Consistent Evidence on Duration Dependence of Price Changes*

Fernando Alvarez
University of Chicago

Katarína Borovičková
New York University

Robert Shimer
University of Chicago

March 16, 2021

Abstract

We consider a discrete time mixed proportional hazard (MPH) model of duration. We prove that the baseline hazard and the frailty distribution of unobserved heterogeneity are nonparametrically identified using multiple-spell data, and use this to develop a GMM estimator of the baseline hazard. Our approach imposes no restrictions on the shape of the baseline hazard or the unobserved frailty distribution, allows for competing risks and spell-specific observable characteristics, and applies to right-censored data. The GMM specification is linear in the baseline hazard, which makes estimation and inference straightforward. We also develop tests of whether the MPH model is the data generating process.

We apply our estimation procedure to the duration of price spells in weekly store data from IRI. Our setup allows us to integrate filters for sale prices into our statistical model. In contrast to most of the existing literature, we find substantial unobserved heterogeneity, accounting for a large fraction of the decrease in the Kaplan-Meier hazard over time. The baseline hazard of regular price changes is mostly flat except for a notable spike near one year duration.

*We are grateful to David Argente, Jaroslav Borovička, Stéphane Bonhomme, Luca Dedola, Mark Gertler, Simon Gilchrist, Yuriy Gorodnichenko, Lars Hansen, Hervé Le Bihan, John Leahy, Simon Mongey, Emi Nakamura, Diego Perez, Edouard Schaal, Alex Torgovitsky, Iván Werning and participants at Berkeley, NYU, and ECB seminars for comments and suggestions. We thank the excellent research assistance of Aleksei Oskolkov and Olivia Bordeu Gazmuri. This material is based, in part, on work supported by the National Science Foundation under grant number SES-1559459 and SES-1559225. An earlier version of this paper circulated under the title "The Proportional Hazard Model: Estimation and Testing using Price Change and Labor Market Data".

version: March 16, 2021 9:13pm -05:00

1 Introduction

We propose and analyze a discrete time mixed proportional hazard (MPH) model for duration data (Cox, 1972; Lancaster, 1979). We develop a simple and robust method for estimating and testing the model using the Generalized Method of Moments (GMM).

We then study the elapsed time between price changes in a large panel of retail prices at the universal product code (UPC) level. Using our method, we find a decreasing hazard of price changes at the product level, but one that is much flatter than is found with commonly-used methods. This reflects the fact that our approach uncovers evidence of much more unobserved heterogeneity.

To better understand the decreasing hazard, we extend our model and estimation method to a competing risks framework with spell-specific observable characteristics, which allows us to distinguish between whether a price spell starts and ends with a price increase or a price decrease. We find that the baseline hazard for spells which start and end with a price increase is initially decreasing and is flat after duration of 6 weeks; the baseline hazard for spells that end with a price decrease are slowly declining; and the baseline hazard for spells that start with a price decrease and end with a price increase is sharply decreasing. We relate these findings to existing theories of price rigidities.

The Discrete Time MPH. We start with a general description of the discrete time MPH model. We assume that the probability that a price changes t periods after the last price change, conditional on not having changed earlier, is equal to θb_t. The *frailty* parameter θ is product-specific and fixed over time for a product. We refer to it as the product type and assume throughout that it is unobserved. The value of b_t is the *baseline hazard*. It is common across products, but can vary arbitrarily with the elapsed time since the last price change. The MPH model assumes that, conditional on the product type θ, the durations of any two spells are independently and identically distributed. Thus the model is completely specified by a value of b_t for each duration t and a distribution G for the frailty parameter θ. Finally, we assume that the hazard θb_t is bounded strictly above zero, which ensures that all spells end in finite time.

We do two standard exercises with the MPH model. First, we compare the shape of the baseline hazard b_t as a function of t with the shape of hazards obtained in theoretical models for a given product. For example, in Calvo (1983), each product has a constant hazard, and in Golosov and Lucas (2007) each product has an increasing hazard. Second, we quantify the importance of unobserved heterogeneity in shaping the Kaplan-Meier survivor function. Let H_t be the hazard of the Kaplan-Meier survivor function, hereafter the Kaplan-Meier hazard, of those products with duration t. Then $H_t = b_t \mathbb{E}[\theta|t]$ where $\mathbb{E}[\theta|t]$ is the mean frailty parameter among those products that have survived with unchanged prices until duration t. Moreover, $\mathbb{E}[\theta|t]$ must be decreasing with duration t, and more so if the distribution among surviving types has higher variance. Thus, heterogeneity in the frailty parameter necessarily pushes down the Kaplan-Meier hazard over time.

