

Now we know that a BNPL platform can benefit consumers both through lending money and through lowering prices by providing firms private data. When will we see out-of-platform firms and in-platforms coexist in a market? The answer depends on the data provision capabilities of BNPL platforms. Recall that the only way we can see both out-of-platform merchants and in-platform merchants is that out-of-platform merchants can offer lower prices.

Proposition 40. *There exist φ_0 and b_0 such that if $\varphi < \varphi_0$ and $k(1-\theta) > b_0$, in-platform merchants can coexist with out-of-platform merchants.*

Proof see appendix.

Two conditions are imposed in Proposition 40. The first condition ensures that data processing ability of the platform is not extremely powerful so that in-platform merchants will not charge lower prices. The intuition is that of φ is big, then in-platform merchants will have too much data and their marginal cost will be very small. The second condition guarantees that the platform can indeed provide benefits to constrained consumers. Recall that if k is bigger, then consumers' budget constraint will be more slack. So the size of k measures the benefit brought to consumers. However, these constrained consumers need to bear higher price owing to the performance fee paid to the platform, which is θ. In order to have some constrained consumers prefer to use the BNPL service, the benefit from borrowing must compensate the loss from performance fee. That is why $k(1-\theta)$ has to be large.

3.4.4.2 No Out-of-Platform Firms

In this part, we discuss a situation where the platform does so well and no firms can survive out of the platform. As discussed at the beginning of Subsection 3.4.4, the necessary and sufficient condition is $c_{mi} < c_{mo}$. The key result for this special case is that even if the platform can be very efficient in data provision, it can still hurt rich consumers' welfare. To show the intuition, we first solve consumers' problem and then characterize firm distribution in this case.

Proposition 41. *There exists a cutoffs W^c_{mp} such that*

1. *If $W < W^c_{mp}$, type W consumer chooses quality $\nu^c_{mi} = \sqrt{\frac{kW}{c_{mi}}}$.*
2. *If $W \geqslant W^c_{mp}$, type W consumer chooses quality $\frac{1}{2c_{mi}}$.*

where $W^c_{mp} = \frac{1}{4c_{mi}}$.

Proof see appendix.

Similarly, unconstrained consumers choose the same quality but constrained consumers can only purchase the most expensive product that they can afford. By the same argument as in Proposition 33, we can derive the firm distribution G_{mp} in this case. For a product with quality $\nu < \frac{1}{2c_{mi}}$, the demand density is $f(W(\nu))$. Single firm's production scale is $\frac{2c_e}{k(1-\theta)W(\nu)}$. So the G_{mp} has density $g_{mp}(\nu) = \frac{k(1-\theta)Wf(W(\nu))}{2c_e}$. For the unconstrained optimal quality $\frac{1}{2c_{mi}}$, total demand is $1 - F(W^c_{mp})$. Single firm's production scale is $\frac{2c_e}{k(1-\theta)W^c_{mp}}$. So G_{mp} has an atom at $\frac{1}{2c_{mi}}$ with mass $\frac{k(1-\theta)W^c_{mp}(1-F(W^c_{mp}))}{2c_e}$. Then we can get the total number of firms in the market in this special case

$$n_{mp} = \int_{\underline{W}}^{W^c_{mp}} \frac{k(1-\theta)Wf(W)}{2c_e} dW + \frac{k(1-\theta)W^c_{mp}}{2c_e}(1 - F(W^c_{mp})) \tag{3.20}$$

Comparing Equation 3.9 and Equation 3.20, we can find two differences. One is the extra coefficient $k(1-\theta)$ at the right hand side in Equation 3.20. The other is the difference between cutoffs W^c_b and W^c_{mp}. The total number of firms increases in $k(1-\theta)$ and the cutoff. The cutoff W^c_{mp} itself is positively related to platforms' data provision ability φ and $k(1-\theta)$. It then implies that in order to increase the number of firms in the market, the platform needs to have strong data provision capabilities but charge a low performance fee. The next proposition summarizes the result.

Proposition 42.

1. *If φ, θ are small and $k(1-\theta) < 1$, then $n_{mp} < n_b$.*
2. *If $\varphi > \alpha$ and $k(1-\theta) > 1$, then $n_{mp} > n_b$.*

Proof see appendix.

The impact of the platform on consumers' welfare can then derived from Proposition 42 directly.

Corollary 43.

1. *If φ, $1-\theta$ are small and $k(1-\theta) < 1$, then the platform only benefits poor consumers.*
2. *If $\varphi > \alpha$ and $k(1-\theta) > 1$, then all consumers' welfare increases with the platform.*

Corollary 43 again shows that the BNPL platform may not benefit all consumers. In general, those poor consumers can enjoy higher utility after the introduction of the platform. The intuition is the same as in Proposition 39. However, if the platform cannot be productive in data provision and charges a high fee, it still hurts rich people. The intuition is slightly different from the one in Proposition 39. Here, all the consumers are shopping via the platform. The difference is that rich people do not care about lending. What they really want is a flatter price schedule. Comparing to the benchmark case, two factors can increase the price schedule in the platform. One is the performance fee. The other is the shrinking number of firms in the market. In particular, as long as θ is big, and φ is small, only a few firms will enter the industry. Both the performance fee and the smaller number of firms increase the cost of firms and then push up the price schedule. Rich people's welfare then decreases as a consequence.

We conclude Subsection 3.4.4.2 with a sufficient condition that shows when the BNPL platform always dominates out-of-platform merchants.

Proposition 44. *Given the platform's data provision ability φ, there always exists a θ_0 such that if $\theta < \theta_0$, all the consumers will choose to use the BNPL platform.*

Proof see appendix.

Proposition 44 tells us that there are two ways that a BNPL platform can beat out-of-platform merchants. One is increasing data provision ability but maintaining a relatively low performance fee. The other is charging a low performance fee and still provides strong data support (φ is big). In sum, all the cost resulted from the platform comes from performance fee. To make constrained consumers better, the platform can lend more (k increases). To make unconstrained consumers better, the platform must provide enough private data so that firms' cost goes down and then prices go down as well.

3.4.5 General Equilibrium

In this subsection, we will formally study the monopolistic platform's decision. Our goal is to check which of the two cases in the partial equilibrium will the platform choose. We first compute the platform's profit and present its objective function.

Suppose a constrained consumer of type W purchases a product with quality ν. Then the revenue the platform can get from products with quality ν is

$$\pi_{pm}(\nu) = \theta p_{mi}(\nu) q_{mi}(\nu) = \frac{\theta(1-\theta)p_{mi}^2}{\nu^4 \phi(D_{mi})} = 2\frac{\theta}{(1-\theta)}c_e \tag{3.21}$$

The second equality uses Equation 3.15. The third equality uses the fact that $p_{mi}(\nu) = kW$ and Proposition 37. Notice that Equation 3.21 shows that the platform's profit from one type of products has nothing to do with consumer's type. It implies that if unconstrained consumers who choose quality ν^* also use the platform, then the profit from the product with quality ν^* should also be $\frac{2\theta c_e}{1-\theta}$. Therefore, the platform chooses performance fee θ to maximize the following objective function

$$\Pi_{pm} = \max_{\theta} 2\frac{\theta}{1-\theta}c_e F(W_m^p) - c_p \tag{3.22}$$

Because each type of consumers provide $2\frac{\theta}{1-\theta}c_e$ to the platform and the measure of consumers using the platform is $F(W_m^p)$, the platform's total revenue should be the first term in the Equation 3.22. The second term is the entry cost. Obviously, the platform faces one trade-off. Does it want higher performance fee by picking a larger θ or a larger number of partner firms by picking a smaller θ? We want to know when $k(1-\theta) > 1$. Because only if this condition is satisfied, the platform can benefit everyone. The result below shows us the answer depends on the platform's lending capability k.

Proposition 45. *There exist a_1, a_2, φ_1, k_1 and k_2 such that if $a \notin [a_1, a_2]$ and $\varphi > \varphi_1$,*

1. $k(1-\theta) < 1$ *if* $k < k_1$
2. $k(1-\theta) > 1$ *if* $k > k_2$

Proof see appendix.

Proposition 45 provides sufficient conditions on when the platform can benefit everyone. The key lies in platform's lending capability k. Actually, we can see even if the platform has extremely strong data provision ability, it may still hurt unconstrained consumers if k is small. Notice that if the platform only provides data and does not make loans ($k = 1$), the existence of platform always lowers the total number of firms in the market. The reason is that the performance fee always increases firms' production scale and private data provision does not change firms' scale. The impact of performance fee on production scale is from the quality shifting effect. High performance fee θ increases the product prices and force consumers to purchase low quality products. Because it is cheaper to produce low quality products, firms' production scale goes up. Private data provision has two effects on firms' production decisions. On one hand, it increases scale because marginal cost goes down. On the other hand, it decreases scale because it flattens the price schedule and makes consumers to purchase products with higher quality. However, it is more costly to produce high quality goods, so firms cut production. It turns out these two effects cancel out perfectly. In sum, without lending, the quality deterioration effect dominates. In order to increase the total number of firms in the market, we need to introduce quality improvement effect. The only way to do this is to increase k. Large k will loosen constrained consumers' budget constraint and let them purchase products with higher quality.

3.5 Model with Two BNPL Platforms

In this section, we present a model with two BNPL platforms competing with each other in a Betrand game. The main question to answer is whether competition between platforms can improve consumers' welfare? Firms' problem and consumers' problem are the same as in the monopoly case. We will omit that and only present platforms' problem in the follows. Moreover, we focus on the equilibrium where in-platform merchants and out-of-platform merchants coexist for the rest of this section.

3.5.1 BNPL Platforms

The two platforms $i = 1, 2$ compete with each other in a Betrand price competition. They choose performance fee θ_i simultaneously. If $\theta_1 < \theta_2$, then platform 1 will get all the merchants who want to join. If $\theta_1 = \theta_2$, then we assume that each merchant chooses one of the platforms to join with probability $\frac{1}{2}$. Similarly, the consumer will also choose to use one of the platforms with probability $\frac{1}{2}$. In other words, two platforms will evenly split consumers and merchants who want to use the services.

In equilibrium, because of competition, both platforms will earn zero profit. That means performance fee θ must satisfy the following condition

$$c_e \frac{\theta}{1-\theta} F(W_d^p) = c_p \tag{3.23}$$

The left hand side of Equation 3.23 is the platform's profit. W_d^p is the wealth cutoff similar to W_m^p such that only consumers of type $W \leqslant W_d^p$ will shop via platforms. The total number of consumers who use platforms will be $F(W_d^p)$. Due to the competition, each platform will get half of the consumers, which is $\frac{1}{2}F(W_d^p)$. The profit from each type of consumers is $\frac{2\theta}{1-\theta}c_e$, which is the same as in Equation 3.22.

3.5.2 Equilibrium and Welfare Implication

By the similar argument in Subsection 3.4.4, the total number of firms in the market is the most important variable to focus on. Only if the introduction of two platforms can increase the total number of firms in the market, then all consumers can benefit from the platforms. The number of firms in the platforms n_{di} and the total number of firms in the market n_d satisfy the following conditions,

$$n_{di} = \int_{\underline{W}}^{W_d^p} \frac{k(1-\theta)Wf(W)}{2c_e} dW \tag{3.24}$$

$$n_d = \int_{\underline{W}}^{W_d^p} \frac{k(1-\theta)Wf(W)}{2c_e} dW + \int_{W_d^p}^{W_d^c} \frac{Wf(W)}{2c_e} dW + \frac{W_d^c(1-F(W_d^c))}{2c_e} \tag{3.25}$$

Equation 3.24 and Equation 3.25 are similar to Equation 3.16 and Equation 3.17. To understand this, notice that here there are still two wealth cutoffs W_d^p and W_d^c to characterize consumers' demand. In particular, if the consumer has endowment $W > W_d^c$, she will be unconstrained. For a firm in a platform i producing goods with quality ν, its production scale is $\frac{2c_e}{k(1-\theta)W(\nu)}$ by Equation 3.15. The demand density is $\frac{f(W(\nu))}{2}$ due to competition. So the density of firms producing goods with quality ν in the market should be $2\frac{k(1-\theta)Wf(W)}{4c_e} = \frac{k(1-\theta)Wf(W)}{2c_e}$. Then we get Equation 3.24 by taking integration. From Equation 3.25, we can see if and only if $k(1-\theta) > 1$, the total number of firms n_d can be larger than n_b in the benchmark case. The following proposition gives sufficient conditions on when $k(1-\theta)$ will be satisfied.

Proposition 46. *There exists c_0 such that if $c_p < c_0$, the inequality $k(1-\theta)$ will be satisfied and all consumers' welfare increases with the platforms.*

Proof see appendix.

Proposition 46 shows that in order to make sure that the platforms can help increase the number of firms in the market, the entry cost must be small. The intuition is that if entry cost decreases, competition requires both platforms to reduce performance fee θ. As a result, $k(1-\theta)$ goes up. Otherwise, platforms need a larger θ to maintain zero profit which then decreases the number of firms in the market. Similar to Proposition 39, if $k(1-\theta) < 1$, then only poor consumers can benefit from the BNPL services. The reason is that they are so constrained and borrowing via the platform can help increase their welfare.

Finally, we will compare this duopoly competition model to the monopoly model and check whether it really improves consumers' welfare. In general, competition between two platforms results in one trade-off. One is lowering the performance fee θ, which obviously benefits the consumers. However, such competition also has a negative impact on private data provision. Because each platform becomes smaller, the private data provision can be less efficient than in the monopoly case. It then increases in-platform merchants' marginal cost, pushes up the in-platform price schedule and lowers poor consumers' welfare. The following proposition then provides sufficient conditions on which effect will be the dominant force and shows when the introduction of competition is beneficial.

Proposition 47. *Denote θ_m and θ_d by the performance fee in the monopoly model and the duopoly model respectively, then*

1. *If $k(1-\theta_m) > 1$, $k(1-\theta_d) > 1$, $a \notin [a_1, a_2]$ and φ is sufficiently large, then all consumers' welfare is higher in the duopoly model.*
2. *If $k(1-\theta_m) < 1$, $k(1-\theta_d) < 1$, $a < a_1$ and c_p is large, then BNPL users get hurt but non-BNPL users get benefit in the duopoly model.*

Proof see appendix.

Proposition 47 discusses the impact of competition between platforms in two scenarios. The first scenario happens when platforms help constrained consumers to purchase higher quality products and increase the number of firms in the market. Notice that if private data provision ability φ is very large, split of partner firms due to competition should not have a big impact on platforms. It then implies that the price schedule in the platform in the duopoly model should be flatter than that in the monopoly model because the performance fee θ_d is smaller. So consumers who are using the BNPL services must get better in the duopoly model. Moreover, due to smaller performance fee, more consumers are willing to use the BNPL services in the duopoly model. By the assumption that platforms have quality improvement effect ($k(1-\theta_d) > 1$), the total number of firms in the market should also be higher in the duopoly model. Therefore, non-BNPL users' welfare increases as well in the duopoly model.

The second result analyzes the scenario where platforms have the quality deterioration effect ($k(1-\theta_m) > 1$, $k(1-\theta_d) > 1$). So their existence can decrease the number of firms in the market. When entry cost c_p is large, the difference between performance fees in two models should be small. Now the difference between slopes of two price schedules in the platform should only depend on data provision. Under the assumption a is small, what really matters is the number of firms in the platform because it is hard to get public data.

