NBER WORKING PAPER SERIES

SURVIVING THE U.S. IMPORT MARKET:
THE ROLE OF PRODUCTION DIFFERENTIATION

Tibor Besedes
Thomas J. Prusa

Working Paper 10319
http://www.nber.org/papers/w10319

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2004

ABSTRACT

We examine the extent that product differentiation affects the duration of US import trade relationships. Applying nonparametric and semiparametric techniques to highly disaggregated product-level data we estimate that the hazard rate is at least 18 percent higher for homogenous goods than for differentiated products. Put another way, the median survival time for trade relationships involving differentiated products is five years as compared to two years for homogenous products.

We find that our results are not only highly robust but often are strengthened under alternative specifications. For instance, if we define trade relationships using industry-level rather than product-level data we find that the hazard rate is 30-35 percent higher for homogenous goods than for differentiated products.

We also find that the survival ranking across product types holds across individual industries. We show that dropping the smallest trade relationships further accentuates the differences among product types. We also control for the possible measurement error in measuring spell lengths and the role of multiple spell relationships and find that in all cases the differences among products types are greater than in our benchmark analysis.

Tibor Besedes
Louisiana State University
Department of Economics
E.J. Ourso College of Business Administration
Baton Rouge, LA 70803
besedes@lsu.edu

Thomas J. Prusa
Department of Economics
New Jersey Hall
Rutgers University
New Brunswick, NJ 08901
and NBER
prusa@econ.rutgers.edu

1 Introduction

The influence of the pioneering work of Krugman (1980) and Helpman and Krugman (1985) has been so profound that nearly all discussions of prominent trade issues such as inter- and intra-industry trade, the home market effect, and imperfect competition involve the implicit assumption that trade in homogenous and differentiated goods is different. In fact, the impact of this work is so great that it is difficult to remember when the difference in homogenous and differentiated goods trade was not one of the self-evident truths of international economics.

Despite the pre-eminence of this view, or perhaps because of it, there has been little effort at identifying how and to what extent trade in homogenous and differentiated goods really is different.[1] An exception is the work of Rauch who has spent the better part of the past decade wrestling with characterizing and understanding how product type influences trade. Much of Rauch's work emphasizes the fact that most homogenous goods, but few differentiated goods, are traded on organized exchanges. This insight underlies Rauch's view that the network/search theory is important for understanding trade in differentiated products (Rauch, 1996, 2001; Casella and Rauch, 2003; Rauch and Trindade, 2002).

In this paper we ask a more basic question; namely, can we empirically document that trade in homogenous goods differs from trade in differentiated products? Although this question is straightforward it has largely been ignored in the literature. The work closest to ours is Rauch (1999) who uses a gravity equation framework to show that network/search measures are more important for trade in differentiated products than for homogenous products.

We take an entirely new approach to the issue. We examine whether being different has any impact on duration of trading relationships. In previous work we have shown that trade relationships are generally very short-lived (Besedeš and Prusa, 2003); in this paper we find that exactly how short depends significantly on the extent of product differentiation. We find that at each point in time the survival function for differentiated products is above the survival function for homogenous products. Moreover, the differences are economically large. Half of all relationships involving homogenous goods survive only one or two years; by contrast, the median survival time for differentiated products is five years.

Intuitively, one would expect that trade relationships involving homogenous products might be quite fragile. For products such as corn, wheat, and oil one can imagine a "world market" where all foreign suppliers ship and all buyers purchase their products. In this case, trade relationships might be very short; relationship-specific factors may not matter and source country may be irrelevant. By contrast, the seminal work of Dixit and Stiglitz (1977)

---

[1]There are a number of papers that indirectly discuss the importance of product type but they are mainly interested in other issues (Helpman, 1987; Hummels and Levinsohn, 1995; Evenett and Keller, 2002; Debaere, 2003). More recently, Dalgin, Mitra and Trindade (2003) examine if product type helps explain inequality and trade.

and Helpman and Krugman (1985) suggests that each variety of differentiated products should be desired by consumers; hence trade relationships for differentiated goods might be very long-lived (i.e., source country matters). More generally, differentiated products are likely to involve relationship-specific investments and as a result one might expect far longer lived relationships.

Using Rauch's (1999) classification scheme we are able to characterize how product type affects duration and our empirical findings provide strong support for the view that trade in differentiated and homogenous goods does differ.[2] Specifically, Rauch's scheme allows us to classify products into one of three types—differentiated, reference priced, and homogenous. We find that differentiated products tend to have the longest survival, followed by reference priced products and then homogenous products. Simply put, the ordering of product types that we find is completely consistent with the implicit assumption made in many standard theoretical models of trade.

Moreover, we show that these results are highly robust to a large number of alterations to our benchmark data. For instance, we show that the findings are not driven by the highly disaggregated nature of our data. We find that as we aggregate from product-level data to industry-level data the estimated differences in survival among product types increase. When using product-level trade data we find that as compared with differentiated products the hazard rate for reference priced goods is 16 percent higher and for homogenous goods 18 percent higher. In comparison, when using industry-level trade data we find that as compared with differentiated products the hazard rate for reference priced goods is 24 percent higher and for homogenous goods 32 percent higher. The industry-level results are a striking confirmation of our product-level results and are remarkably reassuring given the potential for misclassification that exists in a product-level trade dataset that is constructed from literally millions of U.S. Customs declaration forms.

Another potential concern is that our results are driven by differences in the value of trade across product types. Specifically, it seems reasonable to expect that trade relationships with large values of trade to be longer lived, perhaps reflecting the beach-head effect discussed in Baldwin (1988). For instance, all else equal, one might expect relationships with $1 million of trade will survive longer than relationships with $100,000 of trade, which in turn will survive longer than relationships with $1,000 of trade. As it turns out, relationships with large trade values do indeed have longer duration. This raises the intriguing possibility that our estimated product type differences are simply reflecting size differences rather than really telling us something about the impact of product type. It turns out, however, this fear is unfounded. In fact, rather than weakening our results, we find that controlling for size increases the role of product type. That is, we find significantly *greater* differences among product types when we drop the smallest trade relationships. For in-

---

[2]We are agnostic about what precise theory(ies) explains the difference. In the appendix we briefly sketch how both "address" and "nonaddress" models can underlie the difference in duration. We also found Rauch's network/search models quite compelling.

stance, if we limit our analysis to those relationships whose trade value in the first year is greater than $100,000 ($1,000,000) we find that the hazard rate for reference priced goods is 50 percent (150 percent) higher and for homogenous goods 62 percent (180 percent) higher than for differentiated goods. In other words, size does matter and it serves to intensify, not diminish, the importance of product type.

We also examine the extent that our results are influenced by possible incorrect inferences about the end of trade relationships. Our concern is that some relationships experience multiple spells of service and if the time between these spells is short, it is possible that the gap of non-service is mis-measured and hence that the duration is mis-measured. If this is the case, for instance, two short spells (separated by a gap) might really be one longer spell. We consider several alternative methods of controlling for this potential measurement error and find little impact of these adjustments: large and significant differences among product types remain a key feature of the data.

The remainder of the paper is organized as follows. Section 2 discusses the data set used in our study. Section 3 describes the econometric models used in our empirical analysis and section 4 presents the basic empirical results. In section 5 we verify the robustness of our findings. Section 6 concludes.

2   Data

Our analysis is based on US import statistics as compiled by Feenstra (1996).[3] From 1972 through 1988 import products were classified according to the seven digit Tariff Schedule of the United States (TS).[4] Since 1989 imports have been classified according to the ten digit Harmonized System (HS). Given that all products were recoded in 1989, we limit our analysis to the period 1972-1988 in order to avoid potential concordance issues.[5] In simplest terms, for each commodity and year we can identify all of the countries from whom the United States purchased. Over the entire period there is a total of about 22,000 different products in the TS data. On average, in each year we observe import trade for about 10,000 products sourced from about 160 countries.

We are interested in studying the length of time until a country no longer exports a product to the US, an event that we will refer to as a "failure." As a result calendar time is not as important as analysis time, which measures the length or duration that a country exports a product to the US. Hence, for each product and country we use the annual data to create spell data. By this we mean that if we observe trade in product i from country c from

---

[3]Details on the sources of our data are included in the appendix.

[4]We only study imports because there is no concordance between the disaggregated TS import and export codes. In addition, the export data is self-reported making it more likely that exports are misreported.

[5]When the TS codes were replaced with the HS codes multiple TS products were mapped into single HS codes and vice versa; this makes it impossible to discern actual exits during 1988–89. See Schott (2001) for more discussion of the TS-HS concordance problem.

1976–1980, we would say that the ci^{th} *trade relationship* has a spell length of five. Thinking of imports in terms of spells our dataset has 693,963 observations. The observations have a median (mean) spell length of 4 (2.7) years.

Our benchmark analysis is based on the most disaggregated data available, the 7-digit TS data, which means our analysis is at the product-level, not the industry-level. Inferences are based on trade in tangible products rather than aggregate summaries. Until relatively recently such disaggregated data were not widely available and empirical studies were based on industry classifications, such as the Standard International Trade Classification (SITC). We chose to base our benchmark results on the 7-digit TS data primarily because it does the best job of measuring trade in products rather than industries. For instance, the TS data reports information on more than 30 different types of ball bearings, differentiating by specialized application (e.g., automobiles), size, and chemistry. By contrast, if we were to perform our analysis using SITC classifications (as we do later in our robustness checks) we would aggregate all of these ball bearing codes along with other similar, or perhaps not so similar, products to create spells of service. In fact, the dozens of ball bearing codes map into a single SITC category "all, roller or needle roller bearings" (4 digit code=7491) combining ball bearings with roller and needle roller bearings. Further aggregation is possible, to "Non-electric machinery parts" (3 digit code=749) or "Industrial machinery" (2 digit code=74), but doing so makes it increasingly difficult to interpret the results since each classification includes highly disparate products.

For duration analysis we believe that using highly disaggregated data is crucial. First of all, we need the data to be sufficiently disaggregated to allow us to identify the extent of product differentiation. The more aggregated is the data, the more we will be identifying industries rather than products and many different types of products can be sold by a single industry.[6] Hence, industry-level analysis may make it more difficult to identify the role of product types.

In addition, if the products are too broadly defined, we cannot expect to see any source countries exit the data. If we aggregate all imports from each country, at the extreme, we will never observe any exit since the United States purchases some product from each source country every year.

Further, once we began to think of trade data in terms of spells it became apparent that we needed to account for censoring in our analysis. By this we mean that we often do not know whether a trade relationship ended because of failure or for some other reason. In practice this means that we do not know for certain the beginning or ending date for some trade relationships.[7] As it turns out, censoring is pretty common in US import data—about

---

[6]For instance, at the 1-digit SITC level, only two of the ten industries are composed of a single product type.

