IFN Working Paper No. 829, 2010

The Duration of Trade Revisited: Continuous-Time vs. Discrete-Time Hazards

Wolfgang Hess and Maria Persson

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

The Duration of Trade Revisited

Continuous-Time vs. Discrete-Time Hazards*

Wolfgang Hess† Maria Persson‡

April 9, 2010

Abstract

The recent literature on the duration of trade has predominantly analyzed the determinants of trade flow durations using Cox proportional hazards models. The purpose of this paper is to show why it is inappropriate to analyze the duration of trade with continuous-time models such as the Cox model, and to propose alternative discrete-time models which are more suitable for estimation. Briefly, the Cox model has three major drawbacks when applied to large trade data sets. First, it faces problems in the presence of many tied duration times, leading to biased coefficient estimates and standard errors. Second, it is difficult to properly control for unobserved heterogeneity, which can result in spurious duration dependence and parameter bias. Third, the Cox model imposes the restrictive and empirically questionable assumption of proportional hazards. By contrast, with discrete-time models there is no problem handling ties; unobserved heterogeneity can be controlled for without difficulty; and the restrictive proportional hazards assumption can easily be bypassed. By replicating an influential study by Besedeš and Prusa from 2006, but employing discrete-time models as well as the original Cox model, we find empirical support for each of these arguments against the Cox model. Moreover, when comparing estimation results obtained from a Cox model and our preferred discrete-time specification, we find significant differences in both the predicted hazard rates and the estimated effects of explanatory variables on the hazard. In other words, the choice between models affects the conclusions that can be drawn.

Keywords: Duration of Trade; Continuous-Time versus Discrete-Time Hazard Models; Proportional Hazards; Unobserved Heterogeneity.

JEL Classification: C41; F10; F14.

---

* This paper draws heavily on Working Paper 2009:12, Department of Economics, Lund University. Previous versions of that paper were presented at the 2009 ETSG Conference in Rome and at seminars at Lund University and IFN in Stockholm. We would like to thank conference and seminar participants, and in particular Magnus Henrekson and Fredrik Sjöholm, for many valuable comments and suggestions. We are also very grateful to Tibor Besedeš and Thomas Prusa for generously granting us access to their data. Financial support from the Jan Wallander and Tom Hedelius Foundation under research grant numbers H2007-0281:1, P2006-0131:1, P2009-0189:1 and W2009-0352:1, and from the Marianne and Marcus Wallenberg Foundation is gratefully acknowledged.

† Corresponding author: Department of Economics, Lund University, P.O. Box 7082, S-220 07 Lund, Sweden. E-mail: `Wolfgang.Hess@nek.lu.se`; Tel.: +46 46 222 8653.

‡ Department of Economics, Lund University and Research Institute of Industrial Economics (IFN). E-mail: `Maria.Persson@nek.lu.se`.

1 Introduction

In recent years, a new literature focusing on the duration of international trade has emerged. Based on the surprising finding in Besedeš and Prusa (2006a) that US import flows have a remarkably short duration, the question asked is: "which factors determine how long international trade relationships last?" From a policy-oriented point of view this is indeed an important question to ask. Trade will not grow very much if new products stop being exported after only a few years. Therefore, to better understand which factors may help countries increase their trade, and thereby potentially improve economic development, it is important to learn more about what determines the duration of trade flows.

The first paper to offer an answer to this question is Besedeš and Prusa (2006b). In that article, the authors estimate a Cox proportional hazards model, as originally proposed by Cox (1972), and conclude among other things that differentiated products have lower hazard rates than homogeneous goods, and that within each product type, the larger the value of the initial trade flow, the longer the duration. Following the example set by Besedeš and Prusa (2006b), other authors have subsequently used similar Cox approaches to analyze what determines the duration of trade. Papers in this tradition include Besedeš (2008), Brenton, Pierola and von Uexküll (2009), Nitsch (2009), and Fugazza and Molina (2009).

While the use of the Cox model may be a convenient way to analyze what determines the duration of trade, there are some concerns with this approach from an econometric point of view. Specifically, even though trade takes place in continuous time, data on the duration of trade relationships is usually grouped into yearly intervals. Yet, the Cox model is designed to deal with continuous duration times. The purpose of this paper is mainly to show why it is inappropriate to analyze the duration of trade with continuous-time models such as the Cox model, and to propose alternative discrete-time models which are more suitable for estimation. By replicating the study in Besedeš and Prusa (2006b), but using discrete-time models as well as the original Cox model, we also seek to investigate whether the theoretical arguments that can be brought to bear against the Cox model matter in practice.

Briefly, there are three major problems with continuous-time models, such as the Cox model, in this context. First, the predominantly short-lived trade relations combined with the coarse grouping of durations into yearly intervals result in a large number of tied survival times, i.e. spells of trade with exactly the same duration. Continuous-time methods face difficulties in the presence of heavy ties, leading to biased coefficient estimates and standard errors. Second, it is difficult to properly control for unobserved heterogeneity, which, if such heterogeneity is indeed important, will cause spurious negative duration dependence of the estimated hazard function as well as parameter bias. Third, the Cox model imposes the restrictive and empirically questionable assumption of proportional hazards. There are two reasons why the proportional hazards assumption may fail to hold. First, the effect of explanatory variables (covariates) on the hazard may be intrinsically non-proportional. Second, unobserved individual heterogeneity that is not accounted for may cause the impact of observed regressors to depend on duration time, even if the underlying model is of the proportional hazards form. Incorrectly imposing the assumption of proportional hazards will lead to bias in the estimated covariate effects. By contrast, with discrete-time models there is no problem handling ties; unobserved heterogeneity can be controlled for without difficulty even when dealing with very large data sets; and the restrictive proportional hazards assumption can easily be bypassed. In addition, these theoretically more appropriate models are readily implemented using standard statistical software. Typically, the researcher will be able to choose between at least three different discrete-time model specifications: logit, probit, and complementary log-log (cloglog).

Replicating the study in Besedeš and Prusa (2006b), but using discrete-time models as well as the original Cox model, we find empirical evidence in support of the arguments against the Cox model. The large number of ties in the data set does indeed seem to lead to biased estimation coefficients in the Cox model; unobserved heterogeneity plays a significant role and has to be accounted for, and several different tests unambiguously reject the proportional hazards assumption for our specific data set. Comparing the results obtained from a discrete-time probit model with random effects and the Cox proportional hazards model, we find significant differences in both the predicted hazard rates and the estimated effects of covariates on the hazard. This suggests that the choice of hazard model is not innocuous in this context and will have important implications for the conclusions that can be drawn. Since we also find strong empirical evidence in favor of the probit model, we conclude that the mentioned drawbacks of the Cox model are not just purely theoretical issues that can safely be ignored by empirical researchers.

The remainder of the paper is organized as follows. Section 2 provides a brief survey of the existing literature on trade survival from a methodological point of view. Section 3 contains an extensive discussion of the shortcomings of continuous-time models in this context, and an outline of alternative, discrete-time models. Section 4 illustrates the consequences of model selection by replicating the empirical work of Besedeš and Prusa (2006b) with different models, and Section 5 concludes.

2 Previous Research

The literature on the duration of trade, which is still rather young, started with a series of articles by Tibor Besedeš and Thomas Prusa. In Besedeš and Prusa (2006a), these authors use detailed data on US imports for 1972-1988 to estimate descriptive Kaplan-Meier survival functions. Their results, largely confirmed by using data for 1989-2001, suggest that the duration of exports to the US is in general very short.[1] Using the same import data, Besedeš and Prusa (2006b) apply a Cox proportional hazards model, which, unlike the Kaplan-Meier methodology, enables them to include independent explanatory

[1] A similar methodological approach for describing the duration of trade was taken by Besedeš and Prusa (2007), who focus on the extensive and intensive margins of trade.

variables in order to search for explanations for the short trade durations. Some interesting findings include that differentiated products have lower hazard rates than homogeneous goods, and that within each product type, the larger the value of the initial trade flow, the longer the duration. A very similar methodological approach is taken by Besedeš (2008), which, however, employs a stratified Cox approach. Using a stratified Cox model has the advantage of making it possible to, at least to some extent, control for unobserved heterogeneity by allowing for group-specific variation in the baseline hazard.

