

Ins, Outs, and the Duration of Trade*

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

Thomas J. Prusa
Rutgers University and NBER
Department of Economics
New Jersey Hall
New Brunswick, NJ 08901–1248, USA
prusa@econ.rutgers.edu

December 2004

Abstract

We employ survival analysis to study the duration of US imports. Our findings indicate that international trade is far more dynamic than previously thought. We find the median duration of exporting a product to the US is very short, on the order of two to four years. There is negative duration dependence—if a country is able to survive in the exporting market for the first few years it will face a very small probability of failure and will likely export the product for a long period of time. The results hold across countries and industries and are robust to aggregation.

Key words: International Trade, Duration Analysis, Market Entry, Market Exit

JEL classification: F14, F19, C14, C41

*We would like to thank the seminar participants at Rutgers University, Louisiana State University, the Midwest International Trade Meetings, the NBER ITI group and also David Hummels, Simon Evenett, Roger Klein, Johannes Moenius, and Peter Schott for helpful suggestions and comments.

1. Introduction

Positive trade theory usually asks questions which address the “who, what, when, and why” of international trade. One question not often addressed is “how long?” When countries trade, how long do their trade relationships last? Are they exchanging products over long or short periods of time?

To answer these questions we study duration of trade relationships. To our knowledge we are the first to do so and several important findings emerge from the analysis. First, not only is there a remarkable amount of entry and exit in the US import market, but the period of time a country is “in” the market is often fleeting. Some trade relationships are long-lived—a country exports a product for many years. It is far more common to observe short-lived trade relationships wherein a country trades a product for a few years and then stops. More than half of all trade relationships are observed for a single year and approximately 80 percent are observed for less than five years.

Second, because the analysis is performed at a highly disaggregated product level one must carefully account for censored observations. Censoring is a particularly thorny issue with the benchmark product level data because of reclassification of product codes. Once we account for censoring we can estimate survival and hazard rate functions of exporting to the United States. The median duration that a country exports a product to the US is very short, somewhere between two and four years depending on the censoring approach employed. Although duration does vary by source country and region, short-lived relationships characterize trade by most countries. Short relationships are prevalent for both OECD and non-OECD countries although OECD trade relationships exhibit systematically longer survival.

Third, the results are not only robust to aggregation but in fact are strengthened by aggregation. As expected, within the SITC classification, higher levels of aggregation are associated

with longer survival times. Nevertheless, there remain a large number of short spells of service until the data are very highly aggregated. Despite the higher degree of aggregation the median survival time for SITC data is only 2–3 years.

Surprisingly, as we aggregate from the product level to the SITC industry level the estimated probability of survival decreases. This paradoxical result is related to the unique censoring problems only present at the product level. As a result, it is more straightforward to conduct industry level analysis although doing so makes it more difficult to interpret what the relationship represents. Even so the aggregation exercise confirms that the main finding is not an anomaly.

Fourth, the results indicate the presence of negative duration dependence—the conditional probability of failure decreases with duration. The hazard rate in the first year is 33 percent and between year one and year five is an additional 30 percent. Conditional on observing a trade relationship surviving the first five years, the hazard rate for the remainder of observed time is just 7–12 percent. A type of a threshold effect may be present—once a relationship is established and has survived the first few years it is likely to survive a long time.

Fifth, the findings are not sensitive to changes in how we measure trade relationships or define failure. The median duration is often shorter in alternative formulations than in benchmark data. The most significant changes occur when we give greater weight to relationships with large trade values. While increasing duration, 50 percent of trade weighted relationships are observed for less than five years.

The results indicate more is happening at the micro-level than suggested by either existing theory or empirical studies. According to the factor proportions theory, trade is based on factor endowment differences. Since such differences change gradually trade patterns are likewise expected to evolve slowly. Unhappiness with this implication spurred the development of trade models with richer trade dynamics (Krugman, 1979; Dollar, 1990; Grossman and Helpman,

1991). Of these, Vernon's (1966) seminal product cycle theory is probably the best known. Vernon's model generates a particular pattern of trade relationships. Technological leaders develop and export a product until others learn how to manufacture it and enter the market. As technology becomes more standardized, other countries will begin to produce and export the product. If follower countries have relatively low labor costs, they will eventually take over the market and push out the leaders. All models imply a fairly predictable pattern of trade dynamics—they evolve either slowly or in a logical progression from developed to developing countries.¹ None of the models suggest trade relationships would be as fleeting as implied by our results.

To our best knowledge, we are the first to investigate the issue of duration of trade. While we introduce a new methodology (or perspective) of investigating international trade issues, the paper straddles several literatures. We can place our paper in a developing literature taking advantage of newly available, highly disaggregated trade data. Feenstra and Rose (2000) were among the first to use such data. They investigated the product cycle theory by ranking countries based on the first year of exporting a product to the US. Their rankings are consistent with the product cycle theory. Schott (2004) investigates factor-proportions specialization and finds evidence supporting it within products, but not across products. Broda and Weinstein (2004) use the same data to study growth in new varieties of traded products and estimate elasticities of substitution as well as an exact price index.

The paper also makes a contribution to the literature on dynamics of trade. Proudman and Redding (1998) and Redding (2001) investigate changes in the pattern of specialization over time, as does Schott (2004). The paper is also related in some respects to the literature on firm- and plant-level dynamics as reviewed by Tybout (2003). While the firm- and plant-level literature studies dynamics at a highly disaggregated level as do we, it is primarily focused

¹In Grossman and Helpman's (1991) "quality ladder" variant of the product cycle model the leader and follower exchange dominance of exports of a particular good over time, as leaders re-enter the market for a given good by innovating and offering a more advanced version.

on export performance, while we are investigating primarily the importing side of trade. We discuss other significant differences in the conclusion.

Our investigation of duration of trade is stimulated in part by the findings of Haveman and Hummels (2004), Feenstra and Rose (2000), and Schott (2004) who document that in any given year and for any given product, many countries do not trade.² None of the earlier studies examined whether a country's current state (i.e., being "in" or "out" of the market) was a permanent feature. We show market presence is often a transitory phenomenon.

2. Data

The analysis is based on US import statistics as compiled by Feenstra (1996) and augmented by Feenstra, Romalis, and Schott (2002).³ From 1972 through 1988 import products were classified according to the 7-digit Tariff Schedule of the United States (TS). Since 1989 imports have been classified according to the 10-digit Harmonized System (HS).⁴ To avoid potential concordance issues, we use the second period, 1989-2001, as a natural robustness test.⁵ We only study imports because there is no concordance between the disaggregated import and export codes at the TS level. For HS data such a concordance is available, but only at the 6-digit level. In addition, export data are self-reported making it more likely that exports are misreported.

For each product we can identify all countries from which the US imported it in a given year. In each period the US imported a total of about 23,000 different products. On average, in each year we observe import trade for about 10,000 products sourced from about 160 countries

²Feenstra and Rose's approach depends on the fact that many of the zero trade value observations become positive in later years.

³The data are available from <http://data.econ.ucdavis.edu/international/>.

⁴We use TS (HS) interchangeably to denote the classification scheme and the period when it was used, 1972-88 (1989-2001).

⁵Many TS products are mapped into multiple HS codes and vice-versa making it impossible to create a long product-level panel. See Schott (2004) for more discussion of the TS-HS concordance problem.

between 1972 and 1988 and for about 15,000 products from about 180 countries between 1989 and 2001. We will refer to the results based on TS and HS data as the benchmark results. Later we perform a series of robustness checks including aggregating from the product level to the industry level data where we use Standard International Trade Classification (SITC) industry codes to define trade relationships.

2.1. Trade relationships and spells of service

Our interest is to study the length of time until a country ceases to export a product to the US, an event we will refer to as a “failure.” Calendar time is not as important as analysis time, which measures the length of time a country exports a product to the US. For each product and country we use the annual data to create spell data. If the US imports product i from country c from 1976–1980, the ci^{th} trade relationship has a spell length of five.

The benchmark analysis is based on the most disaggregated data available: the analysis is at the product level, not the industry level. Inferences are based on trade in tangible products rather than aggregate summaries. Until recently such disaggregated data were not widely available and empirical studies were based on industry classifications.

As an example of the level of detail, TS data report information on more than 30 different types of ball bearings, differentiating by specialized application (e.g., automobiles), size, and chemistry. At the SITC level, all ball bearing codes are aggregated along with other similar, or perhaps not so similar, products to create spells of service. The dozens of ball bearing codes map into a single SITC category, “Ball, roller or needle roller bearings” (4 digit code=7491). 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.

Using highly disaggregated data for duration analysis is imperative. First, the more aggre-

gated the data, the more the analysis identifies industry trends rather than competitive dynamics at the product level. That country c in industry j has a long duration tells us little about duration of commodity trade or underlying trade dynamics.

Suppose we observe three countries exporting “Industrial machinery” with each having long duration. One might surmise dynamics among the countries to be very similar. It need not be the case. Country c_1 might be uniformly superior: every product it sells could have a long duration. Country c_2 might be a dismal failure: the success (i.e., long duration) of one product could hide a multitude of failures (short-lived spells). Country c_3 could be a classic example of Vernon’s product cycle: the observed long duration at the industry level could reflect a progression from simpler products (ball bearings) to very complicated products (engines). The more aggregated the data the more cautious one should be interpreting the results; a less munificent view might be that aggregated data make it impossible to measure dynamics.

Second, if the products are too broadly defined, we cannot expect to see any source countries exit the market. If we aggregate all imports from each country, we will observe little exit since the United States purchases some product from nearly every source country every year. Higher aggregation may hide a great deal of dynamics and competition in the import market.

For each product we create a panel of countries which export the product to the US. Table 1 provides an example of TS data for a representative product, “Milled Corn (TS=1312000).” The “X”s in the table indicate years in which each country exports the product to the US. There are countries that export corn every year, such as Canada and Portugal. Exporting a product in every observed year creates a 17 year long spell. Another dozen countries have a single spell. Mexico begins exporting corn in 1974, South Korea in 1976, and Peru in 1984, and all three service the US market in every year after entry. The remaining countries with a single spell all export corn for just one year.