[7]There are a number of important issues (including the aggregation issue) involved in applying duration analysis to import data. Most of the issues are not central to the primary focus in this paper (the impact of product differentiation). Besedeš and Prusa (2003) contains an extensive discussion of the import data and

4

half of the trade observations are censored and about 20 percent of spells are censored at one year. The censoring problem comes in two flavors. First, we have no information on trade relationships for the years before the beginning (pre-1972) and after the end of our sample (post-1988). For example, we observe that the US imported corn (TS=1312000) from the Philippines in 1972 and that this relationship was observed for exactly one year. This trade relationship may have begun in 1972 or it may have begun in some prior year. The most appropriate interpretation is that this relationship had a duration of at least one year. Similarly, we also observe that the US imported corn from Peru from 1984 to 1988. Unfortunately, our data does not continue beyond 1988 and hence we cannot be sure how long that spell ultimately lasted. Once again, the most appropriate interpretation is that this relationship had a duration of at least five years. About 10 percent of spells are observed in 1972, while about 22 percent are observed in 1988.[8] This first type of censoring is typical in survival studies and is accounted for in all of our subsequent analysis.

The second type of censoring is somewhat unique and stems from the fact that our analysis is performed at the product-level. US Customs revises product definitions for the tariff codes on an ongoing basis, sometimes splitting a single code into multiple codes and other times combining multiple codes into fewer codes. Unfortunately, information does not exist to allow us to map the old TS codes into the new ones. When a code is changed we no longer observe trade in the old product code. But, is this due to the end of a relationship or does it simply mean that trade stopped due to the redefinition? We choose to be cautious and classify all such "exits" as censored, which means that we interpret reclassified relationships as having duration of at least x years (where x is the number of years where trade in the original code was observed). An analogous problem exists for the new codes. When a code is changed we begin to observe trade in the new product code(s). But, is this really the beginning of a new relationship or does it simply reflect the redefinition? Once again for much of our analysis we choose to be cautious and assume the relationship had a duration of at least y years (where y is the number of years where trade in the new code was observed). About 20 percent of the spells are censored due to reclassification.

This second type of censoring is a peculiar characteristic of our benchmark data. It can be argued that our decision to interpret all product code changes as being censored is overly conservative—one could alternatively classify these product code changes as indicating entry and exit. We believe this alternative classification scheme will undoubtedly understate the true duration and this view is confirmed in Besedeš and Prusa (2003). Thus, we do not pursue this alternative approach and instead present results based on defining relationships using the 5-digit SITC industry-level classification. While the size of our database falls dramatically, we are still left with over 157,000 observations of spells of service. Since

---

its applicability to duration analysis.

[8]Less than two percent of all spells span the entire 1972 through 1988 period.

the SITC classification system is unchanged throughout our sample, industry-level analysis is not plagued by the second type of censoring and, hence, is not subject to errors due to reclassification. As a result only the first type of censoring (which is driven by the beginning and the end of our sample) is present in the SITC analysis. In addition, the SITC level analysis serves to verify whether our results are robust to aggregation since there may be some concern that the 7-digit data is too disaggregated.[9]

Our next major task involved characterizing the extent of product differentiation for each product. We follow Rauch (1999) and classify commodities into three categories: homogenous, reference priced, and differentiated. Rauch classified products that are traded on an organized exchange as homogenous goods. Products not sold on exchanges but whose benchmark prices exist were classified as reference priced; all other products were deemed differentiated products.

Although coarser than one would like, Rauch's classification scheme has several virtues. First and foremost, it is the only classification scheme that we know of that exists at a highly disaggregated level. Rauch classifies products at the 4-digit SITC level and we mapped his codes to the 7-digit TS codes using the concordance found in Feenstra (1996). In a first best world, we would know each product's elasticity of substitution. Unfortunately, such elasticities are generally only available for 2-digit industries and perhaps a handful of 3-digit industries. Given the requirement that we use highly disaggregated data in our analysis, Rauch's scheme is preferred. Moreover, seven of the ten SITC industries are represented by products from each of the three product types, which confirms that Rauch's scheme is fairly rich and is not simply a re-mapping of industry codes. Second, Rauch's scheme makes intuitive sense since it broadly captures what economists mean by product substitutability. Products that are sold on organized exchanges (like corn, oil, wheat, etc.) are exactly those products typically cited as being homogenous. Consumers may neither know nor care about the source of the product that they are purchasing. On the other hand, products like many types of steel and chemicals whose prices are listed in industry guides and trade journals will likely have some unique attributes (e.g., quality may vary by source country) but are essentially substitutable. In this case, consumers will know the source country, but this may only have a small impact on their purchasing decision. In the final category are differentiated products. These are products that not only have many characteristics that vary across suppliers but may even be specifically tailored to the end-user's needs. Automobiles are perhaps the most often cited example of such a good; in fact, most consumer goods (e.g., toys, apparel, cookware) are classified as differentiated. Third, Rauch's classification scheme is quite comprehensive, covering about 98 percent of all US import relationships. There is no selection bias as might be the case if we limited our analysis to only those products where substitution elasticities exist.

---

[9]It can be argued that parsing the data too finely leads to excessive exits as modest changes in a product's specifications or minor mistakes made by Custom's officials can result in the product being assigned to a different TS code. SITC level analysis will be far less prone to such errors.

3 Empirical Approach

Since we are interested in investigating how likely a country is to cease exporting a product to the United States, it is only natural that we approach this issue by using duration analysis methods. We first estimate survival functions across product types by using the nonparametric Kaplan-Meier product limit estimator, and then proceed to model the hazard of US import trade with the Cox proportional hazard model. The goal of this exercise is to investigate whether differences across product types are robust once we include factors that we think affect the duration of trade.

3.1 Nonparametric Estimation

The survivor function $S(t)$ is usually estimated nonparametrically using the Kaplan-Meier product limit estimator. Derivation is as follows. Assume a sample contains n independent observations denoted $(t_i; c_i)$, $i = 1, 2, \ldots, n$, where t_i is the survival time, while c_i is the censoring indicator variable C (taking on a value of 1 if failure occurred, 0 otherwise) of observation i. Assume there are $m < n$ recorded times of failure. Denote the rank-ordered survival times as $t(1) < t(2) < \cdots < t(m)$. Let n_j denote the number of subjects at risk of failing at $t(j)$ and let d_j denote the number of observed failures. The Kaplan-Meier estimator of the survival function is then

$$\hat{S}(t) = \prod_{t(i) < t} \frac{n_j - d_j}{n_j},$$

with the convention that $\hat{S}(t) = 1$ if $t < t(1)$. Given that many of our observations are censored, we note that the Kaplan-Meier estimator is robust to censoring and uses information from both censored and non-censored observations.

In the discrete case, the hazard function can be estimated by

$$\widehat{h}(t_j) = \frac{d_j}{n_j}.$$

3.2 Semiparametric Estimation

The Cox proportional hazard regression model asserts that the hazard rate is

$$h(t, x, \beta) = h_0(t) \exp(x\beta),$$

where x denotes a vector of explanatory variables and the βs are to be estimated. The baseline hazard, $h_0(t)$ characterizes how the hazard function changes as a function of time. The baseline hazard captures individual heterogeneity that is not explained by the covariates.

In the Cox proportional hazard model one subject's hazard is a multiplicative replica of another's; for instance, the ratio of subject j to subject m's hazard is

$$HR(t, x_m, x_j) = h(t, x_j, \beta)/h(t, x_m, \beta) = \frac{h_0(t)\exp(x_j\beta)}{h_0(t)\exp(x_m\beta)} = \frac{\exp(x_j\beta)}{\exp(x_m\beta)}.$$

One particular application of this property is to compare one subject with covariates $\{x_1, x_2, \ldots, x_k\}$ with another with the same covariates, except that x_2 is incremented by 1. Taking the ratio of the hazards of these two subjects, we have

$$h(t, x_1, x_2 + 1, \ldots, x_k, \beta)/h(t, x_1, x_2, \ldots, x_k, \beta) = \exp(\beta_2).$$

Note that this ratio is not a function of t. Hence, exponentiated individual coefficients have the interpretation of the ratio of the hazards for a one-unit change in the corresponding covariate. A particular advantage of the Cox model is that the baseline hazard is left unspecified and is not estimated. This allows us to account for unobserved heterogeneity in a more tractable approach than the Kaplan-Meier estimator and a less restrictive approach than a fully parametric model.

4 Empirical Findings — Benchmark 7-digit TS Data

4.1 Nonparametric Results

We begin by performing some nonparametric estimates using the benchmark 7-digit TS data. The Kaplan-Meier survival function, $\hat{S}(t)$, is graphed in the upper-left panel of Figure 1 for each of the three product types. Survival functions across product types are similar in that each is downward sloping with a decreasing slope. This indicates that regardless of the product type trade relationships face a large probability of failure in their first few years. As discussed in Besedeš and Prusa (2003), the risk declines quite markedly once trade relationships last for 4-5 years. In fact, we find the conditional probability of failing in the first year to be remarkably high, 45 percent for homogenous goods, 41 percent for reference priced goods, and 31 percent for differentiated goods. However, by year 5, the conditional probability of failing for each type has fallen to less than seven percent; after 10 years, the hazard rate for each type has fallen to less than two percent.

In terms of the main research issue in this paper the differences in the survival functions across product types are notable. At each point the survival function for differentiated products is above the survival function for reference priced products, which in turn is above the survival function for homogenous products. Moreover, given the large number of observations in our sample, the differences between the Kaplan-Meier survivor functions are statistically significant.[10]

---

[10]The standard errors are in the range 0.001 to 0.02.

Our findings indicate that differentiated products have the longest survival, followed by reference priced products, and then homogenous products. Differentiated products have a median survival time of five years; reference priced goods and homogenous products have a median survival time of only two years. Put another way, this means that after controlling for censoring *half* of the trade relationships involving homogenous goods fail during the first two years. This is an extraordinarily short life expectancy. Not surprisingly, the chance of observing a spell of more than 10 years is far greater for differentiated products (45%) than for either reference priced goods (32%) or for homogenous products (25%). Given the extraordinary failure rates during the first few years it is impressive that we nevertheless find that such a high percentage of goods survive for more than 10 years. The explanation is not that there are few exits; rather, most exits that are observed in later years are classified as censored as opposed to failures, resulting in low hazard rates in those periods.

4.2 Semiparametric Results

Table 1 contains the Cox proportional hazard estimates using our benchmark data. Our basic estimation model includes the standard regressors that appear in the gravity equation literature (e.g., GDP, language, continuity, distance). While we do not offer a theoretical model of how these variables affect survival, we believe it is reasonable to include the gravity equation variables in our empirical model. The gravity equation states that the bilateral trade between two countries is directly proportional to the product of the countries' GDPs and therefore larger countries will tend to trade more with each other. Given the gravity equation's remarkable job of predicting bilateral trade, we view the gravity-motivated regressors as an exogenous control for the propensity for source countries to supply the US market.

