The Value of Identity: Measuring the Cost of Privacy

Aparajithan Venkateswaran   Brendan Palmer

Johann Kailey-Steiner

Spring 2018, MCM Submission
University of Colorado, Boulder

Abstract

Private information (PI) can reveal a person's personality, ideas, interests, and identity. Social media networks like Facebook are currently using this information to target advertisements and netting large sums of money. However, there is no system in place for the owner of this information to receive financial compensation for his/her information. To explore the value of PI, we created a model that illustrates the trade of PI as a common good.

We present a model that determines the price of PI, treating it as a commodity in a free market. Our model:

- Develops a price point accounting for the risks and benefits associated with sharing PI
- Captures the compounding risk when multiple pieces of information are disclosed
- Considers the forces of supply and demand for PI in a free market to determine the ideal price and quantity of PI that benefits the buyer and seller
- Effectively captures the network effect of data sharing under the assumption that human data is highly correlated and we can infer information about other people from a single person

Our model generalizes the calculation of the price of PI in the context of five domains of PI: Social Media, Financial Transactions, Health Records, General Identification, and Tracking. Within each of these domains our model assumes distinct records of information such as name, birthday, a post, etc. each having unique risk values associated with them.

Incorporating such distinct values makes our model robust as it can be extremely accurate and easily scalable to new domains and records types. Our model also captures the interconnectivity of human information by calculating ranks of individuals in a community and determining how connected a community is.

1 Introduction

In 2017, the average adult in America spent 5 hours and 50 minutes on digital media per day [1]. For a large percentage of that time people are divulging personal information (PI). This could be browser search data, photos, posts, likes, comments, credit card information, etc. When people buy items online they share their names, emails, credit card information, and address'. New features on cellphones allow users to transfer money via text requiring another layer of protection against cyber crime. A person's PI can reveal more than just their identity, including their opinions, interests, friends, family, and financial situation. While criminals try to exploit the person by using their PI, this can also be used in many different types of businesses for profitable gain. For instance, it can be used to study the behavior patterns of people in a society and therefore detect anomalies such as crime and terrorism.

1.1 Problem Summary

- We need to develop a price point for PI that takes into account the risks and benefits involved in sharing data with an unknown third party.
- With the help of the price point, we need to create a pricing structure for PI.
- Using this pricing structure, we need to develop a pricing system that treats PI as a commodity that could be traded.
- The model we develop should also consider the fact that human data is extremely linked and highly correlated. The model should effectively capture the network effects of data sharing.
- We also need to consider the political, cultural and ethical implications of PI being available for sale.

1.2 Our Model

- To create a price point for PI, we took a weighted average approach. We accounted for characteristics (such as education, age, etc.) that are most relevant to each specific facet of PI (social media, finance, general ID, etc.) and factored in the risk associated with people sharing their PI depending on the characteristics.
- Using this price point, we developed a pricing structure that depends on the actual value of each PI record (name, birthday, bank information, etc.). With this pricing structure we turned PI into a commodity and brought in forces of supply and demand for PI under the assumptions of a free market.
- To effectively capture the network effects of data sharing, we used network ranking algorithms (particularly PageRank [2]) to determine how much influence a person has in their society. We factored this into our pricing structure while also keeping in mind how connected the network is.

Since our model works under the assumptions of a free market and obeys the laws of microeconomics, our model can theoretically scale very well to accommodate future technological innovations, or changes in government regulations, in regards to PI.

The rest of this report is organized as follows: Section 2 discusses why we need to consider PI uniquely from personal property and intellectual property; Section 3 discusses in detail the assumptions under which we design our model; Section 4 develops our model from scratch; Section 5 discusses how our model can be effectively put into practice; Section 6 discusses the constraints and limitations of our model and tests our model under dynamic and unpredictable events; Section 7 considers the political, cultural and ethical implications of trading PI in the market; and Section 8 concludes our examination of our model. This report is finally followed by a policy recommendation to the decision makers on how to design policies around allowing PI to be a legal commodity.

2 Background

2.1 The Need to Consider PI Uniquely from PP and IP

When we approach the issue of treating PI as a commodity, it is important to compare PI to personal property (PP) and intellectual property (IP). This will help us identify why we need to treat PI as a separate entity from PP or IP. PI, unlike PP and IP, is a combination of both tangible and intangible data. This makes the handling of PI as a commodity complicated. Much like IP, PI is fundamentally social i.e., it has little intrinsic value of its own. It is the person and the social interactions of the human society that make PI valuable. And just like IP, PI can be easily shared between people making PI extremely vulnerable to theft.

The most important similarity between PI, PP, IP is their profitability. The lucrativeness of PP and IP is clear in present society. The current use of PI for financial gain is less apparent. The best example of this is in social media. Companies like Facebook track people and use their PI for targeted advertising and make a profit larger than $16 per user per year [3].

If PI was treated as a commodity similar to PP or IP, buying and selling PI would be an incredibly profitable business. This is the biggest reason why it is important to estimate the cost of one's privacy.

While doing so, it is also important to consider the political, cultural and ethical issues surrounding the idea of treating PI as a tradeable good. We will discuss this in more detail in Section 7.

3 Assumptions

For the creation of any model, it is necessary to make assumptions to simplify the model into an effective and efficient representation of the problem at hand. Our model is no different and is built upon necessary assumptions regarding policy and personal behavior.

3.1 The Free Market Assumption

The strongest assumption made to develop our model is that PI is legally bought and sold in a free market i.e., there are no government regulations such as taxes and fixed price laws. This also assumes that there are no international trade regulations such as tariffs and import duties, privacy protection regulations, etc. We make this assumption to facilitate the development of our model. Later in this report we will see that our model is in fact consistent with the laws of microeconomics thereby making it possible to relax this assumption and introduce government regulations and international trade.

3.2 The Finiteness of Variables Assumption

Our model assumes that there are only a finite number of facets of PI, which we will henceforth call *domains*. The domains we will consider in this model are Social Media, Finance, Health, General ID, and Tracking. Further, we will also assume that each of these domains has a unique set of records that are also finite. Table 1 lists the different records of each of these domains.

Along with this, our model assumes that the risk and benefits of sharing PI to a third party will depend only on a finite number of *characteristics*, uniquely determined for each person. In our model we will consider only the following characteristics: Connectivity (how socially connected a person is), Age, Education level, Social Class, and Quality of Life. We will later see how we can calculate these values.

Our model also assumes that the factors that affect the risk associated with each domain are also unique. This allows us to more specifically calculate benefits and risks based on the characteristics depending on which are believed to have more of an impact in each domain.