There are a number of countries with two spells. The United States imports corn from

Venezuela in every year but 1981. Ecuador exports corn in 1973 and 1974 (first spell, length 2) and then services the US market for the second time starting in 1979 and continuing through 1988 (second spell, length 10). The Dominican Republic also has two spells, but both of length 1. Colombia and Italy each have three spells, both starting with two short ones (of length 3 and 1; and 1 and 2), before entering the US market for the third time (1981 and 1980) and exporting for the remainder of observed time. Of the remaining countries most have very short spells, with the exception of the Netherlands whose first spell is five years long. Its remaining four spells, however, are all one year long.

As shown in Table 2 there are 693,963 (918,236) observed spells of service at the TS (HS) level. Both TS and HS datasets have a median spell length of 1 year and a mean spell length of about 3 years. We will discuss how aggregation affects the results in more detail in section 5, but we note here that Table 2 also includes summary statistics for trade relationships created from the more aggregated SITC industry level data. In the upper half of the table we see that in the 1972–88 period there are 157,441 observations at the 5-digit SITC level, 98,035 observations at the 4-digit level and just 2,445 observations at the 1-digit level. As expected, aggregating the data diminishes the ability to observe entry and exit—for the 1972–88 period the mean spell length increases from 2.7 years in 7-digit TS data to 3.9 years in the 5-digit SITC data to 8.4 years in the 1-digit SITC data.⁶

In Figure 1 we plot the distribution for each observed spell length for each sub-period. The x -axis plots the observed spell length. The y -axis plots the percentage of observations whose observed spell of service is greater than a given length. The graphs depict very similar trends. In both the 7-digit TS and 10-digit HS data more than half of all spells are just one year long; about 70 percent of the spells are observed for two or fewer years; about 80 percent of the spells

⁶Qualitatively similar trends are seen for the 1989–2001 period (the lower half of the table) although the mean and median spell lengths do increase slightly.

are observed for three or fewer years. In both periods we see that only a small fraction of spells are observed the full length of the panel. The main point is clear: there is a plethora of short spells. In Figure 1 we plot the distribution for alternative measures of trade relationships as well; these other distributions will be discussed later in the paper.

2.2. Multiple spells

As illustrated in Table 1, some trade relationships re-occur, exhibiting what we will refer to as multiple spells of service. A country will service the market, exit, then re-enter the market, and then almost always exit again. Approximately 30% of relationships experience multiple spells of service in the disaggregated product level data. About two-thirds of relationships with multiple spells experience just two spells; less than ten percent have more than three spells.

We begin by treating multiple spells as independent. While the assumption is made primarily in the interest of simplicity, in section 5 we explore alternatives and find the results are likely not overly sensitive to it. To get a sense of why, it is instructive to look at the distribution of spells excluding multiple spells. First let us restrict the analysis to trade relationships with just one spell. While the message is the same whether we look at HS or TS data, for expositional clarity let us look at TS data. As reported in Table 2, restricting ourselves to the “one spell only” results in a median (mean) spell length of 1 (3.2) year. The median is exactly the same as in benchmark data and the mean is just a half a year longer. In the upper graph in Figure 1 we show the entire distribution for both the benchmark and the “one spell only” data. More than half of all spells are just one year long and about 75 percent of the spells are observed for less than four years; further, in both distributions only about 5 percent of the spells are observed for 10 or more years. In other words, the overall distribution of spell lengths does not appear to be distorted by multiple spell observations.

Alternatively, we can restrict the analysis to the first spell only: relationships with just one

spell and the first spell of multiple spell relationships. As was the case in the first comparison, the distribution is very similar to the benchmark (see Figure 1). The similarity among the distributions suggests the independence assumption is a reasonable starting place.

2.3. Censoring

Once we begin to think of data in terms of spells it becomes apparent we need to account for censoring in the analysis. It is often unknown whether a trade relationship ends because of a failure or for some other reason. Consequently, there is uncertainty regarding either the beginning or the ending date (or both) for some trade relationships.

Censoring is common in US import data. In both periods about half of all spells are censored and about 20 percent of spells are censored at one year. The censoring problem comes in two flavors. First, there is no information on trade relationships for the years before the beginning and after the end of the sample. From Table 1 we observe the US imported corn from the Philippines in 1972 and the relationship was observed for exactly one year. It may have begun in 1972 or it may have begun in some prior year. The most appropriate interpretation is it had a duration of at least one year. Similarly, we observe the US importing corn from Peru from 1984 to 1988. Data do not continue beyond 1988 (recorded using the TS classification) and it is impossible to ascertain how long the spell ultimately lasted. Once again, the most appropriate interpretation is a duration of at least five years. For TS data about 10 percent of spells are observed in 1972, while about 22 percent are observed in 1988. For HS data about 20 percent of spells are observed in 1989 and about 28 percent are observed in 2001.⁷ The first type of censoring is typical in survival studies and is incorporated in the subsequent analysis.

The second type of censoring is unique to the product level data. The US Customs revises

⁷For TS data less than two percent of all spells are observed continuously in every year, while less than six percent of spells for HS data are observed in every year.

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, there is no information to allow us to map old product codes into new ones. When a code is changed there is no longer any observed trade in the old code, but is this due to the end of a relationship or does it simply mean trade stopped due to reclassification? Throughout much of the analysis we choose to be cautious and classify all such “exits” as censored. Reclassified relationships are interpreted as having duration of at least x years (where x is the number of years when trade in the original code was observed).

An analogous problem exists for new codes. When a code is changed we begin to observe trade in the new code, but is this really the beginning of a new relationship or does it simply reflect the reclassification? Once again, we choose to be cautious and assume the relationship had a duration of at least y years (where y is the number of years when trade in the new code was observed). About 20 (10) percent of the spells are censored in TS (HS) data due to reclassification. The second type of censoring is a peculiar characteristic of our particular application and we incorporate it into benchmark estimates.

The interpretation of all product code changes as being censored is very conservative—certainly some of the new codes were truly measuring new products and some of the obsolete codes were measuring products that were no longer traded. In benchmark TS and HS data some spells are probably classified as being censored when they were truly associated with either entry or exit. If more information were available, it would be possible to identify such births and deaths.⁸ Unfortunately, this additional information does not exist. Consequently, the benchmark results will overstate the true duration of a typical trading relationship.

⁸Xiang (2004) uses the changes in the verbal description of 4-digit SIC industry codes to identify new products. Even using relatively aggregated industry data (about 450 industries) Xiang describes a time-consuming and painstaking process.

3. Modeling Duration

3.1. Duration models

Let T denote time to a failure event. Since time in the analysis is discrete, we assume T is a discrete random variable taking on values $t_i, i = 1, 2, \dots, n$ with a probability density function $p(t_i) = \Pr(T = t_i), i = 1, 2, \dots, n$ where $t_1 < t_2 < \dots < t_n$. The survival function for a random variable T is given by

$$S(t) = \Pr(T > t) = \sum_{t_i > t} p(t_i).$$

The hazard function is

$$h(t_i) = \Pr(T = t_i | T \geq t_i) = \frac{p(t_i)}{S(t_{i-1})}, \quad i = 1, 2, \dots, n$$

where $S(t_0) = 1$. The survival and hazard functions are related through the following expression

$$S(t) = \prod_{t_i < t} [1 - h(t_i)].$$

3.2. Nonparametric estimation

To estimate the survival and hazard functions we will assume we have n independent observations denoted $(t_i, c_i), i = 1, 2, \dots, n$, where t_i is the survival time and c_i is the censoring indicator variable C of observation i . C_i takes on a value of 1 if failure occurred and 0 otherwise. Assume there are $m \leq n$ recorded times of failure. Denote the rank-ordered survival times as $t_{(1)} < t_{(2)} < \dots < t_{(m)}$. Let n_i denote the number of subjects at risk of failing at $t_{(i)}$ and let d_i denote the number of observed failures. The Kaplan-Meier product limit estimator of the

survivor function is then

$$\widehat{S}(t) = \prod_{t^{(i)} \leq t} \frac{n_i - d_i}{n_i},$$

with the convention that $\widehat{S}(t) = 1$ if $t < t_{(1)}$. The Kaplan-Meier estimator is robust to censoring and uses information from both censored and non-censored observations.

The hazard function is estimated by taking the ratio of subjects who fail to the number of subjects at risk in a given year,

$$\widehat{h}(t) = \frac{d_i}{n_i}.$$

We choose to estimate the hazard function at the observed failure times only.

4. Empirical Results

4.1. 7-digit TS data

4.1.1. Benchmark and modified censoring survival functions

We begin by examining the benchmark 7-digit TS data and report our findings in Tables 2 and 3. Table 2 reports the 1-, 4-, and 12-year survival rates for benchmark data and also for a series of alternative formulations designed to investigate the robustness of the results. Table 3 reports survival rates for benchmark data and also for sub-samples (i.e., by region and industry). The analysis conveys several important lessons about duration of trade.

First and foremost, a very large fraction of relationships fail after only a year or two. For benchmark TS data, only 67 percent of relationships survive one year; 49 percent survive four years; 42 percent survive 12 years (Table 2). An almost identical survival experience is found in HS data. In fact, as we will discuss below, a qualitatively similar experience is seen across all of our runs.⁹ The message is quite clear: the typical US trade relationship is very short-lived.

⁹The alternatives will be discussed in section 5.

The estimated overall survival function, $\widehat{S}(t)$, is graphed in the upper left hand corner of Figure 2 together with the 95% confidence interval. The confidence interval is very tight (im-perceptible), which is not surprising given the size of the data set.¹⁰ The survival function is downward sloping with a decreasing slope. It suggests a declining hazard rate function, as confirmed in the lower left hand corner of Figure 2.