Due to the split of partner firms in the duopoly model, it indicates the number of firms in a single platform $\frac{n_{di}}{2}$ in the duopoly model should be smaller than n_{mi} in the monopoly model. Therefore, the price schedule in the monopoly model will be relatively flatter. We then know BNPL users' welfare is higher in the monopoly model. However, the total number of firms in the monopoly model is actually smaller. The reason is that more constrained consumers are willing to use the platform in the monopoly model: $W_m^p < W_d^p$. Such demand shifting will have the quality deterioration effect. Consumers are purchasing products with worse quality in comparison to the benchmark case. The total number of firms in the market has to decrease in the monopoly model. Therefore, non-BNPL users get benefit in the duopoly model. This result shows that the monopoly platform may only benefit consumers who are in the club at the expense of consumers out of the club.

3.6 Conclusion

This paper explores the impact of BNPL platforms on consumer welfare. We find the introduction of BNPL platforms does not always increase the welfare of all consumers. In particular, rich consumers who do not use the platforms can get hurt. The intuition is that if the lending capability of platforms is not strong, the platform fee can prevent new entrants. Fewer firms competing with each other has a negative impact on the market because less public data is available. Firms out of the platform then have higher production cost. Product prices out of the platform go up. As a consequence, consumers who are not using the platform services get hurt.

The second question that this paper tries to answer is whether platform competition can benefit everyone. The model predicts that under certain circumstances, platform competition can decrease the welfare of those consumers who use the BNPL services. In particular, if the entry cost is relatively high and it is hard to get public data, platform competition will make private data provision less efficient. The intuition is that the platforms are now partnering with fewer firms and it is harder to provide better private data support. This leads to higher production cost for in-platform merchants and higher in-platform prices. As a consequence, consumers who are using the BNPL services get hurt.

Appendix A. Proof

Proof of Lemma 1

Proof. This is the standard consumption problem with Dixit-Stiglitz utility. The first order condition(FOC) for c_i is

$$\rho(b_i c_i)^{\rho-1} b_i = \lambda p_i$$

where λ is the multiplier for budge constraint. Combining the FOC for c_i and c_j, we have

$$\left(\frac{b_i}{b_j}\right)^{\rho}\left(\frac{c_i}{c_j}\right)^{\rho-1} = \frac{p_i}{p_j}$$

Rearranging the above equation, we get

$$c_i = \left(\frac{b_i}{b_j}\right)^{-\sigma} c_j \left(\frac{b_j}{b_i}\right)^{1-\sigma}$$

where $\sigma = \frac{1}{1-\rho}$. Substituting the above relation into the budget constraint, we have

$$W_1 = P^{1-\sigma} p_j^{\sigma} c_j b_j^{1-\sigma}$$

where $P = \left(\int \left(\frac{p_i}{b_i}\right)^{1-\sigma} di\right)^{\frac{1}{1-\sigma}}$. Rearranging it, we get the desired demand function in the lemma.

□

Proof of Lemma 2

Proof. Substituting the demand function in Lemma 1 into the utility function for the bundle, $\left(\int (c_i b_i)^{\rho} \, di\right)^{\frac{\alpha}{\rho}}$, we can get the desired result. □

Proof of Lemma 3

Proof. The first order condition for W_1 is

$$\alpha W_1^{\alpha-1} P^{-\alpha} - \frac{b_0}{p_0} = 0$$

With some algebra, we have

$$W_1 = \alpha^{\frac{1}{1-\alpha}} (\frac{p_0}{b_0})^{\frac{1}{1-\alpha}} P^{-\frac{\alpha}{1-\alpha}}$$

□

Proof of Theorem 6

Proof. To prove the value function is well defined, we verify the Blackwell's condition. Denote the optimization operator in the firm's problem by T and suppose f and g are two value functions such that $f \leqslant g$. Then it is easy to observe that firms with value function g can adopt the same strategy as firms with value function f but achieve a higher payoff because $f \leqslant g$. So $Tf \leqslant Tg$. Because of the death rate δ, the discounting condition will be satisfied automatically. Therefore, the value function is well-defined. Next, we prove that the value function is increasing. Suppose that V is increasing. Then

$$\begin{aligned} TV(\Omega) &= \max_{k,z,\omega} E[p|I_{it}]Ak^d - r_k k - r_z z + r_\omega \omega + \beta(1-\delta)V(\Omega^{'}) \\ &= \varphi P^{\frac{\phi}{\sigma}}(m - \Omega^{-1} - \sigma_\epsilon^2)A^\rho k^{d\rho} - r_k k - r_z z + r_\omega \omega + \beta(1-\delta)V(\Omega^{'}) \end{aligned}$$

where $\phi = \frac{\rho}{1-\rho} - \frac{\alpha}{1-\alpha} > 0$ and $\varphi = (\alpha \frac{p_0}{b_0})^{\frac{1}{\sigma(1-\alpha)}}$.

If $\Omega_2 > \Omega_1$, we can pick $k(\Omega_2) = k(\Omega_1)$, $z(\Omega_2) = z(\Omega_1)$ and $\omega(\Omega_2) > \omega(\Omega_1)$ such that $\Omega'(\Omega_2) = \Omega'(\Omega_1)$. Under this choice, firm's profit is higher because its data sales goes up. So $TV(\Omega_2) > TV(\Omega_1)$ and the value function is increasing.

To prove the value function is concave, first observe that

$$\frac{\partial^2[\tau^2(\Omega + \sigma_\epsilon^{-2})^{-1}]}{\partial \Omega^2} = -\frac{2\sigma_\epsilon^6 \sigma_\eta^2 \tau^2}{((1+\Omega\sigma_\epsilon^2)\sigma_\eta^2 + \sigma_\epsilon^2\tau^2)^3} \leqslant 0$$

So the first constraint is concave in Ω. Additionally, one can check that if m is big enough, then $(m-\Omega^{-1}-\sigma_\epsilon^2)k^d$ will be concave. Take two initial precision levels Ω_1 and Ω_2 as given. Suppose that the corresponding optimal choices are (k_1, z_1, ω_1) and (k_2, z_2, ω_2). The convex combination of these two choices will be feasible if the initial precision is $\lambda\Omega_1 + (1-\lambda)\Omega_2$. From concavity of the constraint, we know $(\lambda\Omega_1+(1-\lambda)\Omega_2)'$ will be bigger than $\lambda\Omega_1' + (1-\lambda)\Omega_2'$. Moreover, the value of $\varphi P^{\frac{\phi}{\sigma}}(m-\Omega^{-1}-\sigma_\epsilon^2)A^\rho k^{d\rho} - r_k k - r_z z + r_\omega \omega$ when the firm uses the convex combination of two choices is bigger than the convex combination of the corresponding values from two choices. So if we assume that the value function V is concave, then we have shown that $TV(\lambda\Omega_1+(1-\lambda)\Omega_2) \geqslant \lambda V(\Omega_1) + (1-\lambda)V(\Omega_2)$. Therefore, the value function is concave. □

Proof of Lemma 5

Proof. By Corollary 4, we only need to compute the conditional expectation of b_i^ρ.

$$E(b_i^\rho|I_{it}) = E(m_i - (\tilde{\theta}_{it} + \epsilon_{it} - E(\tilde{\theta}_{it}|I_{it}))^2|I_{it}) = m_i - \Omega_{it}^{-1} - \sigma_\epsilon^2 \tag{A.1}$$

□

Proof of Theorem 7

Proof. Differentiate the objective function with respect to ω, and we have

$$(1-\delta)V'(\Omega') = \frac{r_\omega}{\zeta} \tag{A.2}$$

To show what types of firms will choose higher forecast precision Ω', we need to check how m and σ_ζ^{-2} can affect $V'(\Omega')$. We'll prove that the value function is supermodular in m and Ω. Similarly, one can show the value function is supermodular in σ_ζ^{-2} and Ω. To begin with, suppose $V_0(\Omega, m)$ is a supermodular function in m and Ω. Then we define $V_1(\Omega, m)$ to be the maximum profit for the firm by replacing the true value function V by V_0. So V_1 is actually the result from the first iteration of the value function iteration algorithm.

By envelop theorem, we know that $\frac{\partial V_1}{\partial m} = I(P)A^{\rho}k^{d\rho} + \frac{\partial V_0(\Omega')}{\partial m}$,where $I(P) = \varphi P^{\frac{\phi}{\sigma}}$ is an increasing function of price index P. If we know how Ω affects k, it will be sufficient to determine whether or not V_1 is supermodular in m and Ω. To verify that, we first rewrite the objective function. Suppose in equilibrium, optimal $\Omega^{'} = \Omega_j$. We can then substitute out ω and get the following equivalent optimization problem

$$\max_{k,z,\Omega_j} E[p|I]Ak^d - r_k k - r_z z + \frac{r_\omega}{\zeta}\left([\tau^2(\Omega+\sigma_\epsilon^{-2}) - 1 + \sigma_\eta^2]^{-1} + Ak^d z^{1-d}\sigma_\xi^{-2} - \Omega_j\right) + \beta(1-\delta)V(\Omega_j) \quad \text{(A.3)}$$

the first order conditions for k and z are

$$I(P)(m - \Omega^{-1} - \sigma_\epsilon^2)A^\rho d\rho k^{d\rho-1} + \frac{r_\omega}{\zeta}Ad(\frac{z}{k})^{1-d}\sigma_\xi^{-2} = r_k \quad \text{(A.4)}$$

$$\frac{r_\omega}{\zeta}A(1-d)(\frac{k}{z})^d\sigma_\xi^{-2} = r_z \quad \text{(A.5)}$$

Using Equation A.5, we can simplify Equation A.4,

$$I(P)(m - \Omega^{-1} - \sigma_\epsilon^2)A^\rho d\rho k^{d\rho-1} + \frac{r_z d}{1-d}\frac{z}{k} = r_k \quad \text{(A.6)}$$

From Equation A.5, we can get that $\frac{z}{k}$ keeps constant as Ω changes. Then by Equation A.6, we know k must increase in Ω because $(m - \Omega^{-1} - \sigma_\epsilon^2)$ increases in Ω. Together with the assumption that V_0 is supermodular in m and Ω, we get V_1 is supermodular in m and Ω. Because supermodularity is preserved under taking limit, the value function V will then be supermodular in m and Ω. It implies that $V'(\Omega')$ increases in m. Therefore, if m is big, the firm will want to get higher forecast precision. Otherwise, firms will only retain lower forecast precision. The result on τ_ξ can be derived in the similar way and will be omitted.

□

Proof of Theorem 8

Proof. We first need to prove that the price index and the number of firms are bounded from above. It will imply that the space of sequences of price indexes, the space of sequences of number of firms and the space of sequences of number of new entrants are compact. We define a correspondence T as follows:

$$T((p_0, p_1, ..), (r_{\omega 0}, r_{\omega 1}, ...), (n_0, n_1, ...), (e_0, e_1, ...)) = (SP, SR, SN, SE) \tag{A.7}$$

where SP is a set of sequences of price indexes solving the following equation

$$\iint V(\Omega_0, m, \sigma_{\xi}^{-2}) dF_m dF_{\xi} = c \tag{A.8}$$

SR is a set of sequences of data prices solving the following equation.

$$\iiint \omega(\Omega, m, \tau_{\xi}) d\lambda_t dF_m dF_{\xi} = 0 \tag{A.9}$$

SN is a set of sequences of number of firms such that clearing the consumption goods market given the sequence of price indexes and SE is a set of sequences of number of new entrants satisfying $n_t = (1-\delta)n_{t-1} + e_t$.

Because the value function and the data trading size ω are continuous, from Equation A.8 and Equation A.9, we know that SP and r_ω must be bounded.

Then we'll show the boundedness of the number of firms. From Equation A.4 the first order condition for capital choice k is

$$\frac{\partial E(p|I)}{\partial k} Ak^d + E(p|I) dAk^{d-1} - r_k + Adr_\omega \tau_\xi (\frac{z}{k})^{1-d} = 0$$

For simplicity, we ignore the subscript for each firm. Substituting in the formula of expected inverse demand function in Lemma 5, we can rearrange the first order condition as follows,

$$A^\rho \varphi P^{\frac{\phi}{\sigma}} k^{d\rho - 1} (m - \Omega^{-1} - \sigma_\epsilon^2) d\rho + Adr_\omega \tau_\xi (\frac{z}{k})^{1-d} = r_k \tag{A.10}$$

where $\phi = \frac{\rho}{1-\rho} - \frac{\alpha}{1-\alpha} > 0$ and $\varphi = (\alpha \frac{p_0}{b_0})^{\frac{1}{\sigma(1-\alpha)}}$. From Equation A.5, we can solve $\frac{z}{k}$:

$$\frac{z}{k} = \left(\frac{r_\omega A(1-d)\tau_\xi}{r_z \zeta} \right)^{\frac{1}{d}} \tag{A.11}$$

Substituting Equation A.11 into Equation A.10, we can solve k

$$k = \left[\frac{A^{\rho} d\rho\varphi(m - \Omega^{-1} - \sigma_{\epsilon}^{2})}{r_k - \Gamma} \right]^{\frac{1}{1-d\rho}} P^{\frac{\phi}{\sigma(1-d\rho)}} \tag{A.12}$$

where $\Gamma = \frac{r_z d}{(1-d)} \left(\frac{r_{\omega} A(1-d)\Xi_m}{r_z \zeta} \right)^{\frac{1}{d}}$. Using Equation A.12 and Lemma 5, we can derive the monopolistic price for each firm,

$$p = \varphi P^{\frac{\phi}{\sigma}} [m - (\tilde{\theta} + \epsilon - \theta)^2] A^{-\frac{1}{\sigma}} \left[\frac{A^{\rho} d\rho\varphi(m - \Omega^{-1} - \sigma_{\epsilon}^{2})}{r_k - \Gamma} \right]^{-\frac{d}{(1-d\rho)\sigma}} P^{-\frac{d\phi}{\sigma^2(1-d\rho)}} \tag{A.13}$$

which simplifies to

$$p = \varphi^{1-\frac{d}{\sigma(1-d\rho)}} A^{-\frac{1}{\sigma(1-d\rho)}} [m - (\tilde{\theta} + \epsilon - \theta)^2] \left[\frac{d\rho(m - \Omega^{-1} - \sigma_{\epsilon}^{2})}{r_k - \Gamma} \right]^{-\frac{d}{\sigma(1-d\rho)}} P^{\frac{\phi}{\sigma}(1-\frac{d}{\sigma(1-d\rho)})} \tag{A.14}$$

Observe that

$$1 - \frac{d}{\sigma(1-d\rho)} = \frac{\sigma - \sigma d\rho - d}{\sigma(1-d\rho)} = \frac{\sigma(1-d\rho)}{\sigma(1-d\rho)} > \frac{\sigma(1-\rho) - d}{\sigma(1-d\rho)} = \frac{1-d}{\sigma(1-d\rho)} > 0 \tag{A.15}$$