3.3 The Legality Assumptions

Our model also assumes that all records contained in the domains are available for sale upon agreement between buyer and seller, meaning that government regulation or outside influences don't force certain *records* from being sold. This is a necessary assumption because it allows us to create a model that follows from a free market scenario where we can consider the forces of supply and demand.

In the current scenario, PI, particularly tracking data, is often given away to corporations for free. The ability to obtain PI without a purchase would not work well under our free market assumption. Therefore, we make the assumption that services cannot use data without legally purchasing the data from its owner in a free market.

3.4 The Compounding Risk Assumption

In creating our model, we included the compounding risk when selling multiple records together. However, we assume that the privacy lost (and thereby extra information gained by the buyer) by selling additional items would compound at the same rate per additional record sold i.e. γ (discussed in detail in Section 4). We believe that the assumption that the compounded risk does not change between different sets of records makes our model more concise without losing accuracy.

3.5 The Future of Technology Assumption

Another assumption important to our model is that technology, as we know it, does not change abnormally as it did in the late 1900's. Such a revolution in technology can drastically change the meaning of PI. Therefore, this is a reasonable assumption because we want our model to predict what is most likely to happen rather than trying to predict the abnormal.

4 Model

First, we will develop a price point for protecting one's PI based on risks and trade-offs. Using this price point, we will develop a pricing structure and system that will turn PI into a commodity. Finally, we will attempt to capture the network effects of sharing PI given that human data is highly linked and correlated.

4.1 A Price Point

To develop a price point, we decided to quantify the complete risk associated with each person having their information leaked. This is particularly hard because we are dealing with subjective risk [4]. So, we modeled the risk to be dependent on the domain of PI in question and the characteristics of the person. We ranked the importance of different characteristics for a particular domain. Then, we assigned a normalized score for each characteristic to quantify the risk associated with that characteristic having that domain of PI leaked. To quantify the characteristics of the person, we used the measures as outlined in Table 3. And thus, we modeled the risk associated with each person having their information from domain D leaked, R_D as:

$$R_D = \langle \vec{\alpha_D}, \vec{c} \rangle \tag{1}$$

where $\langle\ ,\ \rangle$ represents the Euclidean dot product; $\vec{\alpha_D}$ is a vector of risk values for information from D being leaked, $\sum_{i=1}^{n} \alpha_{Di} = 1$, and $\alpha_{Dk} = \frac{r_k}{\sum_i r_i}$ where r_i is the particular risk associated with characteristic c_i; and $\vec{c}$ is the vector values of a each person's characteristics as measured from Table 3. This risk value, quantified by R_D, forms the basis of our price point.

4.2 A Pricing Structure

Now, we look at the actual value a single raw piece of information is worth. This value represents just the worth of the data and does not take into account the risk or profits, and it differs not only from domain D_1 to D_2, but also varies with each record i in D_1. Let us call this raw value $P_{raw}^{(i)}$ to denote that this varies with each record i. From the point of the person selling this piece of information, we also need to account for their risk. And this risk would increase the value of $P_{raw}^{(i)}$. Note that R_D has no units. Therefore, we can define the price at which someone would sell a single piece of information from record i as:

$$P^{(i)} = (1 + R_D)P_{raw}^{(i)} \tag{2}$$

To calculate the total cost of n different pieces of information, our natural instinct would be to add them individually. For instance, let the set of all information being sold be denoted by X. Then

$$P_{tot} = \sum_{i \in X} P^{(i)} \tag{3}$$

However, equation 3 does not account for the fact that we can get more value out of two different pieces of information put together than having them separately. For instance, bank information and social security alone is worth less than having them together. And by divulging multiple pieces of information, a person is putting themselves at more risk. Let us denote this compounding risk factor by γ. By selling the n-th piece of information, this information is worth γ^{n-1} times more than what it would have been worth individually. Naturally, $\gamma > 1$. Now, we can write the total cost as:

$$P_{tot} = \Big(\sum_{j=1}^{|X|} \gamma^{j-1}\Big) \sum_{i \in X} P^{(i)} \tag{4}$$

Notice that equation 4 grows rapidly (the order of $O(\gamma^n)$). To fix that, we redefine $\gamma = 1 + \epsilon$, ϵ is a small positive real number, and we divide out the compounded risk coefficient by the number of items bought, $|X| = n$. Therefore, we have modeled the cost of PI from the perspective of the person selling their data as:

$$P_{tot} = \frac{1}{n} \sum_{j=1}^{n} \gamma^{j-1} \sum_{i \in X} P^{(i)} \tag{5}$$

4.3 A Pricing System

4.3.1 Supply and Demand

Now that we have established a pricing structure that would measure the cost of privacy for each individual, we can now bring PI into the market as a commodity. And by doing so, we can begin considering the forces of supply and demand. The supply curve represents the

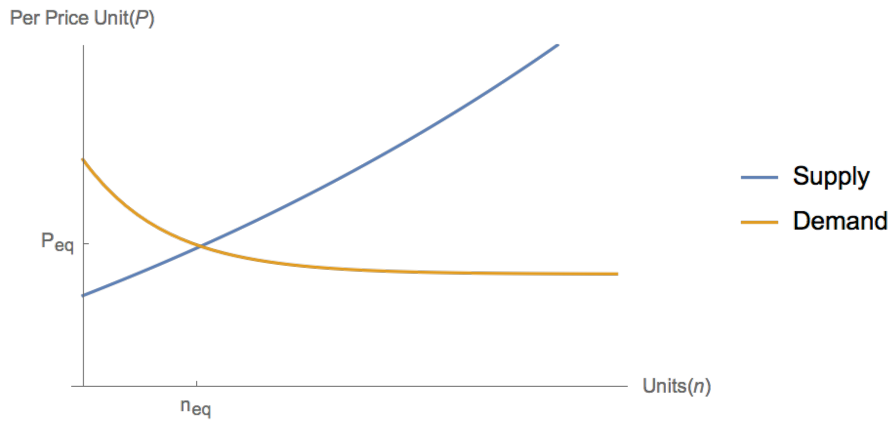

Figure 1: *Approximation of the Supply-Demand curves for PI under a free market*

price per unit of PI at different quantities. For the supply of PI, since the people who are selling their PI are suppliers, we can modify equation 5 to determine the cost per unit:

$$\begin{aligned}\text{Cost per unit} &= \frac{P_{tot}}{n} \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\gamma^{j-1}\sum_{i\in X}P^{(i)}\end{aligned} \tag{6}$$

where n is the number of pieces of PI being sold. Notice that this curve is discrete and difficult to plot considering that each person has their own R_D and it varies with each record i. Also notice that the total price is mainly dependent on the compounding risk γ. So, to make our math cleaner and easier to understand, we can make a generalized assumption here and set all $P^{(i)} = P$, where P is a constant. Therefore, our supply, S, can *approximately* be modeled as:

$$\begin{aligned}S &= \frac{1}{n^2}\Big(nP\Big)\sum_{j=1}^{n}\gamma^{j-1} \\ &= \frac{P}{n}\frac{\gamma^n - 1}{\gamma - 1}\end{aligned} \tag{7}$$

