UNIVERSITY OF ALBERTA
FACULTY OF ARTS
Department of Economics

Working Paper No. 2022-08

Reputation of Quality in International Trade: Evidence from Consumer Product Recalls

Jiatong Zhong
University of Alberta

March 2022

Reputation of Quality in International Trade: Evidence from Consumer Product Recalls

Jiatong Zhong*

This version: August 2021

Click Here for the Most Recent Version

Abstract

This paper examines the impact of exporting countries' reputations for product quality on aggregate trade flows. Using a Bayesian learning model, I construct a measure of exporter reputation in which consumers internalize product recalls as signals of poor quality. Structural estimation of the model finds that reputation is important and especially impactful for toys. The market share elasticity of an exporter's reputation is 2.396 for toys. Improving reputation can increase export value, but reputational change is sluggish. Counterfactual exercises confirm that quality inspection institutions are welfare improving.

Keywords: International trade, reputation, Bayesian learning, quality uncertainty, product recalls

JEL Classification: F14, D83, L15

*Department of Economics, University of Alberta. jzhong5@ualberta.ca

1 Introduction

Vertical product differentiation plays a critical role in explaining production and consumption patterns in international trade. The most popular quality measure in trade is price-adjusted sales, which is estimated assuming consumers have perfect information about product quality. It is often the case, however, that consumers only have access to imperfect product information. While previous works have theorized how quality uncertainty affects trade and consumer welfare, their models typically focus on static equilibrium outcomes (e.g. Bond, 1984; Falvey, 1989; Chisik, 2002), and their empirical investigations are limited to changes after one event, such as implementing quality standards (e.g. Potoski and Prakash, 2009). A dynamic model allows demand to be path-dependent and to adjust slowly to quality signals, both of which are important components of decisions concerning investment into product quality. This paper focuses on the dynamic demand responses to quality signals and evaluates whether the key premises for dynamic quality investment models (i.e. that seller reputation matters) have empirical support in international trade.

When consumers are unsure about quality, they rely on their knowledge of the product, which is referred to as the reputation of sellers in this paper. Capturing reputation empirically is challenging for two reasons. First, reputation is history-dependent, so it needs to be measured dynamically. Second, estimation of dynamic models requires a data set containing events that repeatedly impact or signal product quality as well as the market responses to such events. This paper proposes a measure of reputation for exporting countries constructed by exploiting the cross-country, cross-time variation in product recalls. By quantifying the value of reputation, I evaluate exporter's incentives to improve product quality. I also conduct counterfactual exercises to quantify consumers' welfare gains from having an effective quality inspection institution.

This paper introduces a novel data set that merges product recalls with import flows to reveal how the market responds to informative signals. I scrape recall notifications posted by the Consumer Product Safety Commission (henceforth CPSC) from 1973 to 2015, and use recall date, product descriptions, and country of origin to match the recall incidences to U.S. monthly import data from April 1990 to December 2009.[1] A prominent data pattern revealed in this data set, as illustrated by Figure A.1, is that larger exporters tend to

[1] The Commission provides public access to their recall database through a Recalls Application Program Interface (API).

face more recalls. However, if we examine the trade patterns of each exporting country, we see that an exporter's market share declines immediately after a major recall event hits.[2] An intuitive explanation for this observation is that the volume of trade matters: conditional on the fraction of unsafe products, countries selling more units are more likely to face recalls, so even if recalls have negative impacts on sales, the effect is obscured by sales volume in a micro-econometric analysis. This paper disentangles the impact of recalls from the sales volume and provides a quantitative method to evaluate the impact of bad signals.

In this model, product quality is binary: a product is either safe or unsafe.[3] Each exporting country-product pair represents a variety, and each variety has a different fraction of safe products. Consumers do not know the fraction of safe products for any given variety, and unsafe products look identical to safe products before purchase. However, consumers can use observed recalls to learn about the fraction of unsafe products and form an expectation for each variety's latent quality. Following Board and Meyer-ter Vehn (2013), the expectation of quality formed in each period is the reputation for that variety at that time, and it enters aggregate demand as a product characteristic.

The model is estimated by exploiting the market share responses to recalls as well as the mean and variance of recall events. The parameters that shape the consumer learning process are identified with a convergence property of Bayesian learning. The learning parameters are estimated such that the mean and variance of recalls predicted by the learning outcomes in the final period match the moments from the observed recalls. Reputation is constructed with the learning parameters, quantity of imports, and recalls. The taste for reputation is estimated such that the predicted market shares match the observed market shares as closely as possible. All parameters are estimated simultaneously using generalized method of moments, and as a mathematics program with equilibrium constraints (MPEC).

Using the estimated reputation and preferences, I perform counterfactual exercises concerning both exporters and consumers. For exporters, I calculate the impact of recall events

[2]See Figure A.2 for an example using Hong Kong export of toys.

[3]The term "quality" in this paper is different from the quality commonly used in empirical trade. In empirical trade, quality is measured with unit values or demand residuals (Schott (2004); Hummels and Klenow (2005); Khandelwal (2010)), the term captures an array of product characteristics that are observed by consumers, but not by the econometrician. In this paper, the binary quality is one of the many product characteristics consumers may care about, but it is different from "quality" in empirical trade in the sense that it is unobserved before purchase even for consumers.

on market share and trade value. My estimates suggest that while consumers do not factor reputation into decision making for most products, they weight the reputation for children's toys quite heavily. On average, a 10% improvement in reputation can increase market share by 23.96% for an exporter of toys. However, reputational change is sluggish, especially for small exporters who used to be large exporters. Even for an average large exporter, it takes over 57 years of recall-free presence in the United States to improve its reputation by 10%.[4] For consumers, I examine the value of having a quality inspection institution by simulating a scenario in which the probability of a bad product being recalled is reduced from 90% to 60%. Total welfare losses average 5.98 billion dollars per quarter for consumers of toys. These results suggest that for the importing country, a product quality inspection institution like the CPSC can improve consumer welfare.

This paper is related to the trade literature involving quality uncertainty. The theoretical component uses a learning approach, which adds to two popular methods to model quality uncertainty of imported goods: adverse selection (e.g. Bond, 1984) and reputation premium (e.g. Falvey, 1989).[5] I introduce a dynamic framework featuring quality uncertainty into the international context, which is closer to the recent models of reputation and uncertain product quality developed in the industrial organization literature. The learning model also allows me to evaluate the welfare impact of information disclosure, which complements both long-standing theoretical and recent empirical investigations on this topic (Creane and Miyagiwa, 2008; Jovanovic, 2021). The empirical literature of quality uncertainty in trade primarily examines the effects of national and international quality standards (Swann, Temple and Shurmer, 1996; Potoski and Prakash, 2009). Compared to quality standards, recalls provide more frequent changes we can use to infer reputation. Relative to the customer ratings from online platforms used in empirical industrial organization (for example, Mayzlin, Dover and Chevalier, 2014), this data set contains a wider set of products and more information about exporting countries.[6]

The empirical analysis also contributes to studies using product recalls. Freedman, Kearney and Lederman (2012) used toys recalls from the CPSC to run a difference-in-difference regression, estimating the spillover effect in volume of sales to the producer and

[4]Here, large is defined as in the upper quartile of export quantity.

[5]Other papers that use adverse selection to model quality uncertainty in trade include Donnenfeld, Weber and Ben-Zion (1985); Donnenfeld (1986); Donnenfeld and Mayer (1987).

[6]See Donnenfeld (1986); Falvey and Kierzkowski (1984) for additional empirical works on quality standards. Chen and Wu (2020) uses online reviews as a proxy for reputation of foreign individual sellers.

the industry. Grundke and Moser (2019) examines whether the FDA uses import refusals strategically during recessions under the pressure of protectionism. Basker and Kamal (2020) linked CPSC recall data with firm trade transactions, and found that firms turn to other suppliers when their trade partner is under recall. Jovanovic (2021) used automobile recalls and stock prices to estimate the value of firm reputation, and he concludes that reputation is worth around 8% of the firm's value. Jovanovic (2021) and this paper both attempt to empirically understand the value of reputation taking advantage of the external shock brought by recalls, but the model in Jovanovic (2021) focuses on firms' efforts, while my model focuses more on buyers' reaction. Buyers in Jovanovic (2021) are indifferent to sellers since they are fully insured against any loss from defect, while consumers' preferences towards exporters are center to my empirical analysis.

The model builds on a rich literature studying sellers' reputation when product quality cannot be perfectly observed (See Bar-Isaac, Tadelis et al. (2008) for a detailed survey). It fits into the branch of the literature where sellers have information that is hidden from consumers, and it is most similar to that in Bar-Isaac (2003), sharing the feature of learners updating their belief under Bayes' rule. It borrows the definition of reputation from Board and Meyer-ter Vehn (2013) as they explicitly model signals in a manner close to how product recalls happen. This paper focuses on consumer responses instead of firms' investment in product quality. The empirical literature on sellers' reputation almost exclusively uses data from electronic market places, with Jovanovic (2021) being an exception. My results are consistent with their findings in that sellers can be rewarded for having a good reputation (e.g. Eaton, 2005), although this is not the case for all products.[7] Most empirical works cover one specific good or service (e.g. iPod in Saeedi (2019)), but my study covers many products and studies the impact on exporters instead of individual sellers.

There is a growing literature that incorporates learning in trade models, which mostly concerns how firms learn about foreign markets before entry (Eaton et al., 2009; Albornoz et al., 2012; Holloway, 2017) and how firms building a relationship with foreign suppliers (Rauch and Watson, 2003). Two learning models are popular among trade economists, learning with experimentation featuring firms that start with small transactions before expansion (Albornoz et al., 2012; Rauch and Watson, 2003) and Bayesian learning characterizing how firms obtain information about foreign markets (Eaton et al., 2009; Holloway,

[7]Other papers that have similar conclusions include Livingston (2005); Houser and Wooders (2006); Mayzlin, Dover and Chevalier (2014); Chen and Wu (2020); Saeedi (2019); Jovanovic (2021)

2017).[8] This paper follows the tradition of Eaton et al. (2009) and Holloway (2017), but focuses on the consumers' perception.

The rest of the paper is organized as follows. In the next section I introduce a partial equilibrium model that captures how consumers update their perception of an exporter's reputation in a market using observations of product recalls each period. Section 3 explains the empirical strategy for estimating this model. Section 4 describes the novel data set, and 5 and 6 report the results. Section 7 concludes.

2 A Learning Model for Exporters' Reputations

In this model, I introduce the definition of reputation, how reputation evolves over time, and how consumers respond to it. I assume that firms within an exporting country face perfect competition, and supply inelastically in each period. Consumers make purchase decisions based on prices and the current reputation for each exporting country. After purchasing the product, they observe the quantity sold, recalls, and update the reputation at the end of the period with past reputation and the new signals they observe.

2.1 Consumers' Problem

There is a continuum of consumers indexed by i. In each period t, every consumer consumes one unit of a differentiated product, s, and $y_{i,t}$ units of a homogeneous product. Consumers do not observe the true quality of the differentiated product, but they observe the country-of-origin, j. The differentiated product is either safe or unsafe, characterized by the unobserved quality z that takes value 1 if it is safe, and 0 otherwise. Consumers cannot distinguish between safe and unsafe products before purchase, but they observe the outcome after purchase which factors into their realized utility. I assume a utility function similar to that of Petrin (2002). The utility after purchase and quality revelation is written as

$$u_{ijs,t} = \alpha_0^s \log(y_{i,t}) + \alpha_x^s z_{js,t} + \eta_{js} + \psi_{s,t} + \xi_{js,t} + \epsilon_{ijs,t}.$$

[8] I abstract away from the concept of "experimentation" discussed in Bolton and Harris (1999), which features consumers who strategically make purchase decisions in order to obtain more information. In my context, signals are sent out by a quality inspection institution.