In addition, we include several variables to capture relative cost and competitiveness issues. We use the industry-level tariff rate to control for the ease with which foreign firms can enter the market. Whether higher tariffs increase or decrease the hazard depends largely on whether time series or cross-section variation dominates. For a given product, an increase in the tariff should lead to some foreign firms exiting because higher tariffs raise the costs of servicing the US market. It follows that time series variation in tariffs should lead us to find that higher tariffs raise the hazard. On the other hand, looking across industries higher tariffs mean less competition for those firms currently in the market. As a result both domestic and foreign firms servicing the US market face less risk and hence a lower hazard. It follows that cross section variation in tariffs should lead us to find that higher tariffs lower the hazard. Given that there is relatively little time series variation in tariffs, we expect the cross section effect to dominate which means that higher tariffs should lower the hazard.

In order to capture the impact of cost changes on the hazard we included the change in the *relative* real exchange rate. To construct this measure of relative costs, we began by

9

defining each country's exchange rate so that an increase corresponds to a real depreciation (i.e., foreign currency per US dollar). Then in each year we normalized the exchange rate by the average percentage change of all supplying countries' exchange rates. This gives us a measure of how each country's exchange rate changed relative to its competitors. An increase in this measure reflects that a country's currency has weakened relatively more than its competitors. If one country's currency depreciates relative to other countries' currencies, its firms should become more competitive vis-à-vis other foreign and domestic firms and hence less likely to exit.

Schott (2001) argues that even with data as disaggregated as the 7-digit TS import data, some products are more broadly defined than one would like. In order to control for diversity of the products we calculate the coefficient of variation of unit values for each TS product in each year. We expect that the smaller the coefficient of variation the more homogenous is the product and hence the greater the hazard.

Schott's work also persuaded us to include the relative unit value in our regression. This index is calculated by normalizing each supplier's unit value by the average unit value for that product. Thus, a value of one means the supplier's unit value is the mean price; values greater (smaller) than one mean the supplier charges a higher (lower) price.

We were also concerned that a supplier's market presence may influence the duration. Specifically, we were concerned that for any given product the bigger the supplier the less likely the exit from the market. Simply controlling for size, however, is insufficient as the meaning of "big" varies dramatically across product codes. Therefore, in order to control for product heterogeneity we calculated market share for each country supplying a given product.

Finally, there is the issue of multiple spells—trade relationships that have multiple periods of service separated by a period with no service. Specifically, some trade relationships are observed for a period of consecutive years (spell 1), then are followed by a period of no trade, and then again observed for another service spell (spell 2).[11] We believe that the first failure makes a second failure more likely (higher hazard). On the other hand, it is also possible that the return of the foreign supplier to the market is a positive sign making a second failure less likely. In either case the hazard rate will depend on whether we are observing a second spell and should be controlled for in our estimation. In our base specifications we treat multiple spells as independent. That is, we include spells after the first and use a dummy to control for any impact of higher order spells. We will consider alternative methods for addressing multiple spells in the following section.

In the first column of Table 1 we report our benchmark estimates based on the entire sample. Throughout we present results in terms of hazard ratios. Hence, an estimated hazard rate coefficient less than (greater than) (equal to) 1 is interpreted as implying that

---

[11]About one-quarter of trade relationships experience multiple spells of service and about two-thirds of those experience just two spells. Less than one percent of trade relationships have more than three spells.

the variable lowers (raises) (has no impact on) the risk of failure. All reported estimates include region dummies.[12]

In many respects the estimates are consistent with those in the gravity literature. The estimated effects for the standard gravity variables, however, are generally small. In particular, our estimates imply that:

- Distance increases the hazard rate, but by a fairly small magnitude, and in some specifications is statistically insignificant. With few exceptions (e.g., Australia) the differences in distance between the US and its trading partners would only lead to about a 1 percent difference in hazard rates.

- Common language lowers the hazard, but again the effect is fairly small, about 5 percent. On the other hand, countries that are contiguous with the US (Canada and Mexico) face significantly lower hazard rates—on the order of 30 percent.

- Larger countries (as measured by GDP) face a lower hazard. Given the variance in country GDP, the size of the effect depends significantly on country size. Thinking either in terms of differences across time or countries the estimates imply that a $100 billion increase in GDP lowers the hazard rate by about 2 percent; a $1 trillion increase in GDP lowers the hazard rate by about 20 percent.

- Industries with higher tariffs face a lower hazard. A 1 percentage point increase in an industry's tariff lowers the hazard by approximately 2 percent.

We also find that changes in the relative real exchange rate have a large impact on the hazard rate. We find that a 10 percent depreciation in the real exchange rate (relative to other suppliers) lowers the hazard by about 10 percent. As a country's currency becomes cheaper, its products become more competitive.

We also find that products with higher variation in unit values face a significantly lower hazard rate. At first blush this might seem surprising since it suggests that high and low prices for the same product increase duration. However, the result confirms Schott's (2001) contention that one must be cautious in interpreting even extremely disaggregated TS product codes as identifying an identical product. Rather, it appears that a single code code can capture differentiated products. For instance, the "cotton T-shirt" product might include commodity grade products from China and Bangladesh and also fashion designer products from Italy. Our results suggest that high variation in unit values indicates that the product code contains products with differentiated characteristics.

---

[12]We estimated with and without region dummies and found with few exceptions that the results do not vary significantly across specifications. We also estimated with country dummies which greatly increased computation time without significantly affecting the results.

The next two regressors, relative unit values and product market share, have a small impact on duration. A product priced 20 percent higher than the mean will only have about a one percent higher hazard. Albeit a small effect, the result is consistent with Schott's conjecture discussed above. A supplier's market presence as captured by product market share lowers the hazard, but only by an economically insignificant amount.

We are primarily interested in the product type estimates. Letting differentiated products be the benchmark, we find that reference priced goods have a 16 percent higher hazard and homogenous goods an 18 percent higher hazard. Thus, the estimates strongly support what Figure 1 suggested: namely, product type matters.

In Table 2 we report the results when we consider a more flexible specification where we allow the coefficients on all variables to vary by product type. By in large, the estimates confirm what we learned from the basic specification. Namely, Canada/Mexico, larger countries, high tariff products, and weaker currencies, all have lower hazard rates. We also find that higher variation in unit values lowers the hazard for differentiated products, but increases it for homogeneous goods. This is what standard economic theory would predict. In particular, in a homogenous product market one would be surprised if high priced suppliers would be likely to have long lived spells of service; on the other hand, in a differentiated product market, high priced suppliers could easily have a long service spell if their high prices reflect quality differences.

Most importantly, even after allowing for systematic differences across product types, we continue to find that homogenous products face a significantly higher hazard than differentiated products, on the order of 10 percent. Reference priced goods, however, are not estimated to be significantly different than differentiated goods.

5   Robustness

We next perform several exercises to investigate the robustness of our findings. There are five concerns we explore. First, are the same patterns found in industry-level data? Second, are our results driven by differences in the value of trade across product types? Third, are our results affected by a potential measurement error regarding the end of the spell? Fourth, does the fact that we do not observe trade relationships prior to 1972 influence the results? Fifth, are the differences across product types driven by the distribution of product types across industries?

5.1   Industry-level analysis

As discussed above we can also define trade relationships using industry-level data. In addition to mitigating the censoring problem, industry-level analysis allows us to explore whether the differences across product types are due to the highly disaggregated nature of our data. Our concern is that the TS classification system is too fine—perhaps 30 different

12

ball-bearing codes represents an overly fine parsing of the data and leads us to observe excessive entry and exit. If so, trade relationships might be better measured using the SITC industry classification.

To investigate this concern we calculated spells of service using the 5- and 4-digit SITC industry data. We plot the Kaplan-Meier survival function for each of the three product types in Figure 2 for the 5-digit SITC data (upper-left).[13] As with our benchmark data at each time the survival function for differentiated products is above the survival function for reference priced products, which in turn is above the survival function for homogenous products. The graphs suggest that the differences among the product types are somewhat attenuated as compared to the benchmark data. For instance, the median survival time for all three product types is just two years. Nevertheless, the differences persist and are statistically significant.

In Table 3 we report the Cox proportional hazard estimates. We note that not all of the variables are available when we analyze the data at the SITC level. For instance, because units vary by product industry-level unit values cannot be calculated. Interestingly, the results indicate that once we control for other factors, the differences among product types are greater than in our benchmark data. Looking at the 5-digit SITC results (column two) we see that the ordering of the products is the same as in our benchmark data, but the magnitudes are larger. Specifically, we find that as compared with differentiated products the hazard rate for reference priced goods is 24 percent higher and for homogenous goods 32 percent higher when we use the 5-digit SITC data.

Results for the 4-digit SITC data are similar (column three of Table 3). In particular, the Kaplan-Meier ranking of survival functions is exactly the same as for the 7-digit TS and the 5-digit SITC data. As was the case with the 5-digit data the differences among product types are greater than in our benchmark data.

In Table 4 we report the results when we consider the more flexible specification where the coefficients on all variables vary by product type. The results are a bit stronger than what we saw in our benchmark runs. In particular, homogenous products face a significantly higher hazard than differentiated products. Reference priced goods have a higher hazard than differentiated goods but lower than homogenous goods (i.e., the same ordering as we saw in our benchmark runs).

All in all, we view the industry-level analysis as confirming our main findings: (i) differentiated goods have a significantly lower hazard rate than reference price goods which in turn have a significantly lower hazard than homogenous products; and (ii) the difference in hazard rates between differentiated and reference priced goods is smaller than between differentiated and homogenous goods.

---

[13]The Kaplan-Meier plots are similar for the 4-digit data.

5.2 Does size matter?

The inclusion of market share in our specification controls for the size within each given product market. Our concern here is that our results are driven by differences in the value of trade across product markets. It seems reasonable to expect that spells with large values of trade will be longer lived. For instance, all else equal, one might expect a relationship with $1 million of trade in the first year of the spell to survive longer than an observation with $100,000 of trade in the initial year.[14] If so, the findings reported above might reflect that (i) small valued spells are at greatest risk and (ii) homogenous goods tend to involve small value spells. In other words, the estimates may merely be capturing differences in traded value rather than telling us something about the importance of product type.

Let us look at the support for each hypothesis separately. First, are small valued spells at greatest risk? To address this we graph the distribution of spell lengths in Figure 3 using the benchmark 7-digit TS data. On the x-axis we graph the duration of the spell. We begin by plotting data on all trade observations; this benchmark distribution reflects all trade spells and should be interpreted as an equally weighted distribution as each observation is treated symmetrically (plotted with a solid line). As shown, more than 50 percent of the observations are observed for only one year. More than 70 percent of all spells are observed for less than three years. Overall, fewer than five percent of all trade spells last more than ten years.

One method to get a sense of whether value matters is to weigh each observation by the value of trade in the first year of the spell. By doing so an observation with $1 million of trade has 10 times the impact of an observation with $100,000 of trade. This weighted distribution is plotted with a dashed line in Figure 3. As one can see this measure indicates that larger dollar relationships are indeed longer lived. Weighed by value, 12.4 percent of spells are observed for only one year; 29.9 percent are observed for less than three years—significantly less than the 70 percent attrition that we found in the unweighted sample. Looking at those spells that last more than ten years, long lived spells are about five times more likely when we weigh spells by value than when we equally weigh the observations.