Now, we consider the demand from the perspective of the buyer. With each piece of information the buyer has, let us assume they can make at least a net profit, p_o. We also know that each piece of raw information is worth $P_{max}^{(i)}$. The principle of diminishing marginal utility applies here i.e., the increased value the buyer gets from purchasing the n-th piece of information is not as large as the increased value the buyer gets from purchasing the first piece of PI. Therefore, as n increases, the buyer should be willing to pay less for each PI. However it will reach a limiting value of p_o as the buyer can make at least that much profit from each PI. Therefore, we can model our buyer's profit per unit as an exponential function with limiting value p_o.

$$\text{Profit per unit} = (P_{max}^{(i)})^{-n} + p_o \tag{8}$$

Once again, notice the difficulty in calculating this value for varying i. Therefore, we can simplify the problem by setting $P^{(i)}_{max} = P'$, P' is a constant, and model the demand, D *approximately* as:

$$D = (P')^{-n} + p_o \tag{9}$$

Figure 1 shows how the supply and demand curves interact with each other. Under the assumptions of the free market, trade of PI would take place at the values indicated by the intersection of these curves i.e., n_{eq} would be the quantity of PI traded, each of which will be sold at a price of P_{eq}.

4.3.2 The Self-Worth Factor

We initially assumed that we are dealing with a free market. Since people have control to sell their PI in a free market, we can also account for the fact that people can choose to sell their PI at a difference price depending on their how much they think their data is worth. Note that the new value can either be less than the original price or greater than the original price depending on whether the sellers value themselves higher or lower. So, we can define a new vector $\vec{\omega_D}$ where $\omega_{Dk} = \frac{v_k - r_k}{\sum_i v_i}$. Here, r_k has the same definition as earlier, and v_k is the value of the domain D as perceived by the person under consideration in regards to the corresponding characteristic k. Then we can determine the worth factor as:

$$W_D = \langle \vec{\omega_D}, \vec{c} \rangle \tag{10}$$

Note that W_D can either be negative or positive depending on whether the person values their data lower or higher than the actual risk associated with it. Now, we can redefine our $P^{(i)}$ as:

$$P^{(i)} = (1 + R_D + W_D)P^{(i)}_{raw} \tag{11}$$

This naturally ensures that if a person values their data lower, then the price at which they will sell their data is also lower than the original price determined by equation 2. Similarly, the inverse of that statement also holds true.

Therefore, if people start valuing their data higher, then the supply curve will shift up and if people value their data lower, then the supply curve will shift down.

4.4 Capturing Network Effects of Data Sharing

It is quite obvious that human data is highly linked and information obtained from one individual can be used to infer the behavior and details of other people who are close to them.

We can represent people as a graph with vertices representing people and the edges representing the connection between people. The weight of the edges would represent how highly correlated two people are, a higher value indicating a greater correlation. In practice, we would compute these weights by taking into account the number of interactions between two people.

We can then translate this graph into an adjacency matrix A where a_{ij} represents the correlation between person i and person j. Due to the way we have defined our edge weights, $a_{ij} = a_{ji}$ i.e., A is a symmetric matrix. We set $a_{ii} = 0$ because interactions with oneself does not count as a correlation. We also normalize these values such that $\sum_j a_{kj} = 1$.

Now, we need a measure of how much we can learn about the entire network if we get information from some person k. There are many existing algorithms that can help us calculate this. Perhaps, the most useful of them are the ranking algorithms that determine the rank of a node in such a network. Here, we will use the PageRank algorithm [2], [5] to determine how important or how influential a person k is in a given network using the number of interactions between k and others in the network. If we denote this value by ρ_k, then the value of the PI gets compounded by ρ_k. Therefore we can rewrite the price of each unit of PI, from the perspective of the seller as:

$$P^{(i)} = (1 + R_D + W_D + \rho_k)P^{(i)}_{raw} \tag{12}$$

We immediately notice that this is going to shift the supply curve upwards.

From the buyer's perspective, the buyer gets an increased profit by learning more about the entire network with just a single piece of information. How much more they can learn from this information depends on how connected the network is. We can compute the connectivity of a graph [6] using either the edge-connectivity algorithm [7] or the vertex-connectivity algorithm [8] on the connected subgraph containing person k. It does not matter which algorithm we choose to use because both of these measures are equally effective on average. Let us call this value, computed using either of the two algorithms, κ. Once again the law of diminishing marginal utility says that the greater the amount of information about the network we already have, the smaller the marginal increase we get upon each subsequent purchase of information. So, we can represent κ, along with the diminishing marginal utility as $\frac{\kappa}{n}$. Thus, we can rewrite the profit per unit function and the approximated demand as:

$$\text{Profit per unit} = (P^{(i)}_{max})^{-n} + p_o + \frac{\kappa}{n} \tag{13}$$

$$D = (P')^{-n} + p_o + \frac{\kappa}{n} \tag{14}$$

Subsequently, we notice that the demand curve is also going to shift upwards thus increasing the final equilibrium price per unit under free trade.

5 Solutions

In our model we use the constants $\vec{\alpha_D}$, P_{max} and $\vec{c}$, which will be determined through studies of what determines, risk, worth, and value of PI. In this section, we attempt to estimate these values based on our current studies and surveys.

5.1 Calculating P_{max}

Our P_{max} variable determines the raw value of each record within each domain. We determined the value of each of these variables by researching the raw cost of buying the specific record. Our results are highlighted in Table 1.

Table 1: *P_{max} values for the different record we considered in our model*

Social Media	**Finance**	**Medical**	**General ID**	**Tracking (cost per day)**
Account Data $0.04 [9]	Card Info $0.50 [10] Bank $15 [14]	Disabilities $10 [11] Wellness Check $10 [11] Existing Conditions $10 [11] Sensitive Reports $50 [15]	Name $0.001 [12] Photo $0.001 [12] SSN $1 [10] Birthday $0.001 [12] Zipcode $0.001 [12] Ethnicity $0.001 [12]	Browsing $.067 [13] Location $0.67 [13] Webcam $0.67 [13]

5.2 Calculating $\vec{\alpha_D}$ and $\vec{c}$

In determining the risk and worth for specific individuals in each domain, we took a weighted average of characteristics c_i. The variables c_i will be discussed further on in the paper, as we will first address our $\vec{\alpha}$ value set. We used a set of $\vec{\alpha}$ values, whose inner product with the set of $\vec{c}$ values were taken to find the risk and worth equations for each domain. The only requirements of $\vec{\alpha}$ is that $\sum_i \alpha_i = 1$ and the values must reflect the weight of each characteristic. While $\vec{\alpha}$ values could be changed, our proposed values and corresponding c set for each domain is listed in 2.