η_{js} is the time-invariant preference common across all consumers for a product from a country, which captures time-invariant unobserved characteristics. ψ_{st} captures the time specific demand for product s, for example, higher demand for toys in the last quarter of the year. $\xi_{js,t}$ represents unobserved demand shocks that vary across time, country, and product, but affect all consumers in the same manner, such as retail channels and unobserved variety characteristics. $\epsilon_{ijs,t}$ is the idiosyncratic preference shock that follows an i.i.d. Extreme Value distribution.

In each period, consumers maximize their expected utility by choosing one exporting country to buy one unit of the differentiated product from. Let $\mathcal{H}_t$ denote the information set available to consumers when making a purchase decision. The expected quality of product s from country j is denoted as $x_{js,t} = \mathbb{E}[z_{js}|\mathcal{H}_t]$. We will discuss what is in the information set $\mathcal{H}_t$ and the functional form of the expectation in the next section. Using the law of iterated expectations, we can write consumer's maximization problem as:

$$\begin{aligned} \max_{j\in\mathbb{J}_s} \quad & \mathbb{E}[u_{ijs,t}] = \mathbb{E}\left[\mathbb{E}[u_{ijs,t}|\mathcal{H}_t]\right] \\ & \qquad\qquad = \alpha_0^s \log(y_{i,t}) + \alpha_x^s x_{js,t} + \eta_{js} + \psi_{st} + \xi_{js,t} + \epsilon_{ijs,t} \\ \text{subject to} \quad & y_{i,t} + p_{js,t} \leq I_t, \end{aligned} \tag{1}$$

where I_t is the budget constraint that can be interpreted as income, $p_{js,t}$ is the price for one unit of the differentiated product s from country $j \in \mathbb{J}_s$, and $\mathbb{J}_s$ is the set of exporters who sell product s to the United States. The price of the homogenous product is normalized to 1. The consumer's optimization problem is a standard discrete choice problem as in Petrin (2002), where the expected quality of the differentiated product enters the consumer's decision as a product characteristic. Following Board and Meyer-ter Vehn (2013), I refer to the expected quality $x_{js,t}$ as the reputation for product s from country j at period t, and I will henceforth call it "reputation".[9] In the next section, I will derive the law of motion for reputation.

[9]Note that the definition of reputation is similar to that of the "perfect bad signal" scenario in Board and Meyer-ter Vehn (2013), but the model is different in two ways. First, this model is in discrete time while Board and Meyer-ter Vehn (2013) sets their model in continuous time. More importantly, Board and Meyer-ter Vehn (2013) concerns the firm's investment in efforts and their model includes a productivity shock, but this model abstracts away from the firm's strategy or productivity.

2.2 Reputation Updating

This section begins with a sketch of the probability problem a consumer faces when she infers the expected quality of the product using history of sales, recalls, and country-of-origin. I then derive the reputation updating process from the consumer's rational expectation, and show that reputation can approach the true average quality for each exporting country given sufficient periods of learning.

2.2.1 Deriving the updating process

Consumers do not observe the quality of differentiated products, but they can observe the country-of-origin label. I assume that the fraction of safe product s from an exporting country j is θ_{js}, which consumers do not know fully, but it can be learned about through signals. In particular, consumers' beliefs follow a distribution on support [0,1], and signals change that distribution over time. The true fraction is assumed to be constant over the periods of learning. If the product is unsafe, then there is a probability μ^s that it will be recalled. That probability is product-specific, but common across time and across exporting countries. Figure 1 illustrates the above-described process.

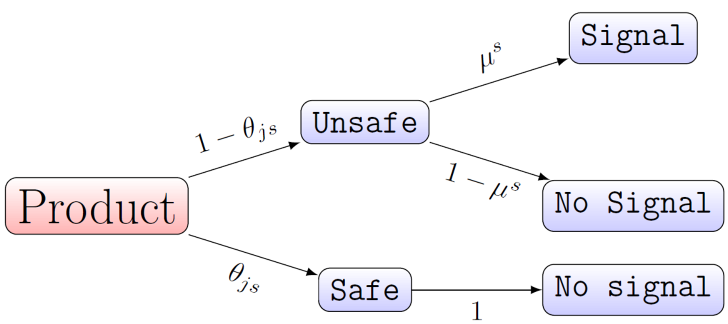

Figure 1: Probability of recall before revelation of quality

I assume that safe products will never result in a recall, which should not be far from reality. Most recalls are triggered *after* one or more hazardous events are reported by consumers or retailers. The CPSC then investigates these reports and if the Commission decides that there is a "substantial product hazard", it will issue a recall. If a retailer or manufacturer voluntarily recalls the product—usually after a consumer complaint—the recall notice will be issued faster.[10] In both cases, recalls are mostly complaint-driven, so

[10]Consumers, government agencies and medical practitioners can voluntarily file reports of product

it is reasonable to assume that recalls are only issued for problematic products.

Although consumers do not know the value of θ_{js}, they form an expectation of its value based on informative signals. Their information set for product s from exporter j at period t is $\mathcal{H}_{jst}$, which contains the history of recalls $\{r_{js,\tau}\}_{\tau=1}^{t-1}$, quantity $\{q_{js,\tau}\}_{\tau=1}^{t-1}$, and reputation $\{x_{js,\tau}\}_{\tau=1}^{t-1}$ at period t. When the realized quality is 0 or 1, the reputation coincides with the expected fraction of true products:[11]

$$x_{js,t} = \mathbb{E}[z_{js}|\mathcal{H}_t] = \mathbb{E}\left[\theta_{js}|\mathcal{H}_{jst}\right].$$

Consumers' expectations form a vector of reputations $\{x_{js,t}\}_{j\in\mathbb{J}^s}$ for different exporters j. Information set $\mathcal{H}_t$ contains all information sets for any country-product pair $\mathcal{H}_{jst}$, but the information necessary to update one country's reputation is only its own history.

Before consumers make the purchase decision in period t, they learn about the probability of getting a safe product if they buy from country j by Bayesian updating their probability assessment using the signals of recalls they receive in the last period. Purchasing from country j is analogous to making a random draw from a pool of size $q_{js,t}$. Given that the true and unobserved fraction of safe products is θ_{js} for a country j, consumers purchased a total of $q_{js,t}(1-\theta_{js})$ units of unsafe products. For each unit of the unsafe product purchased, there is a probability μ that the CPSC will issue a recall. This can be due to consumers being unaware of the product defect or the CPSC's investigation failing to confirm the product's defect after the initial report.

In the derivation following, I suppress product and country indices since the same process applies to all product-country pairs. The signals are sent through standard Bernoulli trials, and we can define the unconditional probability of sending a recall signals for each draw as $\gamma \equiv (1-\theta)\mu$. If we assume that the prior distribution of γ is a Beta distribution, the reputation updating process follows the equations in Proposition 1. The Beta distribution is a conjugate prior distribution for the Bernoulli likelihood function: it means that before and after the update, the distributions of γ are both Beta distributions. This is algebraically convenient for us to compute an expectation before and after learning in a

hazards to the CPSC, while manufacturers, importers, distributors, and retailers have a legal obligation to report the products to the CPSC once they learned the product defects and hazards.

[11] More generally, when the realized quality is a when the product is unsafe and b when the product is safe ($b > a$), expected quality is a linear transformation of the conditional expectation of θ_{js}: $x_{js,t} = \mathbb{E}[z_{js}|\mathcal{H}_t] = a + (b-a)\,\mathbb{E}\left[\theta_{js}|\mathcal{H}_{jst}\right]$. The motion of reputation is a straightforward extension of the current form.

period. [12]

Proposition 1. *When we choose a Beta distribution* $\mathcal{B}(\beta_0, \delta_0)$ *as the prior distribution for* $\gamma \equiv (1-\theta)\mu$*, the reputation update from period* t *to* $t+1$ *follows:*

$$\begin{cases} x_1 = 1 - \dfrac{\beta_0}{\mu(\beta_0 + \delta_0)} & \text{(2a)} \\ \\ x_{t+1} = \dfrac{\beta_t + \delta_t}{\beta_t + \delta_t + q_t} x_t + \dfrac{q_t}{\beta_t + \delta_t + q_t}\left(1 - \dfrac{r_t}{\mu q_t}\right) & \text{(2b)} \end{cases}$$

with β_0 *and* δ_0 *as the initial parameter values for the Beta distribution,* $\beta_t = \beta_0 + \sum_{\tau=1}^{t-1} r_\tau$ *and* $\delta_t = \delta_0 + \sum_{\tau=1}^{t-1} q_\tau - \sum_{\tau=1}^{t-1} r_\tau$*.*

Appendix C.1 shows derivations of the above equations. The intuition for β_0 is the cumulative units of goods ever recalled from a variety before the first period the data set allows econometricians to observe. Similarly, δ_0 is the cumulative units of un-recalled products sold into the United States before the first observation. β_0 and δ_0 absorb the history before the starting period in estimation. β_t is the total cumulative units of recalled products up to period t, and δ_t is the total cumulative units of safe products sold up to t. The summation of δ_t and β_t is the total cumulative units of goods sold before period t.

Equation 2b is in the form of a weighted average of current reputation x_t and new information $1 - \dfrac{r_t}{\mu q_t}$. The first term in equation 2b contains a coefficient of x_t that captures the persistence of reputation. The denominator of coefficient is the cumulative units of goods sold at the end of period t, and numerator is the cumulative units sold before period t, so intuitively, the coefficient captures the "weight of history". When β_0 and δ_0 are small relative to the total quantity sold in past periods, the coefficient is dominated by the fraction of the summation of the units sold up to period $t-1$ over the units sold up to period t. This weight is between 0 and 1, and it increases over time, so it is a term that captures the convergence of reputation.

The second term in equation 2b captures the new information in period t. The coefficient is the fraction of quantity sold in period t in the cumulative units of goods sold at the

[12] In appendix C.2 I include a discussion of using truncated Beta distribution as a prior, for readers who are concerned about the upper limit of the distribution of γ. Appendix C.2 discusses the case when $\gamma \in [0, \mu]$. I show that if β and δ are large enough, the reputation updating procedure can be closely approximated by the one shown using standard Beta as a prior.

end of period t, which is intuitively the "weight of new information". The term $\left(1 - \frac{r_t}{\mu q_t}\right)$ is the expected fraction of safe products in the market in period t.

Equation 2a represents the initial condition. $\frac{\beta_0}{\beta_0 + \delta_0}$ is the fraction of the cumulative sum of recalled products relative to the sum of all units sold before the first observation. Adjusted by the efficiency of the recall $\frac{1}{\mu}$, $\frac{\beta_0}{\mu(\beta_0 + \delta_0)}$ is the expected fraction of unsafe products in the first period.

β_0 and δ_0 must be positive numbers, as implied by intuition, and they are likely in a magnitude comparable to (or larger than) the volume of trade flows observed. The probability of recall (given that a unit of product is bad) is given by parameter μ, and $\mu \in (0, 1]$. μ cannot be zero; otherwise, equation 2a and 2b are not well-defined. Intuitively, the effectiveness of inspection cannot be so bad that recall never happens. Quantity q_t is a positive number that does not go to infinity, and the units of recall r_t are nonnegative and bounded above by q_t in each period. The range of parameters in proposition 1 imposes almost no other restrictions beyond those implied by economic intuition, but they are necessary for the asymptotic property presented in the next section.

2.2.2 Asymptotic property of reputation learning

Bayesian learning is a type of perfectly rational learning. With some restrictions, the expectation converges asymptotically to the true value agents learn about. I will refer to this asymptotic property as "effective learning" henceforth. I will return to this property in the estimation section, as it is useful for identification.

I assume that, conditional on the history $\mathcal{H}_{jst}$, the fraction of safe products θ_{js} and probability of recall for unsafe products μ^s, the expectation of import in period $t+1$ is product-country-specific, but time-invariant. That is, consumers do not learn about the size of market from history. This assumption and the assumption on bounds of parameters are formalized in Appendix C.6 as assumption 1 and 2. Together, they provide sufficient conditions for asymptotic effective learning.