Following the example set by above all Besedeš and Prusa (2006b), other authors have subsequently used similar Cox approaches to analyze what determines the duration of trade. Brenton, Pierola and von Uexküll (2009) use a Cox model to estimate determinants of trade in a data set with 44 exporters and 56 importers over a 21-year period. Nitsch (2009) employs a stratified Cox model when examining the duration of German imports for 1995-2005. Fugazza and Molina (2009) use an extended version of the Cox model where the estimation coefficients are allowed to vary over duration time, and estimate determinants of trade duration among 96 trading countries for the period 1995-2004. Lastly, Brenton, Saborowski and von Uexküll (2009) employ a discrete-time equivalent of the Cox model, namely a cloglog model to look at the duration of export flows from 82 exporters to 53 importers over the period 1985-2005.[2]

3 Continuous-Time or Discrete-Time Methods?

As outlined above, the existing literature aimed at explaining the duration of bilateral trade relationships has largely followed Besedeš and Prusa (2006b) and estimated various versions of continuous-time Cox proportional hazards models. While this may seem to be a good choice, particularly given that the Cox model can be estimated without having to specify a functional form for the baseline hazard, we argue that certain problems arise when the Cox model is applied to the particular area of trade durations. In this section we provide an extensive discussion of these problems, and propose discrete-time duration models that may be used as an alternative.

3.1 Why Cox Should Not Analyze Trade

There are three main reasons why it is inappropriate to apply the Cox model when analyzing the duration of trade relationships:

1. The Cox model is a continuous-time specification, whereas the duration of trade relationships is observed in discrete units of yearly length. As a consequence, many

[2]These authors argue that in the presence of unobserved heterogeneity which is not properly controlled for, the proportional hazards assumption implicitly made in a Cox model will not hold. They therefore employ the discrete-time equivalent of a Cox model, i.e. a cloglog model, where unobserved heterogeneity can much more easily be controlled for. However, as will be further discussed below, this is a risky choice, because a cloglog model also assumes proportional hazards, so if that assumption is intrinsically invalid, the cloglog model will also be a bad choice.

trade relations are observed to be of equal length, and no "natural" way exists to treat such tied duration times within the partial likelihood framework of the Cox model. The presence of ties causes asymptotic bias in both the estimation of the regression coefficients and in the estimation of the corresponding covariance matrix.

2. Unobserved heterogeneity cannot be included without the presence of multiple integrals, which makes estimation difficult, if not impossible. Ignoring unobserved heterogeneity causes – if it is indeed important – parameter bias and spurious duration dependence.

3. The Cox model imposes the rather restrictive assumption of proportional hazards. In other words, the effects of explanatory variables on the hazard rate are assumed to be constant across duration time. This is unlikely to hold for the regressors typically employed to analyze the duration of trade relationships. Incorrectly imposing proportionality will produce misleading estimates of covariate effects.

In the following, these reasons for the inappropriateness of the Cox model are evaluated in more detail.

3.1.1 Tied duration times

International trade is traditionally only observed once a year, even though the underlying trade transactions may take place every day of the year, or, at the extreme, only once a year. This implies that the observed durations of trade will be grouped into yearly intervals. However, the Cox model is based on the assumption that duration times can take on any value on the positive real line and that this value can be observed exactly. In order to carry out the partial likelihood estimation procedure of the Cox model, the recorded duration times need to be ordered chronologically. Then, in the case of intrinsically discrete or grouped survival time data – such as the available data on trade durations – a substantial complication may arise. If there are only a few time intervals or if the time units are large, many trade flows are reported to cease at exactly the same time, and the number of ties becomes high. Then, strictly speaking, continuous-time techniques are inappropriate (see Cox and Oakes, 1984, p. 99). Kalbfleisch and Prentice (1980, p. 75) emphasize that the presence of ties causes asymptotic bias in both the estimation of the regression coefficients and in the estimation of the corresponding covariance matrix. This applies not only to the Cox model but also to fully parameterized continuous-time models.

In the statistics literature, several different approaches to deal with tied survival times in continuous-time hazard models have been developed. One of the most commonly applied procedures of handling ties is a method proposed by Breslow (1974), which is based on a rather simple approximation of the exact marginal likelihood. While computationally undemanding, the Breslow method will be inaccurate if there are many ties in the data set, leading to an increasing asymptotic bias of the parameter estimates as the grouping of duration times becomes more coarse (see e.g. Prentice and Gloeckler, 1978, or Hsieh, 1995). A more accurate approximation of the exact marginal likelihood was proposed by Efron (1977), but this procedure of handling ties is computationally more cumbersome and still inaccurate in the presence of heavy ties. Besides these two standard approaches there are various so-called exact methods of dealing with ties. However, all these methods lead to biased estimates when the true underlying model is in fact a Cox model (see Scheike and Sun, 2007). Thus, to summarize, heavy ties will lead to biased parameter estimates and standard errors, and there is no sufficient way to tackle this problem in a Cox model framework. This issue therefore poses a problem for all the papers in the trade duration literature that employ Cox models.

3.1.2 Unobserved heterogeneity

Accounting for unmeasured heterogeneity (also known as *frailty* in the biostatistics literature) within the partial likelihood framework of the Cox model is computationally burdensome, since it involves multidimensional integrals. In fact, when applying the software package Stata, estimation becomes impossible as the number of frailty components exceeds 11 000. In our empirical analysis we outnumber this limit by a factor of almost three. Even when using more advanced programs, it will be infeasible to estimate a Cox model with random effects when faced with such a huge data set, since it would require the computation of an integral with several thousand dimensions at every iteration of the likelihood maximization process.[3]

Nevertheless, individual heterogeneity cannot be ignored, since it will rarely be the case in practice that all sources of individual variation in the hazard rate are exhausted by the observed explanatory variables included in the model. The biases and spurious duration dependencies caused by ignoring this variation in the hazard have long been known in the literature (see e.g. Salant, 1977, Vaupel, Manton and Stallard, 1979, or Vaupel and Yashin, 1985, for early discussions of this phenomenon).

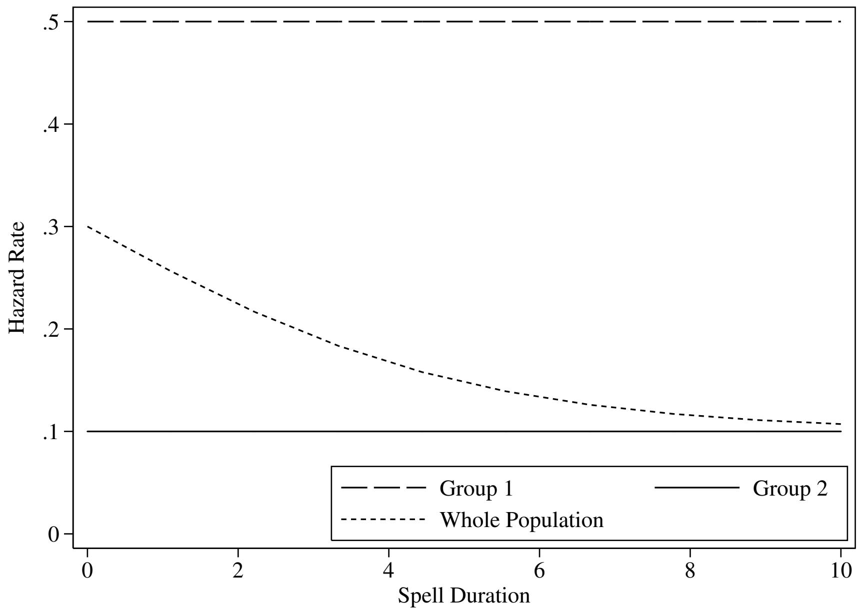
Most strikingly, the presence of unobserved heterogeneity can produce duration dependence patterns for a population as a whole that are entirely different from the patterns of the corresponding sub-populations or individual observations. The reason for this phenomenon is simply a selection process. When investigating a heterogenous population sample over a certain time period, the sample composition is likely to change over time, since the ob-

[3]It should be mentioned, though, that unobserved heterogeneity can be accounted for in the Cox model by allowing the baseline hazard to vary between observations. The respective model can then be estimated using a stratified partial likelihood approach. Encompassing unobserved variation only on a rather crude level, this strategy also suffers from the drawback that the effects of explanatory variables that do not exhibit within-stratum variation cannot be estimated. Stratified Cox models have been employed by Besedeš (2008) and Nitsch (2009). To some extent, frailty can also be accounted for by including dummy variables in the regression function. For example, individual effects that are specific to every exporting country can be modelled in this fashion – this approach was chosen by Besedeš and Prusa (2006b). This, however, is not sufficient if unobserved heterogeneity on a more detailed level (e.g. exporter-product level) is present, and another drawback is that the effects of covariates that are specific to every export country can no longer be estimated.

servations with the highest hazard rates tend to exit first. Hence, the average duration dependence for the whole population will change over time. This dynamic selection process is frequently referred to as *weeding out* or *sorting effect* in the duration literature (see e.g. van den Berg, 2001, for an extensive formal discussion of this mechanism).