Then we can substitute the above inequality into the definition for price index and get,

$$P = \left[\int (\frac{p_i}{b_i})^{1-\sigma} di \right]^{\frac{1}{1-\sigma}} = P^{\frac{\phi}{\sigma}(1-\frac{d}{\sigma(1-d\rho)})} \varphi^{1-\frac{d}{\sigma(1-d\rho)}} A^{-\frac{1}{\sigma(1-d\rho)}} (d\rho)^{-\frac{d}{\sigma(1-d\rho)}} \left[\int \left(\frac{1}{r_k - \Gamma_i} \right)^{\frac{d(\sigma-1)}{\sigma(1-d\rho)}} (m_i - \Omega_i^{-1} - \sigma_{\epsilon}^{2})^{1+\frac{d(\sigma-1)}{\sigma(1-d\rho)}} di \right]^{\frac{1}{1-\sigma}} \tag{A.16}$$

Equivalently, we can rewrite the price index in the following way

$$P = nP^{\frac{\phi}{\sigma}(1-\frac{d}{\sigma(1-d\rho)})} \varphi A^{-\frac{1}{\sigma}} \left[\int \left(\frac{A^{\rho} d\rho\varphi}{r_k - \Gamma} \right)^{\frac{d(\sigma-1)}{\sigma(1-d\rho)}} (m - \Omega^{-1} - \sigma_{\epsilon}^{2})^{1+\frac{d(\sigma-1)}{\sigma(1-d\rho)}} d\lambda dF_m dF_{\xi} \right]^{\frac{1}{1-\sigma}} \tag{A.17}$$

Given bounded price index P and data price r_{ω}, Ω^{-1} is always bounded. We can immediately get that the number of firms in the market n must be bounded. Based on this, we can derive that SN and SE must be bounded and thus compact. Therefore, the correspondence T we defined is a single-valued continuous map. Its domain and range are both compact and convex. By Fan-Glicksberg's fixed point theorem, we can prove the existence of the competitive equilibrium. □

Proof of Theorem 9

Proof. Firstly, we can determine price index P from the zero entry condition in Equation A.8. Because value function in strictly increasing in P, Equation A.8 has only one solution. Suppose n is unique. Then from Equation 2.12, we can solve λ and the solution is unique. From the definition of stationary equilibrium, $e = \delta n$, $F_m^m = nF_m$ and $F_\xi^m = nF_\xi$. From the first order conditions, Equation A.4 and Equation A.5, we know that k and z have unique solutions. Therefore, to finish the proof, we only need to prove that n is unique. Notice that n can be determined from the Equation A.17. Because P and λ are unique, the solution of n must be unique as well. □

Proof of Theorem 10

Proof. Look at the extreme case where m and τ_ξ are the same for all firms and $f(\Omega) = 0$. Then given Ω_0, as long as τ_ξ is relatively big, it is easy to note that $Ak^dz^{1-d}\tau_\xi > \Omega_0$ because $Ak^dz^{1-d}\tau$ increases in τ_ξ. So if there is no data nonrivalry: $\kappa = 1$. Then dynamics must be: 1.) new firms enter with initial precision Ω_0 and purchase data; 2) firms sell data starting at $t = 1$ and reach steady state with precision Ω_1 since then. Now if $\kappa < 1$, we show that it is impossible for firms to first sell data and then purchase data in the next period. Suppose that a firm sells data at $t = 1$ and purchase data at $t = 2$. Then, we know $Ak_1^dz_1^{1-d}\tau_\xi > \Omega_2$ and $Ak_2^dz_2^{1-d}\tau_\xi < \Omega_1$, where we use the fact that firms which buy(sell) data have the same precision $\Omega_2(\Omega_1)$. Recall that $\Omega_1 > \Omega_2$. The Envelope condition of the problem is,

$$r_\omega = (1-\delta)\varphi P^{\frac{\phi}{\sigma}}\Omega^{-2}A^\rho k^{d\rho}\zeta + (1-\delta)r_\omega f'(\Omega) \tag{A.18}$$

So

$$\frac{\Omega_1^2}{\Omega_2^2} = \left(\frac{k_1}{k_2}\right)^{d\rho}\frac{1-(1-\delta)f'(\Omega_2)}{1-(1-\delta)f'(\Omega_1)}\kappa \tag{A.19}$$

It is sufficient to show that $\frac{\Omega_2}{\Omega_1} < \frac{Ak_1^d z_1^{1-d}\tau_\xi}{Ak_2^d z_2^{1-d}\tau_\xi}$ such that the conditions $Ak_1^d z_1^{1-d}\tau_\xi > \Omega_2$ and $Ak_2^d z_2^{1-d}\tau_\xi < \Omega_1$ cannot hold simultaneously. Using Equation A.11,

$$\frac{Ak_1^d z_1^{1-d}\tau_\xi}{Ak_2^d z_2^{1-d}\tau_\xi} = \frac{k_1}{\kappa k_2} \tag{A.20}$$

Combining Equation A.19, we can get the following equivalent condition

$$\frac{\Omega_2}{\Omega_1} < \frac{Ak_1^d z_1^{1-d}\tau_\xi}{Ak_2^d z_2^{1-d}\tau_\xi} \tag{A.21}$$

$$\Leftrightarrow \quad \left(\frac{k_2}{k_1}\right)^{\frac{d\rho}{2}} \left(\frac{1-(1-\delta)f'(\Omega_1)}{1-(1-\delta)f'(\Omega_2)}\right)^{\frac{1}{2}} \left(\frac{1}{\kappa}\right)^{\frac{1}{2}} < \frac{k_1}{\kappa k_2} \tag{A.22}$$

$$\Leftrightarrow \quad \left(\frac{k_1}{k_2}\right)^{1-\frac{d\rho}{2}} > \left(\kappa \frac{1-(1-\delta)f'(\Omega_1)}{1-(1-\delta)f'(\Omega_2)}\right)^{\frac{1}{2}} \tag{A.23}$$

$$\Leftrightarrow \quad \left(\frac{(r_k-\Gamma_2)(m_1-\Omega_1^{-1}-\sigma_\epsilon^2)}{(r_k-\Gamma_1)(m_2-\Omega_2^{-1}-\sigma_\epsilon^2)}\right)^{\frac{1-\frac{d\rho}{2}}{1-d\rho}} > \left(\kappa \frac{1-(1-\delta)f'(\Omega_1)}{1-(1-\delta)f'(\Omega_2)}\right)^{\frac{1}{2}} \tag{A.24}$$

where we use Equation A.12 in Inquality A.24.

By the definition of Γ, we know $\Gamma_2 = \frac{r_z d}{1-d}(r_\omega(1-d)A\tau_\xi)^{\frac{1}{d}} < \frac{r_z d}{1-d}(r_\omega(1-d)A\tau_\xi\kappa^{-1})^{\frac{1}{d}} = \Gamma_1$. LHS of Inequality A.24 is always strictly bigger than 1. Suppose that τ_ξ are big enough or $r_z issmall$. Then RHS of Inequality A.24 will be smaller than 1 because precision Ω will be high and the inequality will be satisfied. Therefore, as long as τ_ξ is relatively big or r_z is small, if firms start to sell, then they always sell in the next periods. This finishes the proof for the extreme case where there is only one type of firms. By continuity, we can always easily derive that if τ is small enough, σ_η is big enough and the dispersions of F_ξ and F_m are small, then the result still hold.

Moreover, we can get a useful lemma from the proof, which can be used in later proofs.

Lemma 48. *Under the conditions as in Theorem 10, in the stationary equilibrium,* $Ak_t^d z_t^{1-d}\tau_\xi > \Omega_{t+1}$*, for any* $t \geqslant 1$.

□

Proof of Proposition 11

Proof. This is proved at the proof of Theorem 7. See Equation A.5 □

Proof of Proposition 12

Proof. We already know that k increases in Ω from Equation A.12. Then from Proposition 11, z must also increase in Ω. □

Proof of Proposition 13

Proof. From Equation A.5, we know that $\frac{z}{k}$ is not affected by m. Then we claim that capital k increases in m. With this claim, we immediately have that z will increase in m as well. To prove the claim, notice that from Equation A.6, we have the term in the LHS must remain constant. Because $d\rho - 1 < 1$, capital investment k must increase. □

Proof of Proposition 14

Proof. From Equation A.5, we know that $\frac{z}{k}$ increases τ_{ξ}. Then we claim that capital k increases in τ_{ξ}. With this claim, we immediately have that z will increase in m as well. To prove the claim, notice that from Equation A.6, we have the term in the LHS must remain decrease. Because $d\rho - 1 < 1$, capital investment k must increase. □

Proof of Proposition 16

Proof. From Lemma 5, we know

$$PPU_i = \varphi P^{\frac{\phi}{\sigma}} c_i^{-\frac{1}{\sigma}} \tag{A.25}$$

Obviously, PPU decreases in k. Then from Proposition 12, we can get that PPU decreases in firm size. □

Proof of Proposition 17

Proof. From Equation A.6, we can immediately have that $\frac{z}{k}$ does not change with respect to m. Then we can rewrite the first order condition for capital as

$$E(p)Adk^{d-1}(1-\frac{d}{\sigma})+\frac{r_z d}{1-d}\frac{z}{k}=r_k \tag{A.26}$$

The first term in the above must remain constant as m changes. We know that k increases in m. So $E(p)$ will decrease in m. □

Proof of Proposition 18

Proof. By Lemma 5 and Equation A.12, we can get the expected price will be proportional to the following formula

$$\mathrm{E}[p_i] \propto P^{\frac{\phi}{\sigma}(1-\frac{d}{\sigma(1-d\rho)})}\left(\frac{m_i-\Omega_i^{-1}-\sigma_\epsilon^2}{r_k-\Gamma}\right)^{-\frac{d}{\sigma(1-d\rho)}}(m_i-\Omega_i^{-1}-\sigma_\epsilon^2) \tag{A.27}$$

The constant term we ignored do not contain τ_ξ. The first part of the proposition is easy. If two firms are both new, both firms have the same forecast precision Ω_0. The only term affecting the expected price is Γ. In Proposition 14, we know $\frac{z}{k}$ increases in τ_ξ. Notice that Γ increases in $\frac{z}{k}$. So Γ increases in τ_ξ. Equation A.27 tells us $\mathrm{E}[p_i]$ decreases in Γ. Therefore, expected price decreases in τ_ξ. This finishes the first part.

To prove the second part of the result, we differentiate the RHS of Equation A.27 with respect to τ_ξ,

$$\frac{\partial p_i}{\partial \tau_\xi} \propto (r_k-\Gamma_i)^{\frac{d}{\sigma(1-d\rho)}-1}(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)^{-\frac{d}{\sigma(1-d\rho)}}\left[-\frac{d}{\sigma(1-d\rho)}(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)\frac{\partial \Gamma_i}{\partial \Xi_m}+(1-\frac{d}{\sigma(1-d\rho)})(r_k-\Gamma_i)\Omega_i^{-2}\frac{\partial \Omega_i}{\partial \tau_\xi}\right] \tag{A.28}$$

where $\frac{\partial \Gamma_i}{\partial \Xi_m}=\frac{1}{d}\frac{\Gamma}{\tau_\xi}>0$. Then we compute $\frac{\partial \Omega_i}{\partial \tau_\xi}$. From Equation A.11, we have

$$\frac{\partial \frac{z}{k}}{\partial \tau_\xi}=\frac{1}{d}\frac{\frac{z}{k}}{\tau_\xi} \tag{A.29}$$

Differentiating both sides of Equation A.10 with respect to τ_ξ, we have

$$\varphi P^{\frac{\phi}{\sigma}}(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)A^\rho d\rho(d\rho-1)k^{d\rho-2}\frac{\partial k}{\partial \tau_\xi}+\frac{r_z}{1-d}\tau_\xi^{-1}\frac{z}{k}+\varphi P^{\frac{\phi}{\sigma}}\Omega^{-2}A^\rho d\rho k^{d\rho-1}\frac{\partial \Omega}{\partial \tau}=0 \tag{A.30}$$

The Envelop condition of the problem is

$$V'(\Omega) = \varphi P^{\frac{\phi}{\sigma}} \Omega^{-2} A^{\rho} k^{d\rho} + \frac{r_{\omega}}{\zeta} f'(\Omega) \tag{A.31}$$

where we have used Equation A.2 and the fact that mature firms have reached their steady state. Using Equation A.2 again to substitute out $V'(\Omega)$, we get

$$r_{\omega} = (1-\delta)\varphi P^{\frac{\phi}{\sigma}} \Omega^{-2} A^{\rho} k^{d\rho} \zeta + (1-\delta) r_{\omega} f'(\Omega) \tag{A.32}$$

Differentiating both sides of the Envelop condition with respect to τ_{ξ},

$$(2\varphi P^{\frac{\phi}{\sigma}} \Omega^{-3} A^{\rho} k^{d\rho} \zeta - r_{\omega} f''(\Omega)) \frac{\partial \Omega}{\partial \tau_{\xi}} = d\rho \varphi P^{\frac{\phi}{\sigma}} \Omega^{-2} A^{\rho} k^{d\rho - 1} \zeta \frac{\partial k}{\partial \tau_{\xi}} \tag{A.33}$$

From Equation A.33, we have

$$\frac{\partial \Omega}{\partial \tau_{\xi}} = \frac{d\rho \varphi P^{\frac{\phi}{\sigma}} \Omega^{-2} A^{\rho} k^{d\rho - 1} \zeta}{2\varphi P^{\frac{\phi}{\sigma}} \Omega^{-3} A^{\rho} k^{d\rho} \zeta - r_{\omega} f''(\Omega)} \frac{\partial k}{\partial \tau_{\xi}} \tag{A.34}$$

For simplicity, let's denote β_k by $\frac{d\rho \varphi P^{\frac{\phi}{\sigma}} \Omega^{-2} A^{\rho} k^{d\rho - 1} \zeta}{2\varphi P^{\frac{\phi}{\sigma}} \Omega^{-3} A^{\rho} k^{d\rho} \zeta - r_{\omega} f''(\Omega)}$. Substituting Equation A.34 in Equation A.30, we have

$$\frac{\partial k}{\partial \tau_{\xi}} = \frac{\frac{r_z}{(1-d)\tau_{\xi}} \frac{z}{k}}{\varphi P^{\frac{\phi}{\sigma}} (m_i - \Omega_i^{-1} - \sigma_{\epsilon}^2) A^{\rho} d\rho (1-d\rho) k^{d\rho - 2} \frac{\partial k}{\partial \tau_{\xi}} - \varphi P^{\frac{\phi}{\sigma}} \Omega^{-2} A^{\rho} d\rho k^{d\rho - 1} \beta_k} \tag{A.35}$$

Substituting Equation A.35 in Equation A.34, we then have

$$\frac{\partial \Omega}{\partial \tau_{\xi}} = \frac{\frac{r_z}{(1-d)\tau_{\xi}} \frac{z}{k}}{\varphi P^{\frac{\phi}{\sigma}} (m_i - \Omega_i^{-1} - \sigma_{\epsilon}^2) A^{\rho} d\rho (1-d\rho) k^{d\rho - 2} \frac{\partial k}{\partial \tau_{\xi}} \beta_k^{-1} - \varphi P^{\frac{\phi}{\sigma}} \Omega^{-2} A^{\rho} d\rho k^{d\rho - 1}} \tag{A.36}$$