Theorem 1. *Given assumptions 1 and 2, learning is effective asymptotically. That is, the expectation converges to the truth when T is large:*

$$x_{js,T} \to \theta_{js}, \quad \textit{as} \quad T \to \infty$$

Proof. See Appendix C.6.[13] □

In each period t, every consumer forms their expectation for product quality from the observed signals $r_{js,t}$ and market size $q_{js,t}$, and then from the menu of reputation and price they make their purchase decision. By aggregating individual purchase decisions, we can compute the countries' market shares using a discrete choice model.

2.3 Equilibrium

Following standard logistic demand assumptions and let the budget constraints hold with equality, the market share of country j in a particular product market s in time t is:

$$\begin{aligned} s_{js,t} &= \int_{\epsilon_{ijs,t}|u_{ijs,t}>u_{ij's,t}\forall j'\neq j\}} d\mathcal{F}_\epsilon(\epsilon) \\ &= \frac{(I_t - p_{js,t})^{\alpha_0^s}\exp(\alpha_x^s x_{js,t} + \eta_{js} + \psi_{st} + \xi_{js,t})}{1 + \sum_{j'=1}^{J_s}(I_t - p_{j's,t})^{\alpha_0^s}\exp(\alpha_x^s x_{j's,t} + \eta_{js} + \psi_{st} + \xi_{js,t})} \end{aligned} \tag{3}$$

subject to constraint:

$$I_t \geq p_{js,t}$$

In equilibrium, the goods market clears. In each period, the United States imports as many units of products from each exporter as demanded by the domestic market. The United States is treated as a supplier as well, and the utility of purchasing from the U.S. is normalized to 1. Since firms are perfectly competitive within an exporting nation, price is determined by country-specific costs and treated as given in this framework.

Formally, the equilibrium definition is:

Definition 1 (Equilibrium with Learning)**.** *An equilibrium in this model is defined as a $J\times S\times T$-by-3 matrix of price, reputation and import flows $[p_{js,t}, x_{js,t}, q_{js,t}]$ with a Bayesian learning motion such that:*

1. *Import Market Clears:*

$$S_{js,t} = s_{js,t}(p_{js,t}, x_{js,t}, \xi_{js,t}; \alpha^s, \mu^s, \beta_0^s, \delta_0^s)$$

[13] This proof is only slightly different from a standard proof of convergence in Bayesian learning.

2. The Bayesian learning motion satisfies:

$$x_{js,t+1} = \frac{\beta_{js,t} + \delta_{js,t}}{\beta_{js,t} + \delta_{js,t} + q_{js,t}} x_{js,t} + \frac{q_{js,t}}{\beta_{js,t} + \delta_{js,t} + q_{js,t}} \left(1 - \frac{r_{js,t}}{\mu^s_{js,t}}\right)$$

where $\beta_{jst} = \beta^s_0 + \sum_{\tau=1}^{t-1} r_{js,\tau}$ *and* $\delta_{jst} = \delta^s_0 + \sum_{\tau=1}^{t-1} q_{js,\tau} - \sum_{\tau=1}^{t-1} r_{js,\tau}$*; and* β^s_0 *and* δ^s_0 *as the initial parameter values.*

3 Empirical Strategy

In the equilibrium with learning, I can observe income I_t, price $p_{js,t}$, total units of sale $q_{js,t}$, quantity of risky products $r_{js,t}$, and market share $S_{js,t}$ from data. For each product s, μ^s, β^s_0, δ^s_0, and the vector of demand function coefficients α^s are parameters that need to be estimated. Price, quantity, the number of recalls, and market share vary across time and varieties, while income varies over time only. Parameters vary across products, but are constant over time and across exporters.

The baseline estimation is done product by product. A product is a commodity classified under a six-digits harmonized system code in the import data. Within each product s, the set of learning parameters $(\mu^s, \beta^s_0, \delta^s_0)$ enters the model non-linearly, and given estimated reputation $\{x_{js,t}\}_{j\in J^s, t\in 1,2...T}$, the vector of demand parameters α^s enters linearly.

There are three main challenges to estimation. First, although the demand equation can be linearized, the system of equations is still non-linear because of the Bayesian learning motion. In addition, the reputation measure $x_{js,t}$ is constructed, so to make sure its value aligns with data, I use a property of Bayesian learning and introduce an additional objective function in estimation. Multiple objective function optimization problem (henceforth MOOP) is common in engineering, but less so in economics, so I borrow a classic method in engineering to transform this problem into a single objective function problem. Finally, price is endogenous in the demand equation, and I use unit shipping cost as an instrumental variable.

The empirical strategy has two parts, though they are estimated simultaneously. These parts correspond to the main challenges in identification. In the first part, I use history of import quantity and recall units to back out the parameters that determine reputation dynamics, exploiting the asymptotic property in theorem 1. This condition implies that after enough periods of learning, the reputations for each country approach the true unobserved fraction of good products.

3.1 Estimating Bayesian Updating Parameters from Recalls and Quantities

Separately identifying the preference for reputation, α_x, and learning parameters μ, β_0, δ_0 requires us to take advantage of a property of learning, because reputation $x_{js,t}$ is constructed. Intuitively, I use the fraction of unsafe products implied by the learning model to predict the mean and variance of recalls, and match the moments to those observed in recall data. Theorem 1 shows that, given enough periods of learning, reputation converges to the true expected quality. I take the vector of reputation in the last period $x_{js,T}$ and use it as a proxy for the unobserved fraction of good products θ_{js}. To ensure that consumers actually learn sufficiently, I only include exporters who have been in the U.S. market for more than 10 quarters. Using J'_s to denote the set of exporters of product s that we have observed for more than 10 periods, we can formulate this criteria as the following likelihood estimation. Given the units of import from each country in each period $q_{js,t}$, the number of unsafe products in the market in period t is:

$$L_{s,t}(\mu^s, \beta_0^s, \delta_0^s) = \sum_{j \in J'_s} q_{js,t} \times x_{js,T}(\mu^s; \beta_0^s, \delta_0^s)$$

I observe the total number of recalled products in each period $R_{s,t} = \sum_{j \in J'_s} r_{js,t}$. For each unsafe product in the market, the probability of being recalled is μ^s. $R_{s,t}$ is the realization in period t of $L_{s,t}$ independent Bernoulli trials with "success" probability μ^s and follows binomial distribution. Given that $L_{s,t}$ is large, we can use a normal distribution $\mathcal{N}(\mu^s L_{s,t}, \mu^s(1-\mu^s)L_{s,t})$ to approximate the binomial distribution, and the log-likelihood function is:

$$\mathcal{L}(R_{st}|\hat{\theta}(\mu^s; \beta_0^s, \delta_0^s), Q_{s,t}) = \sum_{t=1}^{T} log\phi(R_{st}|\hat{\theta}(\mu^s), Q_{s,t}) \tag{4}$$

where $\phi(R_{st}|\hat{\theta}(\mu^s), Q_{s,t})$ is the normal probability density function with mean μL_t and variance $\mu(1-\mu)L_t$.[14] Given learning parameters, reputation can be constructed without price or market share data.

[14] The approximation greatly improves computation efficiency. Computing the likelihood of this binomial distribution is impossibly inefficient since the power exponent is too large.

3.2 Demand Estimation

For each set of value $(\mu^s, \beta_0^s, \delta_0^s)$, reputation can be computed as a given product's characteristics. The rest of the parameters—the preference parameters (α_0^s, α_x^s) constants and fixed effects—are estimated from a standard discrete-choice demand system. I follow Khandelwal (2010) and treat purchasing from the United States as the outside option in the discrete choice. In cases without income heterogeneity, the demand equation can be linearized (see Berry (1994)). The log-linearization of market share equation 3 is:

$$\ln(s_{sj,t}) - \ln(s_{s,US,t}) = c^s + \alpha_x^s x_{js,t} + \alpha_0^s \ln(I_t - p_{js,t}) + \eta_{js} + \psi_{st} + \xi_{js,t}$$

I_t is the average household expenditure on consumption goods per quarter over all observed periods. The coefficient α_0^s is the own price elasticity of the good s. The term $\ln(I_t - p_{js,t})$, given price is involved, is correlated with the unobserved product characteristics. I use unit shipping cost as the price instruments following the argument in Khandelwal (2010).[15]

The definition of market share as a fraction of trade values instead of quantity implies that a small modification of the linearized equation is necessary.[16] The regression equation in the case of homogeneous income is:

$$\ln(S_{js,t}) - \ln(S_{US,s,t}) - \ln(p_{js,t}) = c^s + \alpha_x^s x_{js,t} + \alpha_0^s \ln(I_t - p_{js,t}) + \eta_{js} + \psi_{st} + \epsilon_{js,t} \quad (5)$$

Denote $y_{js,t} \equiv \ln(S_{js,t}) - \ln(S_{US,s,t}) - \ln(p_{js,t})$, and henceforth I will use $y = \{y_{js,t}\}_{s,t}$ to refer the dependent variable constructed from market shares.

By keeping parameters invariant across time and exporters, the framework assumes that consumers only "discriminate rationally". Namely, they differentiate exporters' products based only on the products' current reputations and the signals received in this period. The coefficient α_x^s governs the utility differentiation between a high quality and a low quality product s. The larger α_x^s is, the more consumers value a high quality product over a low one—in other words, consumers care about the quality of that product. As discussed in the introduction, in this empirical exercise, "quality" only concerns the safety of the product.

[15] Khandelwal provided an explanation for the validity of these instruments, see Khandelwal (2010) for details. I have also tried exchange rates and oil price times distance between importer and exporter as instruments, but the first stage test shows that they are not as ideal.

[16] Trade value is a more consistent measure of market share than quantity in trade data, because custom data report quantity both in weight and units.

For example, if α_x is higher in "toys and sports equipment" than "apparels," then we would conclude that consumers care more about safety of toys than clothes. Surely consumers want safe products in both categories, but the harm done to consumers by a toy with lead paint can be more severe than a battery that can overheat. μ is the probability of a recall if the product is of low quality. The arrival rate is determined by product characteristics and how consumers use them. When μ is high, we will consistently see frequent recalls for low reputation countries. When μ is low, fewer products are recalled per period and the variation relative to the mean of recall level is higher.

The residual of regression 5 forms the orthogonality condition necessary for GMM estimation:

$$\mathbb{E}[\xi_{js,t}|h(x_{js,t}, z_{js,t})] = 0$$

where h is a function of the observed exogenous variables and the instrument.[17] $\{z_{js,t}\}_{j\in J^s,t\in T}$ is a vector of exogenous variables and instruments.[18]

3.3 Estimating the Model as One MPEC Problem

The model is estimated as a mathematical program with equilibrium constraints (henceforth, "MPEC") problem. This is a technique widely used in engineering and recently adopted in economics to solve optimization problems with many nonlinear constraints. Dubé, Fox and Su (2012) have shown that MPEC has a significant speed advantage for the estimation of large-dimensional problems with many markets and also improves convergence compared to the nested-fixed point algorithm. By setting the Bayesian learning

[17] In the Nested Fix Point approach (Berry, 1994), the unobserved characteristic ξ_t is calculated by inverting the market share equation 3. The MPEC approach does not require such an inverse and can thus be faster.

[18] In the main estimation, I provided the constraint Jacobian and Hessian matrix to improve computation speed. I have also tried using $h(.)$ as a second order polynomial following Dubé, Fox and Su without providing the Jacobian. The estimation results are similar to that using only simple instruments.
Newey (1990)discusses finding asymptotically efficient instruments for nonlinear models using nonparametric method. He introduced two methods: k-nearest neighborhood and series approximation–which is the polynomial-based instruments. Series approximation is easier to implement in this case because I need to provide constraint Jacobian to speed up computation; and to derive the constraint Jacobian I need the optimal set of instruments to be differentiable. In fact, this set of instruments performs reasonably well in an efficiency comparison. Reynaert and Verboven (2014) ran a simulation estimating a random coefficient model and found that the set of instruments used in Dubé, Fox and Su outperforms pseudo Monte Carlo integration.

procedures as dynamic constraints, the model can be estimated simultaneously as a MPEC problem.