Intuitively, one would assume that unobserved heterogeneity has a negative effect on the hazard rate, since the fraction of observations with relatively low hazard rates in the sample should tend to increase over time. And in fact, Heckman and Singer (1984a) proved that ignoring unobserved heterogeneity will bias the estimated hazards towards negative duration dependence, meaning that a decreasing hazard will appear more rapidly decreasing, while an increasing hazard may appear more slowly increasing, constant, or even decreasing.

Figure 1: Spurious Duration Dependence Resulting from Unobserved Heterogeneity



An intuitive way of illustrating this effect of neglecting unobserved heterogeneity is to consider a sample consisting in equal shares of two sub-samples that differ according to some unobserved characteristic. If we further assume that each observation in the first sub-sample has a constant hazard rate of 50 percent and that each observation in the second sub-sample faces a constant hazard of 10 percent, the average hazard rate for the whole sample will be 30 percent initially. However, as duration increases, observations from the first sub-group will leave the sample at a higher rate than observations from the second sub-group, thereby shifting the shares of the two groups in support of the sub-sample facing the lower hazard rate. Then, as Figure 1 shows, the hazard rate for the whole sample decreases over time. The observed negative duration dependence of the hazard function, however, is merely an artefact of unobserved heterogeneity. Neglecting unobserved heterogeneity may also bias the coefficients of the explanatory variables in the hazard model. In particular, omitting unobserved heterogeneity in continuous-time (mixed) proportional hazards models, such as the Cox model, leads to an underestimation of the proportionate effect of covariates on the hazard (see van den Berg, 2001, for a formal proof).

To summarize, the inability to properly control for unobserved heterogeneity poses another problem for the papers in the literature which use Cox models. While some papers have gone some way toward mitigating the problem by employing stratified Cox models or including country dummies, these solutions have their limitations.

3.1.3 Proportional hazards

Another drawback immanent to the Cox model is that it imposes the individual hazard functions to be proportional. If the assumption of proportional hazards is incorrectly imposed, this will lead to bias in the estimated covariate effects. In Cox regression models the analysis of trade survival is approached through the definition of a hazard function that represents the instantaneous rate at which trade relationships cease, depending on their elapsed duration and other explanatory factors. The impact of explanatory variables is specified as shifting the baseline hazard, which depicts duration dependence, in a proportional fashion. This proportional hazards assumption implies that the effect of covariates on the hazard is restricted to be constant throughout the whole progression of a trade relationship. The notion that the effect of explanatory variables may not be constant across durations is well developed in the literature and several tests exist to examine the proportionality assumption imposed by the Cox model (see e.g. McCall, 1994, and references therein).

There are two reasons why the proportional hazards assumption may fail to hold. First, the effect of explanatory variables on the hazard may be intrinsically non-proportional. For example, the initial trade volume is unlikely to affect the probability that the trade relationship ceases during the first year of service to the same extent as the probability that it ends during the tenth year of service. Second, unobserved individual heterogeneity that is not accounted for will cause the impact of observed regressors to depend on duration time, even if the underlying model is of the proportional hazards form (see Lancaster and Nickell, 1980).

To our knowledge, there is only one study on trade durations where the assumption of proportional hazards is actually tested. For their data on developing countries' export flows, Brenton, Saborowski and von Uexküll (2009) find significant evidence against the validity of the proportional hazards assumption. Since incorrectly imposing the proportional hazards assumption will cause bias in the estimated covariate effects, the authors employ a discrete-time proportional hazards model incorporating random effects to account for unobserved heterogeneity. However, while tackling the effect of omitted regressors, their model does not allow for intrinsic non-proportionality. A different approach is chosen by Fugazza and Molina (2009) who apply an extended version of the Cox model with time-varying coefficients, thereby allowing for intrinsically non-proportional covariate effects. This approach, however, has the disadvantage that a large number of additional parameters has to be estimated. Moreover, as discussed above, it is difficult to properly control for unobserved heterogeneity in a Cox model when dealing with data sets as large as those typically encountered in trade duration studies. Therefore, the other possible reason for the proportional hazards assumption to fail is not taken into account. Hence, the two papers in the literature that have actually approached the issue of the proportional hazards assumption have each only dealt with one of the potential reasons why the assumption may not hold.

3.2 Discrete-Time Duration Models

As outlined above, the continuous-time Cox model is inappropriate to use when analyzing trade data. Is there a better alternative available? We recommend that the researcher uses discrete-time models. The very high proportion of ties, which is typically found at all durations when analyzing trade data, does not constitute a problem for discrete-time duration models. Further, besides their ability to deal with tied failure times, discrete-time models are also preferable for computational reasons. Hazard rate models for grouped duration times can be estimated using conventional regression models for binary response panel data, which are implemented in all common statistical software packages. These models are computationally less demanding than the continuous-time Cox model, especially when analyzing large samples containing heavy ties (which are typical characteristics of the data sets under consideration in the trade duration literature). More importantly, discrete-time duration models can easily be extended to account for unobserved individual heterogeneity, even if the number of observations is large. Finally, by applying discrete-time duration models, we can easily circumvent the rather restrictive proportional hazards assumption. We can choose from different model specifications, implying different degrees of proportionality, while still being able to estimate duration dependence in a nonparametric fashion.

Standard statistical software packages such as Stata or LimDep usually provide three different specifications: cloglog, logit, and probit. While it can be shown that the cloglog model with period-specific intercepts represents the exact grouped-duration analogue of the Cox proportional hazards model (see e.g. Kalbfleisch and Prentice, 1973, or Prentice and Gloeckler, 1978), the logit and probit specifications do not impose this proportionality assumption. The logit model is rather similar to the cloglog model and departs only slightly from proportionality, whereas the probit specification is decidedly non-proportional.[4]

Having pointed out why it is preferable to apply discrete-time hazard models when analyzing grouped duration data, we now introduce these methods in more detail. Let T_i be a continuous, non-negative random variable measuring the survival time of a particular trade relation. In a discrete-time framework, the core of duration analysis is formed by the probability that a particular trade relation terminates in a given time interval $[t_k, t_{k+1})$,

[4]See Sueyoshi (1995) for an extensive discussion of these model specifications in a duration context.

$k = 1, 2, \ldots, k^{max}$, and $t_1 = 0$, conditional on its survival up to the beginning of the interval and given the explanatory variables included in the regression model. This conditional probability is termed the discrete-time hazard rate and formally defined as

$$h_{ik} := P(T_i < t_{k+1} | T_i \geq t_k, \boldsymbol{x}_{ik}) = F(\boldsymbol{x}'_{ik}\boldsymbol{\beta} + \gamma_k), \tag{1}$$

where $\boldsymbol{x}_{ik}$ is a vector of possibly time-varying covariates, γ_k is a function of (interval) time that allows the hazard rate to vary across periods (somewhat loosely, we will refer to γ_k as the grouped-duration baseline hazard, although this is not formally correct in all instances), and $F(\cdot)$ is an appropriate distribution function ensuring that $0 \leq h_{ik} \leq 1$ for all i, k. In our case, the subscript i denotes separate spells of trade relationships ($i = 1, \ldots, n$) for any given exporter-product combination. Since the underlying baseline hazard is unknown in practice, γ_k is usually incorporated in the model as a dummy variable marking the current length of the spell. However, a functional form for γ_k can also be specified in order to reduce the number of parameters in the model.

Working with trade data, the researcher typically observes the value of a country's imports from another country for each year of the observation period. Thus, the first year of a consecutive period where a product is imported from a certain country can be regarded as the starting year of the corresponding trade spell. Equivalently, the last year of this period would be the terminating year of the spell. If a positive import value is observed for the first year of the observation period, the corresponding spell might have started prior to that date, and the exact spell length is unknown. In any empirical application, these left-censored spells should be disregarded in order to avoid any restrictive *a priori* assumptions about the duration dependence of the hazard rate. If a positive import value is observed for the last year of the observation period, the corresponding spell length is also unknown. Such right-censored spells, however, do not constitute any problems for the derivation of the sample likelihood, and all right-censored spells can be included in the analysis.