It is easy to note that $\beta_k < \frac{d\rho\Omega}{2k}$. Using the definition of Γ and Equation A.12, we can derive an upper bound of $\frac{\partial \Omega}{\partial \tau_{\xi}}$,

$$\frac{\partial \Omega}{\partial \tau_{\xi}} \leqslant (r_k - \Gamma) \frac{\frac{\Gamma}{d\tau_{\xi}}}{\frac{2}{d\rho}(1-d\rho)\Omega^{-1} - \Omega^{-2}(m_i - \Omega_i^{-1} - \sigma_{\epsilon}^2)^{-1}} \tag{A.37}$$

Substituting Inequality A.37 in Equation A.28, we have

$$\frac{\partial p_i}{\partial \tau_\xi} \leqslant (r_k - \Gamma_i)^{\frac{d}{\sigma(1-d\rho)}-1}(m_i - \Omega_i^{-1} - \sigma_\epsilon^2)^{1-\frac{d}{\sigma(1-d\rho)}} \frac{1}{d}\frac{\Gamma}{\tau_\xi}\left[-\frac{d}{\sigma(1-d\rho)} + (1 - \frac{d}{\sigma(1-d\rho)})\frac{1}{\frac{2}{d\rho}(1-d\rho)\Omega(m_i - \Omega_i^{-1} - \sigma_\epsilon^2) - 1}\right] \tag{A.38}$$

Whether the term in the bracket is negative or positive depends on the weight $\frac{d}{\sigma(1-d\rho)}$. If d is close to 1, then $\frac{d}{\sigma(1-d\rho)}$ is close to 1 and the term in the bracket will be negative. This concludes the proof.[1]

□

Proof of Theorem 19

Proof. Given output y, the marginal cost of the firm is

$$MC = \frac{r_k}{d} A^{-\frac{1}{d}} y^{\frac{1-d}{d}} \tag{A.39}$$

Using Equation A.12, we can simplify the expression of MC

$$MC = \frac{r_k}{d} A^{-\frac{1-\rho}{1-d\rho}} (d\rho\varphi(m_i - \Omega_i^{-1} - \sigma_\epsilon^2))^{\frac{1-d}{(1-d\rho)}} (r_k - \Gamma)^{-\frac{1-d}{(1-d\rho)}} P^{\frac{\phi(1-d)}{\sigma(1-d\rho)}} \tag{A.40}$$

From Lemma 5, and Equation A.12, we have

$$\mathrm{E}(p) = A^{-\frac{1}{\sigma(1-d\rho)}} \varphi^{1-\frac{d}{\sigma(1-d\rho)}} (d\rho)^{-\frac{d}{\sigma(1-d\rho)}} (m_i - \Omega_i^{-1} - \sigma_\epsilon^2)^{1-\frac{d}{\sigma(1-d\rho)}} (r_k - \Gamma)^{\frac{d}{\sigma(1-d\rho)}} P^{\frac{\phi}{\sigma}(1-\frac{d}{\sigma(1-d\rho)})} \tag{A.41}$$

Notice that $1 - \frac{d}{\sigma(1-d\rho)} = \frac{1-d}{(1-d\rho)}$. Therefore,

$$\mathcal{M} = \frac{\mathrm{E}(p)}{MC} = \frac{r_k - \Gamma}{\rho r_k} \tag{A.42}$$

□

Proof of Proposition 20

[1] A small remark is that as long as m is relatively large, the second term in the bracket is positive.

Proof. From Proposition 11, we know that Γ does not change with respect to Ω. So once a firm enters the market, its mark-up is fixed. By Proposition 14, we know Γ increases in τ_ξ. So mark-ups in high predictive accuracy industries is lower than mark-ups in low predictive accuracy industries. By Proposition 13, we know Γ does not change with respect to m. So mark-ups are the same in premium and non-premium industry. □

Proof of Proposition 21

Proof. We first show that data price r_ω cannot affect new firms' expected profit. Notice that

$$\frac{\partial \mathrm{E}(V(\Omega_0))}{\partial r_\omega} = \mathrm{E}\left(\omega_0 + (1-\delta)\frac{\partial \mathrm{E}(V(\Omega'))}{\partial r_\omega}\right) \tag{A.43}$$

$$\frac{\partial \mathrm{E}(V(\Omega'))}{\partial r_\omega} = \mathrm{E}\left(\omega_1 + (1-\delta)\frac{\partial \mathrm{E}(V(\Omega''))}{\partial r_\omega}\right) \tag{A.44}$$

$$\frac{\partial \mathrm{E}(V(\Omega''))}{\partial r_\omega} = \mathrm{E}\left(\omega_2 + (1-\delta)\frac{\partial \mathrm{E}(V(\Omega''))}{\partial r_\omega}\right) \tag{A.45}$$

where we use the fact that all firms reach their steady state starting from $t = 2$. So we can get the expression of $\frac{\partial \mathrm{E}(V(\Omega_0))}{\partial r_\omega}$

$$\frac{\partial \mathrm{E}(V(\Omega_0))}{\partial r_\omega} = \mathrm{E}\left(\omega_0 + (1-\delta)\omega_1 + \frac{(1-\delta)^2\omega_2}{\delta}\right) \tag{A.46}$$

Meanwhile, the data market clearing condition in stationary equilibrium says that

$$\delta \mathrm{E}\left(\omega_0 + (1-\delta)\omega_1 + \frac{(1-\delta)^2\omega_2}{\delta}\right) = 0 \tag{A.47}$$

Therefore, we have $\frac{\partial \mathrm{E}(V(\Omega_0))}{\partial r_\omega} = 0$.

Now, it suffices to prove that the value function increases in the price index P and decreases in r_z. Because if these two conditions hold, then from equilibrium condition Equation A.8, price index P must increase in r_z. To prove that the value function increases in

P, let's assume an arbitrary continuation value function V_0 which increases in P. Assume that the firm take V_0 as the value function on the right hand side of Equation A.3. Consider two price indexes $P_1 < P_2$. Firm's optimal decision under P_1 is denoted $(k_1, z_1, \Omega_1^{'})$. Then under P_2, the firm can use the same strategy but get higher payoff. So TV_0 increases in P, where T is the same operator in the proof of Theorem 6. Because monotonicity is closed by limit, we derive that the value function V increases in P. With similar method, we can show that the value function V decreases in r_z. □

Proof of Proposition 22

Proof. We utilize the zero entry condition, Equation A.8. Differentiate both sides with respect to r_z. We then have

$$\frac{\partial P}{\partial r_z} = -\frac{\iint \frac{\partial V(\Omega_0, m, \tau_\xi)}{\partial r_z} dF_m dF_\xi}{\iiint \frac{\partial V(\Omega_0, m, \tau_\xi)}{\partial P} dF_m dF_\xi} \tag{A.48}$$

So we need to compute two derivatives: $\frac{\partial V(\Omega_0, m, \tau_\xi)}{\partial r_z}$ and $\frac{\partial V(\Omega_0, m, \tau_\xi)}{\partial P}$. Using Enveope Theorem and the fact that firms will reach steady state starting from $t = 2$, we can get

$$\frac{\partial V(\Omega^{''}, m, \tau_\xi)}{\partial r_z} = -\frac{1}{\delta} z(\Omega^{''}) \tag{A.49}$$

$$\frac{\partial V(\Omega^{''}, m, \tau_\xi)}{\partial P} = \frac{1}{\delta} I^{\prime}(P)(m - \Omega^{''-1} - \sigma_\epsilon^2) A^\rho k^{d\rho}(\Omega^{''}) \tag{A.50}$$

$$\frac{\partial V(\Omega^{'}, m, \tau_\xi)}{\partial r_z} = -z(\Omega^{'}) + (1-\delta)\frac{\partial V(\Omega^{''}, m, \tau_\xi)}{\partial r_z} \tag{A.51}$$

$$\frac{\partial V(\Omega^{'}, m, \tau_\xi)}{\partial P} = I^{\prime}(P)(m - \Omega^{'-1} - \sigma_\epsilon^2) A^\rho k^{'d\rho}(\Omega^{'}) + (1-\delta)\frac{\partial V(\Omega^{''}, m, \tau_\xi)}{\partial P} \tag{A.52}$$

$$\frac{\partial V(\Omega_0, m, \tau_\xi)}{\partial r_z} = -z(\Omega_0) + (1-\delta)\frac{\partial V(\Omega^{'}, m, \tau_\xi)}{\partial r_z} \tag{A.53}$$

$$\frac{\partial V(\Omega_0, m, \tau_\xi)}{\partial P} = I^{\prime}(P)(m - \Omega_0^{-1} - \sigma_\epsilon^2) A^\rho k_0^{d\rho}(\Omega_0) + (1-\delta)\frac{\partial V(\Omega^{'}, m, \tau_\xi)}{\partial P} \tag{A.54}$$

Substituting Equations A.51∼A.54 to Equation A.49 and Equation A.50, we obtain

$$\frac{\partial V(\Omega_0, m, \tau_\xi)}{\partial r_z} = -z(\Omega_0) - (1-\delta)z(\Omega') - \frac{(1-\delta)^2}{\delta} z(\Omega'') \tag{A.55}$$

$$\frac{\partial V(\Omega_0, m, \tau_\xi)}{\partial P} = I'(P)\left[(m-\Omega_0^{-1}-\sigma_\epsilon^2)A^\rho k_0^{d\rho} + (1-\delta)(m-\Omega'^{-1}-\sigma_\epsilon^2)A^\rho k'^{d\rho} + \frac{(1-\delta)^2}{\delta}(m-\Omega''^{-1}-\sigma_\epsilon^2)A^\rho k'^{d\rho}\right] \tag{A.56}$$

From Equation A.11 and Equation A.12, we can get the expression for z

$$z = \frac{1-d}{d}\frac{1}{r_z}\Gamma\left(\frac{d\rho\varphi A^\rho(m-\Omega^{-1}-\sigma_\epsilon^2)}{r_k-\Gamma}\right)^{\frac{1}{1-d\rho}} P^{\frac{\phi}{\sigma(1-d\rho)}} \tag{A.57}$$

Moreover, $k^{d\rho}$ in the denominator in Equation A.48 can be simplified to

$$k^{d\rho} = (r_k-\Gamma)^{-\frac{d\rho}{1-d\rho}}(m-\Omega^{-1}-\sigma_\epsilon^2)^{\frac{d\rho}{1-d\rho}} P^{\frac{d\rho\phi}{\sigma(1-d\rho)}} \left(A^\rho d\rho\varphi\right)^{\frac{1}{1-d\rho}} \tag{A.58}$$

Recall that $I'(P) = \varphi\frac{\phi}{\sigma}P^{\frac{\phi}{\sigma}-1}$. The Equation A.48 can be simplified to

$$\frac{\partial P}{\partial r_z} = \frac{E\left[X\frac{\Gamma}{r_k-\Gamma} + X'\frac{\Gamma'}{r_k-\Gamma'}(1-\delta) + X''\frac{\Gamma''}{r_k-\Gamma''}\frac{(1-\delta)^2}{\delta}\right]}{E[X + X'(1-\delta) + X''\frac{(1-\delta)^2}{\delta}]}\frac{(1-d)\rho\sigma}{r_z\phi}P \tag{A.59}$$

where

$$X = (r_k-\Gamma)^{-\frac{d\rho}{1-d\rho}}\left[(m-\Omega_0^{-1}-\sigma_\epsilon^2)^{\frac{1}{1-d\rho}}\right] \tag{A.60}$$

Notice that X increases in Γ. To get the desired result, we need the following lemma.

Lemma 49. *Suppose $f(x)$ is an increasing function in x. Then* $\mathrm{Cov}(f(x), x) \geqslant 0$.

Proof. Denote f_0 by $\mathrm{E}(f(x))$ and x_0 by $\mathrm{E}(x)$. Then

$$\mathrm{Cov}(f, x) = \mathrm{E}[(f - f_0)(x - x_0)] \tag{A.61}$$

Pick y such that $(f(y) - f_0)(y - x_0) < 0$. If b does not exist, then we are done. We know that

$$\mathrm{Cov}(f,x) = \mathrm{E}[(f-f(y))(x-x_0)] = \mathrm{E}[(f-f(y))(x-x_0-y)] + y(f_0-f(y)) = \mathrm{E}[(f-f(y))(x-y)] - (f(y)-f_0)(y-x_0) \tag{A.62}$$

The first term in Equation A.62 is positive due to monotonicity and $(f(y) - f_0)(y - x_0)$ is negative by construction. Therefore $\mathrm{Cov}(f, x) \geqslant 0$. □

From Lemma 49, we know that $\text{Cov}(X, \frac{\Gamma}{r_k-\Gamma}) \geqslant 0$. Therefore,

$$\frac{\partial P}{\partial r_z} \geqslant \mathsf{E}\left[\frac{\Gamma}{r_k-\Gamma}\right]\frac{(1-d)\rho\sigma}{r_z\phi}P \tag{A.63}$$

where we have used the fact that $\Gamma' > \Gamma$ because $\kappa \leqslant 1$. The desired result is immediate from the above inequality.

□

Proof of Proposition 23

Proof. It suffices to show that $\frac{\partial r_\omega}{\partial r_z} > 0$. Differentiating both sides of the data market clearing condition, Equation A.9, we have

$$\frac{\partial r_\omega}{\partial r_z} = -\frac{\mathsf{E}\left(\frac{\partial\hat{\omega}}{\partial r_z}\right) + \mathsf{E}\left(\frac{\partial\hat{\omega}}{\partial P}\right)\frac{\partial P}{\partial r_z}}{\mathsf{E}\left(\frac{\partial\hat{\omega}}{\partial r_\omega}\right)} \tag{A.64}$$

where $\hat{\omega} = \omega + (1-\delta)\omega' + \frac{(1-\delta)^2}{\delta}\omega''$. Then we will compute three derivatives: $\frac{\partial\hat{\omega}}{\partial r_z}$, $\frac{\partial\hat{\omega}}{\partial P}$ and $\frac{\partial\hat{\omega}}{\partial r_\omega}$.