It strikes us, however, that weighing the observations by value overstates the role played by the largest trade observations. Given that we want to think about duration of trade relationships we believe that it is an overstatement to say that a $1 billion crude oil purchase is 1000 times more important than a $1 million auto parts purchase. We believe that weighing by trade value focuses too much on the biggest observations.

We feel a better way to get at the issue is to equally weigh the observations but filter out small dollar-value observations. In other words, eliminate spells whose trade in the first year is below some minimum level. This allows us to identify the role played by small value observations without overly focusing on the biggest observations. In Figure 3 we plot the distributions using this filtering scheme. We depict two alternative cut-off levels for year

---

[14]We use the CPI to convert the nominal trade values into real 1987 dollars.

one trade, $100,000 and $1 million, but our results are not sensitive to the precise cut-off level used. As seen, imposing higher cut-offs results in higher distributions. This provides further evidence that the small valued trade relationships are at greatest risk.

Let us now look at the support for the second hypothesis. Do homogenous goods tend to involve small value spells? If so, it is possible that the differences in the survival functions depicted in Figure 1 only reflect size. If size were all that mattered, we should find that differentiated goods tend to involve larger trade values. Yet, the opposite is the case: differentiated goods tend to involve the smallest trade values. On average in the first year of the spell homogenous products involve larger value transactions ($4.5 million) than reference priced ($730,000) or differentiated goods ($700,000).[15]

This discussion suggests that rather than weakening our results controlling for size may strengthen our results. That is, if size matters (and it appears that it does) the observed trade values suggest that product type is even more important than our benchmark results suggest. Given this background discussion we plot the Kaplan-Meier survival function for each of the three product types after dropping the small-valued observations (Figures 1 and 2 for the 7-digit TS and the 5-digit SITC data, respectively). In the upper-right (lower-left) panel the results are based on dropping all observations where the value of first year of the spell was less that $100,000 ($1,000,000). Two important insights are gained. First, as Figure 3 suggested, the survivor functions shift up as we progressively drop observations. This means that small spells tend to be shorter-lived. Second—and most relevant for our discussion—the estimates provide *no* evidence that differences among the product types are driven by small observations. In fact, the differences among product types grow as we progressively eliminate the smaller trade observations.

In order to get a numerical assessment of the impact, we re-estimate the Cox proportional hazard model using the two cutoffs for the value of trade in the initial year of the spell. The estimates are reported in columns two and three of Table 1. Comparing these estimates with the benchmark (column one), we find that the impact of common language, contiguity and tariff rates all increase. The impact of distance increases at the $100,000 cut-off level, but at the $1,000,000 cut-off level the estimate is statistically insignificant. The impact of GDP and variation of unit values increases at both cut-offs, but the impact at the $100,000 cut-off is greater than that at the $1,000,000 level. Interestingly, the impact of the change in exchange rate declines as we restrict ourselves to the bigger spells. By contrast, the impact of relative unit values and market presence are largely unchanged as we filter out the smaller observations.

Most importantly, product type dummies indicate that the results are not driven by small value spells. Compared to differentiated goods, homogeneous goods face a 62 percent higher hazard rate at the $100,000 cut-off level, and a 184 percent higher hazard rate at the

---

[15]If we compare the average value over the entire spell we find the same ordering. In addition, we also compare trade values for groupings of countries such as OECD, non-OECD. No matter the grouping we always find that differentiated products involve far smaller trade values than homogenous goods.

15

$1,000,000 cut-off level. Reference priced products face 50-150 percent higher hazard rates as we restrict our sample to larger trade value observations.

When we look at the more flexible specification (Table 2) we continue to see that restricting our analysis to high value spells accentuates the differences among the types. For instance, if we include all observations, differentiated and reference priced goods are found to have similar hazard rates. If, however, we restrict ourselves to relationships with $1,000,000 of trade (in the first year), reference price goods have an 88% higher hazard. Homogenous goods have a 130% higher hazard.

The bottom-line is that size does matter. But, perhaps surprisingly, we find that in terms of duration size magnifies the impact of product type.

5.3   Measurement error

We now examine how our results are affected if we incorrectly infer that a trade relationship ended. In particular, our concern involves trade relationships with multiple spells. If the time between spells is short (what we will refer to as the gap), it is possible that the gap is mis-measured and that interpreting the initial spell as "failing" is inappropriate. In this case, the two spells might be better thought of as one longer spell. And, if this measurement error is related to product type in that homogenous goods are more likely to have short gaps, then our results will misrepresent the role of product type.

To check the sensitivity of our results, we considered three alternative approaches toward the issue of multiple spells. First, we assume that a one-year gap between spells is an error and merge the individual spells and adjust the duration length accordingly. Gaps of two or more years are assumed to be accurate and no merging is done. As an example, suppose the US imports a product from country c in 1973-74, 1976-77, and 1979-80. Without any adjustment this trade pattern is interpreted as three distinct spells, each of length two years. If we assume that all gaps of one year are errors, we would instead observe just one distinct spell with a length of eight years.

The survival functions for the 7-digit TS and the 5-digit SITC data using the gap-adjusted data are shown in the lower right panels of Figures 1 and 2, respectively. The probability of survival for each product type increases at each point in time, yet the differences across product types remain. The same can be inferred from the estimation results reported in Table 1. We find that the impact of product type increases slightly as compared with our benchmark runs. Reference priced goods have an 18 percent higher hazard rate while homogeneous goods have a 19 percent higher hazard rate than differentiated goods. Interestingly, in the more flexible specification (Table 2) we find evidence that product type may not matter. In particular, the point estimate indicates a higher hazard for homogenous goods, but the estimate is insignificant.[16]

---

[16]The point estimate for reference priced goods is lower. Given our sample size, it is reasonable to interpret this as not statistically significant.

Our second approach toward this potential measurement error issue involved limiting the analysis to only first spells of relationships. This alternative restricts the sample to those relationships with just a single spell and also to the first spell of those relationships with multiple service spells. Here we are implicitly saying that our use of a dummy to control for the impact of multiple spells may be insufficient. As seen in column five of Table 1, the results are qualitatively similar to the gap-adjusted approach. That is, we find large differences between product types with homogenous goods facing the largest hazard followed by reference priced goods.

The third approach is more restrictive than the second. Here we addressed the measurement error issue by limiting the analysis to just single spell relationships. Thus, we drop not only the second or higher spell, but all spells of service for trade relationships with multiple spells. As seen in the sixth column of Table 1 the results are very similar to the first spell only specification.

5.4   Inability to observe starting dates prior to 1972

The next possible concern involves the fact that we do not observe the true starting date for those spells that were active in 1972. As mentioned, this limitation afflicts about 10 percent of observed spells. Consider a relationship that is observed in 1972 and is observed for x years. We believe that the most appropriate interpretation is that this relationship had a duration of at least x years, which in practice means that all such spells are treated as censored. Our concern is that our interpretation may bias the estimation since at least some of the spells were already long-lived in 1972.

The most straightforward approach to get a sense of the potential bias from this censoring issue is to limit our analysis to only those spells that begin after 1972. From 1973 onwards, we observe when a spell starts. The Cox proportional hazard estimates are reported in the last column of Table 1. As we saw with the other robustness tests, the product type estimates are not reduced and in fact are somewhat larger than in our benchmark estimates: reference priced (homogeneous) goods have an 18 percent (25 percent) higher hazard rate than differentiated goods.

5.5   Industry effects

The final concern we would like to investigate is whether the distribution of product types across industries drives the results. A vast majority of products in our data set are classified as differentiated, while fairly few are classified as homogeneous. Furthermore, two industries, Machinery (SITC=7) and Miscellaneous Manufactures (SITC=8), are composed entirely of differentiated goods.[17] Could the driving force between our results be the dis-

---

[17]A third industry, SITC=9, is composed of differentiated and homogenous goods only and is not presented.

tribution of product types across industries? That is, are the preceding results found when looking at individual industries or is it driven by variation across industry?

To investigate we re-estimated our basic specification for each 1-digit industry separately.[18] We chose this approach rather than simply including industry dummies because the majority of products are in two industries (SITC=7 and 8) and these two industries happen to contain only differentiated products. This hinders our inability to separately identify industry and product type effects.

The results are reported in Table 5. In the interest of brevity, we only report the estimates for the product type dummies. In the top panel we report estimates for the benchmark 7-digit data. In the lower two panels we report estimates when relationships are defined using the SITC industry data. In the table we report estimates for all industries where we observe trade relationships for all three product types.[19]

The results largely confirm our findings—for five of six industries differentiated goods tend to have lower hazard rates and involve longer-lived spells. There are, however, two important caveats. First, for two of the industries the difference among product types is not statistically significant. The two exceptions are "Beverages and Tobacco" (SITC=1) and "Animals and vegetable oils" (SITC=4). Not only do we observe fairly few trade relationships for these two industries but in addition our sense is that the products in these two industries may not be differentiated in an economically important way. Second, we find that differentiated products have a lower hazard rate than reference priced or homogeneous goods for one industry "Mineral Fuels" (SITC=3). We are not troubled by this result because (i) we do not know what it means for a mineral fuel to be differentiated and (ii) in this industry being differentiated may be undesirable.

All in all, these findings confirm that our main result—the difference across product types—is not driven by the distribution of product types across industries.

6   Conclusion

The main goal of this paper was to investigate whether duration of trade relationships depends on the nature of the product being traded and whether any differences are systematic. The results presented indicate that indeed duration of trade depends on the type of the product being traded. Differentiated products have a median duration years longer than either reference priced or homogeneous goods. The hazard rate for differentiated products is lower than for reference priced products, which in turn is lower than for homogenous products

---

[18]We continue to define trade relationships at a more disaggregated level (7-digit TS, 5- and 4-digit SITC).

[19]At the 7-digit TS level there are no results for "Beverages and Tobacco" (SITC=1) as missing values in the explanatory variables preclude identifying all three product types at this level of disaggregation.

We perform a large number of robustness exercises which show that differences in duration across product types are systematic. Our results also indicate that shorter distance, common language, common border, higher GDP, higher tariffs, and depreciation of the source country's currency all lead to longer durations. Higher variation in unit values, lower relative unit values, and higher market share all lead to lower hazard.

We based our analysis on Rauch's (1999) classification of products. One of the features of this classification is that about 80 percent of our products are deemed to be differentiated. While we have shown that such a dominance of differentiated goods is not driving our results, we wonder whether a finer delineation within the differentiated goods category would shed more light on the extent that differentiation matters. In future research we hope to refine the Rauch classification so as to address this question. We reiterate, however, that Rauch's classification is the best currently available, and the development of a new one would involve significant time and expense.

Our analysis indicates that the survival in US import markets will be longer if a differentiated good is traded. An open question that we did not seek to answer in this paper is whether exporters should focus on differentiated goods. In other words, future work could study whether a country's development experience is related to its movement from homogenous to differentiated products.

Furthermore, we also would like to investigate what is the driving force behind the differences among products illuminated in this paper. The question of whether they are a reflection of consumer preferences, network issues, production, technology, or fixed costs associated with exporting are all important questions that should be addressed in future work.