For each trade spell, the last year in which a positive trade volume was observed, can be recorded. In the following, this terminal time period is denoted k_i, the subscript i indicating that it may differ across spells. Introducing a binary variable, y_{ik}, taking the value one if spell i is observed to cease during the k^{th} time interval, and zero otherwise, the log-likelihood for the observed data is given by

$$\ln \mathcal{L} = \sum_{i=1}^{n} \sum_{k=1}^{k_i} [y_{ik} \ln(h_{ik}) + (1 - y_{ik}) \ln(1 - h_{ik})]. \tag{2}$$

This expression is structurally isomorphic to a standard log-likelihood function for a binary panel regression model with dependent variable y_{ik}.[5] To be able to estimate the model

[5]To obtain consistent parameter estimates from this log-likelihood, spells must be independent, censoring must occur only at interval boundaries, and censoring must not provide any information about T_i beyond that available in the covariates (see e.g. Allison, 1982, Singer and Willett, 1993, or Jenkins, 1995, for excellent surveys on the derivation of the likelihood). To ensure conditional independence between

parameters, a functional form for the hazard rate h_{ik} needs to be specified. The most commonly encountered functional specifications are the normal, logistic, and extreme-value minimum distribution, leading to a probit, logit, or cloglog model, respectively.

Unobserved heterogeneity can be accounted for by including random effects into the binary choice model framework above. If a specific parametric distribution is assumed for the random effects, calculating the marginal log-likelihood function involves only a one-dimensional integral which can be computed numerically by using e.g. Gauss-Hermite quadrature (see Butler and Moffitt, 1982) or simulation methods (see e.g. Train, 2003). In continuous-time duration models, the heterogeneity distribution is often chosen to be Gamma for analytical convenience (see Lancaster, 1979) and theoretical reasons (see Abbring and van den Berg, 2007). In discrete-time duration analysis, instead, the assumption of a Gaussian distribution may be computationally convenient. Under this assumption, the hazard models can be estimated as binary choice models with normal random effects using widely available software packages. In Stata this can be done using the commands `xtlogit`, `xtprobit`, or `xtcloglog` for the logit, probit, or cloglog model, respectively.

The choice of heterogeneity distribution is a widely discussed issue in the duration literature (see e.g. Heckman and Singer, 1984b, for continuous time and Baker and Melino, 2000, for discrete time). In an extensive simulation study of the discrete-time logit model, Baker and Melino (2000) find that misspecification of the heterogeneity distribution can lead to substantial biases in the parameter estimates. However, the parameters in binary choice models such as logit are only identified up to scale, and Mroz and Zayats (2008) show that the biases reported by Baker and Melino (2000) are due to their neglect of this issue. A recent simulation study by Nicoletti and Rondinelli (2009) suggests that choosing a Gaussian heterogeneity distribution, when the true one is Gamma or discrete, does not affect the parameter estimates. Their findings are supported by several empirical studies. Trussell and Richards (1985), Meyer (1990), and Dolton and van der Klaauw (1995) find empirical evidence that the choice of heterogeneity distribution may be unimportant if duration dependence is modelled in a flexible manner. This suggests that applying conventional binary response panel data models with normal random effects is a sensible approach when estimating discrete-time duration models.

4 Empirical Application: The Duration of US Imports

Having outlined the theoretical reasons for not using a Cox model when studying the duration of trade, and described the discrete-time models that could be used as alternatives, we now illustrate the implications of model choice by replicating the study by Besedeš and Prusa (2006b), but adding several discrete-time estimation methods. We briefly present the data used in our estimations, and then provide an extensive review of the results to

spells, in empirical analyses, special care should be taken to account for multiple spells and the dependencies existing among imports of different products from the same country or imports of a particular product from different countries.

see whether or not the choice of estimation method actually matters for the conclusions that can be drawn.

4.1 Data

We use the same data on US imports as was previously analyzed by Besedeš and Prusa (2006b). The data, which was generously provided to us by the authors, record annual US imports between 1972 and 1988 from virtually every trading partner and include information on the value of imports, customs collected, and other relevant factors that might affect the duration of trade. The traded products are classified according to the 7-digit Tariff Schedule of the United States (TSUSA) which amounts to a total of some 20 000 products. A *trade relationship* is then defined as a certain product being imported from one specific exporter. A *trade spell* is defined as a period of time with uninterrupted import of a given product from one specific country. These spells of trade constitute the core units of analysis in this paper, where the *spell duration* is simply calculated as the number of consecutive years with non-zero imports. The number of spells differs from the number of trade relationships (i.e. exporter-product combinations), since any of the trading parties may choose to terminate the trade relationship and revive it at a later point in time. We refer to such reoccurring trade relationships as *multiple spells of service*.

One peculiarity of duration data is the presence of censored survival times. Our annual durations of trade are subject to two different kinds of censoring. First, trade relationships persisting in 1988 have an uncertain end and are thus right-censored.[6] However, since it is known how long these spells have been in progress, right censoring does not impose any problems to model estimation. Second, trade spells observed in 1972 have an unknown starting date and are thus left-censored. This type of censoring is more problematic and cannot be handled straightforwardly without imposing restrictive assumptions on the baseline hazard. To see the difficulties arising in the presence of left censoring, consider the following example. Assume we include left-censored spells in the analysis, then these observations will be assigned a duration of one year in 1972. Many of these spells, however, will have started prior to 1972 and thus have a true duration which is greater than one year at the beginning of our observation period. If there is true negative duration dependence, trade relationships will, by definition, have a lower hazard rate the longer they have been in progress. Thus, assigning duration times to spells that are lower than their true values will lead to an underestimation of the hazard rate at all durations. This is illustrated in Figure 2, which depicts the predicted baseline hazard functions obtained from our data inclusive and exclusive of left-censored spells.

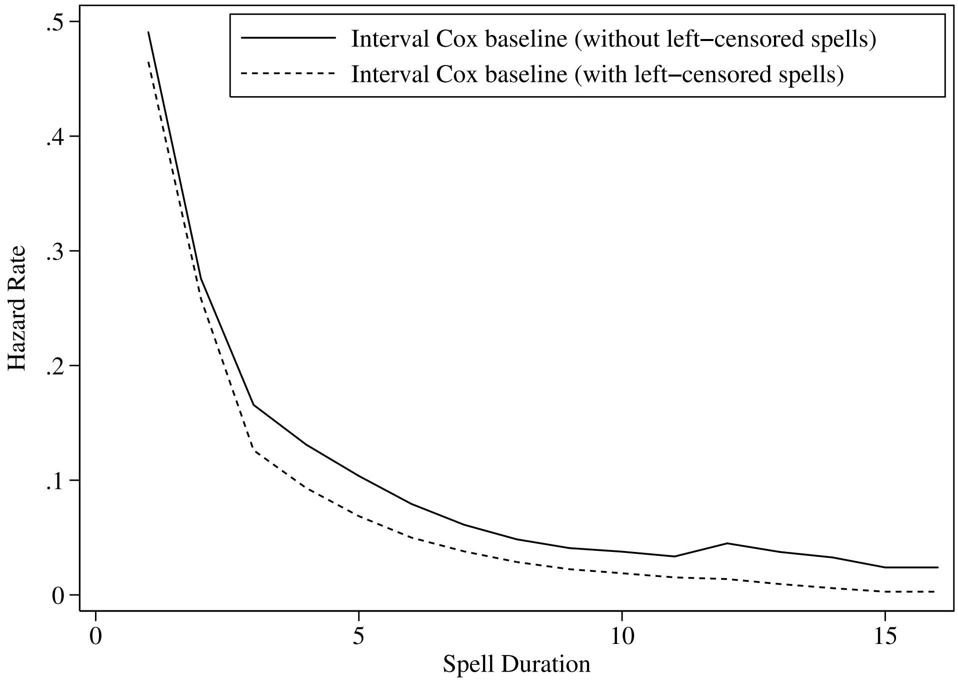
Clearly, left censoring is not innocuous for the purpose of making inference on the baseline hazard. We have thus excluded all left-censored observations from our analysis.[7]

[6]Reclassification of product codes may also lead to right censoring at dates other than the final year of the observation period.

[7]We note that Besedeš and Prusa (2006b) choose to include the left-censored observations in their analysis, and therefore report results from both the original data and the data where the left-censored

As Table 1 shows, these exclusions do not lead to a drastic decrease in the number of spells contained in our data set. Table 1 provides some summary statistics for our data.

Figure 2: Bias in the Estimated Baseline Hazard Arising from Left Censoring



The reduced data set which we use for the following econometric analysis consists of over 300 000 trade relationships accounting for more than 400 000 spells of trade. With 2.35 years, the observed length of trade spells is very short on average, but this is to some extent a direct consequence of the large fraction of right-censored spells (54%). For further details on the data, we refer the reader to Besedeš and Prusa (2006b).