This problem is also a Multiple-Objective Optimization Problem as we have both the GMM objective function and the maximum likelihood function introduced in section 3.1. The MLE adds a layer of complication to the econometrician's problem, but is necessary to pin down the structural parameter μ, β_0, δ_0. I used the epsilon-constraint method for MOOP first introduced by Haimes (1971) to re-write the MLE objective function as an inequality constraint. The epsilon-constraint method keeps one of the objective functions and rewrites the rest into constraints by restricting them within an econometrician-specified range from their optimal values. Before the estimation, the econometrician must run the optimization problem as a single-objective function problem to obtained the objective values for each objective function, which is the "optimal value" mentioned above. Intuitively, there is a trade-off in optimization when there are multiple objective functions. The epsilon-method prioritizes one objective function as long as the secondary objectives are "good enough."[19] The inequality constraint introduced by this method is:

$$|\mathcal{L}(R_t|\hat{\theta}(\mu; \beta_0, \delta_0), Q_t) - \mathcal{L}^*| \leq \epsilon$$

in which $\mathcal{L}^*$ is the maximized value of the log-likelihood function provided by running the constrained optimization with log-likelihood function as the objective function. The value of ϵ is chosen by the econometrician.[20]

Note that I can take advantage of the linear form to greatly reduce the computation time and the number of constraints. Given any guess of $(\mu^s, \beta_0^s, \delta_0^s)$, we can construct $\{x_{js,t}\}_{j \in \mathcal{J}_f, t \in \mathcal{T}}$ to obtain the matrix of independent variable $\tilde{X}$. The solution $\hat{\alpha}$ that minimizes the GMM objective function $g'Wg$ is the standard GMM estimator: $\hat{\alpha}_{gmm} = (\tilde{X}'ZWZ'\tilde{X})^{-1}\tilde{X}'ZWZ'y$ where W is the GMM weighting matrix.[21] The residual $\hat{\xi} =$

[19]Other simple alternatives include using the simple or weighted sum of objective functions. I have tried both and they yield results similar to that of the epsilon-method. The epsilon-method may allow the econometrician to use conventional GMM inference, but in this paper I use bootstrap standard errors instead. In this sense, applying the simple or weighted sum of objective functions may be more straightforward alternatives.

[20]The main challenge with this method is that the value of ϵ is chosen by the econometrician. An ϵ too small will result in a problem with no feasible solution (as constraint not satisfied), and one too large renders the likelihood constraint useless. In my estimation, ϵ is set as 5% of $\mathcal{L}^*$.

[21]I used the identity matrix as the weighting matrix in the estimation. There are, of course, more efficient weighting matrices, but since I use bootstrap standard errors, asymptotic efficiency of the GMM

$y - \tilde{X}\hat{\alpha}$ can thus be specified rather than solved for as in nonlinear demand system (e.g. in a random coefficient specification). This advantage reduces the number of constraints by almost half.

The optimization problem, written as a MPEC problem, is the following:

$$\min_{\beta_0, \delta_0, \mu, g} \quad g' W g$$

subject to:

$$\begin{aligned}
c_1: &\quad x_{t+1} = \frac{\beta_t + \delta_t}{\beta_t + \delta_t + q_t} x_t + \frac{q_t \mu - r_t}{\mu(\beta_t + \delta_t + q_t)} \\
c_2: &\quad Z'\hat{\xi} = g \\
c_3: &\quad |\mathcal{L}(R_t|\hat{\theta}(\mu; \beta_0, \delta_0), Q_t) - \mathcal{L}^*| \leq \epsilon \\
c_4: &\quad \frac{\beta_0}{\beta_0 + \delta_0} \leq \mu
\end{aligned}$$

Constraint c_1 describes the motion of reputation; c_2 is the moment condition, c_3 specifies the likelihood function necessary to pin down μ, and c_4 guarantees that the initial reputation guess does not go beyond [0,1].

In section 3.1, I mentioned that in the construction of c_3, exporters who have been in the U.S. market for fewer than 10 quarters are dropped. They are still included in the MPEC problem, entering in c_1, c_2 and the objective function. This means I still investigate how consumers respond to reputation of exporters who they don't learn much about. Countries that trade with the U.S. only temporarily are excluded from a constraint about learning parameters because they reveal little how consumers learn. If an exporter is not in the market ("no learning"), then its reputation remains unchanged.

Given that we are solving a constrained optimization problem, where learning parameters β_0, δ_0, and μ are identified with the help of constraint c_3, the conventional GMM inference is no longer suitable.[22] I use bootstrap method instead to calculate standard errors. Since this paper uses a panel model and observations within a country overtime are correlated, I use block bootstrap, where I draw countries with replacement instead of

estimator is not a main concern here.

[22]This is an issue specific to my empirical approach. Dubé, Fox and Su (2012) uses MPEC to estimate a standard random coefficient model, and they can calculate the GMM Variance-Covariance matrix in a standard way. Constrain c_3 and the fact that distribution parameters are primarily identified with MLE make standard GMM inference unreliable in this paper.

country-quarter pairs.[23] I then use the set of countries and their observations overtime to construct one bootstrap sample. The bootstrap sample size for each product is 1000.

3.4 Mapping from Variables to Data

Treating the United States as a representative consumer, we can map the variables to data on an aggregate level. I_t maps onto the quarterly average household expenditure on the relevant consumption products. Within each HS6 category, price $p_{js,t}$ maps onto the unit value of the variety (a HS6-exporter pair) in that year; quantity $q_{js,t}$ the number of units, and $r_{js,t}$ a measure of products recalled described below. If no product s from country j is recalled within quarter t, then $r_{js,t} = 0$.

At the time a recall is issued, consumers receive information about certain product from an exporter. Assume that consumers consider the products imported from that country in a window around the recall to be problematic. In the baseline model, I assume that the window is three months after the recall occurs. For example, if a recall for Chinese toys happens in January 2008, all toys imported from China in January, February and March are considered affected by the event. Formally, $r_{js,t}$ can be calculated as:

$$r_{js,t} = \frac{\sum_{m \in t} Q_{js,t,m} \times \mathbb{1}(\mathrm{R}_{js,t,m} + \mathrm{R}_{js,t,m-1} + \mathrm{R}_{js,t,m-2} \neq 0)}{\sum_{m \in t} Q_{jk,t,m}}$$

where m is the subscript for months and t for quarter. If in a single month, multiple recalls for one variety is triggered, we still count the quantity only once in calculation of $r_{js,t}$. As in the previous example, if there is one recall in January and two in February, products imported from these two months are only counted once.

The market share I calculate in the data is the share of trade value. I calculate market share using value instead of units because the unit of the latter is not always consistent. The U.S. import data set reports two different units for some varieties. For example, in 1990, the port of Miami reported 1169 dozen, or 9096 kilograms (shipment weight), of men's suit jackets containing more than 36% wool imported from Colombia. Some exporters only report one of the two units. A common practice in empirical analysis is to keep only the unit that exceeds the other in terms of numbers of units, but an inconvenience introduced by this treatment is that different exporters might use different units within

[23] See Cameron and Trivedi (2005) chapter 11, section 11.6.2 for a detailed discussion.

one product market. Computing market share in terms of the total value of imports—-a unique number for each entry reported in each year with unambiguous units–allows us to avoid the complication of units for reported quantity. It is worth noting that this problem does not affect the estimation of reputation. The fraction of product recalled is the key in computing reputation, so the unit of quantity is irrelevant. The units of recalled products $r_{js,t}$ and import $q_{js,t}$ for a variety are always the same.

4 Data

4.1 Matching Recall Data to U.S. Import Flows

To analyze the impact of informative signals on the market, I created a novel data set that links CPSC recall incidences from 1990 to 2009 to monthly U.S. import data from the Census. I assigned a six-digit harmonized system code (HS6) to the products that are subject to recalls by reading through the descriptions of recall reports, and link recalls to import data by HS codes, country of origin, and recall time. The import data contains quantity, total value of import trade flows, and shipping information by trade partner, by month, and by HS10 product category, which is aggregated to six-digits level before linking.

The reason for aggregating the more informative HS10 categories to HS6, is that recall events are only matched to HS6 level codes. The data appendix has a detailed discussion of the matching process and why it can only be reliably matched to HS6 codes. The data is then aggregated to quarterly HS6 level, and a time period in the analysis will be a quarter henceforth. I need to aggregate monthly data to quarterly data because the computation of units affected by recalls requires one level of aggregation.[24]

The recall data set contains the date of recall, a brief description of the product, the types of hazards it brings, and its manufacturing countries.[25] In addition to the variables I scraped, a CPSC report typically includes images of the products, remedies, the consumer

[24]An alternative to aggregate over time is to aggregate over HS6 products. A major concern to that method is that by aggregating HS6 to, say, HS4, we are implicitly assuming that HS6 products within a HS4 category are perfectly substitutable. This is not true for some HS4 categories. For example, playing cards and game consoles are both HS6 products under category 9504, but they are not substitutable.

[25]In more recent recall events, the CPSC occasionally reported the price and units sold of the products recalled. The price and units sold are only available after October 1, 2010, so I did not use that information in this paper.

contact, and manufacturers' or retailers' names. All incidences have a recall number, recall date, name, type, and description of the product and pictures, but for more dated recall incidences, other information might be missing. To link recalls to import data, a crucial piece of information from the CPSC is the manufacturing countries of the products. As shown in table 5, from 1990 to 2009, 74.3% of the reports recorded at least one manufacturing country. Each report contains a distinct recall ID. It is possible that in one report multiple products are recorded. That is less common in the entries from recent years, but is more likely for recall reports before 2000. In this case, if all the products recalled are from one HS6 category, I treat it as one incidence; otherwise, I record a separate incidence for each HS6 category included under a recall ID. A few reports record multiple exporters under one recall ID. In this case, I treat an incidence as a recall to each exporter.

The matching is done by reading the recall report title and description, so measurement error is possible. Most recall reports are matched to HS6 level, while some are matched to HS4 level. If a report cannot be matched even to HS4 level, it is categorized as "unmatched" and excluded from the data set. For consistency, I only used the incidences matched to HS6 level in this paper. The main difficulty in the matching process is caused by the different information contained in the Harmonized Tariff Schedule and the CPSC recall reports. The HTS schedule is designed for tariff purposes, so the users are customs officers and exporting firms. It specifies the types of goods and often the compositions of goods, which is a piece of information relevant for tariff purposes and known to the producers. The CPSC recall reports, however, provide a description of the end use and appearance of the product so consumers can immediately identify their purchase. For example, a harmonized tariff code may describe one product as "girl's cotton t-shirt, 90% cotton, 10% polyester" while the CPSC will describe the same product as "girl's red cotton t-shirt with Mickey Mouse". Table 5 summarizes the incidences that are matched to six-digits codes, four-digits codes, and unmatched respectively. Only 0.7% of incidences are unmatched due to ambiguity in description. In addition, the mismatch of product description should occur randomly across exporting countries, so match quality does not bias my estimation results.

4.1.1 Macroeconomic Data

Besides the linked trade flow and recall data, I also need a measure for household budget constraint and the market shares of the outside option. To measure a household's budget constraint on products in my data set, it is not desirable to examine U.S. household income or total expenditures since a large share of household expenditures will be on housing, food, transportation and utilities. Instead, I examine relevant categories of consumption goods expenditures by types of products table provided by the Bureau of Economic Analysis using data from the Consumer Expenditure Surveys. The categories I examine are durable and non-durable goods expenditures, excluding food and beverage, motor vehicles, and gasoline.[26] I excluded those categories because the goods in them are not under the administration of the CPSC, so they are irrelevant to this analysis. I construct the quarterly budget as a fourth of the annual expenditure reported by Consumer Expenditure Surveys. All values are then discounted using Consumer Price Indexes from the Federal Reserve Bank of St. Louis, where 1982-1984 are the base years.