Table 2: *$\vec{\alpha_D}$ values for the different records considered in our model*

Characteristic	**Social Media**	**Finance**	**Medical**	**General ID**	**Tracking**
Connectivity	0.6	0	0	0.2	0
Age	0.2	0.2	0.33	0.2	0.33
Social Class	0.2	0.5	0.33	0.2	0.33
Education	0	0.3	0	0.2	0
Quality of Life	0	0	0.33	0.2	0.33

For the variable $\vec{c}$, we created characteristics that factor into the risk value for each domain. The characteristic score is a normalized number, 1 being the most risk/worth possible and 0

Table 3: *Measures for various characteristics*

Characteristic	**Measure**
Popularity/Influence	Percentage of followers on social media
Age	$1 - \frac{\lvert age-(target\ age)\rvert}{age}$
Social Class	Percentile
Education	$1 - \frac{\lvert (education\ level)-(target\ education\ level)\rvert}{education\ level}$
Quality of Life	$1 - \frac{W.H.O\ Ranking}{totalnumberofcountries}$

being the least. Table 1 shows our suggested records for each domain. However modifying the characteristics is simple, given $\vec{\alpha}$ is adjusted accordingly. Table 3 shows how we calculate each of the different values.

5.3 Pricing for Individuals, Groups, and Nations

One advantage of our model is its ability to model pricing for all sizes of groups spanning from individuals to nations. For individuals, characteristic scores are given by comparing that individual's scores to that of the population, as described above. In this case, scores would be given based on the percentile of one person in comparison to the population.

In the case of pricing groups, we first identify communities in a society. There has been a tremendous amount of work done trying to detect communities [16]. We use such a community detection algorithm (SpringRank [17]) to identify groups and sort people into them. We, then, follow the same basic idea as the original model but this time we give the group characteristic percentiles by comparing that group to the population of other groups. Our model uses the average characteristic scores of all the group members and compares that to the scores of other groups to create a percentile.

In the case of nations, we treat the entire nation as one large group. We then directly compare the average characteristic scores of each nation to all other nations, creating a characteristic score, and thus a price for that nation. With this in mind, the most precise way to calculate price for a person is to use the price point with each individuals characteristics. If we choose to establish pricing by groups, precision would be lost when the group size increases owing to the fact that as group sizes increase, the number of outliers (extreme cases) will also increase.

6 Sensitivity Analysis

6.1 Constraints of our Model

Our model faces three large constraints that might affect the computation expense:

(i) The greatest strength of our model is perhaps the fact that it can theoretically be scaled quickly to accommodate for multiple domains of PI each having multiple records, and to add new characteristics. However, this also forms the basis of our first constraint. By introducing additional domains, records, and characteristics, we are also introducing more variables and coefficients that need to be measured and calculated increasing the computational complexity of our model. This especially becomes an issue when we are dealing with large societies like entire nations, or even the entire world.

(ii) In our model, we deal with each domain and characteristic individually and independently. In the trade-off between specificity (accuracy) and computation power, our model chooses specificity. Therefore, it will become expensive to use our model across multiple domains. However, notice that all of the domains and characteristics (of the individual) arise from the same equations making it easier to understand the model while not compromising on specificity.

(iii) Many of the variables and coefficients we use in our model, including R_D and W_D, cannot be determined objectively. They are highly subjective and can vary from person to person. Therefore, when we talk about using our model for entire communities and nations, computing the average across all people does not accurately reflect the nature of the community owing to the presence of a large number of outliers.

(iv) Since, our variables are subjective, it is reasonable to assume that they are not going to remain static. We will see in Section 6.2 how our model handles this dynamic nature. Although our model can capture the dynamism, it does not reflect these changes immediately.

6.2 Variations in Human Decision Making

Human decision-making about the worth of their own data is subjective and prone to change. They could value their data higher or lower than before due to different reasons such as change in the way technology and social media operate; change in government regulations and laws; change in belief about the future and various other subjective beliefs. Our model does not change drastically when this happens.

For instance, when people believe that their data is worth more, $P^{(i)}$ increases thus shifting the supply curve upwards. In the short term, this will increase the equilibrium price P_{eq} and decrease the quantity of PI traded. However, in the long run, to account for their new losses, the demand curve will shift down returning P_{eq} to its original value. This phenomenon is illustrated in Figure 2.

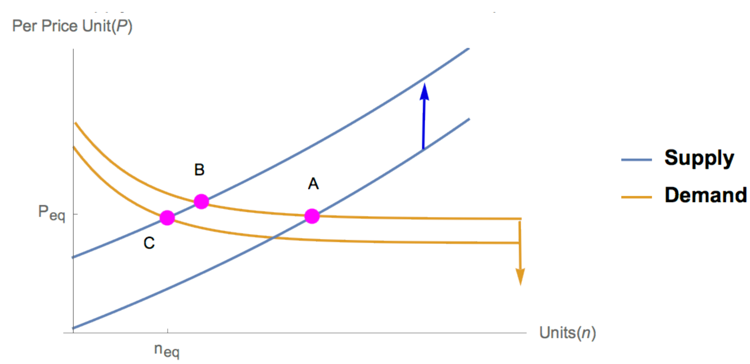

Figure 2: *When self-worth increases, supply increases leading to short term equilibrium B. In the long term, demand decreases going back to C with the original equilibrium price*

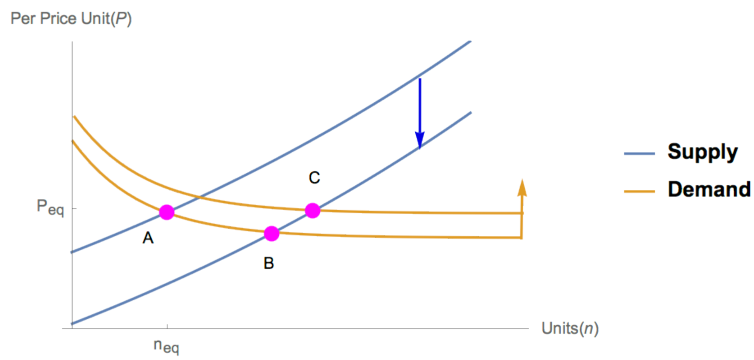

Figure 3: *When self-worth decreases, supply decreases leading to short term equilibrium B. In the long term, demand increases going back to C with the original equilibrium price*

And similarly, when people believe that their data is worth less, $P^{(i)}$ decreases thus shifting the supply curve downwards. In the short term, this will decrease the equilibrium price P_{eq} and increases the quantity of PI traded. However, in the long run, to increase their profits, the demand curve will shift up returning P_{eq} to its original value. This phenomenon is illustrated in Figure 3.