Table 1: Summary Statistics

	Observed spell length in years		Total number	Fraction of spells	Total number of trade	Total number of product
	Mean	Median	of spells	right-censored	relationships	codes
Original data						
	2.79	1	444 378	0.57	335 253	20 351
Data exclusive of left-censored spells						
	2.35	1	414 227	0.54	312 685	20 282

observations have been excluded.

4.2 Model Specifications

To be able to focus strictly on methodological differences, we use exactly the same set of explanatory variables as Besedeš and Prusa (2006b). For a detailed discussion of the explanatory variables included in the regression models, we refer the reader to that article. Since our econometric analysis aims at illustrating the effects of model specification on estimation results, we estimate various discrete-time hazard models (cloglog, logit, probit, and Pareto) as well as a continuous-time Cox proportional hazards model.[8] In all discrete-time models, we specify the baseline hazard in the most flexible possible fashion by means of dummy variables that enable the estimation of period-specific intercepts. This, in turn, allows for unrestricted period-specific changes in the estimated hazard rates. Besides using different hazard models, we also present results from models with and without frailty, i.e. with and without Gaussian random effects for every exporter-product combination.[9] As mentioned previously, estimating the Cox model with random frailty is computationally infeasible for the large data set under consideration, so this model is only included without controls for unobserved heterogeneity.

The continuous-time Cox model as well as the discrete-time cloglog, logit, and probit models are widely applied in duration analysis and do not require further elaboration. The Pareto hazard model, however, deserves a somewhat closer look. In this model, the discrete-time hazard rate – as defined in equation (1) – is parameterized as

$$h_{ik} = 1 - \left(1 + \frac{\xi}{\exp\{-(\boldsymbol{x}'_{ik}\boldsymbol{\beta} + \gamma_k)\}}\right)^{-1/\xi}. \tag{3}$$

The right-hand side of equation (3) describes the distribution function of the generalized log-Burr distribution with shape parameter $\xi \geq 0$ (see Burr, 1942, or Tadikamalla, 1980). As opposed to the conventional cloglog, logit, and probit specifications, the Pareto hazard rate contains a shape parameter, which makes the model considerably more flexible with respect to the imposed effects of covariates on exit probabilities. A particular virtue of the Pareto hazard model is that it contains the cloglog and logit specifications as special cases. To see this, note that the cloglog hazard arises as the limiting case

$$\lim_{\xi \to 0} \left(1 - \left(1 + \frac{\xi}{\exp\{-(\boldsymbol{x}'_{ik}\boldsymbol{\beta} + \gamma_k)\}}\right)^{-1/\xi}\right) = 1 - \exp\{-\exp\{\boldsymbol{x}'_{ik}\boldsymbol{\beta} + \gamma_k\}\},$$

which is then the definition of h_{ik} for $\xi = 0$. The case $\xi = 1$ yields the logistic hazard rate

$$\frac{1}{1 + \exp\{-(\boldsymbol{x}'_{ik}\boldsymbol{\beta} + \gamma_k)\}}.$$

[8]Estimates obtained from the Cox model are reported both for the data with and without left-censored spells. The estimates obtained from the complete data set (*Cox1*) correspond exactly to the results reported by Besedeš and Prusa (2006b) in the first column of their Table 3. However, for reasons of comparability with our discrete-time specifications, we report β-coefficients instead of hazard ratios. Hazard ratios can be obtained from these values as $\Delta h = \exp\{\beta\}$.

[9]Using spell-specific effects instead of "individual"-specific effects has no impact on our results.

In a binary response model context, the use of the generalized log-Burr distribution was first proposed by Prentice (1975, 1976). He has shown that ξ can be consistently estimated together with the other model parameters, which enables discrimination between the cloglog and the logit specification. Hess (2009) has shown that this grouped-duration hazard specification can be linked to the asymptotic distribution of threshold excesses for the underlying continuous duration times.[10] He also illustrates that a larger value of ξ implies a higher degree of non-proportionality for the respective hazard model. Since the Pareto model includes the discrete-time analogue of the Cox model as a special case, it can also be used to test the validity of the Cox model even with coarsely grouped duration data. This simply requires to test the null hypothesis $\xi = 0$ against the one-sided alternative $\xi > 0$.

4.3 Estimation Results

The results obtained from all model specifications can be found in Table 2. Qualitatively, the results are similar for the various estimation procedures, and none of the estimated coefficients changes sign across model specifications. While higher transportation costs increase the termination probability of a spell, a higher GDP of the trading partner, a higher industry level tariff rate, a real depreciation of the exporting country's currency, and a larger coefficient of variation of unit values decrease the hazard. Higher order spells have an increased failure probability and so do trade relationships involving agricultural goods, reference priced products, and homogeneous goods. For a detailed discussion of the estimation results, see Besedeš and Prusa (2006b).

To find the most suitable alternative to the Cox model for the data at hand, we note that the Pareto hazard model and the probit model with frailty are the two model specifications which yield the best fit in terms of log-likelihood values. In the following, when comparing continuous-time and discrete-time models, we will use the probit model with frailty as our preferred specification, in spite of the fact that the Pareto model achieves an even better fit in log-likelihood terms. We focus on the probit model mainly because this makes it possible to explicitly allow for unobserved heterogeneity in the hazard specification.[11] Moreover, the (random-effects) probit model is implemented in many statistical software packages and can be readily applied by the empirical researcher. We will, however, apply the Pareto model as a tool for testing the proportional hazards assumption imposed by the continuous-time Cox model.

[10] This threshold excess distribution is of the generalized Pareto form, whence the name Pareto hazard model.

[11] Unfortunately, we were unable to incorporate random effects into the Pareto model. The reason for this is that the variance of the idiosyncratic error term in a binary response panel data model needs to be fixed in order for the random effects variance to be identified (see Lechner, Lollivier and Magnac, 2008, for details on identification in binary response panel data models with random effects). While this condition holds for the cloglog, logit, and probit specifications, it is not met by the Pareto hazard model.

Table 2: Estimation Results for 1972-1988 7-Digit TSUSA Data

	Models without frailty						Models with frailty		
	Cox1	Cox2	Cloglog	Logit	Probit	Pareto	Cloglog	Logit	Probit
Ad valorem transportation costs (unit=10%)	0.0655	0.0662	0.0820	0.1108	0.0661	0.2055	0.0881	0.1179	0.0694
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GDP (unit=$100 billions)	−0.0558	−0.0942	−0.1049	−0.1226	−0.0680	−0.1883	−0.1149	−0.1354	−0.0745
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tariff rate, 4-digit SITC (unit=1%)	−0.0209	−0.0211	−0.0251	−0.0297	−0.0168	−0.0453	−0.0270	−0.0321	−0.0181
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
%Δ relative real exchange rate (unit=10%)	−0.0994	−0.0989	−0.1239	−0.1472	−0.0654	−0.2086	−0.1321	−0.1556	−0.0694
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Coefficient of variation of unit values	−0.0758	−0.0839	−0.0977	−0.1210	−0.0690	−0.1916	−0.1050	−0.1299	−0.0737
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Multiple spell dummy	0.4023	0.3495	0.3870	0.5150	0.3124	0.9634	0.3328	0.4549	0.2765
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Agricultural goods (dummy)	0.0395	0.0541	0.732	0.1235	0.0735	0.2555	0.1122	0.1634	0.0956
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Reference priced products (dummy)	0.1595	0.1576	0.1877	0.2431	0.1408	0.4147	0.2338	0.2945	0.1690
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Homogeneous goods (dummy)	0.2036	0.2118	0.2638	0.3596	0.2121	0.6479	0.3381	0.4390	0.2552
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Duration dummies	no	no	yes	yes	yes	yes	yes	yes	yes
Exporter dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
ξ, ρ						4.6479	0.1345	0.0957	0.0983
						(0.000)	(0.000)	(0.000)	(0.000)
Observations	1 140 896	943 595	943 595	943 595	943 595	943 595	943 595	943 595	943 595
No. of Spells	444 378	414 227	414 227	414 227	414 227	414 227	414 227	414 227	414 227
Trade relations	335 253	312 685	312 685	312 685	312 685	312 685	312 685	312 685	312 685
Log likelihood	−2 380 457	−2 373 649	−401 167	−400 091	−399 901	−399 020	−400 524	−399 582	−399 456

Note: P-values in parentheses. Models with frailty include exporter-product random effects. The parameter ξ denotes the shape parameter in the *Pareto* model, whereas ρ denotes the fraction of the error variance that is due to variation in the unobserved individual factors (models with frailty only). A trade relation is defined as an exporter-product combination. The number of observations is given by the total number of years with observed trade for all trade relationships. Cox regressions employ the Breslow method for ties. *Cox1* includes left-censored spells.