By Equation A.11, we can get

$$\frac{\partial\frac{z}{k}}{\partial r_z} = -\frac{r_\omega(1-d)}{\zeta d}A(\frac{z}{k})^{1-d}\tau_\xi r_z^{-2} \tag{A.65}$$

$$\frac{\partial\frac{z'}{k'}}{\partial r_z} = -\frac{r_\omega(1-d)}{\zeta' d}A(\frac{z'}{k'})^{1-d}\tau_\xi r_z^{-2} \tag{A.66}$$

$$\frac{\partial\frac{z''}{k''}}{\partial r_z} = -\frac{r_\omega(1-d)}{\zeta'' d}A(\frac{z''}{k''})^{1-d}\tau_\xi r_z^{-2} \tag{A.67}$$

Differentiating both sides of the Envelop conditions, Equation A.32 at $t = 1$ and $t = 2$ with respect to r_z, we can get

$$\frac{\partial\Omega''}{\partial r_z} = \frac{\varphi P^{\frac{\phi}{\sigma}}\Omega''^{-2}A^\rho d\rho k''^{d\rho-1}\zeta''}{2\varphi P^{\frac{\phi}{\sigma}}\Omega''^{-3}A^\rho k''^{d\rho}\zeta'' - r_\omega f''(\Omega'')}\frac{\partial k''}{\partial r_z} \tag{A.68}$$

$$\frac{\partial\Omega'}{\partial r_z} = \frac{\varphi P^{\frac{\phi}{\sigma}}\Omega'^{-2}A^\rho d\rho k'^{d\rho-1}\zeta'}{2\varphi P^{\frac{\phi}{\sigma}}\Omega'^{-3}A^\rho k'^{d\rho}\zeta' - r_\omega f''(\Omega')}\frac{\partial k'}{\partial r_z} \tag{A.69}$$

Denote $\beta' = \frac{\varphi P^{\frac{\phi}{\sigma}}\Omega'^{-2}A^{\rho}d\rho k'^{d\rho-1}\zeta'}{2\varphi P^{\frac{\phi}{\sigma}}\Omega'^{-3}A^{\rho}k'^{d\rho}\zeta'-r_{\omega}f''(\Omega')}$ and $\beta'' = \frac{\varphi P^{\frac{\phi}{\sigma}}\Omega''^{-2}A^{\rho}d\rho k''^{d\rho-1}\zeta''}{2\varphi P^{\frac{\phi}{\sigma}}\Omega''^{-3}A^{\rho}k''^{d\rho}\zeta''-r_{\omega}f''(\Omega'')}$,where $\beta' \leqslant \frac{d\rho\Omega'}{2k'}$ and $\beta'' \leqslant \frac{d\rho\Omega''}{2k''}$. Differentiating both sides of Equation A.10 with respect to r_z and using Equations A.66∼A.69, we have

$$\frac{\partial k'}{\partial r_z} = \frac{\frac{z'}{k'}}{\varphi P^{\frac{\phi}{\sigma}}A^{\rho}d\rho k'^{d\rho-2}\gamma_k''} \tag{A.70}$$

$$\frac{\partial k''}{\partial r_z} = \frac{\frac{z''}{k''}}{\varphi P^{\frac{\phi}{\sigma}}A^{\rho}d\rho k''^{d\rho-2}\gamma_k'} \tag{A.71}$$

where

$$\gamma_k' = \frac{\varphi P^{\frac{\phi}{\sigma}}\Omega'^{-4}A^{\rho}d\rho k'^{d\rho}\zeta'}{2\varphi P^{\frac{\phi}{\sigma}}\Omega'^{-3}A^{\rho}k'^{d\rho}\zeta' - r_{\omega}f''(\Omega')} - (m - \Omega'^{-1} - \sigma_{\epsilon}^2)(1 - d\rho)$$

Notice that $\gamma_k' < \frac{1}{2}\Omega'^{-1}d\rho - (m - \Omega'^{-1} - \sigma_{\epsilon}^2)(1 - d\rho)$. As long as m is big enough, then γ_k' and γ_k'' will be negative. We'll maintain this assumption for the whole paper.

ASSUMPTION 3. *$\frac{1}{2}\Omega'^{-1}d\rho - (m - \Omega'^{-1} - \sigma_{\epsilon}^2)(1 - d\rho) < 0$ and $\frac{1}{2}\Omega''^{-1}d\rho - (m - \Omega''^{-1} - \sigma_{\epsilon}^2)(1 - d\rho) < 0$.*

Differentiating Equation A.10 at $t = 0$ with respect to r_z, we then obtain

$$\frac{\partial k}{\partial r_z} = -\frac{z/k}{\varphi P^{\frac{\phi}{\sigma}}(m - \Omega^{-1} - \sigma_{\epsilon}^2)d\rho(1 - d\rho)k^{d\rho-2}} \tag{A.72}$$

Differentiating the law of motion with respect to r_z, we have

$$\frac{\partial \omega''}{\partial r_z}\zeta'' = A\left(\frac{z''}{k''}\right)^{1-d}\tau_\xi\frac{\partial k''}{\partial r_z} - \frac{\zeta''}{r_\omega d}z'' - (1-f'(\Omega''))\frac{\partial \Omega''}{\partial r_z} \tag{A.73}$$

$$\frac{\partial \omega'}{\partial r_z}\zeta' = A\left(\frac{z'}{k'}\right)^{1-d}\tau_\xi\frac{\partial k'}{\partial r_z} - \frac{\zeta'}{r_\omega d}z' + f'(\Omega')\frac{\partial \Omega'}{\partial r_z} - \frac{\partial \Omega''}{\partial r_z} \tag{A.74}$$

$$\frac{\partial \omega}{\partial r_z}\zeta = A\left(\frac{z}{k}\right)^{1-d}\tau_\xi\frac{\partial k}{\partial r_z} - \frac{\zeta}{r_\omega d}z - \frac{\partial \Omega'}{\partial r_z} \tag{A.75}$$

Therefore,

$$\begin{aligned}\frac{\partial \hat{\omega}}{\partial r_z} &= \left[A\left(\frac{z}{k}\right)^{1-d}\tau_\xi\frac{1}{\zeta}\frac{\partial k}{\partial r_z} - \frac{1}{r_\omega d}z\right] - \left[\frac{1}{\zeta} - \frac{(1-\delta)f'(\Omega')}{\zeta'}\right]\frac{\partial \Omega'}{\partial r_z}\\ &+ \left[A\left(\frac{z'}{k'}\right)^{1-d}\tau_\xi\frac{1}{\zeta'}\frac{\partial k'}{\partial r_z} - \frac{1}{r_\omega d}z'\right](1-\delta) - \left[\frac{1-\delta}{\zeta'} - \frac{(1-\delta)^2 f'(\Omega'')}{\delta\zeta'}\right]\frac{\partial \Omega''}{\partial r_z}\\ &+ \left[A\left(\frac{z''}{k''}\right)^{1-d}\tau_\xi\frac{1}{\zeta''}\frac{\partial k''}{\partial r_z} - \frac{1}{r_\omega d}z''\right]\frac{(1-\delta)^2}{\delta}\end{aligned} \tag{A.76}$$

Then we will prove $\frac{\partial \hat{\omega}}{\partial r_z} < 0$. Notice that $\frac{\partial \Omega'}{\partial r_z}$ and $\frac{\partial \Omega''}{\partial r_z}$ are both negative. It suffices to show

$$A\left(\frac{z'}{k'}\right)^{1-d}\tau_\xi(1-\delta)\frac{\partial k'}{\partial r_z} < \left[\frac{1}{\zeta} - \frac{(1-\delta)f'(\Omega')}{\zeta'}\right]\frac{\partial \Omega'}{\partial r_z}$$

$$A\left(\frac{z''}{k''}\right)^{1-d}\tau_\xi\frac{\partial k''}{\partial r_z}\frac{(1-\delta)^2}{\delta} < \left[\frac{1-\delta}{\zeta'} - \frac{(1-\delta)^2 f'(\Omega'')}{\delta\zeta'}\right]\frac{\partial \Omega''}{\partial r_z}$$

We only prove the first inequality here and the second one can be done in a similar way. Obviously

$$A\left(\frac{z'}{k'}\right)^{1-d}\tau_\xi(1-\delta)\frac{\partial k'}{\partial r_z} - \left[\frac{1}{\zeta} - \frac{(1-\delta)f'(\Omega')}{\zeta'}\right]\frac{\partial \Omega'}{\partial r_z} < A\left(\frac{z'}{k'}\right)^{1-d}\tau_\xi(1-\delta)\frac{\partial k'}{\partial r_z} - \frac{(1-\delta)f'(\Omega')}{\zeta'}\frac{\partial \Omega'}{\partial r_z} \tag{A.77}$$

$$< \left[A\left(\frac{z'}{k'}\right)^{1-d}\tau_\xi - \frac{f'(\Omega')}{\zeta'}\frac{\Omega' d\rho}{2k'}\right](1-\delta)\frac{\partial k'}{\partial r_z} \tag{A.78}$$

By Lemma 48, it is easy to find that the right hand side of the inequality is negative as long as $\frac{f'(\Omega')}{\zeta'}\frac{d\rho}{2} < 1$. So we have proved that $\frac{\partial\hat{\omega}}{\partial r_z} < 0$.

By similar argument, we can find the formula for $\frac{\partial\hat{\omega}}{\partial P}$ and $\frac{\partial\hat{\omega}}{\partial r_\omega}$.

$$\begin{aligned}
\frac{\partial\hat{\omega}}{\partial P} &= \frac{r_z}{r_\omega(1-d)}\frac{z}{k}\frac{\partial k}{\partial P} - \left[\frac{1}{\zeta} - \frac{(1-\delta)f'(\Omega')}{\zeta'}\right]\frac{\partial\Omega'}{\partial P} \\
&+ \frac{r_z}{r_\omega(1-d)}\frac{z'}{k'}\frac{\partial k'}{\partial P}(1-\delta) - \left[\frac{1-\delta}{\zeta'} - \frac{(1-\delta)^2 f'(\Omega'')}{\delta\zeta'}\right]\frac{\partial\Omega''}{\partial P} \\
&+ \frac{r_z}{r_\omega(1-d)}\frac{z''}{k''}\frac{\partial k''}{\partial P}\frac{(1-\delta)}{\delta}
\end{aligned} \tag{A.79}$$

$$\begin{aligned}
\frac{\partial\hat{\omega}}{\partial r_\omega} &= \frac{r_z}{r_\omega(1-d)}\frac{z}{k}\frac{\partial k}{\partial r_\omega} + \frac{r_z}{r_\omega^2}z - \left[\frac{1}{\zeta} - \frac{(1-\delta)f'(\Omega')}{\zeta'}\right]\frac{\partial\Omega'}{\partial r_\omega} \\
&+ \left(\frac{r_z}{r_\omega(1-d)}\frac{z'}{k'}\frac{\partial k'}{\partial r_\omega} + \frac{r_z}{r_\omega^2}z'\right)(1-\delta) - \left[\frac{1-\delta}{\zeta'} - \frac{(1-\delta)^2 f'(\Omega'')}{\delta\zeta'}\right]\frac{\partial\Omega''}{\partial r_\omega} \\
&+ \left(\frac{r_z}{r_\omega(1-d)}\frac{z''}{k''}\frac{\partial k''}{\partial r_\omega} + \frac{r_z}{r_\omega^2}z''\right)\frac{(1-\delta)}{\delta}
\end{aligned} \tag{A.80}$$

Similarly, we can show $\frac{\partial\hat{\omega}}{\partial P} > 0$ and $\frac{\partial\hat{\omega}}{\partial r_\omega} > 0$.

To finish the proof, we only need to show $\mathrm{E}(\frac{\partial\hat{\omega}}{\partial r_z}) + \mathrm{E}(\frac{\partial\hat{\omega}}{\partial P})\frac{\partial P}{\partial r_z} < 0$. By the same way of computing $\frac{\partial\Omega'}{\partial r_z}$ and $\frac{\partial\Omega''}{\partial r_z}$, we can get the following expressions,

$$\frac{\partial\Omega'}{\partial P} = \frac{\frac{\phi}{\sigma}}{Pd\rho}\beta' + \beta'\frac{\partial k'}{\partial P} \tag{A.81}$$

$$\frac{\partial\Omega''}{\partial P} = \frac{\frac{\phi}{\sigma}}{Pd\rho}\beta'' + \beta''\frac{\partial k''}{\partial P} \tag{A.82}$$

$$\frac{\partial\Omega'}{\partial r_\omega} = \frac{(1-\delta)f'(\Omega') - 1}{(1-\delta)[2\varphi P^{\frac{\phi}{\sigma}}\Omega'^{-3}A^\rho k'^{d\rho}\zeta' - r_\omega f''(\Omega')]} + \beta'\frac{\partial k'}{\partial r_\omega} \tag{A.83}$$

$$\frac{\partial\Omega''}{\partial r_\omega} = \frac{(1-\delta)f'(\Omega'') - 1}{(1-\delta)[2\varphi P^{\frac{\phi}{\sigma}}\Omega''^{-3}A^\rho k''^{d\rho}\zeta'' - r_\omega f''(\Omega'')]} + \beta''\frac{\partial k'}{\partial r_\omega} \tag{A.84}$$

We then can find the upper bound for $\frac{\partial\hat{\omega}}{\partial r_z}$ and $\frac{\partial\hat{\omega}}{\partial P}$,

$$\begin{aligned}\frac{\partial\hat{\omega}}{\partial r_z} &< \frac{r_z}{r_\omega(1-d)}\frac{z}{k}\frac{\partial k}{\partial r_z} \\ &+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z'}{k'} - \left[\frac{1}{(1-\delta)\zeta} - \frac{f'(\Omega')}{\zeta'}\right]\beta'\right](1-\delta)\frac{\partial k'}{\partial r_z} \\ &+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z''}{k''} - \left[\frac{\delta}{(1-\delta)\zeta'} - \frac{f'(\Omega'')}{\zeta'}\right]\beta''\right]\frac{(1-\delta)^2}{\delta}\frac{\partial k''}{\partial r_z} \end{aligned} \tag{A.85}$$

$$\begin{aligned}\frac{\partial\hat{\omega}}{\partial P} &< \frac{r_z}{r_\omega(1-d)}\frac{z}{k}\frac{\partial k}{\partial P} \\ &+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z'}{k'} - \left[\frac{1}{(1-\delta)\zeta} - \frac{f'(\Omega')}{\zeta'}\right]\beta'\right]\frac{\partial k'}{\partial P}(1-\delta) \\ &+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z''}{k''} - \left[\frac{\delta}{(1-\delta)\zeta'} - \frac{f'(\Omega'')}{\zeta'}\right]\beta''\right]\frac{\partial k''}{\partial P}\frac{(1-\delta)}{\delta} \end{aligned} \tag{A.86}$$

Moreover, with some simple calculation, we know

$$\frac{\partial k}{\partial r_z} = -\frac{\sigma}{\phi}\frac{\Gamma}{r_k-\Gamma}\frac{1-d}{dr_z}P\frac{\partial k}{\partial P} \tag{A.87}$$

$$\frac{\partial k'}{\partial r_z} > -\frac{\sigma}{\phi}\frac{\Gamma'}{r_k-\Gamma'}\frac{1-d}{dr_z}P\frac{\partial k'}{\partial P} \tag{A.88}$$

$$\frac{\partial k''}{\partial r_z} > -\frac{\sigma}{\phi}\frac{\Gamma''}{r_k-\Gamma''}\frac{1-d}{dr_z}P\frac{\partial k''}{\partial P} \tag{A.89}$$

Suppose $\kappa = 1$ and F_ξ has zero variance. Then

$$\frac{\partial P}{\partial r_z} = \frac{\Gamma}{r_k-\Gamma}\frac{(1-d)\rho P}{r_z}\frac{\sigma}{\phi} \tag{A.90}$$

Then under the same condition, combining Inequality A.85$\sim$ Inequality A.89, it is easy to find that $\frac{\partial\hat{\omega}}{\partial r_z} + \frac{\partial\hat{\omega}}{\partial P}\frac{\partial P}{\partial r_z} < 0$. Therefore, by continuity, if κ is close to 1 and the variance of F_ξ is small, data price r_ω increases in the data analytics cost r_z. □

Proof of Proposition 24

Proof. We compute derivatives: $\frac{\partial \frac{z}{k}}{\partial r_z}$ and $\frac{\partial \Gamma}{\partial r_z}$.