References

[1] Baldwin, Richard, "Hysteresis in Import Prices: The Beachhead Effect," *American Economic Review*, 78 (1988), 773–785.

[2] Besedeš Tibor and Thomas J. Prusa, "On the Duration of Trade," (2003), NBER Working Paper No. 9936.

[3] Dixit, A.K. and J. Stiglitz, "Monopolistic competition and optimum product diversity," *American Economic Review* 67 (1977), 297–308.

[4] Casella, Alessandra and James Rauch, "Overcoming Informational Barriers to International Resource Allocation: Prices and Ties," *Economic Journal*, 113 (2003), 21–42.

[5] Dalgin, Muhammed, Devashish Mitra and Vitor Trindade, "Inequality, Nonhomothetic Preferences, and Trade: a Gravity Approach," (2003), mimeo.

[6] Debaere, Peter, "Relative factor abundance and trade," *Journal of Political Economy*, 111 (2003), 589–610.

[7] Evenett, Simon and Wolfgang Keller, "On Theories Explaining the Success of the Gravity Equation," *Journal of Political Economy* 110 (2002), 281–316.

[8] Feenstra, Robert C., "US Imports, 1972–1994: Data and Concordances," (1996), NBER Working Paper 5515.

[9] Krugman, Paul R., "Scale economies, product differentiation, and the pattern of trade," *American Economic Review* 70 (1980) 950–59.

[10] Helpman, Elhanan, "Imperfect competition and international trade: Evidence from Fourteen Industrial Countries," *Journal of Japanese and International Economies* 1 (1987), 62–81.

[11] Helpman, Elhanan and Paul R. Krugman, *Market Structure and Foreign Trade*. (Cambridge: MIT Press, 1985).

[12] Hummels, David and James Levinsohn, "Monopolistic competition and international trade: Reinterpreting the evidence." *Quarterly Journal of Economics*, 110 (1995), 799–836.

[13] Melitz, Marc, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," Econometrica LXXI (2003), 1695–1725.

[14] Neary, J.P., "Monopolistic competition and international trade theory," (2000), paper prepared for the conference on "The Monopolistic Competition Revolution after Twenty-Five Year."

[15] Rauch, James, "Trade and Search: Social Capital, Sogo Shosha, and Spillovers," (1996), NBER Working Paper 5618.

[16] Rauch, James, "Networks versus Markets in International Trade," *Journal of International Economics*, 48 (1999), 7–35.

[17] Rauch, James, "Business and Social Networks in International Trade," *Journal of Economic Literature* 39 (2001), 1177–1203.

[18] Rauch, James and Vitor Trindade, "Ethnic Chinese Networks in International Trade," *Review of Economics and Statistics* 84 (2002), 116–130.

[19] Salop, Steven, "Monopolistic competition with outside goods," *Bell Journal of Economics*, 10 (1979), 141–56.

[20] Schott, Peter K. "Do Countries Specialize?," (2001), NBER Working Paper No. 8492.

A   Data Appendix

All of the data used in this paper are available from public sources.

Variable	Source
7-digit TS import data	Robert Feenstra's online data collection at http://data.econ.ucdavis.edu/international/
5- and 4-digit SITC import data	http://data.econ.ucdavis.edu/international/
GDP	World Development Indicators (World Bank Statistics)
Consumer Price Index	Bureau of Labor Statistics at http://www.bls.gov/cpi/
Language, Contiguity, Distance	Jon Haveman's online international trade data at http://www.haveman.org/
US Industry level Tariffs	Chris Magee calculated tariff rates; available at http://data.econ.ucdavis.edu/international/
Real Exchange Rates	US Department of Agriculture's Economic Research Service at http://www.ers.usda.gov/Data/exchangerates/
Relative Unit Value (TS level) and Country-Product Market Share	Calculated from 7-digit TS import data from http://data.econ.ucdavis.edu/international/

B   Sketch of models supporting the empirics

In this appendix we use standard models of imperfect competition to gain insight about how the duration of suppliers is affected by the extent of product substitutability. We show that exit is more likely in homogenous product markets than in differentiated product markets. We begin with oligopoly models that take firms' geographical locations as exogenous and emphasize the strategic interaction among these firms (address models). In these models transportation costs (distance) play a key role in differentiating the products. In addition, the strategic interaction among firms can result in firms exiting the market. These models imply the greater is the differentiation between products, the less sensitive are the supplying firms to changes in costs and tariffs.

We also use a standard model of trade with increasing returns and monopolistic competition (a nonaddress model). This model utilizes Dixit-Stiglitz preferences thereby abstracting from strategic issues but nevertheless allowing us to identify the equilibrium number of firms in the market and to show that changes in costs have a greater impact on homogenous product industries than in differentiated product industries.

B.1 Spatial Competition

Begin by considering a version of Hotelling's linear city model applied to trade. In our application we assume that domestic consumers of mass N are uniformly distributed along the $z \in [0, 1]$ interval. A foreign firm is located at each endpoint with firm A located at the left endpoint. Given the nature of our dataset, it is convenient to think of each firm being located in a different foreign country.[20] Hence, domestic consumers can import the product from either country A or B.

In models of spatial competition, transportation costs serve to differentiate the products. For simplicity we will assume that transportation cost are linear in distance shipped.[21] Letting d denote the marginal cost of shipping the product, it follows that the products are less substitutable the larger is d.

The products are produced with constant marginal cost, $c_j > 0$, $j = A, B$. Every consumer wants at most one unit and derives a gross benefit of v from its consumption. For a consumer located a distance x from firm j the total cost of buying is $t_j p_j + dx$, where t_j is one plus the ad valorem tariff. Without loss of generality let firm A have lower effective costs, i.e., $t_A c_A \leq t_B c_B$.

Assuming that all consumers obtain strictly positive surplus by purchasing the good from one of the two countries, one can show that the location of the consumer indifferent between the two firms is

(B.1)
$$z = \frac{3d + t_B c_B - t_A c_A}{6d}.$$

When the firms are symmetric ($t_B c_B = t_A c_A$) $z = 1/2$ and each firm supplies half of the market. For a given value of d firm B's sales decrease as its tariff or its costs increase. In other words, firm B loses market share as it becomes less competitive. Similarly, for a given value of d firm B's sales decrease as firm A's tariff or costs decrease. In other words, firm B loses market share as it's rival becomes more competitive. In addition, note that the larger is d the smaller is the impact of a given change in costs or tariffs. Put another way, tariff and cost changes have a smaller impact on each firms's sales the greater is the product differentiation.

Now, suppose that each firm must incur fixed costs of production, $f_j \geq 0$, to produce the good. As a result firm B will not service the domestic market if

$$t_B c_B > t_A c_A + 3d - (18 f_B d)^{1/2}.$$

---

[20] While our data is extremely disaggregated at the product-level, we only observe the source country of imports. The identity of the specific firm selling the product is not reported.

[21] It easy to demonstrate that a model with quadratic transportation costs yields similar results.

Clearly, if c_B (or t_B) become sufficiently large, firm B will exit the market. Empirically, this means that we might observe firm B active in the market for a period of time and then, after a cost change, exit the market. Note that the greater the product differentiation, the greater the increase in costs before we will observe firm B exiting the market. Taken together these comparative static results suggest that we should find that firms competing in differentiated product industries are less sensitive to tariff and cost changes and hence will experience less exit.

These insights can be extended to the case when there are more than two firms using the model of Salop (1979). In this case it is assumed that consumers are located uniformly on a circle with perimeter 1 and where N firms could potentially service the market. Once again we think of each firm as a foreign firm/country supplying the domestic market, where different consumers prefer to source their imports from different countries.

In the Salop model one can show that the greater is the product differentiation the greater is the profit margin. As was the case in the two-firm linear model, this suggests that we should observe less exit in product markets with greater product differentiation.

B.2 Monopolistic Competition

A standard model of trade with increasing returns to scale and monopolistic competition can also be used to gain insight about how the length of trading relationships varies with the elasticity of substitution.

Let there be N countries and M sectors, where each sector has a large number of product varieties, $n_m < N$. For simplicity (and for consistency with our data), we will assume each country produces only one variety and so we will associate countries with varieties. Domestic consumer preferences are represented by a Cobb-Douglas utility function

$$U = \prod_{i=1}^{M} Q_m^{\gamma_m},$$

where γ_m is the share of expenditures on sector m and Q_m is a composite of symmetric imported product varieties in sector m given by

$$Q_m = \left(\sum_{i=1}^{n_m} x_{mi}^{\frac{\sigma_m - 1}{\sigma_m}} \right)^{\frac{\sigma_m}{\sigma_m - 1}}, \quad \sigma_m > 1,$$

where we have assumed the subutility function has the CES form. The parameter σ is the elasticity of substitution between varieties. The smaller is σ the greater is the extent of product differentiation (Dixit and Stiglitz, 1977).

We can write the budget constraint for a representative domestic consumer as

23

$$\sum_{i=1}^{M} P_i Q_i \leq I.$$

We allow for iceberg transport costs in shipping goods between countries and for import tariffs, so the consumer's price is $P_i = p_i \tau_i d_i$, where p_i is the price the foreign firm earns, τ_i is one plus the ad valorem tariff and d_i is transportation costs.

Each producer (country) i is identified by a good i. The production of a differentiated good i involves a fixed cost f_m and a constant marginal cost c_m. Because of the fixed costs the number of differentiated goods actually being consumed is far less than the number of potential differentiated goods, $n_m < N$. As is usually done, we will assume that n_m is large and that free entry guarantees zero profits.

If we assume symmetry and solve three conditions simultaneously (profit maximization, zero profits, and utility maximization) we can solve for the equilibrium,

(B.2) $$p_m = c_m \sigma_m / (\sigma_m - 1),$$

(B.3) $$x_m = (\sigma_m - 1) f_m / c_m,$$

(B.4) $$n_m = \gamma_m I / (\tau_m d_m f_m \sigma_m)$$

As has been noted elsewhere (Neary, 2000), the price and quantity are functions of costs and the elasticity of substitution. Changes in other parameters (e.g., tariffs and transportation costs) lead to adjustments in the number of foreign firms (countries) only.

Now consider two industries where products in industry j are more differentiated than those in industry i, $\sigma_j < \sigma_i$. From (B.4) it follows that industry i is more sensitive to changes in tariffs, transportation costs, and fixed costs. Therefore, we expect changes in these variables to lead to more exit in homogenous good industries than in differentiated product industries.[22]

---

[22]The substitution parameter also plays a key role in Melitz's (2003) model with heterogenous firms. While Melitz's primary interest is how domestic firms respond to trade, his model implies that changes in tariffs, transportation costs, and fixed costs have greater impact in homogenous good industries than in differentiated product industries.