4.4 Does the Choice of Hazard Model Matter?

In Section 3 we have argued that applying the continuous-time Cox model when analyzing the duration of trade may be inappropriate for various reasons. To illustrate this in the context of US import durations, we have re-estimated the empirical model of Besedeš and Prusa (2006b) using various discrete-time models, as well as the original continuous-time Cox model. In the following, we will discuss in how far tied duration times, unobserved heterogeneity, and the assumption of proportional hazards actually affect the estimation results obtained from the Cox model.

4.4.1 Tied duration times

The first problem discussed was that the large number of ties, particularly at short spell durations, will lead to bias in the Cox model estimates. In our data set, about one third of the spells cease during the first year of service. In Section 3 we have argued that such a large fraction of tied duration times should lead to biased coefficient estimates when applying a Cox model. To see whether this is the case for our data, we can compare the results obtained from the Cox and the cloglog model as reported in columns two and three of Table 2. Since the cloglog model with period-specific intercepts is the exact grouped-duration equivalent of the Cox model, coefficient estimates obtained from these two model specifications should be identical, if the true underlying model were indeed a Cox model. For our data, however, the estimates obtained from the Cox model are altogether smaller in absolute value then their grouped-duration counterparts. Hence, applying a Cox model to analyze the data at hand will lead to incorrectly estimated effects of covariates on spell termination probabilities.[12]

4.4.2 Unobserved heterogeneity

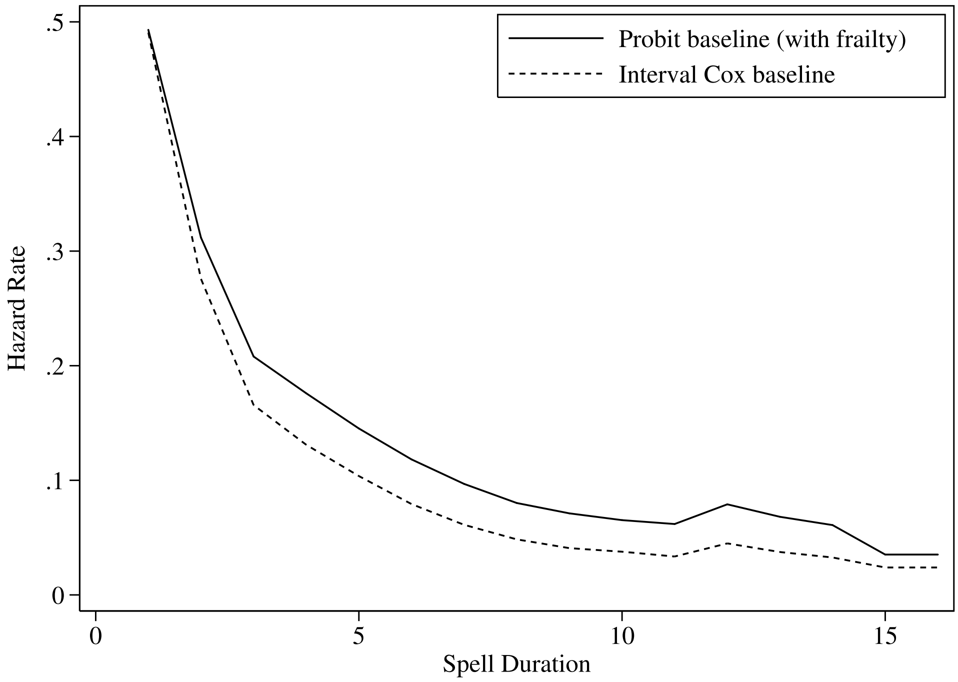
As argued above, when studying such a large set of observations as we do, it will not be possible to allow for random frailty when estimating a continuous-time Cox model. Is this a problem in practice? We note that the likelihood-ratio tests strongly reject the null hypothesis of no latent heterogeneity for all model specifications. Hence, it is not surprising that accounting for unobserved heterogeneity by means of random effects at the exporter-product level increases the respective log-likelihood values for all models where unobserved heterogeneity could be incorporated. The relative importance of unobserved heterogeneity for the different model specifications is indicated by the estimates for ρ given in Table 2. Somewhat loosely, ρ can be interpreted as the fraction of individual variation in the hazard rate that is due to variation in the unobserved factors. This fraction is around ten percent for all model specifications including frailty. In other words, the results strongly suggest

[12] However, we would like to point out that it is likely that the bias results from both tied duration times and the fact that the true underlying model is not a proportional hazards model. The latter issue will be discussed in more detail below.

that the inability to properly control for unobserved heterogeneity in the Cox model is indeed a problem.

Given that there are strong indications that unobserved heterogeneity is present and must be controlled for, can we find any evidence of the implications for the Cox model? According to theory, the neglect of unobserved heterogeneity will lead to spurious negative duration dependence of the estimated baseline hazard. Figure 3 depicts the (interval) baseline hazard functions for both the probit model with frailty and the Cox model.[13]

Figure 3: A Comparison of Predicted Baseline Hazard Functions



As Figure 3 shows, the estimated baseline hazard is about 50 percent in the first year, no matter if we control for unobserved heterogeneity or not. During the following years of service, the effects of the weeding out process – as described in Section 3 – become noticeable, and the exit probability is underestimated if we do not account for unobserved heterogeneity. For the fifth period, the estimated hazard rates are 10 percent for the Cox model without frailty and 15 percent for the probit model including random effects. Due to the weeding out mechanism, the relative differences between the predicted hazard rates become larger as spell duration increases, and by period eleven the hazard rate estimate

[13]Since the probit specification does not imply the proportional hazards assumption, it does not – strictly speaking – possess a baseline hazard. The term "baseline hazard rate" refers in this case to the probability of failure in a given interval if the aggregator function, $\boldsymbol{x}'\boldsymbol{\beta}$, takes on the value zero. Although the baseline hazard is not directly estimated in a Cox model, estimates of the baseline survivor function for different duration times, $S(t_k)$, can be obtained *ex post*. These estimates can then be used to calculate the respective interval baseline hazards as $h_k = 1 - S(t_{k+1})/S(t_k)$.

from the frailty model is twice as large as the respective estimate from the conventional Cox model (6 percent vs. 3 percent).

To summarize, our findings suggest that it is indeed important to control for unobserved heterogeneity, implying that the inability to do so in the Cox model is a serious setback that can generate misleading conclusions regarding the termination probabilities of trade spells. Specifically, not only do the models with controls for unobserved heterogeneity have a better fit, but likelihood-ratio tests also clearly indicate that controlling for unobserved heterogeneity is actually important, and we find direct evidence of spurious negative duration dependence.

4.4.3 Proportional hazards

As argued above, there are two reasons why the proportional hazards assumption may fail to hold. First, the effect of explanatory variables on the hazard may be intrinsically non-proportional. Second, unobserved individual heterogeneity that is not accounted for will cause the impact of observed regressors to depend on duration time, even if the underlying model is of the proportional hazards form (see Lancaster and Nickell, 1980).

Before focusing on this issue in our own empirical application, we note that the only existing study on trade durations where the assumption of proportional hazards is actually tested is Brenton, Saborowski and von Uexküll (2009). For their data on developing countries' export flows, they employ tests which are based on the results derived from a conventional continuous-time Cox model. They argue that the neglect of unobserved heterogeneity in the Cox model causes their tests to reject the proportional hazards assumption, which brings them to use a discrete-time proportional hazards model with spell-specific random effects in their analysis. While this may remove one of the potential sources for the proportional hazards assumption to fail, we note that their testing strategy does not allow them to distinguish between intrinsic non-proportionality and non-proportionality caused by the failure to control for unobserved heterogeneity. Therefore, it is possible that the proportional hazards assumption may fail to hold even when controlling for unobserved heterogeneity.

To remedy this, in the following, we will allow for unobserved heterogeneity when testing the proportional hazards assumption. In doing so we are able to separate the effect of unobserved heterogeneity from "true" non-proportionality. For the data at hand, we have performed three different tests of the proportional hazards assumption. The results of these tests are reported in Table 3.

First, we perform a test based on the Schoenfeld (1982) residuals derived from the continuous-time Cox model with results given in the second column of Table 2. We test the proportional hazards assumption for each explanatory variable individually and for the model as a whole. The upper panel of Table 3 reports the respective test results, disregarding results on the effect of dummy variables. For all explanatory variables, except the tariff rate, the null hypothesis of a constant effect on the hazard rate can be rejected on all common significance levels. Thus, it is not surprising that the global test for proportional hazards also strongly rejects the null.