Discrete choice models allow consumers to have an outside option. Following the approach in Khandelwal (2010), the outside option here is to purchase from the United States. Using the annual production data reported in the NBER-CES Manufacturing Industry Database, the U.S. value of sales is calculated as the difference between the value of shipment and the U.S. export value in that year.

4.2 Selecting Products to Estimate a Learning Model

The linked recall data set contains many products, but not all of them are suitable for estimating a learning model. There are two criteria that they need to satisfy: first, recalls are frequent enough that learning can plausibly happen, and second, the product is not durable.

The first criteria is straightforward: if a product only has a couple recalls over almost twenty years, then consumers do not have enough signals for learning to be meaningful. There will be almost no variation in their reputations even if these products are included in the estimation. Thus I keep only products that have at least 25 recall observations over the years, which is the 90th percentile of the 144 products that have at least one recall in

[26] The categories I included are furnishings and durable household equipment, recreational goods and vehicles, other durable goods (like jewelry, books, luggage and phones), clothing and footwear, and other non-durable goods (recreational items, household supplies, stationary). Some non-durable goods in "other non-durable goods" categories are also excluded. They are "pharmaceutical and other medical products" and "tobacco".

the data set.[27] Applying this criteria leaves me thirteen products.[28]

I limit the estimation to non-durable goods for both empirical and theoretical concerns. Among the 13 frequently recalled products, some varieties have units values far exceed the average quarterly household expenditure, which is around \$1000 across the years.[29] These products tend to be expensive durable goods that consumers do not repeatedly purchase, at least not within a year or a quarter. It is then not appropriate to include these products in the estimation of this particular learning model. I drop all the goods with a large fraction of high unit value observations, and that are intuitively non-durable, which leaves me six products: toys, cotton sweaters, sweaters of man-made fabric, battery, lamps and hair dryers. Table 2 summarizes the descriptive statistics of variables in the industries included in the empirical analysis.

In the following section, I will present the parameter estimates and discussion of results for toys. Given that the model and data have variation across countries, products, and time, presenting results for one good helps us to focus on the cross-exporters and cross-time variation. The discussion illustrates mechanics and properties of the model. Once we have clarified the more subtle implications of the model, we will discuss the cross-product variation. I use toy as an example because it is the most frequently recalled product.[30] It also can cause serious health consequences in children, so consumers tend to value safety in this product.

5 Results in the Toy Industry

5.1 Reputation Formation

The update of reputation depends on learning parameters μ, β_0, δ_0 and the history of sales and recalls. Section 2 defines μ as the probability for a bad toy to be recalled. β_0

[27]Here, the recall observation is not an incidence, but a quarter-variety pair. If toys from Spain have recalls in January and March 2007, that will only count as one observation at 2007 Q1 in the product selection process. It will, however, count as two incidences, and it affects how we calculate the fraction of products recalled.

[28]They are toys, cotton sweaters, sweaters of man-made fabric, battery, lamps, hair dryers, ovens, cradles, stoves and ranges, snow mobile, baby trolley, and equipment for outdoor games.

[29]Unit values of ovens imported from United Kingdom, for example, exceed \$1000 for 35 out of 79 quarters in my observations.

[30]Toys have 837 recall incidences over the years, followed by snowmobiles and golf carts, which have 136 recalls.

and δ_0 are initial values of distribution parameters that shape consumers' prior beliefs. Intuitively, β_0 is the units of toys ever recalled and δ_0 is the total units of un-recalled toys sold to the United States before April 1990.

I estimate the probability of recall μ using variation in units of recalls and quantity of imports. Intuitively, keeping the true fraction of unsafe products constant, if μ is close to 1, the model predicts more recalls with relatively small variance within each exporter, because exporters will see consistent recalls (or the absence of them). In the contrasting case when μ is close to 0, the model predicts few recalls with small variance because there is close to no recalls, and when μ is close to 0.5, some recalls but with larger variance. By fitting predicted recalls to actual recalls, μ can be identified as detailed in section 3.1. The initial distribution parameters β_0 and δ_0 are selected using levels and variation of constructed reputation. The ratio between β_0 and δ_0 can vertically shift the predicted reputation. The magnitudes of β_0 and δ_0 governs the impact of the early periods of learning: intuitively, if β_0 and δ_0 are too small (relative to trade flows), the recalls in the first few periods will have a drastic impact on reputation, and if they are too large, reputation will not change much over 20 years.[31]

When we estimate β_0 and δ_0, I normalized observed quantity, units of recalls, and initial guesses of these two parameters. This is a necessary step if we want to accurately estimate β_0 and δ_0, because they are orders of magnitudes larger than all other parameters. Recall that the intuition of β_0 is the cumulative recalled and safe products before observation, respectively. Given the volumes of trade we commonly observe in a quarter, this number can go up to millions, even billions in some observations. Without normalization, the values of β_0 and δ_0 are different from other parameters by a magnitude of 10^6 or even 10^9, making it hard for the algorithm to search through value space for an optimized solution.[32] Normalization is a simple solution to this problem, but how do we deal with trade data where, within one product, some countries sell billions of units (e.g. China, Mexico),

[31] Consider an example in which an exporter sells 1000 units to the U.S. every quarter, 10% of them are defects, and $\mu = 1$. Supposing $\beta_0 = 3$ and $\delta_0 = 5$, after one period of update, we have $\beta_1 = 1003$ and $\delta_1 = 1005$. The change in reputation induced by recalls from the first period is $\frac{-100}{1000+5+3} \approx -0.099$, but in the second period will be $\frac{-100}{1000+1005+1003} \approx -0.033$. Changes in reputation vary widely if the initial guess β_0 and δ_0 are too small. On the contrary, if β_0 and δ_0 are close to infinity, reputation will be constant, as the "weight of history" is so large that new information has almost no impact. I set the initial values for β_0 and δ_0 to have a comparable magnitude with trade flows, so the reputation variation in the first few periods are not too drastic.

[32] Imagine taking steps of 10^{-6} and starting with an initial guess of 1.4×10^6: how many steps will the algorithm need to take to find the local optimal at, for instance, 3×10^6?

and some only hundreds? In addition, volumes of trade vary by orders of magnitude across products. To tackle these issues, I choose a different unit of normalization for each product. The unit is chosen such that the median observation is a number between 1 to 10, so observations from both the largest exporters and the smallest are normalized to a number that can be accurately calculated.

In addition to the variance of recalls, the model can distinguish between μ and initial distribution parameters β_0 or δ_0 by comparing the changes in reputations when a recall breaks out. Consider the cases of a low μ or a high β_0: both can lead to a lower initial value of reputation and shift the reputation downwards. Reputation is more responsive to recalls if μ is low because when bad products are unlikely to be recalled a recall become more alarming.

To visualize how learning parameters change reputation, I chose one of the varieties to illustrate how estimated reputation changes when learning parameters change, while keeping observed recalls and import quantity as given.[33] figure A.3 uses toys imported from Hong Kong to illustrate the vertical shift when β_0 or δ_0 are reduced by half. We can see that β_0 and δ_0 move initial value of reputation in different directions, but as new information accumulates, reputation approaches the estimated value. Value change of β_0, the cumulative recalled products before the first period of observation, has a more persistent impact than a similar change of δ_0, as recalls are rare and it is harder for new information to "cover" the history.

figure A.6 take estimated β_0 and δ_0 as well as observed quantity as given, and plot how reputation changes when μ, the probability of a defect being recalled, is 0.9 and 0.6 respectively. In addition, I add a scenario where recall is simulated to be consistent with $\mu = 0.6$. We can see that when μ decreases from 0.9 to 0.6, and when both quantity and recalls are kept as observed in data, the initial reputation decreases and reputation does not catch up overtime. Reputation declines as μ decreases because if consumers expect a lower probability for defects to get recalled, they will treat each recall more seriously. Unlike the case when β_0 and δ_0 change, when μ decreases, given the number of recalls observed in data, reputation will not converge to the value estimated in model because the implied fraction of defect has changed. If we replace the number of recalls observed in data

[33] This simulation is different from the simulation I will discuss in section 5.4, although both exercises illustrate how reputation changes with learning parameters. The purpose of this simulation is merely illustrating data pattern that I can use for identification, so recalls are taken as given. The simulation in section 5.4 also simulates recalls to understand the value of information for consumers.

with recalls simulated from observed quantity, estimated reputation, and $\mu = 0.6$, then we can see that reputation will converge to estimated reputation overtime.

Table 1: Parameter Estimates for Toys

Parameter Estimates			
Description	*Parameter*	*Estimate*	*S.E.*
Recall probability given product is low quality	μ	0.9257	(0.3578)
Sum of recalled units before 1990 (millions)	β_0	488.315	(328.06)
Sum of safe products before 1990 (millions)	δ_0	404.038	(114.44)
Preference for Reputation	α_x	5.5688	(2.255)
Coefficient of log(budget-price)	α_0	8.547	(0.663)

Descriptive Statistics of Reputation in the Last Period				
	Mean	*Std. Dev.*	*Min*	*Max*
All Countries	0.4263	0.0742	0.0931	0.9341
Highest reputation quartile	0.4871	0.1217	0.4102	0.9341
Lowest reputation quartile	0.4003	0.0519	0.0931	0.4088
Conditions of Learning				
Periods of Learning	29.154	28.02	1	79
Initial Reputation	0.4088	-	-	-
Number of Countries	149	-	-	-

Note: μ is robust to different initial guesses. I chose 10 guesses spacing equally between 0.1 and 1: all return the same estimate. Initial guess for β_0 is 10 times the average units of recalled products; and for δ_0 10 times the average units of goods sold. Standard errors presented in the table are bootstrap standard errors with 1000 bootstraps.

Panel 1 of table 1 presents the estimates for learning parameters. Panel 2 summarizes the reputation across exporters and time, constructed using the estimates in panel 1. Panel 3 lists periods of learning, the average number of quarters a country exports to the United States, initial reputation, reputation in the first period calculated using estimates in panel 1, number of exporters ever selling to United States since 1990, and number of exporter-quarter pair in this industry. If all exporters stay through the 79 quarters in data set, we can hypothetically have $149 \times 79 = 11771$ observations. Instead, many exporters only export to the United States for a few quarters, so there are only 4376 exporter-quarter pairs in the data. After dropping some exporters who have only exported for a couple years to U.S., we have 3436 observations left to estimate μ, β_0, δ_0.

Panel 2 of table 1 shows the summary statistics of estimated reputation across time and exporters. In the last period, the exporters of toys with best reputations are Mexico and Canada, corresponding to the maximum reputation 0.934 and 0.878; and the minimum corresponds to China. Canada has recalls in only two quarters of the 20 years in my observation, while China has at least one recall in 76 out of the 79 quarters. Both countries export to the United States in all periods, and they export in large quantity. Most exporters—127 out of 149—have never had a product recalled by the CPSC. The consistent presence of Mexico and Canada in the U.S. market and large exports make them stand out among the exporters who have always been safe.

In the preferred specification, I use the unit freight cost as the instrument for price. Table 6 shows that unit freight cost passes the "rule of thumb" test for instrument relevance for most non-durable goods, and it is strong for toys.[34] Exchange rate, though intuitively should be correlated with price, has a weak correlation. This is not surprising given how volatile exchange rate is over time and how big the variation is across currencies. Unit freight cost and oil price times distance have the same channel: cost of transportation enters the "cost, insurance, freight"(CIF) value in the import data set. When we include both, one of the two instruments will appear to be not-correlated, thus keeping only one is sufficient. Table 7 shows that including additional instruments does not change the results much. Note that different instrument specifications in table 7 are estimated from a two-step procedure instead of the one-step MPEC estimates reported in table 1. The two-step procedure takes the reputation constructed using learning parameters estimated in GMM, and runs regression 5. Although it is not the preferred specification since it cannot estimate all parameters simultaneously, it has significant speed advantage and I use it to illustrate alternative regression specifications.[35] Comparing the results in column 4 in table 7 and table 1, we can see that the two-step procedure provides point estimates similar to the one-step MPEC estimates. Column 4 and 5 in table 7 show that adding additional instruments hardly change the point estimates or F-statistics.