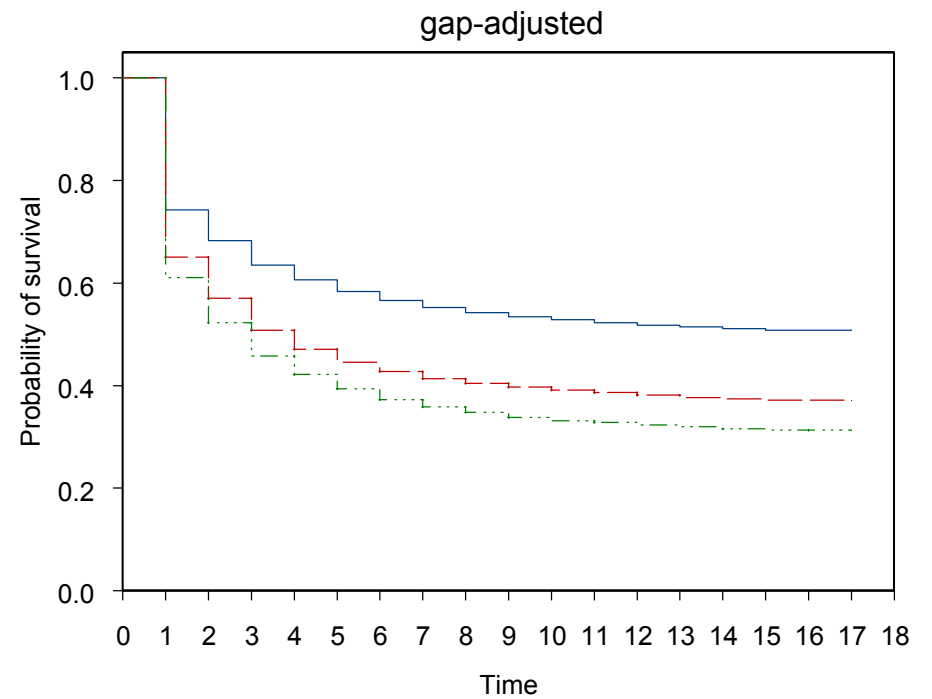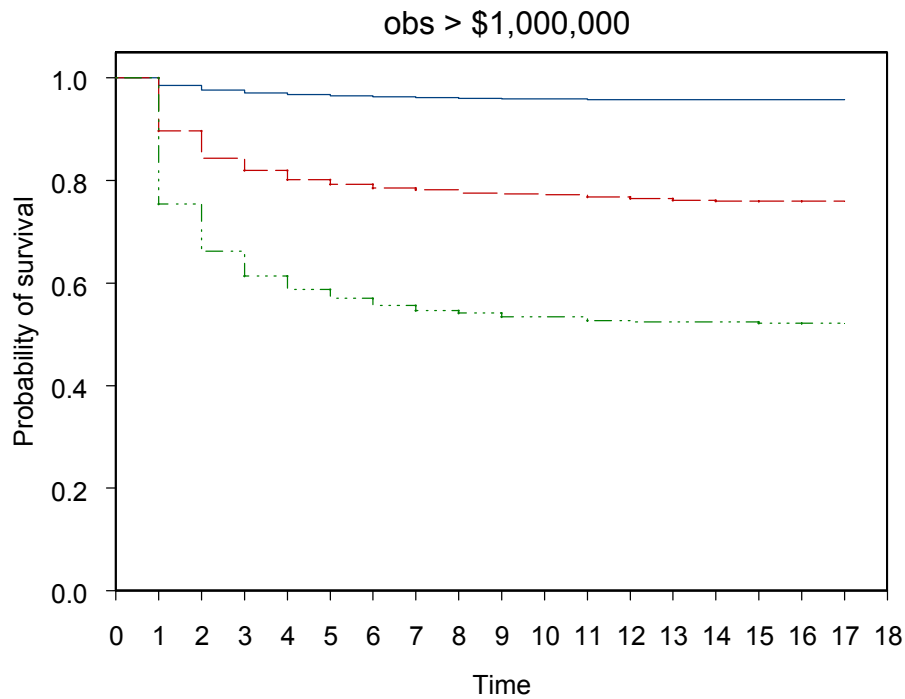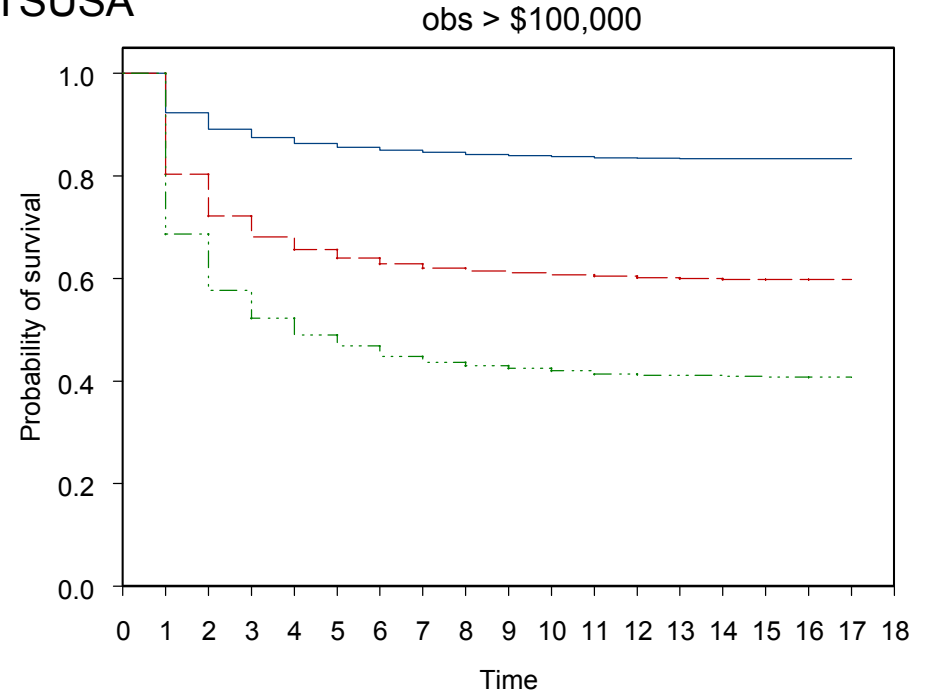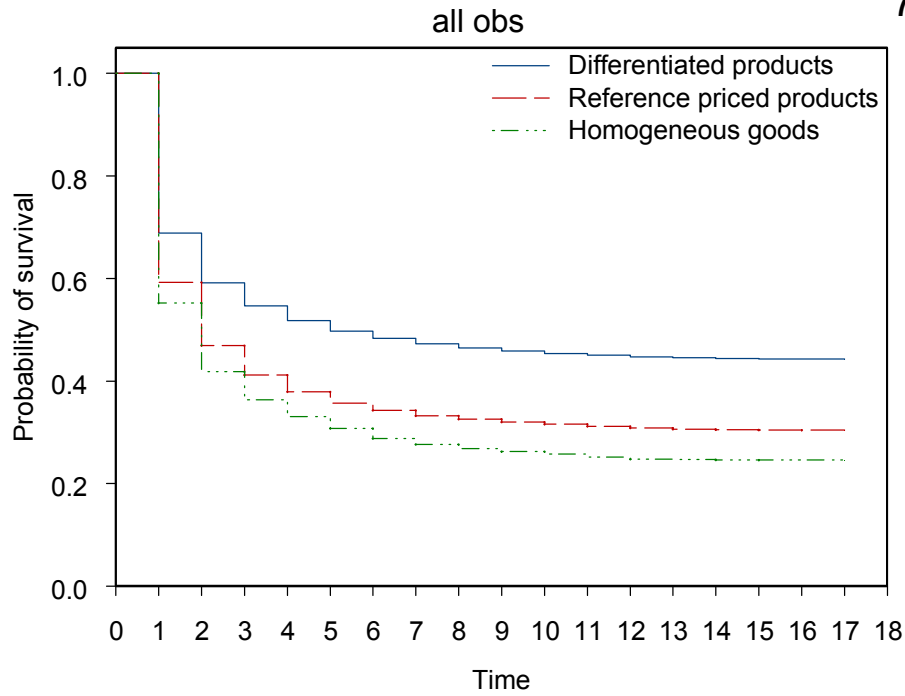Figure 1 - Survival Function for Rauch's Product Classification
7-digit TSUSA