The second important finding is the sharp decline of the risk of failure. It is quite high in the early years, but then rapidly falls once a trade relationship survives a threshold duration. As shown, a large number of relationships fail over the first four years, especially in the first year when the hazard rate is 33 percent for TS data. However, after about four or five years failure becomes a lot less common. The hazard rate between year one and year five is an additional 30 percent. The hazard rate for the remaining twelve years is just 12 percent.¹¹ Countries exporting to the US face a large conditional probability of failure in the early stage of their trade relationship and a much smaller one after surviving a few years.

Both the survival and hazard functions indicate negative duration dependence—the conditional probability of failure decreases as duration increases. There is a type of a threshold effect. Once a relationship is established and has survived the first few years it is likely to survive a long time. The finding is similar to that documented by Pakes and Ericson (1998) for retail trade establishments.¹²

It is somewhat surprising the Kaplan-Meier estimated probability of exporting a product for more than 17 years is so high (0.41) given that less than two percent of all trade relationships span the entire sample (Figure 1). The explanation for the seemingly inconsistent finding is the prevalence of censoring at the product level. Many relationships observed to end are censored

¹⁰Given the narrowness of the estimated confidence intervals we do not plot them in any other figures.

¹¹In TS data more than 50 percent of observed spells of service fail within the first four years, but over the next 13 years about 7 percent of spells fail.

¹²Pakes and Ericson interpret their results as supporting Jovanovic's (1982) theory of learning.

and are not classified as failures in the benchmark results.

The impact of censoring due to product code changes can be identified if we estimate the Kaplan-Meier survival function using the modified censoring approach already discussed. Under the alternative scheme all changes and reclassifications in TS codes are interpreted as births and deaths which leads to much more entry and exit in the data. We report the results in Table 2 and plot the survival curve in Figure 3.¹³ Under the alternative approach duration is even shorter than in the benchmark case: the median duration falls to just two years as compared with four years in benchmark data; the 75th percentile is just six years.

As depicted in Figure 3 the probability of survival is substantially lower during the initial years under modified censoring; by year four it is about 20 percentage points smaller. As was the case with benchmark data, the hazard rate falls sharply and there is little additional failure after four or five years. The probability of exporting a product for more than 17 years under modified censoring is only 0.18—less than half of that of the benchmark—but still considerably higher than the observed two percent of trade relationships that span the entire sample.

4.1.2. Region and industry survival rates

In Table 3 we report survival rates broken out by individual regions and industries for both the benchmark and modified censoring data. For all regions a substantial fraction of trade relationships quickly end in failure: 20–50 percent of relationships fail in the first year. In each region a substantial number of additional relationships fail in the next several years. By the end of year four, no more than two-thirds and as little as 29 percent of relationships have survived. While there are clear differences across regions in the severity of the initial wave of failures the pattern of negative duration dependence is seen for each region.¹⁴

¹³In the modified scheme products traded in 1972 or 1988 are still classified as censored.

¹⁴The survival rates across regions and industries are all statistically different from one another. Given the large sample size this result is not surprising.

The analysis at the individual region level suggests duration of trade can broadly be thought of as a North versus South story, with Northern sources having higher survival rates. With benchmark data Northern regions (as represented by North America, Asia, and Western Europe) have a median duration of at least six years; for each Northern region about a half of spells survive at least twelve years.

Quite a different story emerges for Southern regions. They have a median duration of no more than two years—half of all trade relationships will fail within the first two years. By year 12, no more than one-third of spells survive. The difference between Northern and Southern regions is also observed under the modified censoring scheme, where once again Northern regions have uniformly higher median survival times.¹⁵

To confirm the North-South differences, we divided the countries into two groups, OECD and non-OECD, and estimated the survival function for each group. On average, the estimated survival for OECD countries is about 10 percentage points higher than for non-OECD countries. While the survival pattern is qualitatively similar across regions, Northern countries clearly fare better.¹⁶

Looking at survival functions across industries broadly similar trends appear again—a wave of initial failures followed by a significant decrease in the hazard. At least 25 percent of spells fail in year one for every industry. In benchmark data, with the exception of two industries, the median survival time is three years or less.

In two key industries, Machinery (SITC 7) and Miscellaneous Manufactures (SITC 8), the estimated median duration is far longer, more than 11 years. An explanation for the distinct difference between these two industries and the others lies in the high prevalence of product

¹⁵Using the benchmark data we estimated the survival functions for individual regions. In the interest of conserving space, and since the shapes of the survival functions are similar to that of the survival function for all data, we chose not to present the graphs. Table 3 is sufficient to illustrate the differences across regions and industries.

¹⁶Due to space limitations, we do not present the estimates. Estimates for OECD and non-OECD are available upon request.

reclassification in them. As shown in Table 3 under the modified censoring scheme the two industries have a survival experience very similar to the other industries.

The higher than average incidence of product code changes might reflect more innovation and shorter product cycles; a dynamism which results in numerous TS codes changes. If so, then the modified censoring scheme does a better job capturing the actual dynamics as the code changes probably reflect true entry and exit (i.e., births and deaths). On the other hand, it might reflect other factors that are not associated with economically meaningful entry and exit. Code changes might simply be a way to mitigate the impact of GATT negotiated tariff reductions or expand the scope of the Multifibre Agreement. In this case, the benchmark censoring scheme is more appropriate and the two industries are outliers. A resolution of the issue is beyond the scope of this paper but certainly merits further analysis.

4.2. 10-digit HS data

Are dynamics uncovered using TS data unique to the earlier 1972–88 period? Even if the earlier period is characterized by short-lived trade relationships the more recent period may not be. Has the world changed in any way since the 1972–88 period?

We use the 10-digit HS level data to investigate these questions. Reassuringly, none of the results change substantially in the more recent period. Given they are qualitatively the same we refrain from a lengthy discussion of the 1989–2001 results and only provide a short summary. First, the main result—duration of trade is fleeting—still holds. The median duration at the 10-digit HS level is only four years, just as was the case for the 7-digit TS data.

Second, as depicted in Figure 2 the survival and hazard functions for HS data demonstrate negative duration dependence. The probability of failing in the first year is 34 percent. The hazard rate between year one and year five is an additional 30 percent. The hazard rate for the remaining eight years is just 7 percent.

Third, censoring due to product code changes again significantly affects the estimated survival. As was the case with TS data once we implement the modified censoring approach the median duration is just two years.

Fourth, the estimated survival functions across regions and industries, particularly in terms of which regions or industries perform better are very similar to that found using TS data.¹⁷

5. Robustness of Results

At the product level trade relationships are quite short: the estimated median survival is between two and four years. We now examine whether the findings are robust to how we measure trade relationships and spells.

5.1. Aggregation concerns — SITC level analysis

We begin by exploring whether the findings are possibly due to the highly disaggregated nature of the data. Thirty different ball-bearing codes may represent an overly fine parsing of the data leading to the observation of excessive entry and exit. If so, a trade relationship might be better measured using SITC industry classifications.

We calculated spells of service using SITC industry (revision 2) definitions ranging from the 5-digit to the 1-digit level. From Table 2 we see that for both periods investigated the total number of products (or industries) declines from almost 23,000 at the product level to about 1,700 at the 5-digit level industry level to 10 at the 1-digit industry level. For TS data the number of observed spells of service declines from almost 700,000 to about 157,000 at the 5-digit level to some 2,500 at the 1-digit level. For HS data the number of observed spells of trade declines from more than 900,000 to about 156,000 at the 5-digit level to about 2,200 at the 1-digit level.

¹⁷Results are available upon request.

The extent of aggregation is extraordinary. There are 93% fewer product categories at the 5-digit SITC level than at either the TS or HS level. Thinking in terms of trade relationships (a country-product pair), on average when there is an aggregation in a trading relationship, some 12–14 spells are aggregated.¹⁸ With such extensive aggregation, we expected to observe a significant increase in the estimated probability of survival. Interestingly, the results indicate otherwise.

Results for the SITC classifications are given in Table 2—the median observed spell length is two or fewer years in the 1- through the 4-digit SITC data for both 1972–88 and 1989–2001 periods. The Kaplan-Meier estimates are depicted in Figure 3. The industry level analysis not only confirms our product level findings but also provides an important new insight into the complicated censoring issue.

First, within the SITC classification, aggregation works exactly as expected. For instance as shown in Table 2 and Figure 3 higher levels of aggregation are associated with higher survival rates. The survival function for the 4-digit SITC lies above the 5-digit SITC, the 3-digit lies above the 4-digit SITC, and so on. Although aggregation creates longer spells the estimate survival probabilities remain remarkably low. Only about one-third of 5-digit SITC observations survive four years. There remains a large number of short spells of service until the data are very highly aggregated. The median survival time is two years for the 3-digit, 4-digit, and 5-digit SITC data for the 1972–88 period. Aggregation has a more appreciable impact for the 1989–2001 period but median survival time remains short.

Second, in comparison with the median survival times for either comparable 7-digit TS or 10-digit HS data the brevity of duration times for SITC data is surprising. For TS data, for instance, despite the higher degree of aggregation the 2- through 5-digit SITC data all have a

¹⁸Not all country-product pairs are aggregated when we move to industry level analysis. For instance, some countries supply only one product within a SITC industry. In these cases the country-product pair uniquely maps into country-industry pair.

shorter median survival time than the comparable product level data. Only the 1-digit SITC data have a higher survival function than the comparable product level data. For HS data, the 4- and 5-digit SITC data have a shorter median duration than the comparable product level data.

In other words, aggregation from products to industries lowers survival. Taken at face value, it is an odd result. There is a logical explanation. The SITC classification system (revision 2) is unchanged throughout the sample. As a result only the first type of censoring (which is driven by the beginning and end of the sample) is present in the SITC analysis. It may be more appropriate, therefore, to compare results based on the SITC classification with modified censoring data. As shown in Figure 3, when we make that comparison we get the expected result: survival experience for product level data is lower than for the 5-digit industry level data and aggregation increases estimated survival time.

SITC data make it clear both approaches toward handling censoring in product level data are imperfect. In an ideal world we would be able to truly identify all births, deaths, and censored product changes. Since we cannot do so we are left with two imperfect measures: (i) benchmark product level data, which overstate the true survival experience (since all product code changes are censored) and (ii) modified censoring data, which understate the true survival experience (since all product code changes are interpreted as births and deaths).