[34] Stock, Wright and Yogo (2002) suggests that F-statistics$<$10 should raise concerns of weak instruments in the GMM estimation. Choosing an instrument that works for all industries is challenging, and unit freight cost is the best-performing instrument among those commonly used in the literature.

[35] In addition to run time concerns, changing number of instruments is not a trivial exercise for the MPEC problem, because I provide Jacobian matrix to speed up computation in the MatLab codes. Each additional instrument specification requires a different version of Jacobian. The direction of change in demand coefficients should be the same between two-step procedures and the one-step MPEC, so to illustrate this point the two-step procedure suffice. The biggest difference between one step MPEC and two steps procedure is that the latter has much smaller standard errors.

5.2 Demand response to reputation

Panel 1 of table 1 displays α_x and α_0, the market share responses to reputation and the natural logarithm of budget minus product price. These two parameters reveal how sensitive consumers are to reputation and price. A positive coefficient for reputation implies that it is rewarding for exporters to maintain or aim for higher reputation. The higher the coefficient is, the more consumers are concerned about reputation in this product.

The coefficient of reputation implies that the "reputation elasticity of market share" is 2.396 for toys.[36] If an average exporter of toys can increase reputation by 10%, it can expect to increase its market share by 23.96%. This is a somewhat big change, but given that reputation is history-dependent, it will take the average exporter many periods of safe presence in the U.S. market to achieve that.

To illustrate how long it will take an exporter to improve reputation, I take the reputation in the last period, and predict how long it will take for each exporter to increase reputation by 10% in two scenarios. The first scenario assumes that in each future quarter an exporter sells the same quantity into the United States, which is equal to the average quarterly quantity from the second quarter of 1990 to the last quarter of 2009. I run the reputation updating procedures with no recalls until the reputation reaches target level. An average large exporter of toys who is among the upper quartile in export quantities will need to have a safe presence for 57.5 years consecutively to improve its reputation by 10%. It will take even longer for small exporters because the information update is slow when consumers see few new units in the market. Even for the largest exporter of toys, China, catching up is difficult. It will take 227 years of flawless presence in the United States for its reputation to catch up with that of Mexico, the exporter currently enjoying the best reputation in the U.S. market.

In the second scenario, reputation growth rate polarizes when the simulation includes demand responses. I relax the assumption of sales volume in the previous simulation, allowing market share to change as reputation improves while fixing total units of sales in the U.S. I simulate 10 million agents for 1000 quarters, each choosing an exporter in every

[36] Reputation elasticity of market share is the percentage change of market share induced by one percentage change of reputation. Use σ to denote the reputation elasticity, it is calculated as: $\sigma = \frac{d \ln s}{d \ln x} = \frac{d \ln s}{dx} \cdot \frac{dx}{d \ln x} = \alpha_x \cdot \frac{1}{1/x} = \alpha_x \cdot \bar{x}$, where $\bar{x}$ is the average reputation. The change in market share is relative to the U.S. market share since that is the outside option.

period, and aggregate their choices to market shares. However, most exporting countries have market shares way smaller than 10^{-7}, so simulated market shares will match actual shares of these countries poorly. Instead of using the full set of countries, I focuses on 12 exporting countries who are the top 10% in terms of export quantity.[37] The largest exporter, China, now takes only 262 quarters instead of 908 quarters to catch up with Mexico because its market share increases with reputation. Mexico can improve its reputation to perfection in 1483 quarters, which is much longer compared to 69 quarters in the simulation with the first scenario. Mexico's market shares decline since its reputations cannot improve as fast as China's, and the decline in market share further stagnates improvement of reputations. All other countries in the simulation improve reputation by less than 10% within 1000 quarters, implying that it takes longer to improve reputation for most exporters when market share can change. Investing in quality inspection is more beneficial to large exporters, as they have initial and ongoing advantages in reputation improvement.

5.3 Predictive Power of the Model

To evaluate how well the model predicts out-of-sample market shares, I split my data set into two parts. I estimate the model using the first part of the data (training data), and construct out-of-sample predictions using estimated parameters and variables from the second part of the data (test data). I then correlate the observed outcomes in test data with the predicted outcomes. The division follows a roughly 70-30 split, where the training data contains observations from the first quarter of 1990 to the last quarter of 2003 (55 quarters), and the test data contains observations from the last six years in my observations (24 quarters).

When constructing out-of-sample predictions, there are two caveats. The first is the construction of reputation. Unlike traditional out-of-sample predictions, we cannot use the learning parameter $(\hat{\beta}_{0,t}, \hat{\delta}_{0,t})$ estimated from data up to period t directly to calculate the distribution of initial beliefs from periods $t+1$ to $t+n$, because β_0 and δ_0 determine consumers' belief before they make any observations. By period $t+1$, even the underlying data generating process is exactly the same in period 1 to t and period $t+1$ to $t+n$,

[37]Since the top 10% exporters can change from quarter to quarter, the final set is the union of all the top 10% countries in each period. These 12 countries together export 92.99% of foreign toys in the United States.

consumer's prior beliefs should include information from the first t periods. With this intuition in mind. the prior belief, or "initial" belief, in the out-of-sample prediction is the reputation in the final period estimated using training data.

The second caveat is how to handle time-varying characteristics in the model. The model controls for time fixed-effects, but we do not know the proper values of quarter-fixed-effects in the test periods. One straightforward solution is to take a within transformation by subtracting country averages across all periods from each observation. Formally, the prediction is:

$$y_{j\tau} - \bar{y}_{\tau} = (X'_{j\tau} - \bar{X}'_{\tau})\hat{\alpha}_t + \psi_j + u_{j\tau} - \bar{u}_{\tau}$$

where $\tau = t+1, t+2, ..., t+n$, $\hat{\alpha}_t$ is the coefficient estimated using observations up to period t, and the average of each variable is taken over all countries in that period of time, or $\bar{x}_{\tau} = \frac{1}{J}\sum_{j=1}^{J} x_{j,\tau}$.

The corresponding predicted outcome variable, after adjusting for reputations and time-fixed-effects, will be a demean variable $\hat{y}_{j\tau} - \bar{\hat{y}}_{\tau}$, where $y_{js,t} \equiv \ln(S_{js,t}) - \ln(S_{US,s,t}) - \ln(p_{js,t})$ as specified in equation 5. I then compare it to the data-equivalent of the de-mean variable, subtracting the average outcome variable across countries in each period from the outcome variable.

While time fixed-effects are dropped, I keep (most of) the country fixed-effects. In out-of-sample predictions, I keep only countries that are present for at least one period in training data, because predictions without valid estimation results are not meaningful. That means I exclude all new export partners the US forms with the rest of the world in the test periods when calculating out-of-sample predictions.

Table 3 presents the correlation and root-mean-square-errors(RMSE) between the predicted outcome variable and observed outcome variable. Panel 1 presents the results from the training data, which is taken from the first period of my data set (1990 q1) to the last quarter of 2003. I present the correlation of predicted and observed y in the training data set, to provide a direct comparison between estimation fit and out-of-sample fit.

Panel 1 of table 3 shows that the model fits data reasonably well in the training data, with correlation higher than 0.8 across all products. Out-of-sample predictions in Panel 2 perform only slightly worse than the estimation fit.

It is worth noting that the test periods include year 2007 and 2008, when many recall incidences occurred. This means that this model performs well even when periods of unusually frequent and high-profile recalls are excluded.

5.4 Discussion: Quantifying the Value of Information

Every year, the Consumer Product Safety Commission submits a budget request to the Congress. The budget request for fiscal year 2019, for example, is 123.5 million dollars. From a policy maker's perspective, it is meaningful to ask how important a quality inspection institution like the CPSC is to domestic consumers. The model answers this question from an information perspective.

Consider two scenarios, one in which the inspection institution can catch and recall unsafe products more effectively than the other. Under the more effective scenario ("high inspection accuracy"), if a product is unsafe, it will be caught with 90% chance while in the other scenario ("low inspection accuracy"), that probability is 50%. Note that a low μ does not mean noisier signals: recalls still only signal unsafe products, but the signals are rarer. I measure welfare changes using compensating variation, that is, the changes in income to make consumers indifferent between having high and low inspection accuracy. I assume that in both scenarios, the size of market and the underlying fraction of unsafe products are the same. Let x^L denote the reputation in the low accuracy scenario and x^H in high accuracy scenario. The compensating variation $\text{cv}_{s,t}$ satisfies:

$$\alpha_0 \log(I_t - p_{j^*s,t}) + \alpha_x^s x^H_{j^*s,t} + \eta_{j^*} + \psi_t = \alpha_0 \log(I_t - p_{j's,t} + \text{cv}_{s,t}) + \alpha_x^s x^L_{j's,t} + \eta_{j'} + \psi_t$$

Note that, here, j^* is the exporter that consumer chooses in high inspection accuracy scenario, and j' is the exporter chosen in the other scenario. j^* and j' need not be the same. I assume that when the quality of signal is low, consumers are aware of it and incorporate that knowledge in learning.

I take the underlying fraction of unsafe products as given, and simulate the recall events and consumer learning under high inspection accuracy ($\mu = 0.9$) and low accuracy ($\mu = 0.6$), then I compute the compensating variation for consumers of toys.[38] To simulate recall events, I assume the last period reputation is the best proxy for the true unobserved fraction of bad products. Taking quantity imported in the United States as given, the

[38] I am not aware of any empirical work that specifies the effectiveness of CPSC recalls, so there is no obvious benchmark for this exercise. I pick a high μ as it is close to the estimated value of μ in toys, and a low μ that is roughly one bootstrap standard error below the point estimate.

number of bad products $L_{js,t}$ is the product of reputation estimates in the final period and the quantity of imports. Each unit of bad product has probability μ of being recalled, so the total number of products recalled roughly follows a normal distribution with mean $\mu L_{js,t}$ and variance $\mu(1-\mu)L_{js,t}$. After generating number of products recalled for each exporter in each quarter, I can run the reputation updating following equation 2 to construct $x^L_{j's,t}$ or $x^H_{j's,t}$ under each scenario.

The simulation generates 12,000 agents with individual preferences drawn from an i.i.d. Extreme Value distribution. It provides two measures of welfare: the total compensating variation for the U.S. market and the average compensating variation for each purchase.[39] For each exporter j in each quarter t, a set of simulated agents choose their products (that set can be empty). The compensating variation for an average consumer who buys from country j, multiplying the total units of products the U.S. imports from country j in period t, gives us the country-time specific simulated compensating variation. Total compensating variation in each period is the summation of simulated compensating variation from all exporters present in that period. Average compensating variation is calculated as the total compensating variation divided by the units of product s imported into the U.S. market in period t, which is equivalent to the average of simulated agents' compensating variation weighted using the import share of the countries agents choose to buy from.[40] Total compensating variation is driven by both the change in average compensating variation—a channel that reveals the impact of information—and the changes in demand.[41] While total compensating variation highlights the magnitude of impact, the average compensating variation excludes the impact of import quantity, so it can better reveal the model mechanisms.

The welfare loss per purchase averages around \$6.54 per quarter when inspection accuracy is low.[42] If we consider the volume of purchase in toys, however, the total welfare

[39] Extreme value distribution takes location parameter $\mu = 0$ and scale parameter $\sigma = 1$.

[40] The equations to calculate total and average compensating variation from the simulation are the following. Let $\mathrm{cv}(j^i)_{i,t}$ denote the compensating variation for individual i who chose exporter j^i to purchase from in period t, and q_{j^i} denote the quantity imported from exporter j^i. Total $\mathrm{CV}_t = \dfrac{1}{12000}\displaystyle\sum_{i=1}^{12000} \mathrm{cv}(j^i)_{i,t} \times q_{j^i}$ and Average $\mathrm{CV}_t = \dfrac{\text{Total CV}_t}{\sum_{j\in J} q_{j,t}}$

[41] The total quantity demanded is not explicitly modeled in this framework as I focus on changes in market share, the demand *relative to* your competitors given the number of consumers.