Figure 2 - Survival Function for Rauch's Product Classification

5-digit SITC rev. 2
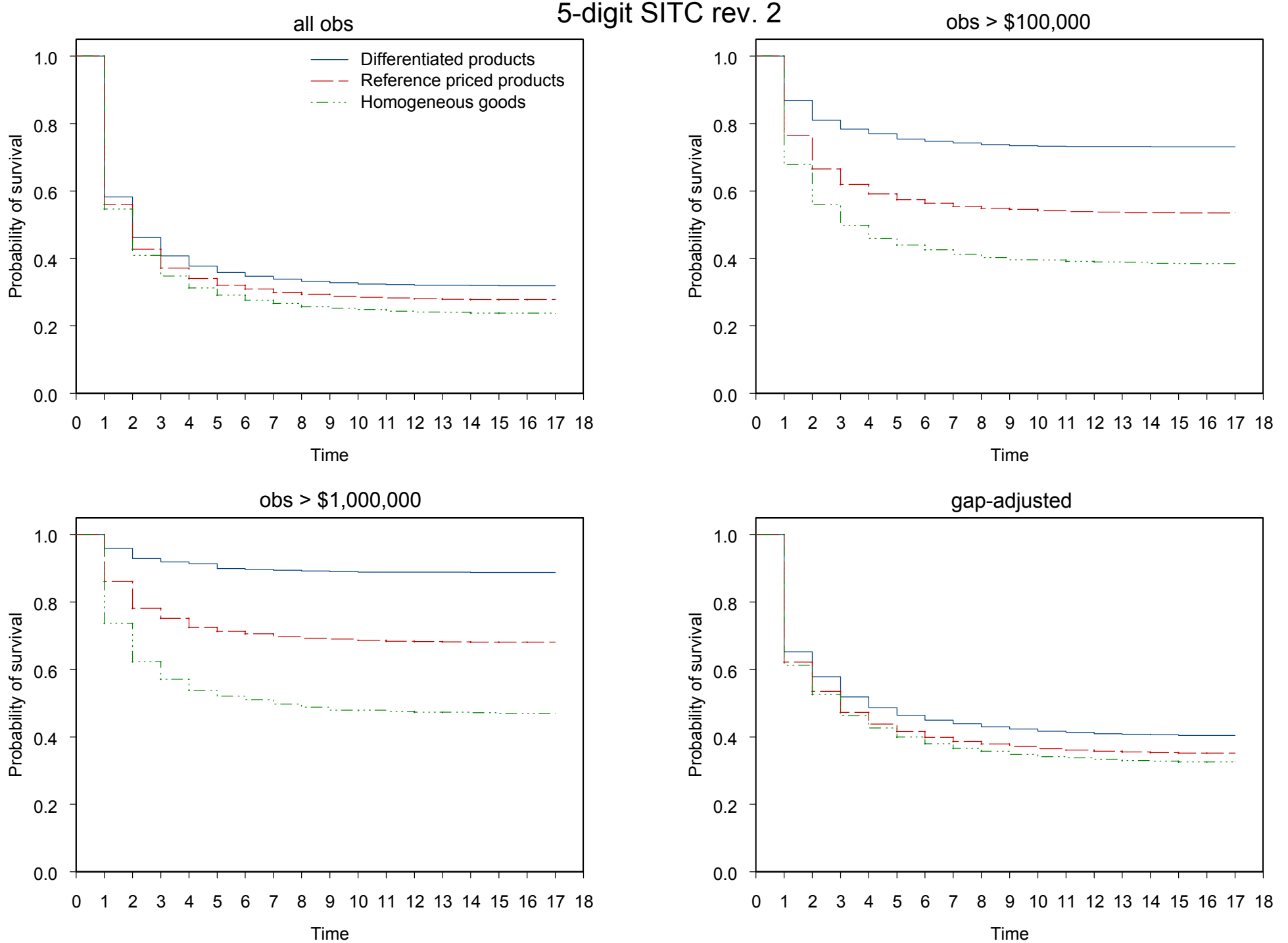
Figure 3 - Distribution of Spell Lengths
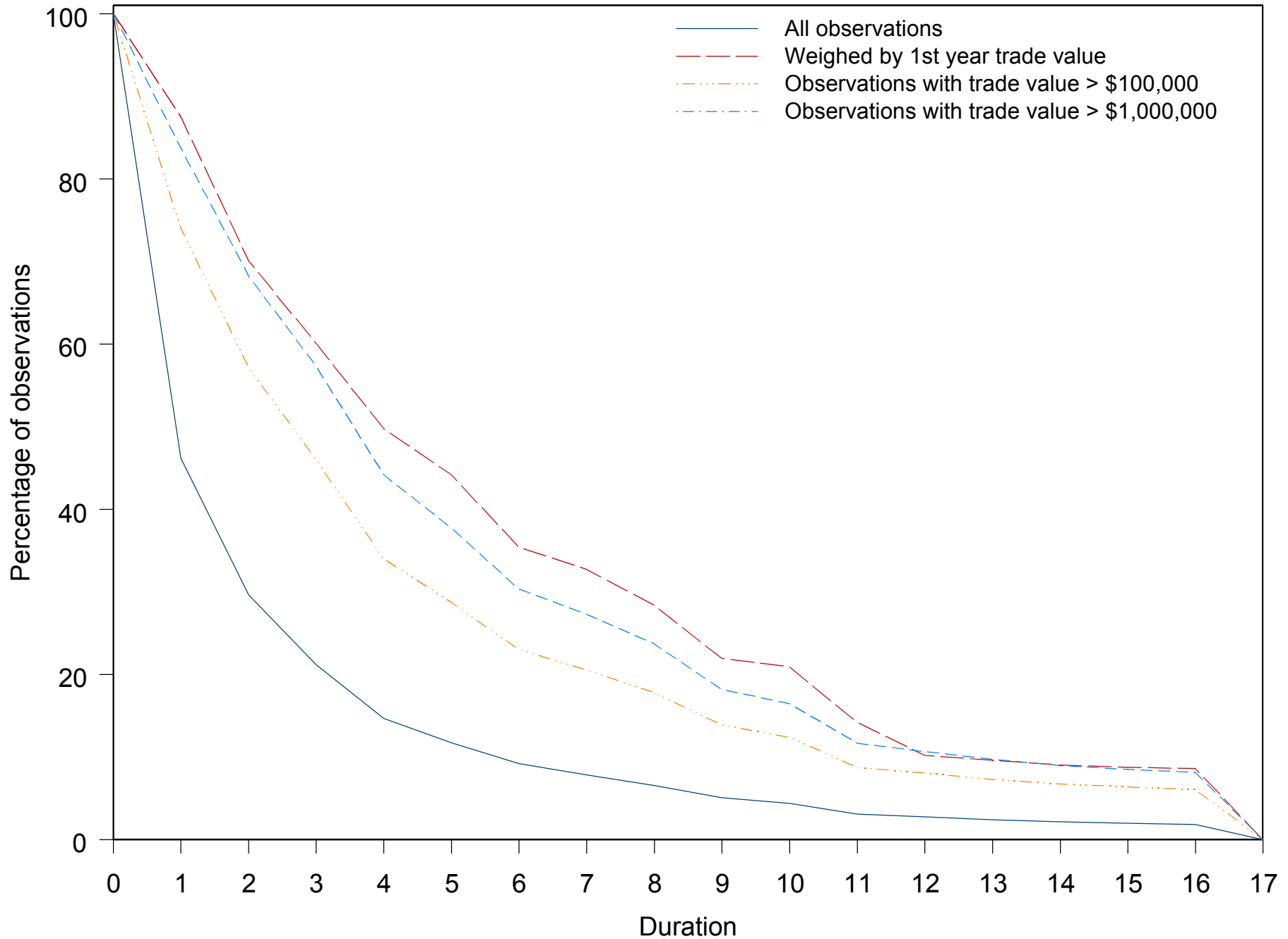
Table 1: TS7 Benchmark and Robustness

	Benchmark	Obs>$100,000	Obs>$1,000,000	Gap-adjusted	First spell only	Single spell relationships	No 1972 spells
Distance	1.00819	1.01383	1.01765	1.01116	1.01388	1.02042	1.00559
(unit = 1,000 kilometers)	(0.000)	(0.005)	(0.115)	(0.000)	(0.000)	(0.000)	(0.000)
Language Dummy	0.94650	0.75806	0.72536	0.93629	0.90335	0.90590	0.96720
(=1 if English)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Contiguous with USA	0.71997	0.52327	0.43649	0.64441	0.64858	0.53858	0.74299
(=1 if share border with USA)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GDP	0.98020	0.96393	0.96913	0.97438	0.97082	0.96421	0.98005
(unit = $100 billion)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
4 digit SIC tariff rate	0.98160	0.94699	0.88849	0.97961	0.97745	0.97633	0.98127
(unit = 1 percentage point)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Percentage change in relative real exchange rate	0.90263	0.90318	0.94450	0.87347	0.92725	0.88464	0.90154
(unit = 10 percentage points)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Coefficient of variation of unit values	0.92058	0.86905	0.91679	0.91211	0.92138	0.88997	0.90007
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Relative unit values	1.00423	1.00862	1.01140	1.05000	1.00518	1.00571	1.00410
(unit = 10 percentage points)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Country-product market share	0.99823	0.99656	0.99045	0.87641	0.99810	0.99962	0.99919
(unit = 1 percentage point)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.088)	(0.000)
Dummy (=1) for spell >= 2	1.70209	2.91827	3.14798	1.72615			1.44327
	(0.000)	(0.000)	(0.000)	(0.000)			(0.000)
Dummy (=1) reference priced	**1.16100**	**1.50266**	**2.53224**	**1.18204**	**1.25144**	**1.22272**	**1.18509**
	(0.000)	**(0.000)**	**(0.000)**	**(0.000)**	**(0.000)**	**(0.000)**	**(0.000)**
Dummy (=1) homogeneous goods	**1.18098**	**1.62029**	**2.83880**	**1.19078**	**1.31844**	**1.30567**	**1.25349**
	(0.000)	**(0.000)**	**(0.000)**	**(0.000)**	**(0.000)**	**(0.000)**	**(0.000)**
Observations	982,418	315,278	117,278	982,418	759,336	631,006	783,364
No. Subjects	365,808	68,217	20,912	314,673	266,251	200,076	333,320
Est. LogL	-1,912,610	-125,443	-17,152	-1,384,010	-1,156,250	-709,864	-1,897,040

p values in parentheses

Region Dummies included but not reported

Table 2: TS7 Benchmark and Robustness -- Product Type Estimates

		Benchmark	Obs>$100,000	Obs>$1,000,000	Gap-adjusted	First spell only	Single spell relationships	No 1972 spells
Differentiated	Distance	1.00635	1.00353	0.96505	1.00887	1.01192	1.01688	1.00328
	(unit = 1,000 kilometers)	(0.000)	(0.551)	(0.084)	(0.000)	(0.000)	(0.000)	(0.020)
	Language Dummy	0.94944	0.76278	0.64723	0.94322	0.90949	0.91436	0.96783
	(=1 if English)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
	Contiguous with USA	0.70603	0.49861	0.45841	0.62755	0.63792	0.52648	0.72378
	(=1 if share border with USA)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	GDP	0.97600	0.95465	0.95843	0.96920	0.96549	0.96135	0.97598
	(unit = $100 billion)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	4 digit SIC tariff rate	0.98128	0.94558	0.88681	0.97923	0.97674	0.97553	0.98128
	(unit = 1 percentage point)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Percentage change in relative real exchange rate	0.89539	0.89495	0.98382	0.86799	0.91554	0.87490	0.89501
	(unit = 10 percentage points)	(0.000)	(0.000)	(0.626)	(0.000)	(0.000)	(0.000)	(0.000)
	Coefficient of variation of unit values	0.89383	0.77488	0.72670	0.88085	0.88256	0.85068	0.87696
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Relative unit values	1.00423	1.00956	1.01268	1.05033	1.00527	1.00585	1.00408
	(unit = 10 percentage points)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Country-product market share	0.99858	0.99710	0.99390	0.92216	0.99851	1.00006	0.99944
	(unit = 1 percentage point)	(0.000)	(0.000)	(0.005)	(0.000)	(0.000)	(0.810)	(0.001)
	Dummy (=1) for spell >= 2	1.80832	4.01020	9.28593	1.84561			1.56169
		(0.000)	(0.000)	(0.000)	(0.000)			(0.000)
Reference priced	Distance	1.01846	1.02746	1.01959	1.02329	1.02622	1.03820	1.01711
	(unit = 1,000 kilometers)	(0.000)	(0.000)	(0.233)	(0.000)	(0.000)	(0.000)	(0.000)
	Language Dummy	0.94557	0.80247	0.75711	0.92430	0.89003	0.87859	0.97110
	(=1 if English)	(0.000)	(0.000)	(0.016)	(0.000)	(0.000)	(0.000)	(0.053)
	Contiguous with USA	0.79329	0.54445	0.33448	0.73211	0.72607	0.63123	0.83288
	(=1 if share border with USA)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	GDP	0.98848	0.96737	0.95526	0.98407	0.98239	0.97007	0.98837
	(unit = $100 billion)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	4 digit SIC tariff rate	0.98058	0.95929	0.89675	0.97734	0.97238	0.97368	0.97890
	(unit = 1 percentage point)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Percentage change in relative real exchange rate	0.92726	0.90651	0.91287	0.89279	0.97789	0.93458	0.92163
	(unit = 10 percentage points)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
	Coefficient of variation of unit values	1.00347	0.98935	1.01335	1.00787	1.04802	1.02881	0.96965
		(0.627)	(0.570)	(0.757)	(0.344)	(0.000)	(0.018)	(0.000)
	Relative unit values	1.00435	1.00549	1.00222	1.05300	1.00526	1.00591	1.00428
	(unit = 10 percentage points)	(0.000)	(0.000)	(0.609)	(0.000)	(0.000)	(0.000)	(0.000)
	Country-product market share	0.99802	0.99611	0.98919	0.83735	0.99786	0.99889	0.99904
	(unit = 1 percentage point)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.019)	(0.000)
	Dummy (=1) for spell >= 2	1.42979	2.26204	2.85765	1.45127			1.16617
		(0.000)	(0.000)	(0.000)	(0.000)			(0.000)
Homogeneous	Distance	1.00845	1.01681	1.01598	1.01398	1.01492	1.02840	1.00638
	(unit = 1,000 kilometers)	(0.040)	(0.035)	(0.225)	(0.004)	(0.006)	(0.000)	(0.119)
	Language Dummy	0.87592	0.70691	0.83665	0.83619	0.80976	0.78269	0.93198
	(=1 if English)	(0.000)	(0.000)	(0.092)	(0.000)	(0.000)	(0.000)	(0.030)
	Contiguous with USA	0.71100	0.53033	0.39045	0.62641	0.61020	0.45643	0.77558
	(=1 if share border with USA)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	GDP	1.00072	1.00035	1.00830	1.00036	1.00242	0.99577	1.00041
	(unit = $100 billion)	(0.689)	(0.931)	(0.217)	(0.868)	(0.385)	(0.339)	(0.819)
	4 digit SIC tariff rate	0.98767	0.93991	0.91005	0.98490	0.98366	0.97911	0.98440
	(unit = 1 percentage point)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Percentage change in relative real exchange rate	0.92768	0.92305	0.94554	0.89169	0.95107	0.90947	0.92889
	(unit = 10 percentage points)	(0.000)	(0.000)	(0.047)	(0.000)	(0.000)	(0.000)	(0.000)
	Coefficient of variation of unit values	1.08638	1.17029	1.14651	1.11388	1.16430	1.16914	1.04500
		(0.000)	(0.000)	(0.013)	(0.000)	(0.000)	(0.000)	(0.015)
	Relative unit values	1.00308	1.01152	1.01081	1.02787	1.00412	1.00342	1.00308
	(unit = 10 percentage points)	(0.000)	(0.000)	(0.030)	(0.011)	(0.005)	(0.078)	(0.001)
	Country-product market share	0.99722	0.99548	0.98677	0.77237	0.99593	0.99592	0.99957
	(unit = 1 percentage point)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)	(0.523)
	Dummy (=1) for spell >= 2	1.21852	1.45754	1.55374	1.24762			0.91961
		(0.000)	(0.000)	(0.000)	(0.000)			(0.002)
	Dummy (=1) reference priced	0.98419	1.11509	1.88295	0.93935	0.91780	0.85553	1.03732
		(0.504)	(0.135)	(0.004)	(0.021)	(0.004)	(0.000)	(0.127)
	Dummy (=1) homogeneous goods	1.11119	1.54068	2.33298	1.01436	0.96978	0.93645	1.25702
		(0.014)	(0.000)	(0.000)	(0.772)	(0.583)	(0.384)	(0.000)
	Observations	982,418	315,278	117,278	982,418	759,336	631,006	783,364
	No. Subjects	365,808	68,217	20,912	314,673	266,251	200,076	333,320
	Est. LogL	-1,912,080	-125,108	-16,988	-1,383,550	-1,155,860	-709,628	-1,896,390