Second, we perform a global test for the validity of the Cox model based on the grouped-duration Pareto specification. Since $\xi = 0$ corresponds to the grouped-duration Cox model, the Pareto hazard model can be used to test the validity of the Cox model. This simply requires to test the null hypothesis $\xi = 0$ against the one-sided alternative $\xi > 0$ by means of a conventional Wald test. The result of this test is reported in the middle panel of Table 3. Clearly, the test provides strong evidence against the validity of the Cox model. As shown in the sixth column of Table 2, the estimated value of ξ is almost five for the data at hand, which suggests that the effect of covariates on exit probabilities is decidedly non-proportional (see Hess, 2009, for a detailed discussion of the proportionality effects implied by this model specification).

Table 3: Tests of the Proportional Hazards Assumption

	Test statistic t	Distribution of t	P-value
Tests based on Schoenfeld residuals from the continuous-time Cox model			
Ad valorem transportation costs	84.64	$\chi^2(1)$	0.000
GDP	143.82	$\chi^2(1)$	0.000
Tariff rate, 4-digit SITC	4.59	$\chi^2(1)$	0.032
%Δ relative real exchange rate	359.13	$\chi^2(1)$	0.000
Coefficient of variation of unit values	18.20	$\chi^2(1)$	0.000
Global test	3422.14	$\chi^2(102)$	0.000
Wald test based on the shape parameter estimate from the Pareto model			
Global test	43.51	Standard normal	0.000
Wald and LR tests based on the grouped-duration Cox model with frailty			
Ad valorem transportation costs	13.78	Standard normal	0.000
GDP	2.43	Standard normal	0.015
Tariff rate, 4-digit SITC	−9.51	Standard normal	0.000
%Δ relative real exchange rate	−12.17	Standard normal	0.000
Coefficient of variation of unit values	−1.83	Standard normal	0.067
Joint LR test	439.40	$\chi^2(5)$	0.000

Note: The null hypothesis for all tests is H_0: *Proportional hazards.* For detailed information on these tests, see Schoenfeld (1982), Hess (2009), and McCall (1994).

Third and last, we test the validity of the Cox model while at the same time allowing for unobserved heterogeneity. For this purpose we have estimated a discrete-time cloglog model with unobserved heterogeneity, where we have allowed the effects of explanatory variables (except dummies) to vary over time. Specifically, we have specified $\boldsymbol{\beta}_k = \boldsymbol{\beta} + \boldsymbol{\delta}k$, where $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are five-dimensional vectors. Testing the proportional hazards assumption in this case reduces to testing the hypothesis $\boldsymbol{\delta} = \mathbf{0}$.[14] Such a test was first proposed by

[14]Note that, using this specification, we only allow the effect of covariates to vary linearly with time.

Cox (1972) for continuous-time models ignoring unobserved heterogeneity. McCall (1994) extended the test procedure to the grouped-duration case where unobserved heterogeneity is explicitly allowed for. He also provides simulation results suggesting that the test is insensitive to the specification of the frailty distribution. The results obtained from this test are reported in the lower panel of Table 3. Looking at the individual test results, the assumption of a proportional effect on the hazard can be rejected on the 1% significance level for three out of five explanatory variables. The joint test for $\boldsymbol{\delta} = \mathbf{0}$ can be rejected on all common significance levels.

Since all our tests reject the proportional hazards assumption for the data at hand, we complete the discussion of proportional hazards by investigating the consequences of incorrectly imposing proportionality. For this purpose we compare the coefficient estimates obtained from the cloglog and probit models with frailty. These are reported in columns seven and nine of Table 2. Since the parameters in these types of models are only identified up to a scale factor, the coefficient estimates obtained from these two models are not directly comparable.[15] However, if both models were equivalent with respect to the estimated effects of covariates on the hazard, the ratio of any two corresponding coefficient estimates should be a constant factor. Clearly, this is not the case for the two models considered here. The distortion becomes most obvious when comparing the estimated effects of *transportation costs* and *exchange rate* obtained from the two different specifications. While the coefficient on transportation costs obtained from the probit model is 0.0694, the corresponding value for the cloglog model is 0.0881. The respective effects of the exchange rate on the hazard are estimated to be -0.0694 and -0.1321. Thus, while transportation costs and exchange rate alterations have an exactly identical effect (in absolute terms) in the probit model, the impact of exchange rate movements is estimated to be 50% larger in absolute terms than the effect of transportation costs when proportionality is imposed. In general, when using a cloglog model instead of a probit model, the estimated covariate effects are relatively larger for factors that decrease the hazard and relatively smaller for factors that increase the hazard. The parameter ratios ($\beta^{\text{cloglog}}/\beta^{\text{probit}}$) range from 1.42 to 1.90 in the former case and from 1.17 to 1.38 in the latter.

Thus, in summary, our tests reject the assumption of proportional hazards even when unobserved heterogeneity is accounted for, implying that the effects of the included independent variables are in themselves non-proportional. This implies that the use of a cloglog model, where the assumption of proportional hazards is also made, is in fact not a solution to the problem even though this enables the researcher to control for unobserved heterogeneity. Further, we find evidence that incorrectly making the assumption of proportional hazards causes distortions in the estimated covariate effects.

This is convenient, since it keeps the number of additional parameters at a viable level. It also leads to a very conservative test, since we only reject the proportional hazards assumption if we detect parameter variation of a linear form.

[15]See Amemiya (1981) for a detailed discussion of this issue.

5 Summary and Conclusions

This paper takes as its starting point the fact that – following the influential contributions by Besedeš and Prusa (2006b) – the literature on the duration of trade has employed Cox proportional hazards models to estimate the determinants of trade durations. The main purpose of the paper is to show why it is inappropriate to analyze the duration of trade with continuous-time models such as the Cox model, and to propose alternative discrete-time models which are more suitable for estimation. By replicating the study in Besedeš and Prusa (2006b), but using discrete-time models as well as the original Cox model, the paper also seeks to investigate whether the theoretical arguments against the Cox model matter in practice.

We have discussed three major problems with continuous-time models when applied to large trade data sets. First, such models face problems in the presence of heavy ties, leading to biased estimation coefficients and standard errors. Second, it is very difficult to properly control for unobserved heterogeneity, which results in spurious negative duration dependence of the estimated hazard function, as well as biased estimation coefficients. Third, the Cox model imposes the restrictive and empirically questionable assumption of proportional hazards. Incorrectly imposing this assumption will cause biases in the estimated covariate effects. There are two reasons why the proportional hazards assumption may fail to hold. First, the effect of explanatory variables on the hazard may be intrinsically non-proportional. Second, unobserved individual heterogeneity that is not accounted for will cause the impact of observed regressors to depend on duration time, even if the underlying model is of the proportional hazards form.

By contrast, with discrete-time models there is no difficulty in dealing with ties; unobserved heterogeneity can easily be controlled for; and one does not have to assume proportional hazards, even though it is possible to do so. In addition, these theoretically more appropriate models are readily implemented using standard statistical software packages.

Replicating the study by Besedeš and Prusa (2006b), applying discrete-time models in addition to the original Cox model, we have found empirical evidence in support of all the theoretical arguments raised against using the Cox model. We have shown that the numerous ties in the trade duration data lead to incorrectly estimated effects of covariates on spell termination probabilities when applying a Cox model. This is a potential problem for all the surveyed papers in the literature that apply Cox models, i.e. all studies except Brenton, Saborowski and von Uexküll (2009).

We have also found that it is indeed important to control for unobserved heterogeneity, implying that the difficulties to do so in the Cox model is a serious setback that can generate misleading conclusions regarding the termination probabilities of trade spells. Specifically, not only do the models with controls for unobserved heterogeneity have a better fit, but likelihood-ratio tests also clearly indicate that controlling for unobserved heterogeneity is actually important, and we find direct evidence of spurious negative duration dependence. Again, this issue poses problems for all the papers employing a Cox model, even though we note that some, such as Besedeš and Prusa (2006b), Besedeš (2008), and Nitsch (2009) have taken steps to tackle the problem by either including exporter-specific dummy variables or by employing a stratified Cox model.

We have, lastly, found that the assumption of proportional hazards is rejected even when unobserved heterogeneity is accounted for. In addition, we find direct evidence that incorrectly making the assumption of proportional hazards causes biases in the estimated covariate effects. Two other papers in the literature have in some way approached the issue of proportional hazards, but they have each only dealt with one of the two potential reasons why the assumption may not hold. On the one hand, Brenton, Saborowski and von Uexküll (2009) employ a discrete-time cloglog model incorporating random effects to account for unobserved heterogeneity. While tackling the effect of omitted regressors, their model does not allow for intrinsic non-proportionality. On the other hand, Fugazza and Molina (2009) apply an extended version of the Cox model with time-varying coefficients, thereby allowing for intrinsically non-proportional covariate effects. Solving the problem of intrinsic non-proportionality, their approach does not allow them to control for unobserved heterogeneity, so here the other reason for the proportional hazards assumption to fail is not dealt with. Since our testing procedure clearly suggest that unobserved heterogeneity should be controlled for *and* that the proportional hazards assumption still does not hold, we argue that neither approach is recommendable.