[42] All dollars are converted into 1982-1984 dollars using CPI, and later quarters are discounted using discount factor 0.995.

loss can average 5.98 billion dollars per quarter. Panel 1 of figure A shows the total compensating variation for toys. The welfare loss from lower inspection accuracy comes from lower mean utility when μ is low and also higher chance of landing an unsafe product.

Two Mechanisms of Welfare Changes

To fully understand the sources of welfare differences under two scenarios, figure A decomposes the two channels through which welfare changes, and I call them the "mean value difference" and "defect surprise." $V(H)'$ denotes the maximized consumer utility after revelation of the product quality in the high μ (strong inspection) scenario, and $V(H)$ is the mean utility before revelation of quality in the high μ scenario. $V(L)'$ denotes the utility after revelation of quality in the low μ (weak inspection) scenario, and $V(L)$ the mean utility before revelation in the low μ scenario. The following equation describes the decomposition:

$$\overbrace{V(H)' - V(L)'}^{\text{Gains from having higher } \mu} = \underbrace{\overbrace{[V(H)' - V(H)]}^{\text{Loss from recall (high } \mu)} - \overbrace{[V(L)' - V(L)]}^{\text{Loss from recall (low } \mu)}}_{\text{``defect surprise''}} + \underbrace{[V(H) - V(L)]}_{\text{Mean utility differences}}$$

When the probability of bad products getting recalls is lower, consumers have a more pessimistic reputation assessment: when observing the same number of recalls, consumers expect the actual fraction of bad products to be higher. As a result, I fix the actual fraction of bad products in the market, and simulate the number of recalls so that they are consistent with the recall probability μ in each scenario. The mean utility differs because prices will be different for the consumers who choose another exporter in the alternative scenario, and reputation changes for all consumers. In addition, consumers who get a unsafe product will take a utility reduction after revelation, and I call this damage "defect surprise." A positive "defect surprise" suggests that utility loss from recall is less damaging when inspection is more effective. Under the weak inspection scenario, consumers will observe fewer recalls but treat each one with greater caution because they know the probability of recall is lower. It will take them longer to approach a more accurate estimate of true fraction of defect products. As a result, under weak inspection consumers will be "surprised" by a defect product more, which incurs a cost illustrated as a curve *above* horizontal axis in figure A.

figure A.6 illustrates that when μ is low, the reputation estimates are lower initially because consumers have a more pessimistic prior, but eventually the reputation will catch up and approach the "true fraction of good products" specified in the simulation. Exactly how long it will take for a country's reputation to converge back to the true fraction, however, depends on the quantity of import from that country. figure A.7 shows that the reputation of China converges quickly because China is a large exporter throughout the years, but for Mexico the discrepancy remains large till the late 1990s when the quantity of toys sold to the U.S. increases. Lower inspection accuracy decreases reputations for all exporters, but the damages are more severe and persistent for exporters who sell fewer units.

The mean value difference can have ambiguous impact. A positive mean value difference, illustrated as when the curve is above horizontal axis, means that the expected utility is higher when inspection is strong. However, weaker inspection can sometimes increase consumers' expected utility. When μ decreases (inspection becomes weak), larger exporters' reputations reduce less compared to smaller exporters.[43] A marginal consumers may choose to buy from a large exporter when inspection becomes weak. If large exporters happen to sell cheaper products and the reduction in consumer expenditures compensates the reduction in reputation, then the saving can lead to higher mean utility. As we can see from figure A though, when inspection is weak the reduction of reputation usually creates a loss in utility that far exceeds the price differences.

The summation of "defect surprise" and "mean value difference" is the costs to having less effective inspection calculated in compensating variation. In the rare case when the benefit of cost-saving outweighs the higher risk of getting a defect, it is possible that better inspection is not welfare-improving. However, the simulation suggests that it is unlikely, and under most scenarios better inspection improves consumer welfare.

Simulation also reveals that smaller exporters benefit more from a highly effective inspection institution. figure A.8 compares market shares when inspection accuracy is high and low. All exporters lose market shares when μ is low because purchasing from any exporter is perceived to be riskier now and consumers prefer the outside option. After several periods however, the market share recovers and the lowest reputation exporter—China—even have a small gain in market share towards the second half of the observed periods. This is seemingly surprising, until we realize that the low reputation exporter (China) also happens to be the largest exporter. A downward shock in inspection effectiveness hurt

[43] Since these two hypothetical scenarios cannot co-exist, there are no actual switchers. The marginal consumers "switch" in the sense that they will choose differently under the alternative scenario.

reputation of all exporters, but larger exporters recover faster. Given the large import volume from China, the gap between reputations before and after the shock closes much earlier for China than other exporters, so it actually gains a temporary advantage: not from better reputation, but from resilience to information quality shock.

6 Results across Industries

In the previous section, I use toys as an example to illustrate model mechanisms and what they can do in terms of welfare and counterfactual analysis. This section discusses estimation results for other products, revealing heterogeneity in consumers' concern of safety across products. The results carry interesting policy implications for any exporter trying to improve quality and for domestic institutes like the CPSC who may need to budget quality inspection expenditures across types of products.

Table 4 shows the difference across products in term of consumers' preferences for reputation. Column 1 and 2 show the coefficients from the MPEC estimation, and column 3 and 4 show the corresponding market share elasticities. Toys, unsurprisingly, is the product that consumers have the strongest preference for safety, with a market share elasticity of reputation of 2.396. It is also the only product that has a coefficient estimate that is statistically significant. Overall, I did not find evidence that the market punishes any product other than toys for having a bad reputation. The market reaction cannot be fully explained by the nature of hazards caused by product defects. Table 8 lists the most frequent hazards for each type of products, and we can see that the most frequent hazards for toys either can be fatal (choking) or can cause long-term distress for users (lead paint). However, CPSC recalls products that threaten consumer safety, so hazards tend to be pretty severe. It is worth noting that the majority of the apparels recalled by the CPSC are clothes for children, although the harmonized system code category may not distinguish products for adults from products for children. [44] The market responses I measure for cotton sweaters or sweaters made from man-made fabric include clothes for adults, which can potentially attenuate the results even the market for infant clothes responses to reputation. Consumers' preference may not only reflect the types of hazards, but also to whom hazards may occur: the same hazard can be far more damaging when it happens to a vulnerable child, which can explain the larger coefficient estimates for toys.

[44] For example, category 620520 describe the product as "Men's or boys' shirts of cotton"

Even we find no evidence for market responses to recalls in products other than toys, the consumer welfare gains from toys alone exceeds the budgets used by a quality inspection institution like the CPSC. If importers or exporters decide to invest in quality inspection, they should prioritize products primarily used by children, since consumers seem to have strong preferences for safe products in these categories. However, reputation improvement can take decades even for large exporters.[45] For most exporters of products used by children, improving reputation can increase their market share. That may not be the case for exporters of other consumption goods, so exporters may have weaker incentives to invest in quality control, and choose to compete through lower prices.

7 Conclusion

This paper analyzes the effect of an exporting country's reputation on import trade flows. It defines an exporter's reputation as the expected probability of getting a high quality product in a market, and it evolves as consumers observe more signals. This paper tackles the challenge of identifying intangible and unobserved reputation in two ways: constructing a data set in which I can observe shocks that affect reputation, and modeling channels in which reputation affects consumers' decisions. The model in this paper can be generalized to estimate consumers learning of any signals in trade, for example, how the market reacts to a scandal that is widely cover in traditional and social media, like the Volkswagen diesel emission scandal.[46]

This paper is a step towards understanding the role of consumers' learning in international trade. There are at least three directions of future research. First, this model uses Bayesian learning with perfect memory. Future work can incorporate imperfect memory models to explore how the reputation dynamic changes. Second, this paper focuses on estimating the learning dynamic for goods that are purchased frequently. Durable goods likely have a different information acquisition dynamic that we can explore. Third, this model abstracts away from exporting firms' decision on investing in quality improvement. Given that reputation matters to market shares of some products, incorporating the producer's decision can be an interesting next step.

[45] It is generally hard for small exporters to improve reputation, but it is especially hard for small exporters who used to be large. More developed Asian exporters (like Hong Kong and South Korea) have displayed this pattern for products like toys.

[46] See BBC News (2015) for a detailed report.

References

Albornoz, Facundo, Héctor F Calvo Pardo, Gregory Corcos, and Emanuel Ornelas. 2012. "Sequential exporting." *Journal of International Economics*, 88(1): 17–31.

Bar-Isaac, Heski. 2003. "Reputation and Survival: learning in a dynamic signalling model." *The Review of Economic Studies*, 70(2): 231–251.

Bar-Isaac, Heski, Steven Tadelis, et al. 2008. "Seller reputation." *Foundations and Trends in Microeconomics*, 4(4): 273–351.

Basker, Emek, and Fariha Kamal. 2020. "Recall and Response: Relationship Adjustments to Adverse Information Shocks." *Working Paper*.

BBC News, Russell Hotten. 2015. "Volkswagen: The scandal explained."

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *The RAND Journal of Economics*.

Board, Simon, and Moritz Meyer-ter Vehn. 2013. "Reputation for quality." *Econometrica*, 81(6): 2381–2462.

Bolton, Patrick, and Christopher Harris. 1999. "Strategic experimentation." *Econometrica*, 67(2): 349–374.

Bond, Eric W. 1984. "International trade with uncertain product quality." *Southern Economic Journal*, 196–207.

Cameron, A Colin, and Pravin K Trivedi. 2005. *Microeconometrics: methods and applications.* Cambridge university press.

Chen, Maggie X, and Min Wu. 2020. "The value of reputation in trade: Evidence from Alibaba." *The Review of Economics and Statistics*, 1–45.

Chisik, Richard. 2002. "Reputational comparative advantage and multinational enterprise." *Economic Inquiry*, 40(4): 582–596.

Commission, The Consumer Product Safety. 2016. "Products Under the Jurisdiction of Other Federal Agencies and Federal Links." Accessed: 2016-12-1.

Commission, United States International Trade. 1987-2012. "HTS Archive." *https://www.usitc.gov/tata/hts/bychapter/index.htm*, Accessed: 2016-11-26.

Creane, Anthony, and Kaz Miyagiwa. 2008. "Information and disclosure in strategic trade policy." *Journal of International Economics*, 75(1): 229–244.

Daalhuis, Adri B. Olde. 2018. "NIST Digital Library of Mathematical Functions." *http://dlmf.nist.gov*.

Donnenfeld, Shabtai. 1986. "Intra-industry trade and imperfect information about product quality." *European Economic Review*, 30(2): 401–417.

Donnenfeld, Shabtai, and Wolfgang Mayer. 1987. "The quality of export products and optimal trade policy." *International Economic Review*, 159–174.

Donnenfeld, Shabtai, Shlomo Weber, and Uri Ben-Zion. 1985. "Import controls under imperfect information." *Journal of International Economics.*

Dubé, Jean-Pierre, Jeremy T Fox, and Che-Lin Su. 2012. "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation." *Econometrica*, 80(5): 2231–2267.

Eaton, David H. 2005. "Valuing Information: Evidence from Guitar Auctions on eBay." *Journal of Applied Economics and Policy*, 24(1): 1–19.

Eaton, Jonathan, Marcela Eslava, Cornell J Krizan, Maurice Kugler, and James Tybout. 2009. "A search and learning model of export dynamics." *Penn State University mimeo.*

Falvey, Rodney E. 1989. "Trade, quality reputations and commercial policy." *International Economic Review*, 607–622.

Falvey, Rodney E, and Henryk Kierzkowski. 1984. *Product quality, intra-industry trade and (im) perfect competition.* Graduate Institute of International Studies.