Given survival at the product level must be shorter than at the industry level, the estimates in Figure 3 indicate the true product level survival lies between the two measures. The SITC results imply the true median survival experience at the product level to be two to four years and is probably closer to two years. In either case, the evidence is clear: duration is very brief.

The results presented in this section cannot be overstated. It is very surprising that trade as measured at the 5- through 2-digit SITC level is so brief. After all, the extent of aggregation from the disaggregated product level data to the 5-digit SITC level is extraordinary. There are about 13 times as many product codes at the product level and 4–6 times as many spells of

service than at the 5-digit level. Yet, the estimated survival rates for the 5- through 3-digit SITC level data is not very different than the benchmark product level data. Except for the highly aggregated SITC level data there is a multitude of short spells of service however the data are examined, whether looking at all spells or looking across regions and industries. Regardless of how aggregated the data are, trade is surprisingly fleeting.

5.2. Multiple spells

About 30% of trade relationships experience multiple spells of service. In the above analysis we have assumed duration is independent of the spell number. We now investigate whether this assumption biases the findings.

We considered several alternative approaches toward the issue of multiple spells. First, we simply limit the analysis to relationships with a single spell only. In Figure 1 we plotted the distribution for trade relationships with a single spell. As noted earlier, there is very little difference between distributions for single spell and benchmark product data, especially for TS data. The Kaplan-Meier estimated survival function for single spell data is depicted in Figure 4. The survival function for single spell data has a similar pattern as benchmark data: high hazard in the first few years followed by a leveling off of the survival function. In spite of the similar pattern, single spell data do exhibit significantly higher survival; for TS data the survival for single spell relationships is about 21 percentage points higher than for benchmark data in year 17.

The similarity of spell length distributions (Figure 1) sharply contrasts with the differences in Kaplan-Meier estimates. The censoring approach taken in benchmark data plays a role. A greater fraction of observations are censored in single spell data which shifts the Kaplan-Meier estimates up. 74 percent survive one year in single spell data as compared with 67 percent in benchmark data. Given that the SITC results indicate that benchmark data overstate the amount

of censoring and that the modified censoring approach may yield a more appropriate measure of the true survival experience, we re-estimate single spell data using the modified censoring approach and present the Kaplan-Meier estimates in Figure 4. The median survival time is now 3 years as compared with 2 years when all observations are included. While single spell data have a higher survival experience, the effect is not as great as in the benchmark censoring approach: about 7 percentage points higher than the benchmark in year 17 (for TS data).

The second approach we explored was to limit the analysis to first spells—relationships with just one spell and the first spell of relationships with multiple spells. Since the results are generally similar to the single spell results they will not be explicitly discussed.

5.3. Measurement error

Another way we addressed the issue of multiple spells was to consider the possibility some of the reported multiple spells are due to a measurement error. Specifically, if the time between spells is short, it may be that the gap is mis-measured and interpreting the initial spell as “failing” is inappropriate. It may be more appropriate to interpret the two spells as one longer spell. To allow for such misreporting, we assume a one-year gap between spells is an error, merge individual spells, and adjust spell length accordingly. Gaps of two or more years are assumed to be accurate and no merging is done.

The spell length distribution for gap-adjusted data is depicted in Figure 1. The adjustment shifts the distribution but short spells remain the norm. In Table 2 we report the summary statistics for the gap-adjusted data. In comparison with benchmark data, the average spell length is less than a year longer. The 1-, 4- and, 12-year survival rates in gap-adjusted data are about 7–9 percentage points higher than in benchmark data.

The Kaplan-Meier estimated survival function for gap-adjusted data is depicted in Figure 4. As expected, the survival function exhibits less early failure and more failure in later times. The

hazard rate for the last twelve years is just 12 percent in benchmark data but 14 percent in gap-adjusted data. As was the case with single spell and first spell alternatives, using the modified censoring approach reduces the difference between gap-adjusted and benchmark data.

The results suggest the independence assumption likely leads to an underestimate of duration although the magnitude of the bias is fairly small for most scenarios. If one believes the modified censoring approach best captures the true survival experience, the independence assumption does not have a significant impact under any scenario.

5.4. Trade weighted analysis

The next issue we explore involves putting more weight on the higher valued trade relationships. Benchmark data—and all of the preceding analysis—is unweighted. Small and large trade value relationships receive equal weight. If short spells involve small values and long spells involve large values, an unweighted distribution might overstate the brevity of spells. We compute a weighted distribution where each observation is weighed by the value of trade in the first year of the relationship. The alternative formulation de-emphasizes low-value spells and gives more importance to high-value spells.

In Figure 1 we plot the spell length distribution under the weighted alternative. There is a significant shift in the distribution, far greater than any of the other alternatives. In benchmark TS and HS data more than half the spells are observed for just a single year. In the weighted analysis only about 15 (25) percent of spells are observed for a single year for TS (HS) data.

Lower value trade relationships tend to be short lived. Even under the weighted scheme there remain a surfeit of short spells—more than 50 percent of spells are observed for four or fewer years for both TS and HS data.

In Figure 4 we present the Kaplan-Meier estimated survival function for the weighted TS and HS data. The impact on survival is far greater than under any of the alternatives previously

considered. The impact on duration is more significant than aggregating to the 1-digit industry level or the restriction to single spell relationships.

6. Concluding Comments

The findings indicate countries tend to trade over short intervals of time. We have shown it is a surprisingly robust result. Trade relationships remain short when we change the way relationships are measured, controlling in different ways for multiple spells and censoring issues. There is a large number of short trade relationships in every industry and virtually every country the US trades with. Trade is of short duration whether one looks at highly disaggregated product level data or moderately aggregated industry level data. Trade relationships become long lasting only once the data are aggregated to the 1-digit SITC level, representing 10 major industries. We do not believe that the single-digit level of aggregation is sufficient to capture all differences between products traded in world markets.

Our goal was to investigate the nature of duration of trade. The question of why trade relationships are so short is beyond the scope of this paper and is a matter of our ongoing research agenda. It does seem to us, however, that answers go beyond standard trade theory. Standard trade models, in particular, seem ill-equipped to analyze duration of trade.

The Heckscher-Ohlin factor endowments model and the Ricardian comparative advantage model would suggest trade to be very persistent. While factor endowments and comparative advantage can and do change over time, it is difficult to believe they change as rapidly as suggested by the results presented. Even more, the great extent of multiple spells of service would suggest that changes in factor endowments and comparative advantage go back and forth. Both models would suggest there may be a range of products over which countries would flip back and forth between exporting and importing, but it would be a very limited range of products. About

30 percent of trade relationships experience multiple spells of service. One is hard pressed to explain such dynamic behavior with Heckscher-Ohlin and Ricardian models. Nor is it clear the gravity equation can explain the results. It would suggest once countries start trading, the relationships would be very persistent and there would be little turnover.

We intend to explore several possible explanations in the near future. The product cycle model seems a natural candidate. It claims countries trade dominance in export markets based on their development stage, with more developed countries exporting more advanced products first and developing countries replacing them as they adopt technology and take advantage of their low labor costs. According to the quality ladder variant of the model developed countries may restore their dominance by supplying a higher quality variety of the product. The product cycle model could be able to explain both the extent of turnover in international trade and the fact that countries may return to the same market they had previously exited.

One difficulty for the product cycle model would be the observation that once developing countries enter a market, developed countries do not leave and still experience systematically longer trade relationships. It is possible by the time developing countries enter, the market is very competitive and it is difficult to establish a long-term presence. Developed countries may also be able to establish a name for themselves and maintain their longer presence through branding. The period of time we investigate could be too short to observe developed countries abandoning a sufficient number of products to identify the full extent of the product cycle, but it should be possible to identify parts of it.

Another potential explanation we intend to pursue is a model of trade and search costs. The idea is that US firms search for appropriate partners in the world market. Such searching is costly and uncertain. Depending on the quality of the search, US buyers may immediately find good partners with whom they will do business with over long periods of time. Or, they could go through several unsuccessful short relationships before they find a good match.

Finally we note that the issues investigated in this paper are related but separate from the literature on plant- and firm-level dynamics as reviewed by Tybout (2003). While we use import data, the majority of research on plant- and firm-level dynamics uses export data. The difference is not innocuous. We observe a country's exports of a product to a single market, while they observe plant or firm exports to the whole world market. By using import data we are not focusing on the question of why firms or countries export, but rather the length of their exports to a specific market, albeit the largest market in the world.

We observe trade at the country level for a multitude of products, while this related literature explores the details of plant- and firm-level data sets. The extent of dynamics of trade uncovered here could be interpreted as a lower bound on the dynamics (and an upper bound on survival) of trade at the plant- and firm-level. If the median duration of trade relationships at the country level is between two and four years, it must mean the median duration of trade at the firm level can be at most two to four years and is very likely even smaller. Considering the median duration of trade does not change significantly as data are aggregated, the results are even more surprising in the light of their implications for firm-level dynamics.

Such a conclusion is reached ignoring the last important difference between our work and the plant- and firm-level dynamics literature: product coverage. Our data are recorded at the product level, while the firm-level data are obviously recorded at the firm- or plant- level. An exporting firm could be producing and exporting a single product, or several, or even change its product mix over time while continuing to export. Our data identify exports of a particular product, regardless of which firm(s) produced it and regardless of whether it is the first product exported by the firm or one in a series of new products the firm produces as it benefits from its export experience. Whether such product switching has an impact on the firm's performance and whether product switching is a consequence of its participation in world markets is beyond the scope of our data. In the light of the aggregation results, we speculate product switching

plays a small role for if it did duration would be longer at the industry level. While the plant- and firm-level dynamics literature can draw conclusions about the importance of sunk entry and exit costs, we can do so only cautiously and with a number of caveats. This is a matter of our ongoing research.