6.3 Massive Data Breach

When there is a massive data breach, two things are bound to happen:

(i) The buyers (people who buy and store PI) are going to face a loss

(ii) The sellers are going to value their PI at a higher level because sharing their PI makes them more vulnerable than they expected

Effect (i) is going to shift the demand curve downwards because of reduced profits. Effect (ii) is going to shift the supply curve upwards because people start believing that their PI is worth more. The net result of these effects is that the quantity of PI traded will will fall. However, it will be difficult to predict what will happen to the new P_{eq} without assigning

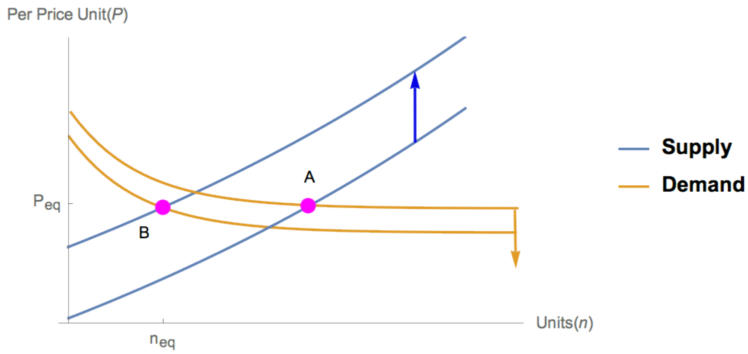

Figure 4: *When there is a massive data breach, supply will shift up because people value their PI more. Demand will shift down because buyers are facing a loss. Therefore, there is a net reduction in the quantity of PI being traded.*

concrete values to our variables. Figure 4 shows how the supply and demand curves shift. Section 7 discusses who is to be held responsible in such a scenario in detail.

6.4 Change in Government Regulations

So far, we have only assumed our trade to take place in a free market. In reality, no legal trade takes place in a free market i.e., the government will regulate the trade of PI by levying taxes on both the parties; and there will be tariffs imposed on international trade of PI.

Since our model is consistent with the principles of microeconomics, our model can easily adapt to trade of PI under government regulations. In short, the only thing that would change is that under government regulations, trade would happen at a different equilibrium price and quantity traded. We discuss how our model would change in detail in the Appendix A.

6.5 Generational Similarities and Differences

It was difficult to find information that would help quantify the risks and benefits people have when sharing their PI to a third party. So, to quantify this, we decided to look at the number of people on the internet categorized by age to calculate the generational trends in risk-benefit ratio. A research done by Pew [18] says that nearly 80% of the people who use social media are of age 40 or less. From this we can conclude that younger generations receive more benefits from sharing PI. Another research done by Pew [19] says that nearly 86% of internet users have taken measures to "remove or mask their digital footprints". It does not say anything about the age distribution of this part of the population. Given that most of the internet users are younger, it is safe to assume that the younger generations are more concerned about the risks of sharing PI than the older generations. Since both risks and benefits are higher for the younger generations, these data suggest that across generations, everyone's perceptions of the risk-benefit ratio is nearly the same.

Without considering the fact that technology had a greater effect on the younger generation, and using the data we collected earlier, it would seem that as generations age, both R_D and W_D in equation 11 reduce (as current older generations have lower values for both). This seems to suggest that as time passes, the supply curve for a particular generation would shift downwards. However, taking into account that the younger generations had greater exposure to technology, their R_D and W_D would stay more or less the same as time passes i.e., there would not be an appreciable change in the supply curve.

7 Political and Ethical Issues

7.1 Privacy Risks in a Community

The interconnectivity of people within a community brings up a number of ethical issues when PI is treated as a commodity. If a highly connected person agrees to sell his/her PI to a buyer, he/she in turn is selling little bits of other people's PI. This begs the question, "How should this extra data be dealt with by the buyer? Also, should the buyer need consent from every person who's data they are indirectly receiving?"

To answer this question we modify our model in equation 12 to include an indicator variable I_k that determines whether PI from a particular person should be available for trade. To do that let us first identify the number of people with whom person k is connected with.

$$c(i,j) = \begin{cases} 1, \text{if } a_{ij} > \delta \\ 0, \text{otherwise} \end{cases} \tag{15}$$

where δ is a small positive number that represents a threshold value for when two people i and j can be said to be "connected". This variable $c(i,j)$ tells us whether the data between person i and j are correlated.

We can also define a new variable $I(A_i)$ that indicates whether person A_i has given consent for their PI to be used or interpreted from other's PI.

$$I(A_i) = \begin{cases} 1, \text{if } A_i \text{ has given consent} \\ 0, \text{otherwise} \end{cases} \tag{16}$$

For small communities, where there are a large number of connections between people, we can determine the average number of connections a person k has as:

$$I_{avg}^{(k)} = \frac{\sum_j c(k,j) I(A_j)}{\sum_j c(k,j)} \tag{17}$$

Using this we can determine our I_k as:

$$I_k = \begin{cases} 1, \text{if } I_{avg}^{(k)} > \epsilon \\ \infty, \text{otherwise} \end{cases} \tag{18}$$

where ϵ represents the threshold at which the community decides as the majority i.e., the minimum number of people required to ensure that it is okay for the community to disclose information.

While dealing with entire nations, where the number of connections does not matter, we can calculate I_{avg} for the entire population as the expectation that $I(A_i) = 1$:

$$I_{avg} = E(I(A_i) = 1) \tag{19}$$

Then we can modify our $P^{(i)}$ as:

$$P^{(i)} = I_k(1 + R_D + W_D + \rho_k)P^{(i)}_{raw} \tag{20}$$

So, if the majority of the population does not give consent, the price of the item blows up to ∞ thereby making it impossible for anyone to buy it. Otherwise, the price does not change.

Since we have taken into account the choice of the majority of the community, our model assumes that in case of shared risks, the community takes the responsibility for the privacy of the entire community by determining the majoring fraction ϵ.