Freedman, Seth, Melissa Kearney, and Mara Lederman. 2012. "Product recalls, imperfect information, and spillover effects: Lessons from the consumer response to the 2007 toy recalls." *Review of Economics and Statistics*, 94(2): 499–516.

Grundke, Robert, and Christoph Moser. 2019. "Hidden protectionism? Evidence from non-tariff barriers to trade in the United States." *Journal of International Economics*, 117: 143–157.

Haimes, YV. 1971. "On a bicriterion formulation of the problems of integrated system identification and system optimization." *IEEE transactions on systems, man, and cybernetics*, 1(3): 296–297.

Holloway, Isaac R. 2017. "Learning via sequential market entry: Evidence from international releases of US movies." *Journal of International Economics*, 104: 104–121.

Houser, Daniel, and John Wooders. 2006. "Reputation in auctions: Theory, and evidence from eBay." *Journal of Economics & Management Strategy*, 15(2): 353–369.

Hummels, David, and Peter J Klenow. 2005. "The variety and quality of a nation's exports." *American Economic Review*, 704–723.

Jacod, Jean, and Philip Protter. 2004. *Probability essentials.* Springer Science & Business Media.

Jovanovic, Boyan. 2021. "Product recalls and firm reputation." *American Economic Journal: Microeconomics (Forthcoming).*

Khandelwal, Amit. 2010. "The long and short (of) quality ladders." *The Review of Economic Studies*, 77(4): 1450–1476.

Livingston, Jeffrey A. 2005. "How valuable is a good reputation? A sample selection model of internet auctions." *Review of Economics and Statistics*, 87(3): 453–465.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. 2014. "Promotional reviews: An empirical investigation of online review manipulation." *The American Economic Review*, 104(8): 2421–2455.

Newey, Whitney K. 1990. "Efficient Instrumental Variables Estimation of Nonlinear Models." *Econometrica*.

Petrin, Amil. 2002. "Quantifying the benefits of new products: The case of the minivan." *Journal of Political Economy*, 110(4): 705–729.

Pierce, JR, and PK Schott. 2009. "VConcording US Harmonized System Codes Over Time." *Journal of Official Statistics*.

Potoski, Matthew, and Aseem Prakash. 2009. "Information asymmetries as trade barriers: ISO 9000 increases international commerce." *Journal of Policy Analysis and Management*, 28(2): 221–238.

Rauch, James E, and Joel Watson. 2003. "Starting small in an unfamiliar environment." *International Journal of Industrial Organization*, 21(7): 1021–1042.

Reynaert, Mathias, and Frank Verboven. 2014. "Improving the performance of random coefficients demand models: the role of optimal instruments." *Journal of Econometrics*, 179(1): 83–98.

Saeedi, Maryam. 2019. "Reputation and adverse selection: theory and evidence from eBay." *The RAND Journal of Economics*, 50(4): 822–853.

Schott, Peter K. 2004. "Across-product versus within-product specialization in international trade." *The Quarterly Journal of Economics*, 647–678.

Stock, James H, Jonathan H Wright, and Motohiro Yogo. 2002. "A survey of weak instruments and weak identification in generalized method of moments." *Journal of Business & Economic Statistics*, 20(4): 518–529.

Swann, Peter, Paul Temple, and Mark Shurmer. 1996. "Standards and trade performance: the UK experience." *The Economic Journal*, 1297–1313.

Tables

Table 2: Summary Statistics

Variables	Statistics	Toys	Sweaters[a]	Sweaters[b]	Battery	Lamps	Hair Dryers
Market share	**Mean**	0.0152	0.00948	0.0103	0.0155	0.00866	0.0499
	Median	0.000121	0.001044	0.0004127	0.00081	0.000206	0.00163
	Max	0.936	0.403	0.442	0.467	0.406	0.619
	Min	7.77×10^{-8}	8.90×10^{-8}	1.82×10^{-7}	4.28×10^{-6}	3.09×10^{-6}	1.67×10^{-5}
Price	**Mean**	38.6	27.4	33.2	47.3	69.2	23.9
	Median	7.31	18.01	19.51	14.07	28.22	16.04
	Max	734	130	159	670	562	160
	Min	0.002637	1.19	1.084	0.06495	0.4885	1.506
Quantity (in millions)	**Mean**	14.3	0.823	0.528	1.09	0.407	0.484
	Median	0.02985	0.06729	0.0148	0.01575	0.003556	0.006247
	Max	1704	59.1	31.8	39.8	25.5	8.21
	Min	2×10^{-6}	3×10^{-6}	2×10^{-6}	4×10^{-6}	3×10^{-6}	1.8×10^{-5}
Trade Value (in millions)	**Mean**	32.6	12.5	8.22	5.24	3.47	3.54
	Median	0.2498	1.044	0.2992	0.1996	0.08098	0.1118
	Max	3063	1042	507	195	200	53.9
	Min	2.52×10^{-4}	2.51×10^{-4}	2.52×10^{-4}	1.26×10^{-3}	1.26×10^{-3}	1.29×10^{-3}
Units of Recall (in millions)	**Mean**	8.04	0.0517	0.0161	0.164	0.0726	0.0438
	Median	0	0	0	0	0	0
	Max	1575	59.1	29.5	39.8	21.7	5.7
	Min	0	0	0	0	0	0
Ratio of Recall	**Mean**	0.0341	0.00442	0.00341	0.00876	0.00561	0.0167
	Median	0	0	0	0	0	0
	Max	1	1	1	1	1	1
	Min	0	0	0	0	0	0
US market share	**Mean**	0.125	0.109	0.155	0.518	0.642	0.367
	Median	0.1143	0.05369	0.1317	0.423	0.6364	0.3616
	Max	0.301	0.346	0.458	0.963	0.784	0.617
	Min	7.82×10^{-20}	3.43×10^{-3}	0.01473	0.3032	0.5261	0.2715

Note: a: Sweaters made of cotton, HS6=611020. b: Sweaters made of man-made fabric, HS6=611030.
Source of trade data is the monthly U.S. Census import data. Recalls come from the CPSC recall database. U.S. manufacturing data comes from NBER-CES data set. All summary statistics are reported from the quarterly data set aggregated from monthly data. The variables, from top to bottom, represent: 1) market share calculated from import values. 2) row reports unit value of import. 3) quantity imported to the U.S. in the unit that reports a larger number of quantity. 4) value of trade in current USD. 5) quantity of recalled products in the same unit as import quantity in 2). 6) ratio of recall to import quantity. 7) U.S. market share.

Table 3: Model Fit Test

	Toys	**Cotton sweaters**	**Sweaters, MMF**	**Batteries**	**Hair dryers**	**Lamps**	**Mean**
			Panel 1: Within sample estimation fit				
Correlation	0.8122	0.9191	0.9052	0.8127	0.8401	0.8867	0.8627
RMSE	2.2139	1.2567	4.2995	2.1962	2.5155	2.0693	2.4252
Obs	1415	659	3473	4873	3986	2110	
			Panel 2: Out-of-sample predictions fit				
Correlation	0.7793	0.7985	0.8072	0.7643	0.7724	0.8157	0.7896
RMSE	2.9447	1.8963	5.2749	2.9013	3.0528	2.5581	3.1047
Obs	794	275	899	2126	2082	1004	

Note: this table presents correlations and root-mean-square errors of the demean predictors. Panel 1 presents the within sample correlation between fitted and observed outcome variables. The sample size include only the "training" periods Panel 2 presents the out-of-sample correlation between observed outcomes and outcomes predicted using parameters estimated with training data. Number of observations for the out-of-sample predictions depends on how many trade partners the U.S. has in the test periods, so it varies substantially across products. In the third column, "Sweaters, MMF" stands for "Sweaters of man-made fabrics."

Table 4: Preference estimates across industries, non-durable goods

Products	**Coefficient**		**Elasticity**		**Obs**
	Reputation	**log(expenditure-price)**	**Reputation**	**Price**	
Toys	5.5688	8.547	2.396	-0.335	4376
	(2.255)	(0.663)			
Sweater, man-made fabric	3.5666	83.743	0.899	-2.55	6099
	(8.8703)	(11.436)			
Sweater, cotton	-3.8106	109.96	-3.7978	-2.783	7006
	(3.1886)	(25.244)			
Battery	-2.2197	15.4548	-2.199	-0.678	2232
	(1.7356)	(1.927)			
Lamps	-1.4709	13.3738	-1.456	-0.915	3039
	(1.3913)	(1.1216)			
Hair dryers	-1.1643	61.3095	-0.13	-1.3509	950
	(4.7026)	(235.1176)			

Notes: Bootstrap standard errors in parentheses. Bootstrap sample size is 1000.

Department of Economics, University of Alberta
Working Paper Series

2022-07: When Opportunity Knocks: China's Open Door Policy and Declining Educational Attainment – Jiang, X., Kennedy, K., **Zhong, J.**
2022-06: Asymmetric Impact of Real Effective Exchange Rate Changes on Domestic Output Revisited: Evidence from Egypt – **Sharaf, M.,** Shahen, A.
2022-05: Country-Based Investing with Exchange Rate and Reserve Currency – **Galvani, V.**
2022-04: The Mean-Variance Core of Cryptocurrencies: When More is Not Better – **Galvani, V.,** Faychuk, V.
2022-03: Outliers and Momentum in the Corporate Bond Market – **Galvani, V.,** Li, L.
2022-02: Higher Education Expansion and Crime: New Evidence from China – **Liu, X.,** Wang, C., Yan, Z., Zhao, Y.
2022-01: A Cautionary Tale of Fat Tails – **Dave, C.,** Dressler, S., Malik, S.
2021-14: Procurement Auctions for Regulated Retail Service Contracts in Restructured Electricity Markets – **Brown, D., Eckert, A.,** Olmstead, D.
2021-13: Socioeconomic and Demographic Disparities in Residential Battery Storage Adoption: Evidence from California – **Brown, D.**
2021-12: Market Structure, Risk Preferences, and Forward Contracting Incentives – **Brown, D.,** Sappington, D.
2021-11: Information and Communication Technologies and Medium-Run Fluctuations – **Brianti, M.,** Gati, L.
2021-10: Are Bank Bailouts Welfare Improving? **– Shukayev, M.,** Ueberfeldt, A.
2021-09: Information, Belief, and Health Behavior: Evidence from China **–** Lei, X., Song, G., **Su, X.**
2021-08: Stay at Home if You Can: COVID-19 Stay-at-Home Guidelines and Local Crime – Díaz, C., **Fossati, S.,** Trajtenberg, N.
2021-07: Financial Frictions, Equity Constraints, and Average Firm Size Across Countries – Bento, P., **Ranasinghe, A.**
2021-06: Should Canada's Capital Gains Taxes be Increased or Reformed? – **McMillan, M.**
2021-05: Financial Shocks, Uncertainty Shocks, and Monetary Policy Trade-Offs – **Brianti, M.**
2021-04: Valuing Elementary Schools: Evidence from Public School Acquisitions in Beijing – **Su, X.,** Yu, H.
2021-03: Colonel Blotto's Tug of War – **Klumpp, T.**
2021-02: Stockpiling and Shortages (the "Toilet Paper Paper") – **Klumpp, T.**
2021-01: When Social Assistance Meets Market Power: A Mixed Duopoly View of Health Insurance in the United States – **Ranasinghe, A., Su, X.**
2020-15: First to $15: Alberta's Minimum Wage Policy on Employment by Wages, Ages, and Places – **Fossati, S., Marchand, J.**
2020-14: Load-Following Forward Contracts – **Brown, D.,** Sappington, D.
2020-13: Loss of Life and Labour Productivity: The Canadian Opioid Crisis – **Cheung, A., Marchand, J.,** Mark, P.
2020-12: Equilibrium Indeterminacy and Extreme Outcomes: A Fat Sunspot Ta(i)l(e) - **Dave, C.,** Sorge, M.