Thus, considering each of the potential problems with the Cox model in our specific empirical application, in all cases we find evidence in support of the arguments against the Cox model. This implies that the problems discussed with continuous-time models are most likely not just purely theoretical issues that can safely be ignored by empirical researchers. Instead, researchers that choose to use a Cox model for their analysis of trade durations will run a serious risk of reaching wrong conclusions regarding both the predicted termination probabilities of trade relationships and the estimated effects of explanatory variables on the hazard. Considering that the theoretically more appropriate discrete-time methods are easily implemented using standard statistical software packages, there are good reasons to use one of these models. Altogether, our findings suggest that non-proportional hazard specifications – such as logit or probit – including random effects should be preferred over the proportional Cox and cloglog specifications.

References

ABBRING, J. H. and VAN DEN BERG, G. J. (2007), "The unobserved heterogeneity distribution in duration analysis", *Biometrika*, vol. 94, pp. 87–99.

ALLISON, P. D. (1982), "Discrete-time methods for the analysis of event histories", *Sociol Methodol*, vol. 13, pp. 61–98.

AMEMIYA, T. (1981), "Qualitative response models: a survey", *J Econ Lit*, vol. 19, pp. 1483–1536.

BAKER, M. and MELINO, A. (2000), "Duration dependence and nonparametric heterogeneity: a Monte Carlo study", *J Econometrics*, vol. 96, pp. 357–393.

BESEDEŠ, T. (2008), "A search cost perspective on formation and duration of trade", *Rev Int Econ*, vol. 16, pp. 835–849.

BESEDEŠ, T. and PRUSA, T. J. (2006a), "Ins, outs and the duration of trade", *Can J Econ*, vol. 39, pp. 266–295.

BESEDEŠ, T. and PRUSA, T. J. (2006b), "Product differentiation and duration of US import trade", *J Int Econ*, vol. 70, pp. 339–358.

BESEDEŠ, T. and PRUSA, T. J. (2007), "The Role of Extensive and Intensive Margins and Export Growth", *NBER Working Paper No. 13628.*

BRENTON, P., PIEROLA, M. D. and VON UEXKÜLL, E. (2009), "The Life and death of trade flows: understanding the survival rates of developing-country exporters", in R. Newfarmer, W. Shaw and P. Walkenhorst (eds.), *Breaking Into New Markets: Emerging Lessons for Export Diversification*, pp. 127–144, Washington D.C.: The World Bank.

BRENTON, P., SABOROWSKI, C. and VON UEXKÜLL, E. (2009), "What explains the low survival rate of developing country export flows", *World Bank Policy Research Working Paper No. 4951.*

BRESLOW, N. (1974), "Covariance analysis of censored survival data", *Biometrics*, vol. 30, pp. 89–99.

BURR, I. W. (1942), "Cumulative frequency functions", *Ann Math Stat*, vol. 13, pp. 215–232.

BUTLER, J. S. and MOFFITT, R. (1982), "A computationally efficient quadrature procedure for the one-factor multinomial probit model", *Econometrica*, vol. 50, pp. 761–764.

COX, D. R. (1972), "Regression models and life-tables (with discussion)", *J R Stat Soc (Series B)*, vol. 34, pp. 187–220.

Cox, D. R. and Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman & Hall/CRC.

Dolton, P. and van der Klaauw, W. (1995), "Leaving teaching in the UK: a duration analysis", *Econ J*, vol. 105, pp. 431–444.

Efron, B. (1977), "The efficiency of Cox's likelihood function for censored data", *J Am Stat Assoc*, vol. 72, pp. 557–565.

Fugazza, M. and Molina, A. C. (2009), "The determinants of trade survival", *HEID Working Paper No. 05/2009.*

Heckman, J. J. and Singer, B. (1984a), "Econometric duration analysis", *J Econometrics*, vol. 24, pp. 63–132.

Heckman, J. J. and Singer, B. (1984b), "A method for minimizing the impact of distributional assumptions in econometric models for duration data", *Econometrica*, vol. 52, pp. 271–320.

Hess, W. (2009), "A flexible hazard rate model for grouped duration data", *Working Paper No. 2009:18, Department of Economics, Lund University.*

Hsieh, F. Y. (1995), "A cautionary note on the analysis of extreme data with Cox regression", *Am Stat*, vol. 49, pp. 226–228.

Jenkins, S. P. (1995), "Easy estimation methods for discrete-time duration models", *Oxford B Econ Stat*, vol. 57, pp. 129–137.

Kalbfleisch, J. D. and Prentice, R. L. (1973), "Marginal likelihoods based on Cox's regression and life model", *Biometrika*, vol. 60, pp. 267–278.

Kalbfleisch, J. D. and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: Wiley.

Lancaster, T. (1979), "Econometric methods for the duration of unemployment", *Econometrica*, vol. 47, pp. 939–956.

Lancaster, T. and Nickell, S. (1980), "The analysis of re-employment probabilities for the unemployed", *J R Stat Soc (Series A)*, vol. 143, pp. 141–165.

Lechner, M., Lollivier, S. and Magnac, T. (2008), "Parametric binary choice models", in L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, pp. 215–245, Berlin: Springer.

McCall, B. P. (1994), "Testing the proportional hazards assumption in the presence of unmeasured heterogeneity", *J Appl Econom*, vol. 9, pp. 321–334.

Meyer, B. D. (1990), "Unemplyment insurance and unemployment spells", *Econometrica*, vol. 58, pp. 757–782.

Mroz, T. A. and Zayats, Y. V. (2008), "Arbitrarily normalized coefficients, information sets, and false reports of "biases" in binary outcome models", *Rev Econ Stat*, vol. 90, pp. 406–413.

Nicoletti, C. and Rondinelli, C. (2009), "The (mis)specification of discrete duration models with unobserved heterogeneity: a Monte Carlo study", *Bank of Italy, Temi di Discussione (Working Papers) No. 705.*

Nitsch, V. (2009), "Die another day: duration in German import trade", *Rev World Econ*, vol. 145, pp. 133–154.

Prentice, R. L. (1975), "Discrimination among some parametric models", *Biometrika*, vol. 62, pp. 607–614.

Prentice, R. L. (1976), "A generalization of the probit and logit methods for dose response curves", *Biometrics*, vol. 32, pp. 761–768.

Prentice, R. L. and Gloeckler, L. A. (1978), "Regression analysis of grouped survival data with application to breast cancer data", *Biometrics*, vol. 34, pp. 57–67.

Salant, S. W. (1977), "Search theory and duration data: a theory of sorts", *Q J Econ*, vol. 91, pp. 39–57.

Scheike, T. H. and Sun, Y. (2007), "Maximum likelihood estimation for tied survival data under Cox regression model via EM-algorithm", *Lifetime Data Anal*, vol. 13, pp. 399–420.

Schoenfeld, D. (1982), "Partial residuals for the proportional hazards regression model", *Biometrika*, vol. 69, pp. 239–241.

Singer, J. D. and Willett, J. B. (1993), "It's about time: using discrete-time survival analysis to study duration and the timing of events", *J Educ Stat*, vol. 18, pp. 155–195.

Sueyoshi, G. T. (1995), "A class of binary response models for grouped duration data", *J Appl Econom*, vol. 10, pp. 411–431.

Tadikamalla, P. R. (1980), "A look at the Burr and related distributions", *Int Stat Rev*, vol. 48, pp. 337–344.

Train, K. E. (2003), *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press.

Trussell, J. and Richards, T. (1985), "Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure", *Sociol Methodol*, vol. 15, pp. 242–276.

van den Berg, G. J. (2001), "Duration models: specification, identification, and multiple durations", in J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics, Vol. 5*, pp. 3381–3460, Amsterdam: North-Holland.

Vaupel, J. W., Manton, K. G. and Stallard, E. (1979), "The impact of heterogeneity in individual frailty on the dynamics of mortality", *Demography*, vol. 16, pp. 439–454.

Vaupel, J. W. and Yashin, A. I. (1985), "Heterogeneity's ruses: some surprising effects of selection on population dynamics", *Am Stat*, vol. 39, pp. 176–185.