7.2 Consequences of Data Breaches

Another important question is, "When a person sells their PI to a buyer, whatever entity that may be, are they trusting the buyer to keep their information safe from a data breach and/or misuse? Who is liable for a data breach? If a seller's data is stolen from the buyer, should the seller be reimbursed?"

The way we have modeled our price $P^{(i)}$ makes it easier for us to answer this question. When we modeled our price, the first thing we took into account was the risk factor. This was defined as the risk associated with that piece of information getting leaked. Under a free market, trade takes place at an acceptable price to both parties, buyer and seller i.e., the seller of PI is already being compensated for the risk associated with the leaking. Since the seller is already reimbursed for this risk, it does not make sense for the buyer to reimburse the people whose data was stolen.

Misuse of data, on the other hand, is a little bit tricky. Although we have basked in the freedom of an unregulated market all this while, this is where government regulations play a huge role. Government regulations and contracts should be enforced such that the buyer agrees not to intentionally misuse the data.

7.3 Political Issues

Introducing a brand new type of commodity to the market comes with new challenges for the government. The first challenge would be to decide whether privacy of information is a fundamental human right. Given that private information can be misused to the gains of the criminal and the loss of the person whose information is being used, privacy of information is vital for a person to survive in the current world. In many cases, facets of PI constitute the identity of a person. Denying the right to own one's privacy is denying them to right to have an identity. This is why we believe that privacy of information should be a fundamental human right.

In addition, the transaction of PI has the potential of being incredibly profitable. Its profitability along with its difficulty to regulate could make the PI industry one that is constantly being taken advantage of and corrupted. How, if at all, should the buying and selling of PI be regulated by the government? This is an important question to be raised given that we have so far assumed a free market. This question was also raised in the previous subsection. We continue along that line of thought and concur that there should be certain laws and regulations that prevent intentional misuse of PI by the buyer. However, the government should not interfere too much with the trade of PI. There should be set laws that make sure that the PI of children younger than 13 should not be available for trade and children younger than 18 need to have their parent/guardian's consent and guidance before they can participate in the trade of PI. This is to ensure that buyers do not exploit children.

Should international PI trade be considered treason or tariffed? What laws should be put in place for how buyers could use PI? How can employers use the PI of employees? These are especially difficult questions that are very subjective and depend highly on the situation of the nation or person. Therefore, these are questions that would have to be dealt with by lawmakers to ensure the safety of a potential PI market.

8 Conclusion

8.1 The Importance of PI

One's personal information is one's identity. It can reveal elements of one's personality and interests as well as one's day-to-day life. As people spend more and more time on digital devices, they build up a PI profile that is valuable to both the owner of the PI and businesses that could use it for profit.

We have to deal with PI separately because, unlike IP and PP, it is a mixture of both tangible and intangible data. With the growth of digitization, the amount of PI is also increasingly growing with platforms like social media and electronic transactions. And with that is the growing risk of cyber crime. Noticing the potential in PI, we need to treat it independently with care because it raises many ethical and political questions, such as "Is privacy a human right?".

8.2 Our Model

Our model takes into account the risk and benefits of PI becoming a common good. All buyers and sellers have unique characteristics that determine the price of a unit of PI. Each particular type of information has a specific risk of disclosure associated with it that factors in to the final value of the PI. Along with this, multiple records of information may be bought/sold at the same time. Our model compounds the prices of individual records when bought together because sharing multiple pieces of information put the seller at a greater risk. Using these factors in our model allowed us to develop supply and demand curves under a free market assumption. The intersection of these curves gives the ideal price for transaction of information.

Our model also took into consideration the fact that human data is highly linked by assigning a rank to each person. This rank would compound the original value of PI. We also used the connectivity of the network of humans to account for the increased information the buyer gets from buying a single unit of PI.

8.3 Our Strengths

The greatest strength of our model is that it is very flexible. It can be customized for any situation with any types of individuals. Our model can theoretically be scaled to multiple domains and records of PI. We have also talked about how we can scale our model well to communities and nations by using community detection algorithms and determining the average characteristics of the community.

Another big strength of our model is that it works really well under a free market assumption. And since under this assumption, our model obeys the laws of microeconomics, we can easily adapt our model to suit real life situations where there are government regulations and international trade.

8.4 Our Weaknesses

Our biggest strength is also perhaps our greatest weakness. Our model says that every new domain that is added requires a new set of values calculated for characteristics, and risk. This can be hard to determine and time-consuming as each one of these values is determined independently.

In addition, our model's compounding risk factor is the same for any two pieces of compounded information. While this assumption simplifies our model, the truth is far from that.

Another weakness of our model is that many of the variables we defined are subjective and can be very difficult to quantify. This makes our model less suitable for practical purposes. Another constraint of our model is that while it can capture the dynamic nature of human decision making, it does not reflect these changes immediately.

If these weaknesses can be fixed in a way that maintains the current benefits, our model would truly be robust.

8.5 Future Research

While our model has given a comprehensive description of PI trade under the assumptions of a free market, there are certainly many weaknesses. One of them is that the variables used in our model are subjective and difficult to quantify. Future research that can find an objective quantification of our subjective variables can immensely make our model more suitable for actual practice.

Another area of research would be understanding what would truly happen if we were to model PI trade from scratch without the free market assumption. It would be interesting to see if accounting for government regulations at every step in the modeling process gives rise to a significantly different model. And it would be interesting to compare these two models and see which one is better for practice.

9 To the Government Regarding the Cost of Privacy

For the attention of the Chairman of the Department of Commerce

Dear Chairman,

One of the biggest emerging industries is the usage of private information (PI) for targeted advertising. Given the ability to use PI for great economic gain, it is reasonable to stipulate that PI carries value. Given companies' ability to profit from the use of PI, we believe that one should be able to sell their PI as a commodity. However, unlike personal property (PP) and intellectual property (IP) the owner of PI is not compensated for the use of his/her property under current regulations. And, unlike PP and IP, we need to treat PI with utmost care because PI represents the identity of the individual.

In order to treat PI as a commodity, we believe that it must be traded in a free market, with each individual being able to choose what kind of PI they sell. Treating PI as a commodity, we created a pricing structure that incorporates the risk and benefits associated with a person disclosing their information to a third party. We factored in the value of different kinds of PI within different domains (such as social media, general identification, etc.). For each domain, we considered different characteristics (of an individual) that directly correlate to the risk and benefits of sharing that PI. In particular, we considered following domains: social media, medical, financial, general identification, and tracking. As more information is gathered on an individual, their privacy is increasingly at risk. To compensate for this, we derived the price per item to increase for every additional item bought.