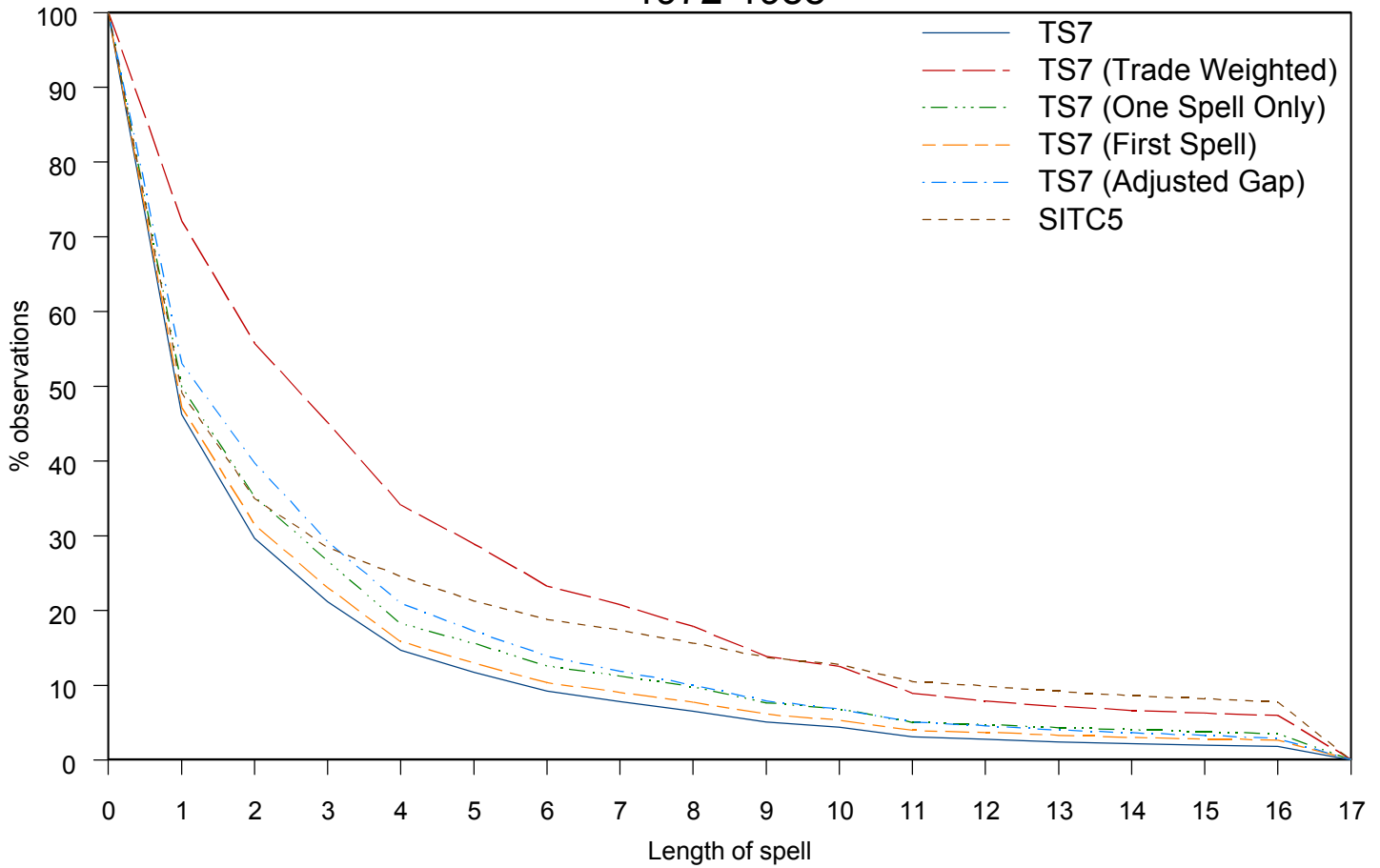
References

- [1] Broda, C. and D.E. Weinstein (2004), “Globalization and the Gains from Variety,” NBER Working Paper 10314.
- [2] Dollar, David (1990), “Technological innovation, capital mobility, and the product cycle in North-South trade,” *American Economic Review* 76(1):177–90.
- [3] Feenstra, R.C. (1996), “US Imports, 1972–1994: Data and Concordances,” NBER Working Paper 5515.
- [4] Feenstra, R.C. and A.K. Rose (2000), “Putting Things in Order: Trade Dynamics and Product Cycles,” *Review of Economics and Statistics*, 82:369–82.
- [5] Feenstra, R.C., J. Romalis and P.K. Schott (2002), “US Imports, Exports, and Tariff Data, 1989–2001,” NBER Working Paper 9387.
- [6] Grossman, G. and E. Helpman (1991), *Innovation and Growth in the Global Economy*, Cambridge: MIT Press.
- [7] Haveman, J. and D. Hummels (2004), “Alternative Hypotheses and the Volume of Trade: The Gravity Equation and the Extent of Specialization,” *Canadian Journal of Economics*, 37:199–218.
- [8] Jovanovic, B. (1982) “Selection and the evolution of industry,” *Econometrica*, 50:649–70.
- [9] Krugman, Paul R. (1979) “Increasing returns, monopolistic competition, and international trade,” *Journal of International Economics*, 9:469–79.
- [10] Pakes Ariel and R. Ericson (1998), “Empirical Implications of Alternative Models of Firms Dynamics,” *Journal of Economic Theory*, 79:1–45.
- [11] Proudman, J. and S. Redding (1998), “Persistence and Mobility of International Trade,” Chapter 2 in *Openness and Growth*, Bank of England, 1998.
- [12] Redding, S. (2001), “Specialization Dynamics,” *Journal of International Economics*, 58:299–334.
- [13] Schott, Peter K. (2004), “Across-Product versus Within-Product Specialization in International Trade”, *Quarterly Journal of Economics*, 119(2):647–678.
- [14] Tybout, James R. (2003), “Plant- and Firm-Level Evidence on ‘New’ Trade Theories,” in *Handbook of International Trade*, E. Kwan Choi and James Harrigan (Eds.), Malden, MA: Blackwell Publishing.

- [15]Vernon, R. (1966) “International Trade and International Investment in the Product Cycle,”
Quarterly Journal of Economics, vol. 80, no. 2, May.
- [16]Xiang, C. (2004), “New goods and the relative demand for skilled labor,” *Review of
Economics and Statistics*, forthcoming.