p values in parentheses

Region Dummies included but not reported

Table 3: SITC Industry Analysis

	Benchmark (TS7)	SITC5	SITC4
Distance	1.00819	1.00766	1.00146
(unit = 1,000 kilometers)	(0.000)	(0.000)	(0.518)
Language Dummy	0.94650	0.95391	0.95196
(=1 if English)	(0.000)	(0.000)	(0.000)
Contiguous with USA	0.71997	0.48881	0.38268
(=1 if share border with USA)	(0.000)	(0.000)	(0.000)
GDP	0.98020	0.96861	0.95578
(unit = $100 billion)	(0.000)	(0.000)	(0.000)
4 digit SIC tariff rate	0.98160		
(unit = 1 percentage point)	(0.000)		
Percentage change in relative real exchange rate	0.90263	0.95772	0.96517
(unit = 10 percentage points)	(0.000)	(0.000)	(0.000)
Coefficient of variation of unit values	0.92058		
	(0.000)		
Relative unit values	1.00423		
(unit = 10 percentage points)	(0.000)		
Country-product market share	0.99823		
(unit = 1 percentage point)	(0.000)		
Dummy (=1) for spell >= 2	1.70209	1.29094	1.34945
	(0.000)	(0.000)	(0.000)
Dummy (=1) reference priced	**1.16100**	**1.23962**	**1.19855**
	(0.000)	**(0.000)**	**(0.000)**
Dummy (=1) homogeneous goods	**1.18098**	**1.31709**	**1.36802**
	(0.000)	**(0.000)**	**(0.000)**
Observations	982,418	499,675	319,141
No. Subjects	365,808	116,009	67,149
Est. LogL	-1,912,610	-774,805	-407,848

Region Dummies included but not reported

Table 4: SITC Industry - Product Type Estimates

		Benchmark (TS7)	SITC5	SITC4
Differentiated	Distance	1.00635	1.00845	1.00255
	(unit = 1,000 kilometers)	(0.000)	(0.000)	(0.299)
	Language Dummy	0.94944	0.95288	0.94069
	(=1 if English)	(0.000)	(0.000)	(0.000)
	Contiguous with USA	0.70603	0.46495	0.35228
	(=1 if share border with USA)	(0.000)	(0.000)	(0.000)
	GDP	0.97600	0.93482	0.90265
	(unit = $100 billion)	(0.000)	(0.000)	(0.000)
	4 digit SIC tariff rate	0.98128		
	(unit = 1 percentage point)	(0.000)		
	Percentage change in relative real exchange rate	0.89539	0.96982	0.97928
	(unit = 10 percentage points)	(0.000)	(0.000)	(0.000)
	Coefficient of variation of unit values	0.89383		
		(0.000)		
	Relative unit values	1.00423		
	(unit = 10 percentage points)	(0.000)		
	Country-product market share	0.99858		
	(unit = 1 percentage point)	(0.000)		
	Dummy (=1) for spell >= 2	1.80832	1.31226	1.36489
		(0.000)	(0.000)	(0.000)
Reference priced	Distance	1.01846	1.01268	1.00824
	(unit = 1,000 kilometers)	(0.000)	(0.000)	(0.015)
	Language Dummy	0.94557	0.95988	0.97124
	(=1 if English)	(0.000)	(0.038)	(0.254)
	Contiguous with USA	0.79329	0.55822	0.43667
	(=1 if share border with USA)	(0.000)	(0.000)	(0.000)
	GDP	0.98848	0.98564	0.97298
	(unit = $100 billion)	(0.000)	(0.000)	(0.000)
	4 digit SIC tariff rate	0.98058		
	(unit = 1 percentage point)	(0.000)		
	Percentage change in relative real exchange rate	0.92726	0.93162	0.93502
	(unit = 10 percentage points)	(0.000)	(0.000)	(0.000)
	Coefficient of variation of unit values	1.00347		
		(0.627)		
	Relative unit values	1.00435		
	(unit = 10 percentage points)	(0.000)		
	Country-product market share	0.99802		
	(unit = 1 percentage point)	(0.000)		
	Dummy (=1) for spell >= 2	1.42979	1.21389	1.27996
		(0.000)	(0.000)	(0.000)
Homogeneous	Distance	1.00845	1.00599	1.00071
	(unit = 1,000 kilometers)	(0.040)	(0.175)	(0.893)
	Language Dummy	0.87592	0.94335	0.95753
	(=1 if English)	(0.000)	(0.105)	(0.313)
	Contiguous with USA	0.71100	0.56005	0.51471
	(=1 if share border with USA)	(0.000)	(0.000)	(0.000)
	GDP	1.00072	0.99927	0.99743
	(unit = $100 billion)	(0.689)	(0.741)	(0.386)
	4 digit SIC tariff rate	0.98767		
	(unit = 1 percentage point)	(0.000)		
	Percentage change in relative real exchange rate	0.92768	0.91653	0.91903
	(unit = 10 percentage points)	(0.000)	(0.000)	(0.000)
	Coefficient of variation of unit values	1.08638		
		(0.000)		
	Relative unit values	1.00308		
	(unit = 10 percentage points)	(0.000)		
	Country-product market share	0.99722		
	(unit = 1 percentage point)	(0.000)		
	Dummy (=1) for spell >= 2	1.21852	1.21978	1.30011
		(0.000)	(0.000)	(0.000)
	Dummy (=1) reference priced	**0.98419**	**1.12351**	**1.06722**
		(0.504)	**(0.000)**	**(0.050)**
	Dummy (=1) homogeneous goods	**1.11119**	**1.21651**	**1.20794**
		(0.014)	**(0.000)**	**(0.000)**
	Observations	982,418	499,675	319,141
	No. Subjects	365,808	116,009	67,149
	Est. LogL	-1,912,080	-774,420	-407,576

p values in parentheses

Region Dummies included but not reported

Table 5: Product Type Effect (within each 1 digit industry)

TS7 Data

	SITC=0	SITC=1	SITC=2	SITC=3	SITC=4	SITC=5	SITC=6
Dummy (=1) reference priced	1.15724		0.99029	0.81701	1.01956	1.15638	1.02592
	(0.000)		(0.782)	(0.038)	(0.874)	(0.000)	(0.015)
Dummy (=1) homogeneous goods	1.12250		1.23185	0.77275	1.11415	1.24701	1.21653
	(0.000)		(0.000)	(0.013)	(0.249)	(0.001)	(0.000)
Observations	60,668		17,898	6,060	4,310	72,688	336,430
No. Subjects	20,654		7,001	2,193	1,728	23,399	125,307

SITC5 Data

	SITC=0	SITC=1	SITC=2	SITC=3	SITC=4	SITC=5	SITC=6
Dummy (=1) reference priced	1.18885	0.93920	1.18070	0.83029	1.17367	1.23992	1.14980
	(0.000)	(0.472)	(0.000)	(0.025)	(0.260)	(0.000)	(0.000)
Dummy (=1) homogeneous goods	1.28220	1.06183	1.26161	0.64195	1.16755	1.53650	1.28076
	(0.000)	(0.589)	(0.000)	(0.000)	(0.169)	(0.000)	(0.000)
Observations	45,600	6,549	36,204	5,245	3,263	46,361	137,756
No. Subjects	11,469	1,446	10,170	1,624	1,089	11,509	31,276

SITC4 Data

	SITC=0	SITC=1	SITC=2	SITC=3	SITC=4	SITC=5	SITC=6
Dummy (=1) reference priced	1.06292	0.99323	1.23944	0.79860	1.07511	1.21607	1.13399
	(0.106)	(0.947)	(0.000)	(0.013)	(0.628)	(0.000)	(0.000)
Dummy (=1) homogeneous goods	1.22083	1.17009	1.38782	0.61029	1.11979	1.36894	1.27158
	(0.000)	(0.198)	(0.000)	(0.000)	(0.294)	(0.022)	(0.000)
Observations	34,568	5,748	25,955	4,621	3,098	30,453	86,289
No. Subjects	7,604	1,189	6,800	1,385	984	6,476	17,475

Full specification results available upon request