To enhance the trade of PI, we developed a pricing system that factored in the forces of supply and demand, just like the trade of any other goods. Based on the risk and benefits of each individual's PI, we theorized a supply and demand model to determine the ideal price for any transaction of PI.

Due to the fact that our model takes into account the current characteristics of an individual, our model does not effectively capture the dynamic nature of human decision making in real time. To account for this, we suggest that PI be sold on contract periods of disclosed amounts of time. At the end of each contract period, we propose buyers and sellers the opportunity to renegotiate their contracts and price PI at a new value based on their newly perceived characteristics.

Treating PI as a commodity would present various complications to government regulation. The trade of PI has the potential of becoming an incredibly lucrative industry that would be difficult to regulate. Hence, poor regulation could lead to the corruption of the PI business. Using our model, we propose certain policies and laws for your consideration.

Firstly, we suggest that privacy of information should be considered a fundamental human right because private information constitutes to one's identity and denying them the right to their PI is the same as denying the right to their identity. Therefore, businesses can only use PI if consent is given by the rightful owners. This issue becomes more complicated when one person's PI contains information about others. For social media services where this factor is most relevant, it would be necessary to clearly state that one's information and behavior can be inferred by another's. Our model tries to fix that problem by taking into account what the majority of the population thinks is right.

Secondly, to facilitate the free trade of PI, no service, including existing ones like Facebook, can use PI of an individual who has not consented. This includes tracking user behavior. This policy goes hand in hand with the first suggestion that privacy be made a human right.

Thirdly, to prevent the exploitation of minors, we propose that children younger than 13 should not be allowed to take part in the trade of PI and; children of ages 13 to 18 require the consent of a parent or legal guardian, who is to monitor their activities, to participate in the trade of PI.

Fourthly, for the trade of PI, a legal contract must be drawn up that outlines the price, contract term, and acceptable use of PI. The contract must have a clause saying that the buyer will not intentionally misuse the PI. To handle unintentional misuse such as data breaches, our model already compensates the seller for such a risk by factoring it into the price of PI. With risk properly accounted for, this does not make the buyer liable in the event of a massive data breach.

Fifthly, the government must regulate PI in a similar way to international trade. We suggest that tariffs be charged on international trade of PI. However, unlike physical goods, PI has a large security risk and therefore, we suggest that all international trade of PI be examined by the government for sensitive material. If the PI being sold is determined to concern national security, appropriate action should be taken against the offense.

Lastly, lawmakers should also consider the ramifications and ethicality of the resale of PI. We suggest the resale of PI be legal, as long as there is a contractual agreement between buyer and seller. With that being said, the new buyer is required to follow all laws regarding the use of the purchased PI.

We sincerely hope that you consider our proposals for establishing the cost of privacy and take appropriate action to treat PI as a commodity.

Sincerely,

Aparajithan Venkateswaran
Brendan Palmer
Johann Kailey-Steiner

References

[1] "Us adults now spend 12 hours 7 minutes a day consuming media." https://www.emarketer.com/Article/US-Adults-Now-Spend-12-Hours-7-Minutes-Day-Consuming-Media/1015775. Accessed: 2018-02-10.

[2] S. Brin and L. Page, "The pagerank citation ranking: Bringing order to the web." http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf, 1998. Accessed: 2018-02-10.

[3] "You won't believe all the personal data facebook has collected on you." https://www.inc.com/larry-kim/you-wont-believe-all-the-personal-data-facebook-has-collected-on-you.html. Accessed: 2018-02-10.

[4] C. I. Howarth, "The relationship between objective risk, subjective risk and behaviour," *Ergonomics*, vol. 31, no. 4, pp. 527–535, 1988.

[5] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine." http://infolab.stanford.edu/~backrub/google.html, 1998. Accessed: 2018-02-10.

[6] "Connectivity (graph theory)." https://en.wikipedia.org/wiki/Connectivity_(graph_theory). Accessed: 2018-02-10.

[7] "k-edge-connected graph." https://en.wikipedia.org/wiki/K-edge-connected_graph. Accessed: 2018-02-10.

[8] "k-vertex-connected graph." https://en.wikipedia.org/wiki/K-vertex-connected_graph. Accessed: 2018-02-10.

[9] "How much money does facebook make from you?." http://www.telegraph.co.uk/technology/2016/11/03/how-much-money-does-facebook-make-from-you/. Accessed: 2018-02-10.

[10] "Here's how much your social security number is worth on the dark web." http://fortune.com/2016/08/03/social-security-dark-web/. Accessed: 2018-02-10.

[11] "Your medical record is worth more to hackers than your credit card." https://www.reuters.com/article/us-cybersecurity-hospitals/your-medical-record-is-worth-more-to-hackers-than-your-credit-card-idUSKCN0HJ21I201 Accessed: 2018-02-10.

[12] "Data dump." https://news.vice.com/en_ca/article/j5dkmd/how-much-is-is-your-personal-data-really-worth. Accessed: 2018-02-10.

[13] "A bite of me." https://www.kickstarter.com/projects/1461902402/a-bit-e-of-me. Accessed: 2018-02-10.

[14] "Heres how much your personal information is selling for on the dark web." https://www.experian.com/blogs/ask-experian/

heres-how-much-your-personal-information-is-selling-for-on-the-dark-web/. Accessed: 2018-02-10.

[15] "Here's how much thieves make by selling your personal data online." http://www.businessinsider.com/heres-how-much-your-personal-data-costs-on-the-dark-web-2015-5. Accessed: 2018-02-10.

[16] "Community detection in graphs." https://arxiv.org/abs/0906.0612. Accessed: 2018-02-10.

[17] "A physical model for efficient ranking in networks." https://arxiv.org/abs/1709.09002. Accessed: 2018-02-10.

[18] "Social media fact sheet." http://www.pewinternet.org/fact-sheet/social-media/. Accessed: 2018-02-10.

[19] "Anonymity, privacy, and security online." http://www.pewinternet.org/2013/09/05/anonymity-privacy-and-security-online/. Accessed: 2018-02-10.

[20] "Privacy, consumers, and costs." https://epic.org/reports/dmfprivacy.html#1a. Accessed: 2018-02-10.

A Introducing Government Regulations

Our original model assumes that trade happens in a free market. Here, we analyze what would happen when there are government regulations such as taxes and fixed price laws.

A.1 Taxes

A.1.1 Taxes on Sellers

When the government imposes taxes on sellers, the sellers are going to increase the price at which they will trade to compensate for the taxes. Therefore, the supply curve will shift up increasing P_{eq} and reducing n_{eq}. In the long term, the buyers will choose to buy less in order to bring the equilibrium price back to the original value. Figure 5 illustrates this.