Figure 1 - Distribution of Spell Lengths

1972-1988



1989-2001

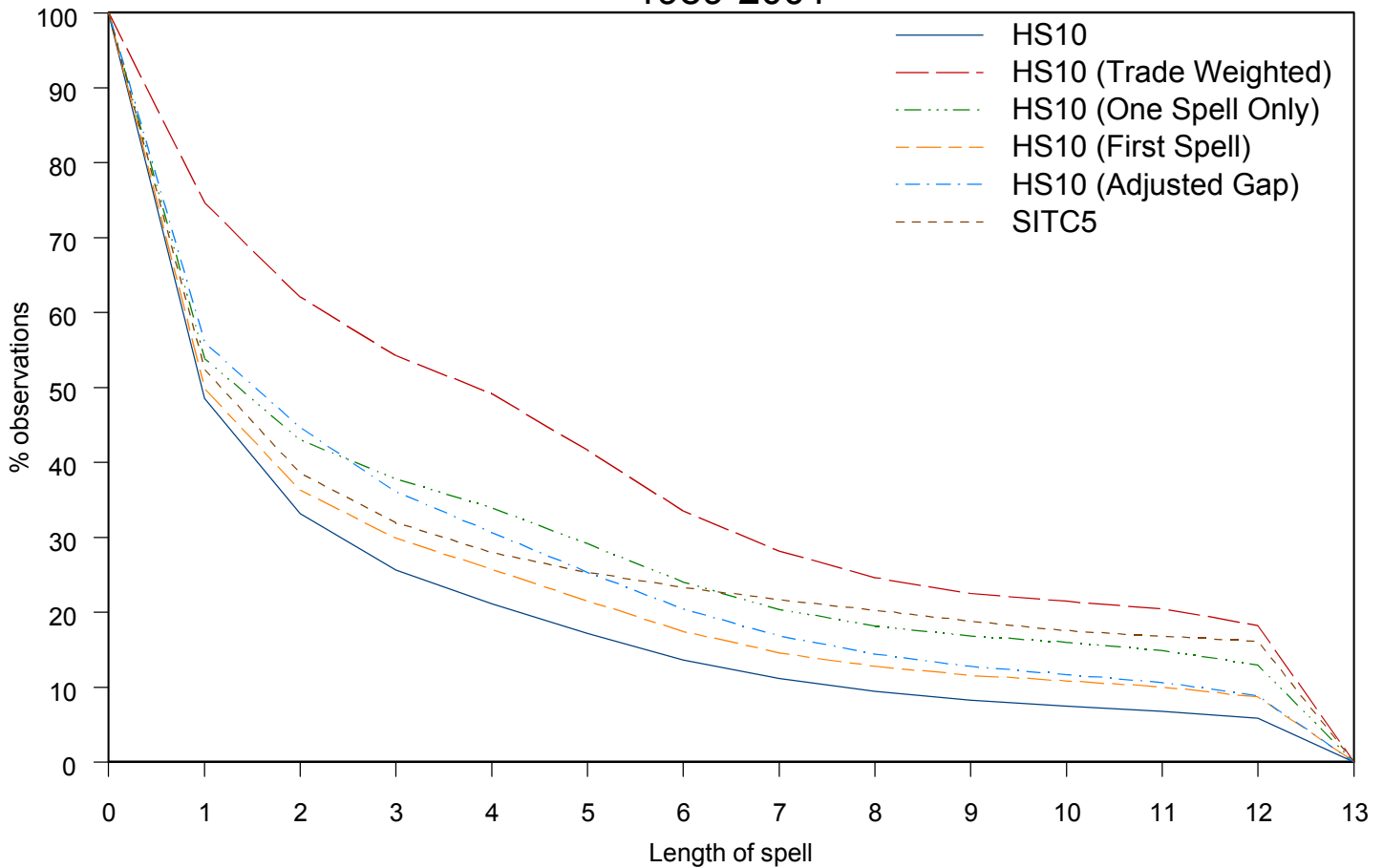
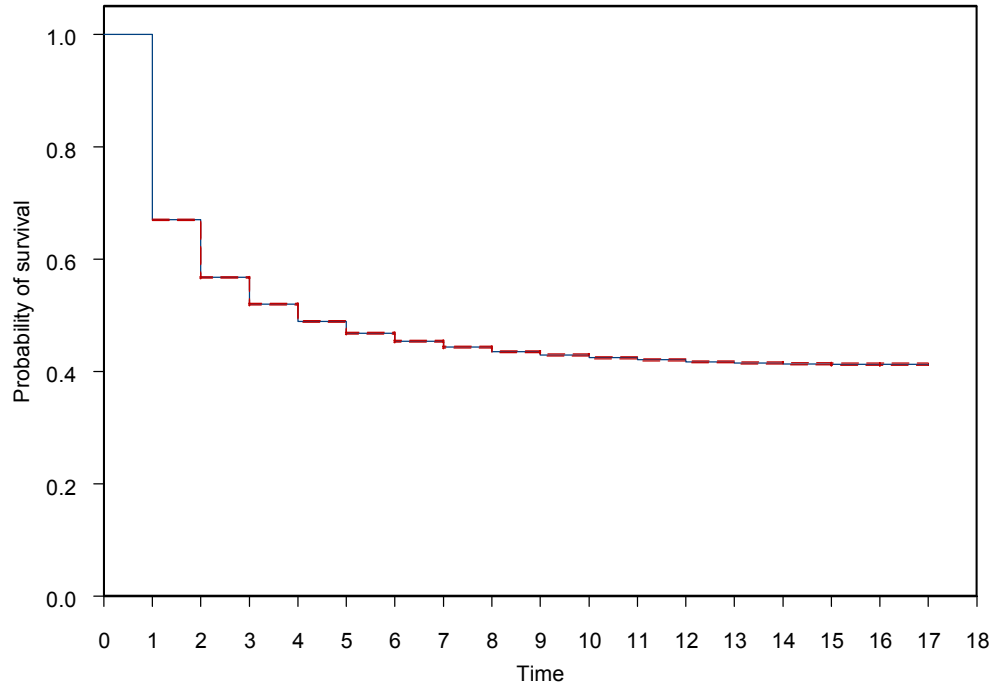
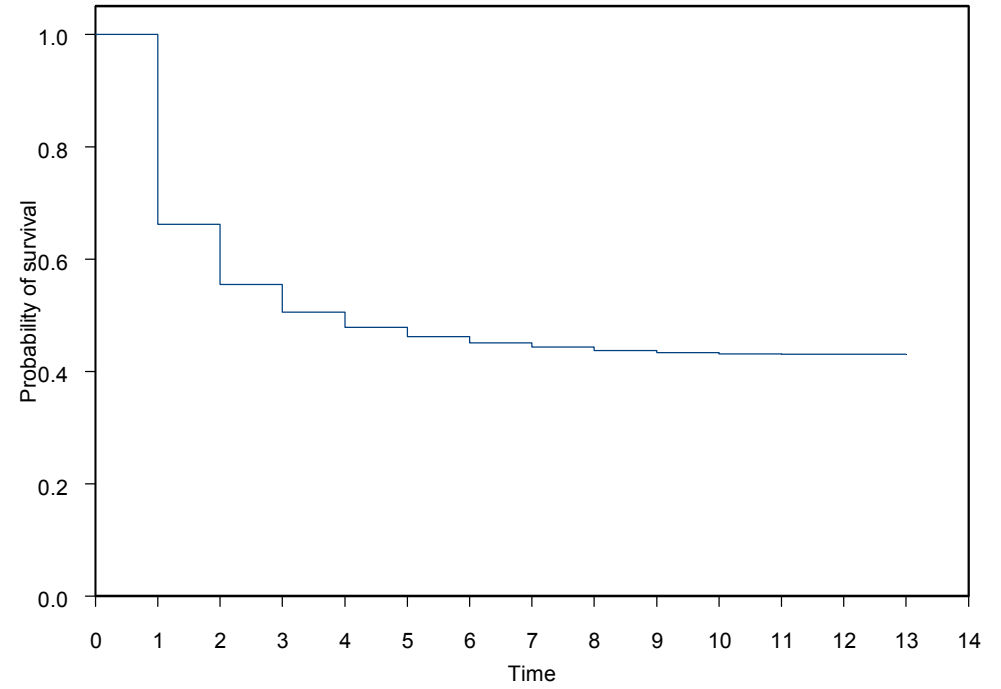