$$
\begin{aligned}
\frac{\partial \frac{z}{k}}{\partial r_z} &= \frac{1-d}{d}\left[-\frac{1}{d}\frac{\Gamma}{r_z^2}+\frac{1}{d}\frac{\Gamma}{r_\omega}\frac{\partial r_\omega}{\partial r_z}\frac{1}{r_z}\right] && \text{(A.91)}\\
&= \frac{1-d}{d^2 r_z}\left[-\frac{\Gamma}{r_z^2}+\frac{\Gamma}{r_\omega}\frac{\partial r_\omega}{\partial r_z}\right] && \text{(A.92)}\\
&= \frac{\Gamma}{d r_z^2}\left[\epsilon_{r_\omega,z}-1\right] && \text{(A.93)}
\end{aligned}
$$

where $\epsilon_{\omega,z}$ is the elasticity of r_ω with respect to r_z.

$$
\begin{aligned}
\frac{\partial \Gamma}{\partial r_z} &= \frac{d}{1-d}\left[\frac{z}{k}+r_z\frac{\partial \frac{z}{k}}{\partial r_z}+r_z\frac{\partial \frac{z}{k}}{\partial r_\omega}\frac{\partial r_\omega}{\partial r_z}\right] && \text{(A.94)}\\
&= \frac{z}{k}\frac{1}{1-d}\left[\epsilon_{r_\omega,z}-(1-d)\right] && \text{(A.95)}
\end{aligned}
$$

Therefore, we only need to prove that $\epsilon_{r_\omega,z} < 1-d$. Note that a natural upper bound for $\epsilon_{r_\omega,z}$ is $-\frac{r_z}{r_\omega}\frac{\mathrm{E}(\frac{\partial \hat{\omega}}{\partial r_z})}{\mathrm{E}(\frac{\partial \hat{\omega}}{\partial r_\omega})}$. Then we derive an upper bound for $-\mathrm{E}(\frac{\partial \hat{\omega}}{\partial r_z})$ and a lower bound for $\mathrm{E}(\frac{\partial \hat{\omega}}{\partial r_\omega})$.

$$
\begin{aligned}
\frac{\partial \hat{\omega}}{\partial r_z} &= \frac{r_z}{r_\omega(1-d)}\frac{z}{k}\frac{\partial k}{\partial r_z}-\frac{1}{r_\omega d}z\\
&+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z'}{k'}-\left[\frac{1}{(1-\delta)\zeta}-\frac{f'(\Omega')}{\zeta'}\right]\beta'\right](1-\delta)\frac{\partial k'}{\partial r_z}-\frac{1}{r_\omega d}z'(1-\delta)\\
&+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z''}{k''}-\left[\frac{\delta}{(1-\delta)\zeta'}-\frac{f'(\Omega'')}{\zeta'}\right]\beta''\right]\frac{(1-\delta)^2}{\delta}\frac{\partial k''}{\partial r_z}--\frac{1}{r_\omega d}z''\frac{(1-\delta)^2}{\delta} && \text{(A.96)}
\end{aligned}
$$

By the same way of computing $\frac{\partial \hat{\omega}}{\partial r_z}$, we can compute $\frac{\partial \hat{\omega}}{\partial r_\omega}$

$$
\begin{aligned}
\frac{\partial \hat{\omega}}{\partial r_\omega} &= \left[\frac{r_z}{r_\omega(1-d)}\frac{z}{k}\frac{\partial k}{\partial r_\omega}+\frac{r_z}{r_\omega^2}z\right]-\left[\frac{1}{\zeta}-\frac{(1-\delta)f'(\Omega')}{\zeta'}\right]\frac{\partial \Omega'}{\partial r_\omega}\\
&+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z'}{k'}\frac{\partial k'}{\partial r_\omega}+\frac{r_z}{r_\omega^2}z'\right](1-\delta)-\left[\frac{1-\delta}{\zeta'}-\frac{(1-\delta)^2 f'(\Omega'')}{\delta\zeta'}\right]\frac{\partial \Omega''}{\partial r_\omega}\\
&+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z''}{k''}\frac{\partial k''}{\partial r_\omega}+\frac{r_z}{r_\omega^2}z''\right]\frac{(1-\delta)^2}{\delta} && \text{(A.97)}
\end{aligned}
$$

Using Equation A.83 and Equation A.84, we can get

$$\begin{aligned}
\frac{\partial\hat{\omega}}{\partial r_\omega} &> \left[\frac{r_z}{r_\omega(1-d)}\frac{z}{k}\frac{\partial k}{\partial r_\omega}\right]+\frac{r_z}{r_\omega^2}z \\
&+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z'}{k'}-\left[\frac{1}{(1-\delta)\zeta}-\frac{f'(\Omega')}{\zeta'}\right]\beta'\right]\frac{\partial k'}{\partial r_\omega}(1-\delta)+\frac{r_z}{r_\omega^2}z'(1-\delta) \\
&+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z''}{k''}-\left[\frac{\delta}{(1-\delta)\zeta'}-\frac{f'(\Omega'')}{\zeta'}\right]\beta''\right]\frac{(1-\delta)^2}{\delta}\frac{\partial k''}{\partial r_\omega}+\frac{r_z}{r_\omega^2}z''\frac{(1-\delta)^2}{\delta} \qquad \text{(A.98)}
\end{aligned}$$

Moreover, with some simple calculation, we know

$$\frac{\partial k}{\partial r_z} = -\frac{(1-d)r_\omega}{r_z}\frac{\partial k}{\partial r_\omega} \qquad \text{(A.99)}$$

$$\frac{\partial k'}{\partial r_z} > -\frac{(1-d)r_\omega}{r_z}\frac{\partial k'}{\partial r_\omega} \qquad \text{(A.100)}$$

$$\frac{\partial k''}{\partial r_z} > -\frac{(1-d)r_\omega}{r_z}\frac{\partial k''}{\partial r_\omega} \qquad \text{(A.101)}$$

Denote x and y by

$$\begin{aligned}
x &= \left[\frac{r_z}{r_\omega(1-d)}\frac{z}{k}\frac{\partial k}{\partial r_\omega}\right]+\left[\frac{r_z}{r_\omega(1-d)}\frac{z'}{k'}-\left[\frac{1}{(1-\delta)\zeta}-\frac{f'(\Omega')}{\zeta'}\right]\beta'\right]\frac{\partial k'}{\partial r_\omega}(1-\delta) \\
&+ \left[\frac{r_z}{r_\omega(1-d)}\frac{z''}{k''}-\left[\frac{\delta}{(1-\delta)\zeta'}-\frac{f'(\Omega'')}{\zeta'}\right]\beta''\right]\frac{(1-\delta)^2}{\delta}\frac{\partial k''}{\partial r_\omega} \qquad \text{(A.102)}
\end{aligned}$$

$$y = \frac{r_z}{r_\omega^2}z+\frac{r_z}{r_\omega^2}z'(1-\delta)+\frac{r_z}{r_\omega^2}z''\frac{(1-\delta)^2}{\delta} \qquad \text{(A.103)}$$

Combining Equation A.96∼ Equation A.101, we have

$$\frac{-E(\frac{\partial\hat{\omega}}{\partial r_z})}{E(\frac{\partial\hat{\omega}}{\partial r_\omega})} \leqslant \frac{r_\omega}{r_z}\left((1-d)\frac{x}{x+y}+\frac{1}{d}\frac{y}{x+y}\right) \qquad \text{(A.104)}$$

Similarly, we can deduce that

$$\frac{\partial k}{\partial P} = \frac{\phi}{\sigma}\frac{r_\omega d}{P}\frac{r_k-\Gamma}{\Gamma}\frac{\partial k}{\partial r_\omega} \quad (A.105)$$

$$\frac{\partial k'}{\partial P} > \frac{\phi}{\sigma}\frac{r_\omega d}{P}\frac{r_k-\Gamma'}{\Gamma'}\frac{\partial k'}{\partial r_\omega} \quad (A.106)$$

$$\frac{\partial k''}{\partial P} > \frac{\phi}{\sigma}\frac{r_\omega d}{P}\frac{r_k-\Gamma''}{\Gamma''}\frac{\partial k''}{\partial r_\omega} \quad (A.107)$$

$$\frac{\partial k}{\partial r_\omega} = \frac{r_z}{(1-d)(1-d\rho)r_\omega}\frac{1}{r_k-\Gamma}z \quad (A.108)$$

$$\frac{\partial k'}{\partial r_\omega} > \frac{r_z}{(1-d)(1-d\rho)r_\omega}\frac{1}{r_k-\Gamma'}z' \quad (A.109)$$

$$\frac{\partial k''}{\partial r_\omega} > \frac{r_z}{(1-d)(1-d\rho)r_\omega}\frac{1}{r_k-\Gamma''}z'' \quad (A.110)$$

Let's assume all the firms share the same τ_ξ and $\kappa = 1$. Then, we can get the following inequality

$$\epsilon_{r_\omega,z} \leqslant (1-d)\frac{x}{x+y} + \frac{1}{d}\frac{y}{x+y} - \frac{d\rho(1-d)x}{x+y} \quad (A.111)$$

So it suffices to prove

$$\left[\frac{1}{d} - (1-d)\right] y < d\rho(1-d)x \quad (A.112)$$

Using the upper bounds on β' and β'' and assume $\delta < \delta_0$, we deduce that

$$\frac{r_z}{r_\omega(1-d)}\frac{z}{k} - \left[\frac{1}{1-\delta} - f'(\Omega')\right]\beta' > \frac{r_z}{r_\omega(1-d)}\frac{z}{k} - (1-f'(\Omega'))\frac{\Omega' d\rho}{2k'} \quad (A.113)$$

$$> \frac{1}{2}\frac{r_z}{r_\omega(1-d)}\frac{z}{k} = \frac{1}{2}\frac{\Gamma}{dr_\omega} \quad (A.114)$$

where we have used the Lemma 48, Equation A.11 and the fact $\kappa = 1$ in deriving the second inequality. Then we can derive a lower bound for x, using Equations A.105$\sim$ A.110

$$x \geqslant \frac{1}{2}\frac{\Gamma}{dr_\omega}\frac{r_\omega}{(1-d)}\frac{1}{(1-d\rho)(r_k-\Gamma)}y = \frac{1}{2}\frac{\Gamma}{r_k-\Gamma}\frac{1}{d(1-d\rho)(1-d)}y \quad (A.115)$$

Then, by Inequality A.112, we only need to show

$$\frac{1}{d} - (1-d) < \frac{1}{2}\frac{\Gamma}{r_k - \Gamma}\frac{\rho}{(1-d\rho)} \tag{A.116}$$

which will be satisfied if d and ρ are close to 1. Therefore, we have finished the proof.

□

Proof of Proposition 25

Proof. We first differentiate both sides of Equation A.4, Equation A.11 and Equation A.32 with respect to r_z. The following derivatives can be derived

$$\frac{1-(1-\delta)f'(\Omega)}{(1-\delta)}\frac{\partial r_\omega}{\partial r_z} = \varphi\frac{\phi}{\sigma}P^{\frac{\phi}{\sigma}-1}\Omega^{-2}A^\rho k^{d\rho}\zeta\frac{\partial P}{\partial r_z} + \left[r_\omega f''(\Omega) - 2\varphi P^{\frac{\phi}{\sigma}}\Omega^{-3}A^\rho k^{d\rho}\zeta\right]\frac{\partial \Omega}{\partial r_z} + \varphi d\rho P^{\frac{\phi}{\sigma}}\Omega^{-2}A^\rho k^{d\rho-1}\zeta\frac{\partial k}{\partial r_z} \tag{A.117}$$

$$1 = \frac{1-d}{\zeta}A\left(\frac{z}{k}\right)^{-d}\tau_\xi\frac{\partial r_\omega}{\partial r_z} - \frac{(1-d)dr_\omega}{\zeta}A\left(\frac{z}{k}\right)^{-1-d}\tau_\xi\frac{\partial\frac{z}{k}}{\partial r_z} \tag{A.118}$$

$$\varphi P^{\frac{\phi}{\sigma}}(m_i - \Omega_i^{-1} - \sigma_\epsilon^2)A^\rho d\rho(1-d\rho)k^{d\rho-2}\frac{\partial k}{\partial r_z} = \frac{d}{1-d}\frac{z}{k} + \frac{r_z d}{1-d}\frac{\partial\frac{z}{k}}{\partial r_z} + \varphi\frac{\phi}{\sigma}P^{\frac{\phi}{\sigma}-1}(m_i - \Omega_i^{-1} - \sigma_\epsilon^2)A^\rho d\rho k^{d\rho-1}\frac{\partial P}{\partial r_z} +$$

$$\varphi P^{\frac{\phi}{\sigma}}\Omega^{-2}A^\rho d\rho k^{d\rho-1}\frac{\partial\Omega}{\partial r_z} \tag{A.119}$$

where Equation A.117 is from the Envelop condition, Equation A.118 is from Equation A.11 and Equation A.119 is from Equation A.4.

Using Equation A.118, we can get

$$\frac{d}{1-d}\left[\frac{z}{k} + r_z\frac{\partial\frac{z}{k}}{\partial r_z}\right] = -\frac{z}{k} + \frac{r_z}{(1-d)r_\omega}\frac{\partial r_\omega}{\partial r_z}\frac{z}{k} \tag{A.120}$$

Substituting Equation A.120 into Equation A.119, we have

$$\varphi d\rho P^{\frac{\phi}{\sigma}}\Omega^{-2}A^\rho k^{d\rho-1}\zeta\frac{\partial k}{\partial r_z} = \frac{k\zeta\Omega^{-2}}{(1-d\rho)(m_i - \Omega_i^{-1} - \sigma_\epsilon^2)}\left[-\frac{z}{k} + \frac{r_z}{(1-d)r_\omega}\frac{\partial r_\omega}{\partial r_z}\frac{z}{k} + \varphi\frac{\phi}{\sigma}P^{\frac{\phi}{\sigma}-1}(m_i - \Omega_i^{-1} - \sigma_\epsilon^2)A^\rho d\rho k^{d\rho-1}\frac{\partial P}{\partial r_z} + \varphi P^{\frac{\phi}{\sigma}}\Omega^{-2}A^\rho d\rho k^{d\rho-1}\frac{\partial\Omega}{\partial r_z}\right] \tag{A.121}$$

Substituting Equation A.121 into Equation A.117, we have

$$\left[(1-\delta)^{-1} - f'(\Omega) - \frac{z\zeta\Omega^{-2}r_z}{(1-d)(1-d\rho)(m_i - \Omega_i^{-1} - \sigma_\epsilon^2)r_\omega}\right]\frac{\partial r_\omega}{\partial r_z} = \beta_{pz}\frac{\partial P}{\partial r_z} + \beta_{oz}\frac{\partial\Omega}{\partial r_z} - \frac{z\zeta\Omega^{-2}}{(1-d\rho)(m_i - \Omega_i^{-1} - \sigma_\epsilon^2)} \tag{A.122}$$

where

$$\beta_{pz} = \varphi\frac{\phi}{\sigma}P^{\frac{\phi}{\sigma}-1}\Omega^{-2}A^{d\rho}k^{d\rho}\zeta\frac{1}{1-d\rho} \tag{A.123}$$

$$\beta_{oz} = r_\omega f''(\Omega) - 2\varphi P^{\frac{\phi}{\sigma}}\Omega^{-3}A^{\rho}k^{d\rho}\zeta + \frac{\varphi\Omega^{-4}A^{\rho}d\rho\zeta}{(1-d\rho)(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)k^{d\rho}P^{\frac{\phi}{\sigma}}} \tag{A.124}$$