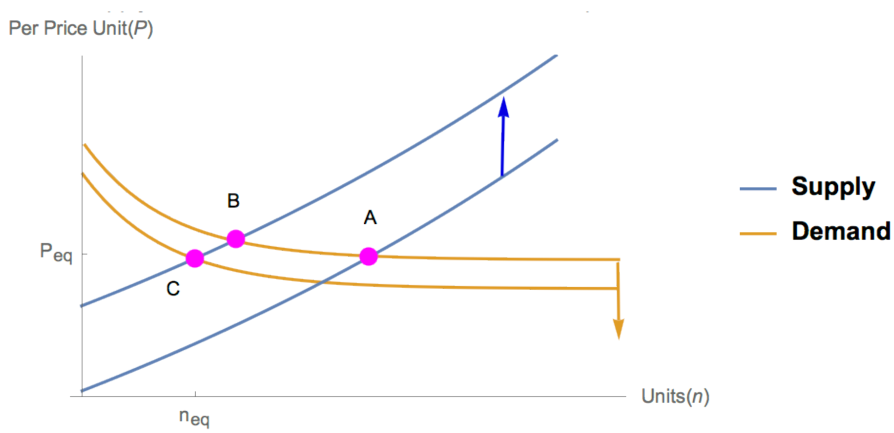

Figure 5: *How levying a tax on sellers change our original model*

A.1.2 Taxes on Buyers

When the government imposes taxes on buyers, the buyers are going to decrease the quantity they buy to compensate for the increased in taxes. Therefore, the demand curve will shift down decreasing P_{eq} and reducing n_{eq}. In the long term, the sellers will choose to sell more in order to bring the equilibrium price back to the original value. Figure 6 illustrates this.

A.2 Fixed Price Laws

A.2.1 Maximum Price Law

If the maximum price determined by the government is higher than the equilibrium price, then there will be no change to our model.

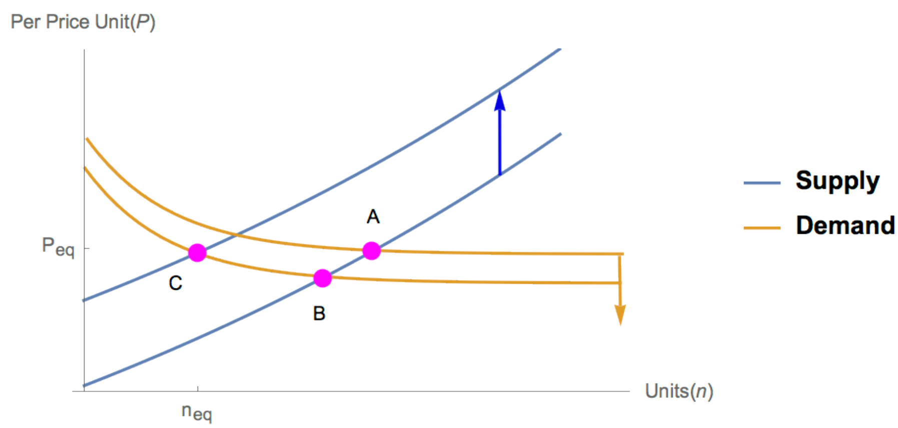

Figure 6: *How levying a tax on buyers change our original model*

If the maximum price is lower than the original equilibrium price, then the new equilibrium price will shift to the maximum price determined. This will also come with a reduction in the quantity of PI being traded. Figure 7 shows this phenomenon.

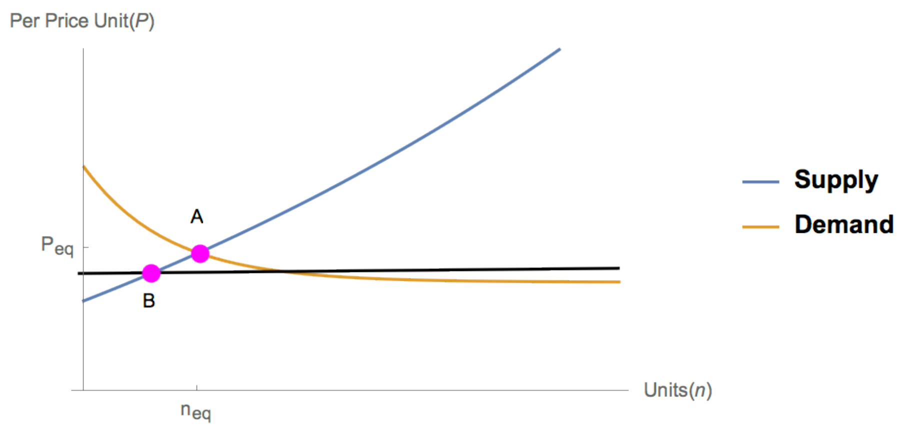

Figure 7: *How imposing a maximum fixed price can change our model*

A.2.2 Minimum Price Law

If the minimum price determined by the government is lower than the equilibrium price, then there will be no change to our model.

If the minimum price is hihger than the original equilibrium price, then the new equilibrium price will shift to the minimum price determined (which is higher). This will also come with a reduction in the quantity of PI being traded. Figure 8 shows this phenomenon.

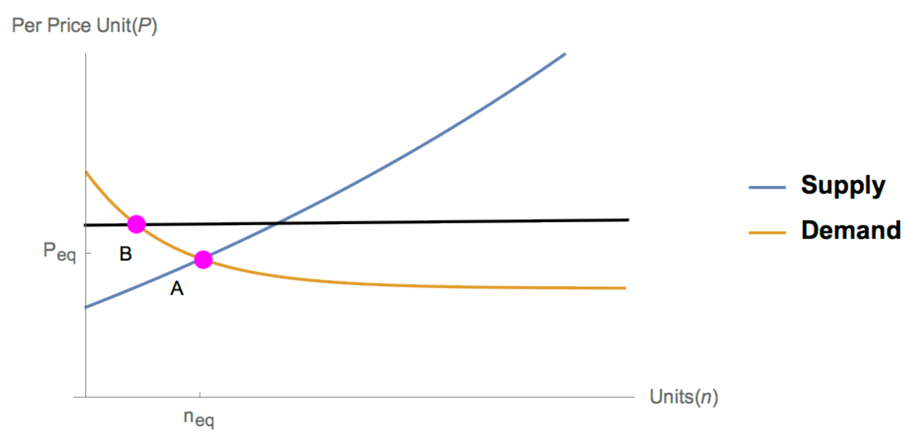

Figure 8: *How imposing a minimum fixed price can change our model*