Figure 2 - Overall Survival and Hazard Functions

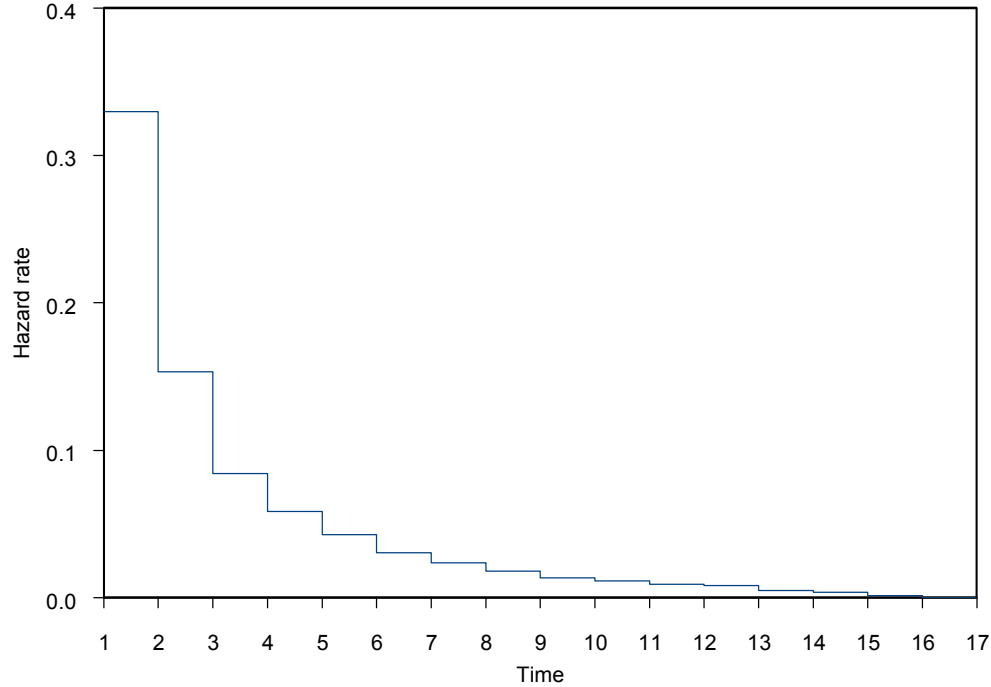
1972-1988



1989-2001



1972-1988



1989-2001

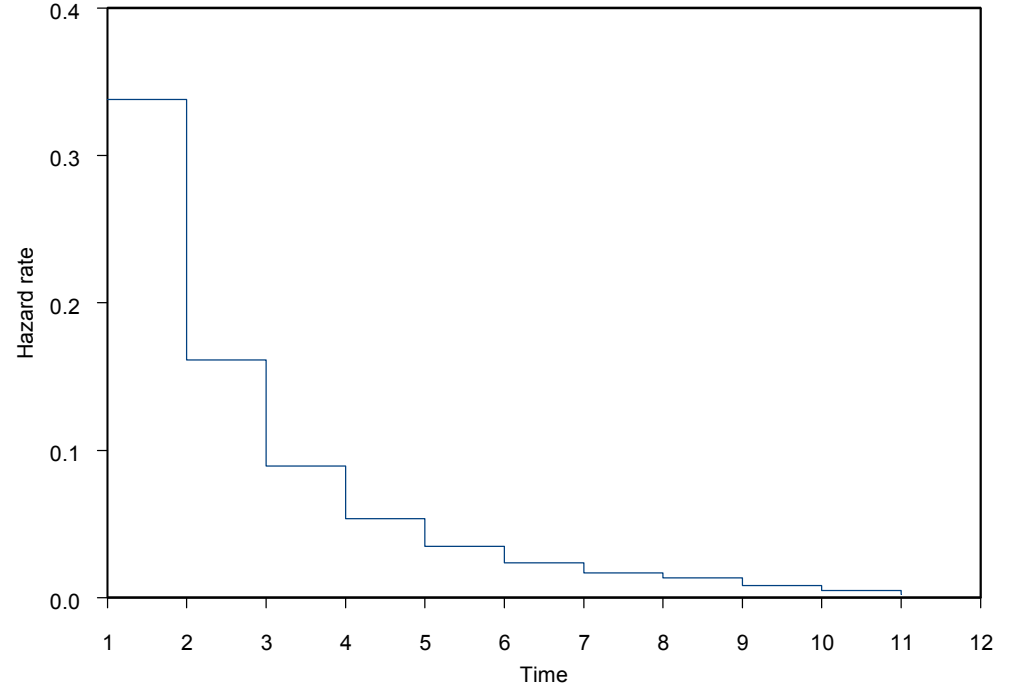
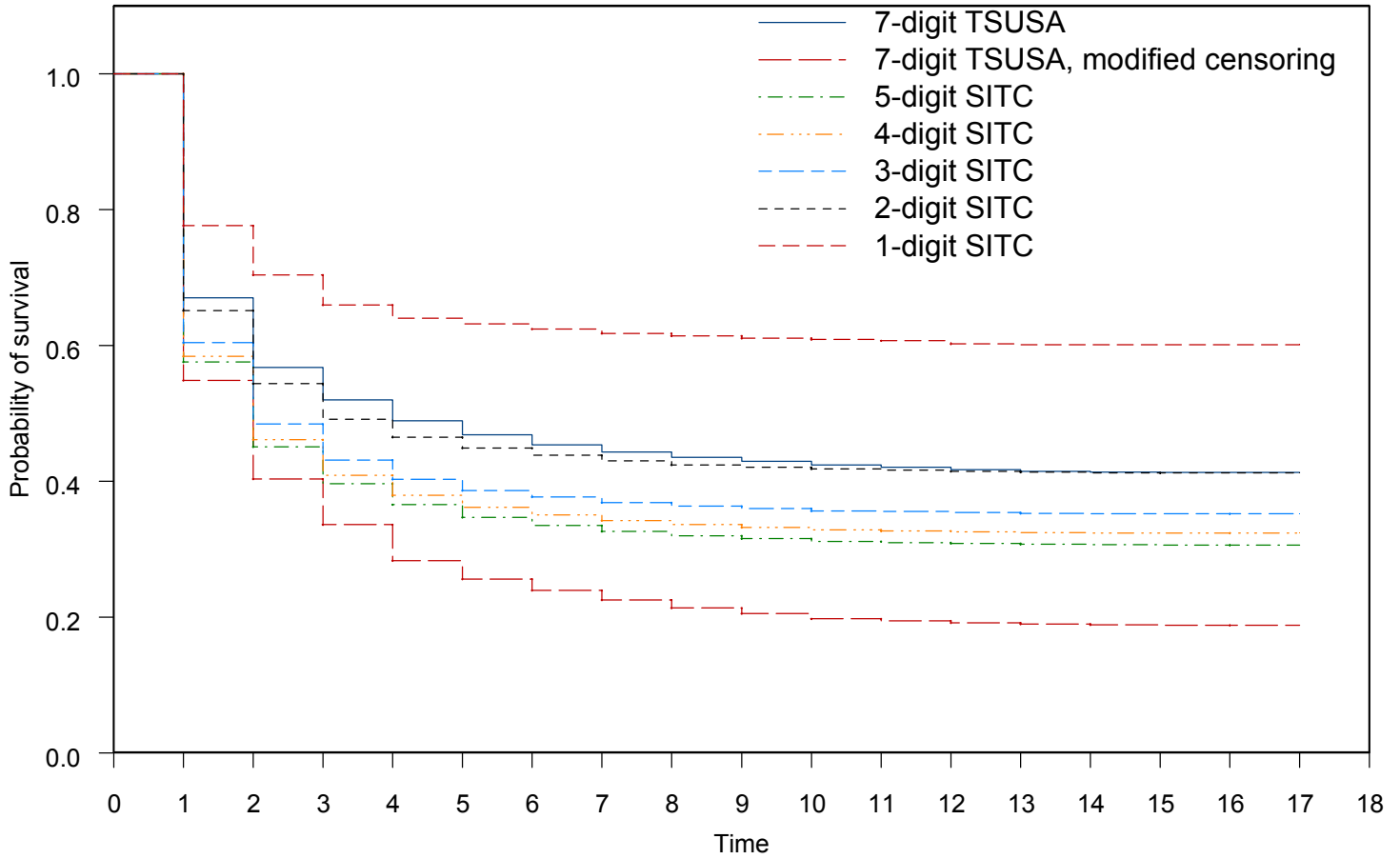


Figure 3 - Product and Industry Level Data

1972-1988



1989-2001

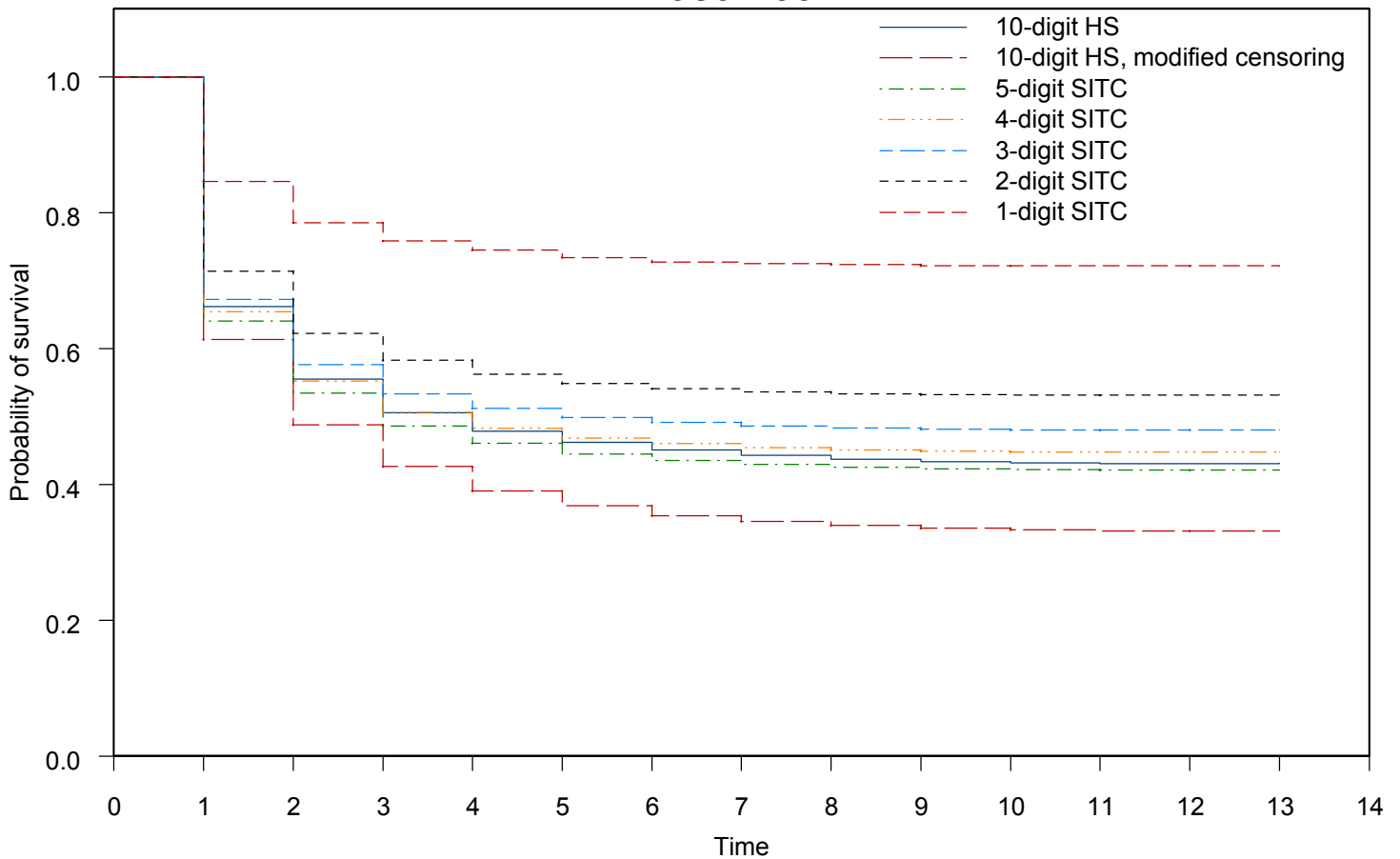
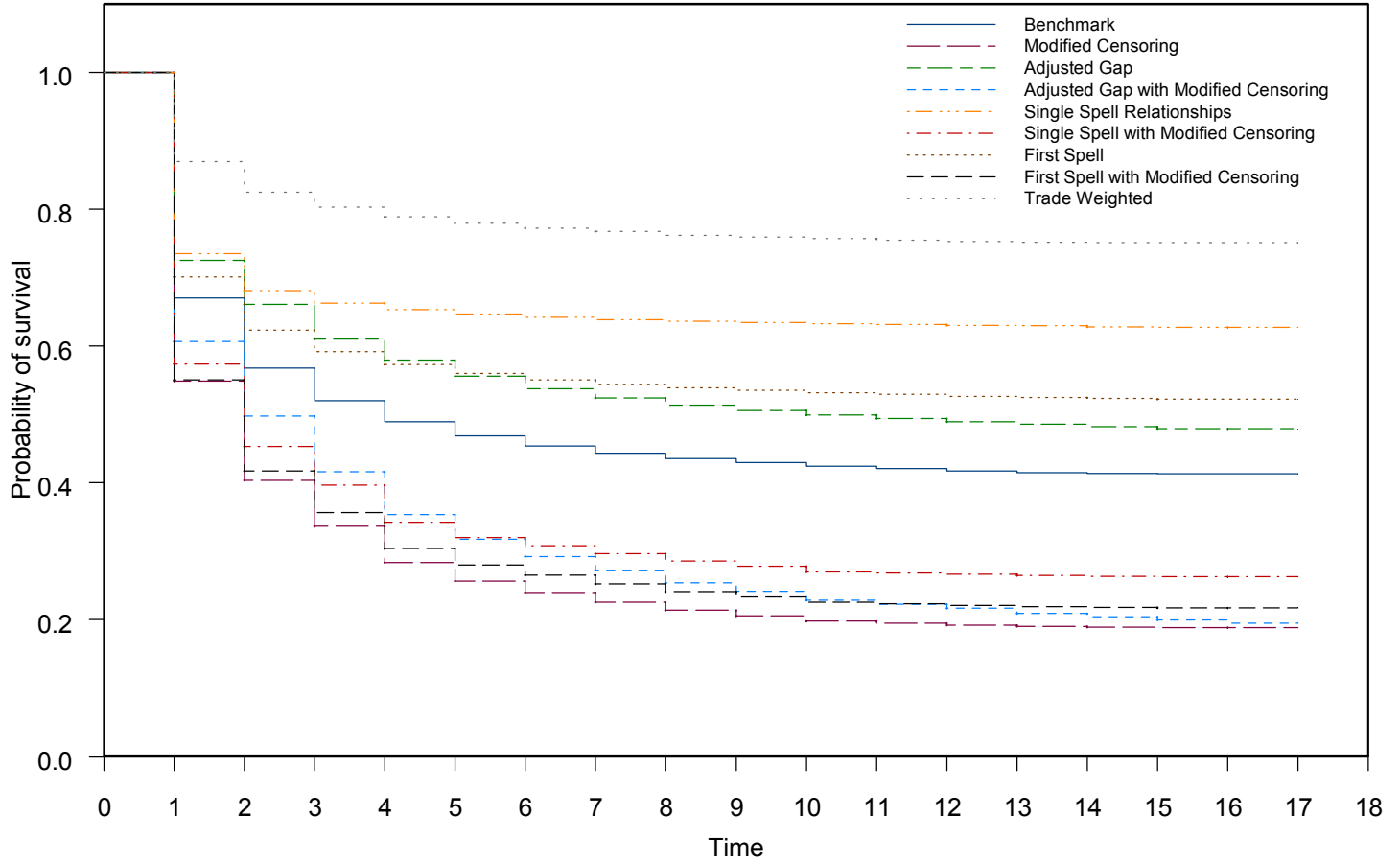