It is easy to check $\beta_{oz} < 0$ under model's assumptions. Using Equation A.32 and Equation A.12, we can deduce that

$$\frac{r_\omega}{\zeta} = (1-\delta)\frac{r_k-\Gamma}{(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)d\rho}\Omega^{-2}k + (1-\delta)\frac{r_\omega}{\zeta}f'(\Omega) \tag{A.125}$$

Substituting Equation A.125 into the LHS of Equation A.122, we have

$$(1-\delta)^{-1} - f'(\Omega) - \frac{z\zeta\Omega^{-2}r_z}{(1-d)(1-d\rho)(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)r_\omega} = \frac{\zeta}{r_\omega}\frac{\Omega^{-2}k}{(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)d\rho}\left[(r_k-\Gamma) - \frac{\Gamma\rho}{1-d\rho}\right] \tag{A.126}$$

Using Equation A.12, β_{pz} and $\frac{z\zeta\Omega^{-2}}{(1-d\rho)(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)}$ in Equation A.122 can be simplified to

$$\beta_{pz} = \frac{\phi}{\sigma}\frac{r_k-\Gamma}{(1-d\rho)P}\frac{\zeta\Omega^{-2}k}{(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)d\rho} \tag{A.127}$$

$$\frac{z\zeta\Omega^{-2}}{(1-d\rho)(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)} = \frac{(1-d)\rho}{1-d\rho}\frac{\Gamma}{r_z}\frac{\zeta\Omega^{-2}k}{(m_i-\Omega_i^{-1}-\sigma_\epsilon^2)d\rho} \tag{A.128}$$

Substituting Equation A.127 and Equation A.128 into Equation A.122, we have

$$(r_k-\Gamma)\frac{1}{r_z}\left[\epsilon_{r_\omega,z} - \frac{\phi}{\sigma}\frac{1}{1-d\rho}\epsilon_{p,z}\right] + \frac{\rho}{1-d\rho}\Gamma\frac{1}{r_z}\left[1-d-\epsilon_{r_\omega,z}\right] = \beta_{oz}\frac{\partial\Omega}{\partial r_z} \tag{A.129}$$

The second term at the LHS is positive by the proof of Proposition 23. It is easy to find that the first term at the LHS is negative if d and ρ are large enough. Therefore, if τ_ξ is large, LHS is positive because the second term dominates the first term. It implies that $\frac{\partial\Omega}{\partial r_z} < 0$ for firms in high predictive accuracy industries. On the contrary, if τ_ξ is small, LHS is negative. It implies that $\frac{\partial\Omega}{\partial r_z} > 0$ for firms in low predictive accuracy industries. This concludes the proof. □

Proof of Proposition 30

Proof. We first compute the total derivative of $V(\Omega, P, r_\omega, r_z)$ with respect to r_z

$$\frac{dV(\Omega, P, r_\omega, r_z)}{dr_z} = \frac{\partial V(\Omega, P, r_\omega, r_z)}{\partial r_z} + \frac{\partial V(\Omega, P, r_\omega, r_z)}{\partial P}\frac{\partial P}{\partial r_z} + \frac{\partial V(\Omega, P, r_\omega, r_z)}{\partial r_\omega}\frac{\partial r_\omega}{\partial r_z} + \frac{\partial V(\Omega, P, r_\omega, r_z)}{\partial \Omega}\frac{\partial \Omega}{\partial r_z} \tag{A.130}$$

By the proof of Proposition 21, we know that the first two terms decrease in τ_ξ and is positive when τ_ξ is small enough. By Proposition 23 and Proposition 25, we know $\frac{\partial r_\omega}{\partial r_z} > 0$ and $\frac{\partial \Omega}{\partial r_z} > 0$. So for firms with small τ_ξ, the sign of the derivative is always positive. Now, we look at the case where τ_ξ is large. Notice that

$$\zeta\omega = Ak\left(\frac{z}{k}\right)^{1-d}\tau_\xi + f(\Omega) - \Omega < z\frac{r_z\zeta}{r_\omega}\frac{1}{1-d} \tag{A.131}$$

where we have used Equation A.11. Then we can derive an upper bound for ω:

$$\omega < \frac{r_z}{r_\omega(1-d)}z \tag{A.132}$$

Hence, we have

$$\frac{r_\omega}{r_z}\frac{\partial V(\Omega, P, r_\omega, r_z)}{\partial r_\omega} < \frac{r_z}{\delta r_\omega(1-d)}z\frac{\partial r_\omega}{\partial r_z} = \frac{z\epsilon_{r_\omega,z}}{(1-d)} < \frac{z}{\delta} \tag{A.133}$$

where we have used the proof of Proposition 24:($\epsilon_{r_\omega,z} < 1-d$) in the last inequality. Substituting Inequality A.133 into Equation A.130, we then have

$$\frac{dV(\Omega, P, r_\omega, r_z)}{dr_z} < \frac{z}{\delta}\left[-1 + \frac{\epsilon_{r_\omega,z}}{1-d}\right] + \frac{\partial P}{\partial r_z}\left[\frac{\partial V(\Omega, P, r_\omega, r_z)}{\partial P} - \frac{\omega \mathrm{E}(\frac{\partial\hat{\omega}}{\partial P})}{\delta \mathrm{E}(\frac{\partial\hat{\omega}}{\partial r_\omega})}\right] + \frac{\partial V(\Omega, P, r_\omega, r_z)}{\partial \Omega}\frac{\partial \Omega}{\partial r_z} \tag{A.134}$$

Then we can observe that the first two terms decrease in τ_ξ. If τ_ξ is big, then the sum of first two terms are negative. The last term is also negative when τ_ξ is big. Therefore, for firms with large τ_ξ, $\frac{dV(\Omega,P,r_\omega,r_z)}{dr_z} < 0$.

□

Proof of Proposition 31

Proof. Substituting Equation 3.8 into Equation 3.7, we can calculate type ν firm's profit

$$\pi(\nu) = \frac{p_b(\nu)^2}{2\nu^4\phi(\alpha n_b)} - c_e \tag{A.135}$$

Because we allow free entry of firms in the market, in the equilibrium, π_ν must be zero. Equating π_ν to 0, we can solve for $p_b(\nu)$. □

Proof of Proposition 32

Proof. Combining Equation 3.5 and Proposition 31, we can first get $\nu_b^u = \frac{1}{2c_b}$. This is the unconstrained optimal. Then the price of the product with quality ν_b^u is

$$p_b(\nu)(\nu_b^u) = \frac{1}{4c_b} \tag{A.136}$$

So in order to purchase type ν_b^u goods, a consumer needs at least endowment $W_b^c = \frac{1}{4c_b}$. If the consumer has fewer endowment, then her budget constrained must be binding. Then from price schedule $p_b(\nu)$, we can calculate her choice of quality $\nu_b^c = \sqrt{\frac{W}{c_b}}$. □

Proof of Proposition 33

Proof. Firstly, the demand for the goods with quality ν_b^u is $1 - F(W_b^c)$. According to Equation 3.8, single firm's supply of type ν_b^u products is

$$q_b(\nu_b^u) = \frac{\frac{1}{2}\nu^4 q^2 \phi(\alpha n_b)}{\nu_b^{u2}\phi(\alpha n_b)} = \frac{4\frac{1}{2}\nu^4 q^2 \phi(\alpha n_b)^3}{\phi(\alpha n_b)} = \frac{2c_e}{W_b^c} \tag{A.137}$$

By market clearing, the measure of firms producing ν_b^u type of goods must be

$$\frac{W_b^c(1 - F(W_b^c))}{2c_e} \tag{A.138}$$

For goods with quality ν lower than ν_b^u, denote the consumer's type who purchase it by $W(\nu)$. Then the demand density will be $f(W(\nu))$. Again, from Equation 3.8, we know that single firm's supply of type ν products is

$$q_b(\nu) = \frac{2c_e}{W(\nu)} \tag{A.139}$$

By market clearing, the density of firms producing ν type of goods must be

$$g_b(\nu) = \frac{W(\nu)f(W(\nu))}{2c_e} \tag{A.140}$$

□

Proof of Theorem 34

Proof. It is sufficient to prove that n_b is uniquely determined by Equation 3.9. We first differentiate the RHS of Equation 3.9 by W_b^c,

$$\frac{\partial}{\partial W_b^c}\left[\int_{\underline{W}}^{W_b^c} \frac{Wf(W)}{2c_e}dW + \frac{W_b^c(1-F(W_b^c))}{2c_e}\right] = \frac{1}{2c_e}(1-F(W_b^c)) > 0 \tag{A.141}$$

Then we differentiate W_b^c with respect to n_b,

$$\frac{\partial W_b^c}{\partial n_b} = -\frac{\alpha}{8\sqrt{2c_e}}\phi^{-\frac{2}{3}}\phi' = \frac{a\alpha}{8\sqrt{2c_e}}\left[(1+\alpha n_b)^{\frac{4}{3}} + a(1+\alpha n_b)^{\frac{1}{3}}\right]^{-\frac{3}{2}} \tag{A.142}$$

Equation A.141 and Equation A.142 imply that the right hand side of Equation 3.9 increases in n_b. Moreover, as n_b increases to ∞, the derivative of RHS of Equation 3.9 with respect to n_b monotonically decreases to 0. Therefore, the left hand of Equation 3.9 can only cross the RHS of Equation 3.9 once from below. n_b is unique. □

Proof of Proposition 35

Proof. Suppose data processing ability a increases. W_b^c increases as well because $\frac{1}{2}\nu^4q^2\phi(\alpha n_b)$ decreases. We know n_b increases in W_b^c. Therefore, n_b increases in a. If n_b increases, $\frac{1}{2}\nu^4q^2\phi(\alpha n_b)$ will decrease even more. The price schedule becomes flatter. It implies that all consumers' welfare should go up. The same argument applies when α increases. □

Proof of Proposition 36

Proof. We only derive the price schedule in the platform. Using the free entry condition and Equation 3.15, we have

$$\pi_{mi}(\nu) = \frac{(1-\theta)^2 p_{mi}^2}{2\nu^4 \phi(D_{mi})} - c_e = 0 \tag{A.143}$$

Then p_{mi} can be calculated easily from the above equation. □

Proof of Proposition 37

Proof. We first derive the consumers' unconstrained optimal choice ν_m^u. By the same argument as in the proof of Proposition 32, it is easy to deduce that

$$\nu_m^u = \frac{1}{2c_{mo}} \tag{A.144}$$

Then we can compute W_m^c. The price of ν_m^u type of products is

$$p_{mo}(\nu_m^u) = \frac{1}{4c_{mo}} \tag{A.145}$$

So $W_m^c = p_{mo}(\nu_m^u) = \frac{1}{4c_{mo}}$. Then in order to determine what types of consumers prefer to use BNPL platforms, we compute consumers' maximal welfare both from in-platform purchase and out-of-platform purchase. Suppose that a consumer purchases from an out-of-platform merchant. Her budget constraint must be binding and then we can compute her welfare

$$u_{mo} = \nu_m^c - W = \sqrt{\frac{W}{c_{mo}}} - W \tag{A.146}$$

Her choice of quality $\nu_m^c = \sqrt{\frac{W}{c_{mo}}}$ is directly from the budget constraint.

Suppose that a consumer purchases via the platform. Her unconstrained optimal choice ν_{mi}^u is

$$\nu_{mi}^u = \frac{1}{2c_{mi}} < \nu_m^u \tag{A.147}$$

The price of type ν_{mi}^{u} products is

$$p_{mi}(\nu_{mi}^{u}) = \frac{1}{4c_{mi}} \tag{A.148}$$

It implies that there exists a cutoff $W_{mi}^{c} = p_{mi}(\nu_{mi}^{u}) = \frac{1}{4c_{mi}}$ such that for consumers of type $W \geqslant W_{mi}^{c}$, they purchase products with quality ν_{mi}^{u} when borrowing via the platform. Their utility is

$$u_{mi}^{u} = \frac{1}{4c_{mi}} \tag{A.149}$$

For consumers of type $W < W_{mi}^{c}$, they will be constrained even if they borrow via the platform. Their choice of products then have quality $\nu_{mi}^{c} = \sqrt{\frac{kW}{c_{mi}}}$. Their utility is

$$u_{mi}^{c} = \sqrt{\frac{kW}{c_{mi}}} - kW \tag{A.150}$$

Let's compare the utility in two markets:

$$u_{mo} - u_{mi}^{u} = \sqrt{\frac{W}{c_{mo}}} - W - \frac{1}{4c_{mi}} \tag{A.151}$$

$$u_{mo} - u_{mi}^{c} = \sqrt{\frac{W}{c_{mo}}} - \sqrt{\frac{kW}{c_{mi}}} + (k-1)W = \sqrt{W}\left[\sqrt{\frac{1}{c_{mo}}} - \sqrt{\frac{k}{c_{mi}}} + (k-1)\sqrt{W}\right] \tag{A.152}$$

Equation A.152 shows us that $u_{mo} > u_{mi}^{c}$ if and only if $W > \frac{1}{(k-1)^2}\left[\sqrt{\frac{k}{c_{mi}}} - \sqrt{\frac{1}{c_{mo}}}\right]^2$. Denote $W_{m}^{p} = \frac{1}{(k-1)^2}\left[\sqrt{\frac{k}{c_{mi}}} - \sqrt{\frac{1}{c_{mo}}}\right]^2$. Then we will know if $W > W_{m}^{p}$, consumers prefer out-of-platform merchants. Otherwise, consumers prefer to use the BNPL services. Finally, we will prove $W_{m}^{p} < W_{mi}^{c}$.

$$
\begin{aligned}
& W_{mi}^c > W_m^p \\
\iff \quad & \frac{1}{4c_{mi}} > \frac{1}{(k-1)^2}\left[\sqrt{\frac{k}{c_{mi}}} - \sqrt{\frac{1}{c_{mo}}}\right]^2 \\
\iff \quad & \frac{1}{2}\sqrt{\frac{1}{c_{mi}}} > \frac{1}{k-1}\sqrt{\frac{k}{c_{mi}}} - \frac{1}{k-1}\sqrt{\frac{1}{c_{mo}}} \\
\iff \quad & \sqrt{\frac{c_{mi}}{c_{mo}}} > \sqrt{k} - \frac{1}{2}(k-1)
\end{aligned}
$$

The right hand side of the last inequality decreases in k and has upper bound 1. Because $c_{mi} > c_{mo}$, so the inequality always hold. Moreover, Equation A.152 shows us if $\frac{1}{c_{mo}} > \frac{k}{c_{mi}}$, then BNPL platforms cannot exist because out-of-platform merchants always provide higher utility. □

Proof of Proposition 38

Proof. We only need to prove the case where $\nu \leqslant \nu_{mi}^c$. The other two cases can be derived in the same way as in Proposition 33. Firstly, single firm's production scale is

$$
q_{mi}(\nu) = \frac{(1-\theta)p_{mi}}{\nu^4\phi(D_{mi})} = \frac{(1-\theta)kW}{(\frac{kW}{c_{mi}})^2\phi(D_{mi})} = \frac{2c_e}{k(1-\theta)W} \tag{A.153}
$$

where we use the fact that $p_{mi} = kW$ and $\nu_{mi}^c = \sqrt{\frac{kW}{c_{mi}}}$. Consumer's demand has density $f(W)$. Then we deduce that $g_m(\nu) = \frac{k(1-\theta)Wf(W)}{2c_e}$ by market clearing condition. □