Figure 4 - Alternative Treatments, Product Level Data

1972-1988



1989-2001

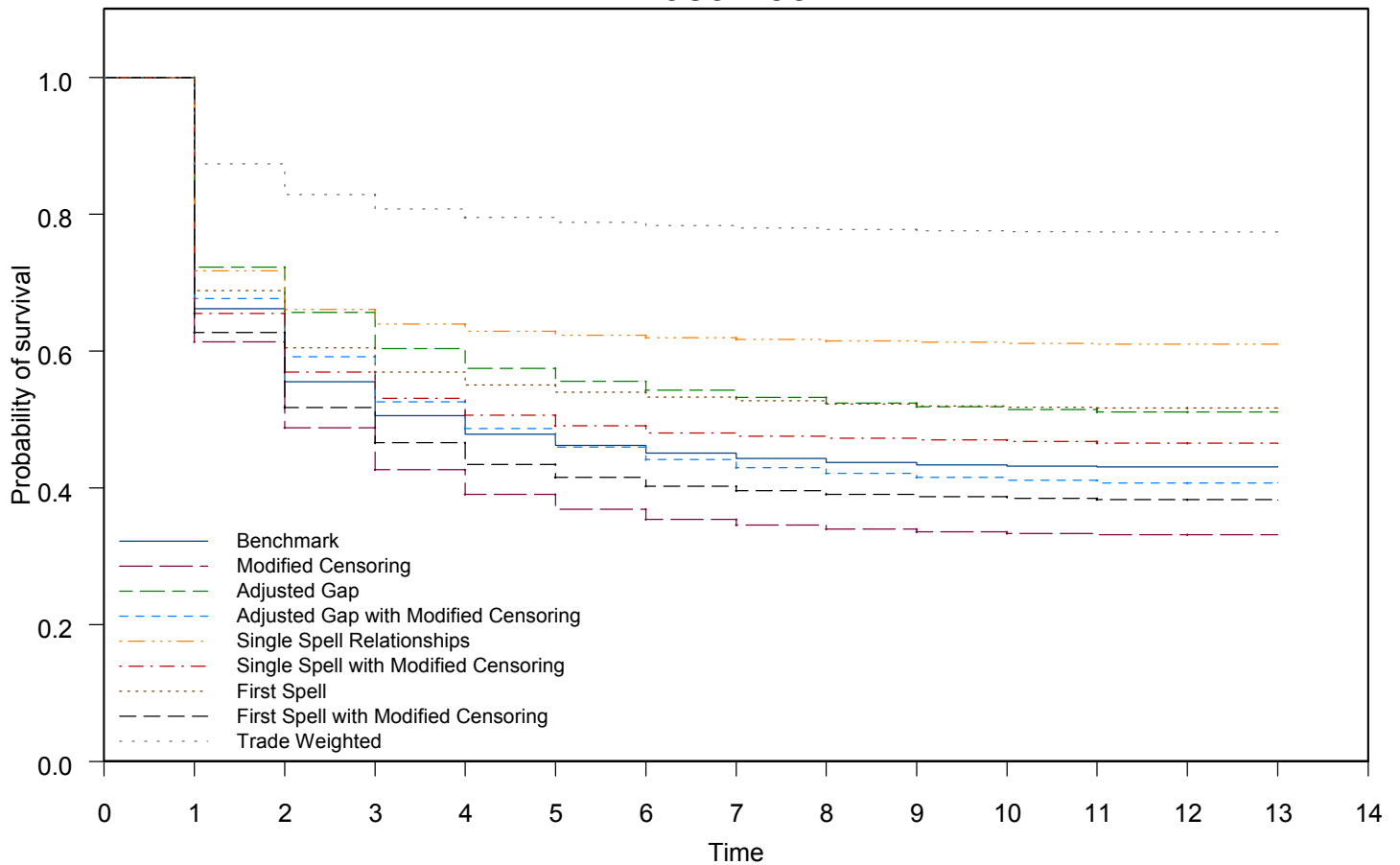


Table 1 - Example of data

CORN (1312000)

	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	No. Years In	# Spells
Canada	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	17	1
Portugal	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	17	1
Mexico			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	15	1
South Korea					X	X	X	X	X	X	X	X	X	X	X	X	X	13	1
Peru													X	X	X	X	X	5	1
Kenya																X		1	1
Honduras											X							1	1
St Kitts Nev							X											1	1
Hong Kong													X					1	1
Philippines	X																	1	1
Belgium-Lux.	X																	1	1
Greece												X						1	1
Spain	X																	1	1
East Germany								X										1	1
Venezuela	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	16	2
Ecuador		X	X					X	X	X	X	X	X	X	X	X	X	12	2
Japan											X	X	X		X			4	2
Argentina			X											X	X			3	2
Guatemala									X				X					2	2
Dominican Rp			X									X						2	2
India							X										X	2	2
China								X				X						2	2
Colombia		X	X	X		X				X	X	X	X	X	X	X	X	12	3
Italy	X				X	X			X	X	X	X	X	X	X	X	X	12	3
United Kingdom							X	X			X	X					X	5	3
Thailand							X			X						X	X	4	3
Taiwan			X								X						X	3	3
Germany				X											X		X	3	3
El Salvador													X		X		X	3	3
France			X								X	X		X			X	5	4
Netherlands	X	X	X	X	X			X				X		X		X		9	5
Brazil	X			X						X	X		X	X			X	7	5

Table 2 - Summary Statistics

1972-1988	Observed		Estimated			Total Number of Spells	Total Number of Product Codes
	Spell Length (years)		KM Survival Rate				
	Mean	Median	1 year	4 year	12 year		
<i>Benchmark data</i>							
TS7	2.7	1	0.67	0.49	0.42	693,963	22,950
<i>Industry-level aggregation</i>							
SITC5	3.9	1	0.58	0.37	0.31	157,441	1,682
SITC4	4.2	2	0.58	0.38	0.33	98,035	827
SITC3	4.7	2	0.60	0.40	0.35	43,480	253
SITC2	5.5	2	0.65	0.46	0.41	15,257	69
SITC1	8.4	5	0.78	0.64	0.60	2,445	10
<i>TS7 Alternatives</i>							
Modified Censoring	2.7	1	0.55	0.28	0.19	693,963	22,950
Trade Weighted	4.7	3	0.87	0.79	0.75	693,963	22,950
First Spell	2.9	1	0.70	0.57	0.53	495,763	22,950
One Spell Only	3.2	1	0.74	0.65	0.63	365,491	22,950
Gap Adjusted	3.3	2	0.73	0.58	0.49	593,450	22,950
<i>1989-2001</i>							
<i>Benchmark data</i>							
HS10	3.1	1	0.66	0.48	0.43	918,236	22,782
<i>Industry-level aggregation</i>							
SITC5	4.1	2	0.64	0.46	0.42	156,110	1,664
SITC4	4.4	2	0.65	0.48	0.45	92,566	768
SITC3	4.7	2	0.67	0.51	0.48	40,068	237
SITC2	5.3	2	0.71	0.56	0.53	14,373	68
SITC1	7.8	10	0.85	0.73	0.72	2,292	10
<i>HS10 Alternatives</i>							
Modified Censoring	3.1	1	0.61	0.39	0.33	918,236	22,782
Trade Weighted	5.5	4	0.87	0.80	0.77	918,236	22,782
First Spell	3.5	1	0.69	0.55	0.52	620,177	22,782
One Spell Only	4.2	2	0.72	0.63	0.61	415,851	22,782
Gap Adjusted	3.9	1	0.72	0.58	0.51	768,048	22,782

Table 3 - Estimated Kaplan-Meier survival rate

	TS7-benchmark			TS7-mod. censoring		
	1 year	4 year	12 year	1 year	4 year	12 year
Total	0.67	0.49	0.42	0.55	0.28	0.19
Regions						
North America	0.78	0.64	0.57	0.68	0.43	0.33
Asia	0.71	0.55	0.48	0.58	0.30	0.20
Western Europe	0.71	0.53	0.46	0.59	0.32	0.23
South America	0.62	0.41	0.32	0.49	0.22	0.13
Eastern Europe	0.58	0.36	0.27	0.45	0.18	0.09
Oceania	0.60	0.37	0.28	0.49	0.22	0.13
Africa	0.52	0.29	0.19	0.39	0.15	0.08
Middle East	0.57	0.36	0.29	0.44	0.19	0.12
SITC (rev 2) 1-digit industries						
0-Food	0.58	0.36	0.29	0.52	0.28	0.20
1-Beverages/Tobacco	0.65	0.47	0.41	0.54	0.28	0.20
2-Crude Materials	0.54	0.31	0.24	0.50	0.25	0.17
3-Mineral Fuels	0.58	0.38	0.28	0.53	0.27	0.19
4-Animal and Vegetable Oils	0.51	0.29	0.21	0.48	0.24	0.15
5-Chemicals	0.62	0.41	0.35	0.56	0.30	0.22
6-Manufactured Materials	0.63	0.44	0.37	0.54	0.29	0.20
7-Machinery	0.72	0.56	0.52	0.58	0.33	0.21
8-Miscellaneous Manufactures	0.73	0.58	0.50	0.55	0.26	0.17