Proof of Proposition 39

Proof. The first case is easy. If $k(1-\theta) > 1$, $n_m > n_b$. Out-of-platform prices are lower. So all consumers of type $W > W_m^p$ get benefits. Moreover, products offered in the platform also have lower price than the benchmark case. Therefore, all consumers welfare increases. Then let's look at the second case. Firstly, out-of-platform prices are now higher because total number of firms decreases. It immediately implies that consumers who are not using BNPL services get hurt. For consumers who are using the service, we compare the welfare difference,

$$u_{mi}^c - u_b = \sqrt{\frac{kW}{c_{mi}}} - \sqrt{\frac{W}{c_b}} - (k-1)W = \sqrt{W}\left[\sqrt{\frac{k}{c_{mi}}} - \sqrt{\frac{1}{c_b}} - (k-1)\sqrt{W}\right] \tag{A.154}$$

Obviously, there exists a cutoff $W_m^b = \frac{1}{(k-1)^2}\left[\sqrt{\frac{k}{c_{mi}}} - \sqrt{\frac{1}{c_b}}\right]$ such that consumers of type $W < W_m^b$ get benefit. Consumers of type $W > W_m^b$ get hurt. □

Proof of Proposition 40

Proof. We first show that $c_{mi} > c_{mo}$ given the assumptions in the proposition.

$$\frac{c_{mi}}{c_{mo}} = \sqrt{\frac{\phi(D_{mi})}{\phi(D_{mo})}}\frac{1}{1-\theta} \tag{A.155}$$

If $\varphi \to 0$, we have

$$\frac{\phi(D_{mi})}{\phi(D_{mo})} \to 1 \tag{A.156}$$

So there exists φ_0 such that if $\varphi < \varphi_0$, $\frac{c_{mi}}{c_{mo}} > 1$. Then we need to ensure that the platform can indeed exist. The condition for this is

$$\frac{k}{c_{mi}} > \frac{1}{c_{mo}} \tag{A.157}$$

which is pointed out in the proof of Proposition 37. So there exists b, such that as long as $k(1-\theta) > b_0$, the condition will be satisfied. □

Proof of Proposition 41

Proof. Similar to the proof of Proposition 32, we can first derive unconstrained consumers' optimal choice of quality must be $\frac{1}{2c_{mi}}$. The cost of this product is then $\frac{1}{4c_{mi}}$. So $W_{mp}^c = \frac{1}{4c_{mi}}$. For constrained consumers, their budget constraint is binding, so their choices are $\sqrt{\frac{kW}{c_{mi}}}$. □

Proof of Proposition 42

Proof. To prove the first claim, we can look at an extreme case. Suppose $\varphi = \alpha$. Then assuming $W_b^c = W_{mp}^c$. Then we can directly deduce that if $k(1-\theta) < 1$, $n_{mp} < n_b$. It then leads to $W_b^c > W_{mp}^c$. This will result in bigger n_b and smaller n_{mp}. So we will have $n_b > n_{mp}$ in this case. Now, we still need to verify that the platform price schedule is indeed flatter than that out-of-platform. Under the assumption $\varphi = \alpha$, we can firstly find that $\phi(D_{mi}) < \phi(D_{mo})$. That means, as long as θ are k are both small, the condition $c_{mi} < c_{mo}$ will be satisfied.

For the second case, if $k(1-\theta) > 1$ and $\varphi > \alpha$, we know $W_b^c < W_{mp}^c$ if $n_b = n_{mp}$. It then indicates that n_b must be smaller than n_{mp} because the right hand side of Equation 3.9 is strictly smaller than the right hand side of Equation 3.20. □

Proof of Proposition 44

Proof. We only need to prove that $c_{mi} < c_{mo}$. We already know that $\frac{\phi(D_{mi})}{\phi(D_{mo})} < 1$. Therefore, there must exist a θ_0, such that if $\theta < \theta_0$, $c_{mi} < c_{mo}$. □

Proof of Proposition 45

Proof. Because $\frac{\partial \phi(D_{mo})}{\partial \theta} = \frac{a\alpha}{(1+D_{mo})^2} = \frac{a\alpha}{(1+\alpha n_m)^2}$, it implies that

$$\lim_{\alpha \to 0} \frac{\partial \phi(D_{mo})}{\partial \theta} = \lim_{\alpha \to \infty} \frac{\partial \phi(D_{mo})}{\partial \theta} = 0$$

In the following proof, we only consider the case where α is close to 0. The other case can be done in a similar way. By Equation 37, We have

$$\lim_{\alpha \to 0, \varphi \to \infty} \frac{\partial W_m^p}{\partial \theta} = -\frac{2^{\frac{3}{4}} c_e^{-\frac{1}{4}}}{k-1} \sqrt{\frac{k W_m^p}{1-\theta}} \tag{A.158}$$

The first order condition for θ is

$$F(W_m^p) = \theta(1-\theta) f(W_m^p) \frac{\partial W_m^p}{\theta} \tag{A.159}$$

Substituting Equation A.158 into Equation A.159, we have

$$\sqrt{k(1-\theta)} = \frac{F(W_m^p)c_e^{\frac{1}{4}}(k-1)}{\theta f(W_m^p)\sqrt{W_m^p}2^{\frac{3}{4}}} \tag{A.160}$$

Now we derive a positive lower bound for θ. Because the platform must own non-negative profit, we have

$$\frac{\theta}{1-\theta}F(W_m^p) > c_p \tag{A.161}$$

Denote the measure of consumers by n_c. We then have

$$\frac{\theta}{1-\theta} > \frac{c_p}{2c_e n_c} \tag{A.162}$$

$$\Leftrightarrow \quad \theta > \frac{c_p}{2c_e n_c + c_p} \tag{A.163}$$

From Equation A.159 and Inequality A.163, we know there must exist k_1 such that if $k < k_1$, $\sqrt{k(1-\theta)} < 1$. Because W_m^p and $F(W_m^p)$ are bounded, then there exist k_2 such that if $k > k_2$, $\sqrt{k(1-\theta)} > 1$. □

Proof of Proposition 46

Proof. By Equation 3.23, we can get

$$\theta = \frac{\frac{c_p}{2c_e}}{\frac{c_p}{2c_e} + F(W_d^p)} \tag{A.164}$$

Then we have the following expression for $k(1-\theta)$

$$k(1-\theta) = \frac{kF(W_d^p)}{\frac{c_p}{2c_e} + F(W_d^p)} \tag{A.165}$$

Obviously, if c_p is very small. □

Proof of Proposition 47

Proof. We first compare the slopes of price schedules in both models. Denote p_{di} by the price schedule in the platform in the duopoly model. Then similar to the proof of Proposition 31, we know $p_{di} = c_{di}\nu^2$, where $c_{di} = \frac{\sqrt{2c_e\phi(D_{di})}}{1-\theta_d}$. D_{di} is the private data provided in the platforms. So

$$\frac{c_{di}}{c_{mi}} = \sqrt{\frac{\phi(D_{di})}{\phi(D_{mi})}}\frac{1-\theta_m}{1-\theta_d} \tag{A.166}$$

If platforms' data provision ability φ is very large, then $\phi(D_{di})$ is very close to $\phi(D_{mi})$. It implies that the price schedule in the duopoly model should be flatter than the price schedule in the monopoly case because $\theta_m > \theta_d$. So we know consumers who are using the platforms must be better in the duopoly model. If public data provision ability a is either very small or very large, then the marginal type of consumers who are using the platform must decrease in performance fee θ. Therefore, $W_d^p > W_m^p$ by fixing the number of firms. Then from Equation 3.16 and Equation 3.24, we can immediately get $n_{di} > n_{mi}$. Moreover, because n_m increases in W_m^p and n_d increases in W_d^p, from Equation 3.17 and Equation 3.25, we can get the conclusion that $n_d > n_m$. So consumers who are not using the platforms also get better in the duopoly case.

Then we prove the second result. Suppose c_p is large. Then it implies that the difference between θ_d and θ_m is small. Because a is small, large total number of firms cannot decrease firms' marginal cost to a large. D_{di} and D_{mi} mainly depends on n_{di} and n_{mi}. If we fix $n_{mi} = n_{di}$, then from Equation A.166, we have $c_{di} > c_{mi}$. This implies $W_m^p > W_d^p$. Therefore, $n_{mi} > n_{di}$. It again implies that $c_{di} > c_{mi}$. Hence, consumers who are using the platforms get hurt in the duopoly model. However, because $k(1-\theta_m)$ and $k(1-\theta_d)$ are smaller than 1, from Equation 3.17 and Equation 3.25, we know total number of firms n_d and n_m decreases in W_d^p and W_m^p respectively. Therefore, we can reach the conclusion that $n_d > n_m$. Consumers who are not using the platforms get better in the duopoly case. □

Bibliography

Abis, S., & Veldkamp, L. (2021). *The changing economics of knowledge production* [Working Paper], Columbia University.

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, *14*(4), 218–56.

Affirm. (2021). *Welcome to the affirm investor forum*.

Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, *60*, 323–351.

Alcazar, J., & Bradford, T. (2021). *The appeal and proliferation of buy now, pay later: Consumer and merchant perspectives* [kcFED Payments System Research Briefing], Federal Reserve of Kansas City.

Allen, F., Gu, X., & Jagtiani, J. (2021). A survey of fintech research and policy discussion. *Review of Corporate Finance*, *1*, 259–339.

Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. *Journal of Monetary Economics*, *97*, 71–87.

Benton, M., Meier, S., & Sprenger, C. (2007). *Overborrowing and undersaving: Lessons and policy implications from research in behavioral economics* [Federal Reserve Bank of Boston Discussion Paper 07-4].

Berg, T., Burg, V., Gombovic, A., & Puri, M. (2019). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies*, *33*, 2845–2897.

Berg, T., Fuster, A., & Puri, M. (2021). *Fintech lending* [Annual Review of Financial Economics (forthcoming)], Swiss Finance Institute.

Bergemann, D., & Bonatti, A. (2019). Markets for information: An introduction. *Annual Review of Economics*, *11*, 85–107.

Bergemann, D., Bonatti, A., & Gan, T. (2022). The economics of social data. *RAND Journal of Economics*, *53*(2), 263–296.

Bergemann, D., Bonatti, A., & Smolin, A. (2018). The design and price of information. *American Economic Review*, *108*(1), 1–48.

Bergemann, D., Brooks, B., & Morris, S. (2015). The limits of price discrimination. *American Economic Review*, *105*(3), 921–57.

Bowers, D. (2020). Top 50 research & data analytics industry report in the u.s. 2020.

Bronnenberg, B. J., Dubé, J.-P., Gentzkow, M., & Shapiro, J. M. (2015). Do pharmacists buy bayer? informed shoppers and the brand premium. *Quarterly Journal of Economics*, *130*, 1669–1726.

Buchak, G., Matvos, G., Piskorski, T., & Seru, A. (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, *130*(3), 453–483.

Caillaud, B., & Jullien, B. (2003). Chicken and egg: Competition among intermediation service providers. *RAND Journal of Economics*, *34*(2), 309–328.

Consortium, C. C. (2012). Challenges and opportunities with big data.

Crouzet, N., & Eberly, J. C. (2021). *Understanding weak capital investment: The role of market concentration and intangibles* [Working Paper].

Davis, S. J., & Haltiwanger, J. C. (2021). *Dynamism diminished: The role of housing markets and credit conditions* [Working Paper], National Bureau of Economic Research.

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, *26*, 297–308.

Eeckhout, J., & Veldkamp, L. (2021). *Data and market power* [Working Paper].

Farboodi, M., Mihet, R., Philippon, T., & Veldkamp, L. (2019). Big data and firm dynamics. *AEA Papers and Proceedings*, *109*, 38–42.

Farboodi, M., & Veldkamp, L. (2021). *A growth model of the data economy* [Working Paper].

Farboodi, M., & Veldkamp, L. (2022). *Data and markets* [Accepted, Annual Review of Economics].

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*, 37–54.

Fuster, A., Goldsmoth-Pinkham, P., Ramadorai, T., & Walther, A. (2021). Predictably unequal? the effects of machine learning on credit markets. *Journal of Finance*, *77*(1), 5–47.

Fuster, A., Plosser, M., Schnabl, P., & Vickery, J. (2019). The role of technology in mortgage lending. *Review of Financial Studies*, *32*(5), 1854–1899.

Gaur, V., Kesavan, S., Raman, A., & Fisher, M. L. (2007). Estimating demand uncertainty using judgmental forecasts. *Manufacturing & Service Operations Management*, *9*, 351–595.

Goldstein, I., Jiang, W., & Karolyi, G. A. (2019). To fintech and beyond. *The Review of Financial Studies*, *32*(5), 1647–1661.

Heidhues, P., & Kőszegi, B. (2010). Exploiting naïvete about self-control in the credit market. *American Economic Review*, *100*(5), 2279–2303.

Ichihashi, S. (2020). Online privacy and information disclosure by consumers. *American Economic Review*, *110*, 569–595.

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, *110*, 2819–2858.

Jorgenson, D. W. (1963). Capital theory and investment behavior. *American Economic Review*, *53*, 247–259.

Jorgenson, D. W. (2001). Information technology and the u.s. economy. *American Economic Review*, *91*(1), 1–32.

Jovanovic, B., & Nyarko, Y. (1996). Learning by doing and the choice of technology. *Econometrica*, *64*(6), 1299–1310.

Jovanovic, B., & Rob, R. (1987). Demand-driven innovation and spatial competition over time. *Review of Economic Studies*, *54*(1), 63–72.

Kirpalani, R., & Phillippon, T. (2020). *Data sharing and market power with two-sided platforms* [Working Paper], National Bureau of Economic Research.

Kirpalani, R., & Philippon, T. (2021). *Data sharing and market power with two-sided platforms* [Working Paper].

Kwon, S. Y., Ma, Y., & Zimmermann, K. (2021). *100 years of rising corporate concentration* [SAFE Working Paper].

Loecker, J. D., Eeckhout, J., & Unger, G. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics*, *135*, 561–644.

Maggio, M. D., & Yao, V. (2021). Fintech borrowers: Lax-screening or cream-skimming. *Review of Financial Studies*, *34*(10), 4565–4618.

Ridder, M. D. (2021). *Market power and innovation in the intangible economy* [Working Paper], LSE.

Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, *1*(4), 990–1029.

Romer, P. M. (1990a). Are nonconvexities important for understanding growth. *American Economic Review*, *80*(2), 97–103.

Romer, P. M. (1990b). Endogenous technological change. *Journal of Political Economy*, *98*(5), S71–S102.

Uhlig, H. (1996). A law of large numbers for large economies. *Economic Theory*, *8*, 41–50.

Ziegler, T., Shneor, R., Wenzlaff, K., Suresh, K., de Camargo Paes, F. F., Mammadova, L., Wanga, C., Kekre, N., Mutinda, S., Wang, B. W., Closs, C. L., Zhang, B., Forbes, H., Soki, E., Alam, N., & Knaup, C. (2021). *Fintech lending* [The 2nd Global Alternative Finance Market Benchmarking Report].