Estimation and Testing. In our methodological contribution, we adapt and extend known results from the continuous time MPH model to our discrete time setup. This allows us to frame estimation and testing as a GMM problem. First, we assume that the MPH model is correctly specified only for a range of durations $\{\underline{T}, \ldots, \bar{T}\}$. Then we turn the proof of identification in Honoré (1993) into a large set of moment conditions for the vector of baseline hazards $\boldsymbol{b} \equiv \{b_{\underline{T}}, \ldots, b_{\bar{T}}\}$. The conditions are linear in $\boldsymbol{b}$. We prove global identification of the parameters of interest, which implies that our GMM estimator is consistent. Importantly, the moment conditions give us a computationally straightforward and robust nonparametric estimator of the baseline hazard. Finally, we prove that if $\bar{T} > \underline{T} + 1$, the number of linearly independent moments conditions is greater than the number of parameters in the vector $\boldsymbol{b}$, so the model is overidentified.

We then extend the basic framework to allow for competing risks and spell-specific observable characteristics and again propose a GMM estimator for this richer model.

Our estimator allows that each product i is observed by the researcher for c^i periods, which is arbitrarily correlated with i's unobserved type θ. Depending on the censoring time c^i and the random realized duration of spells, the researcher observes K^i spells for product i, with the last spell right-censored. We apply our results to estimate the $\bar{T} - \underline{T} + 1$ baseline hazards in the vector $\boldsymbol{b}$.

By casting estimation and testing as a GMM problem, inference becomes straightforward. In particular, estimating the baseline hazards $\boldsymbol{b}$ and obtaining their standard errors is computationally feasible for data sets of any size and for any number of periods $T = \bar{T} - \underline{T} + 1$, since the moment conditions are linear in the parameter vector $\boldsymbol{b}$. This method is much simpler and more robust than commonly-used maximum likelihood alternatives. In particular, it is simpler in that linear GMM is guaranteed to find the minimum of the objective function, even with a large data set and large number of periods T. Also it is robust, in that it does not require us to assume a functional form for the frailty distribution.

Furthermore, we implement two tests of the underlying assumptions of the MPH model. First is the Sargan-Hansen J-test for over-identifying restrictions. Second, in the model without competing risks, we test a vector of inequalities describing dynamic sorting condition, i.e. that the average surviving type, defined as $\mathbb{E}[\theta|t]$, is decreasing in the duration t. We implement the multiple inequality test using the statistic developed by Chen and Szroeter (2014).

Data. For our benchmark estimation, we use the IRI weekly store data, which record weekly average revenue and weekly quantities. We define a product as a combination of a store and UPC code. This is a large data set, covering 30 categories of mostly packaged products, e.g. razor blades, coffee, beer, and frozen pizza. We define a "price spell" as the time between two price changes of a product. We also explore a data set from Cavallo (2018) which, while much smaller in size (250,000 products), has daily frequencies and arguably much less measurement error.

Results. Figure 1 shows that the baseline hazard b_t drops sharply for the first ten weeks before it starts to flatten. In contrast, the Kaplan-Meier hazard H_t is steeply decreasing throughout the first 40 weeks. Both hazards show a jump at durations around one year. As a result, the average type $\mathbb{E}[\theta|t] = h_t/b_t$, which we normalize to 1 at $t = 2$, steadily decreases to about 0.2 after one year. The patterns for the baseline hazard is common in most product categories, and the one for the average type holds for essentially all categories.

We implement the J-test for the overidentifying restrictions on the baseline hazard and we reject them at any conventional significance level with our data set. While visually the average type seems to be decreasing, we reject the hypothesis that it is decreasing at all durations due to very tight standard errors.

We then estimate a richer model of price changes. We assume that there are four different baseline hazards, distinguishing between whether a price spell starts and ends with a price increase or price decrease, and use the competing risks framework with observable characteristic to estimate them. We find that the baseline hazard for two consecutive price increases is decreasing until duration of 6 weeks, is flat between 7 and 45 weeks, and then shows a spike at around one year. This is consistent with models of price plans where a firm switches costlessly between prices of a given price plan but faces some rigidity for switching the plans. The baseline hazard for two consecutive price decreases is mildly decreasing until duration of 60 weeks. We interpret these two hazards as representing regular price changes. Even though the J-test rejects the MPH structure for these two risks, rejection is much milder than in the baseline model. We also examine whether there is a systematic violation of moment conditions for these two hazards, failing to detect any, leading us to conclude that the MPH framework useful tool to study the data.

We interpret the remaining two hazards, one associated with a price increase following a price decrease and vice-versa, as representing temporary price changes, including sales. We find that these hazards are steeply declining. The J-test strongly rejects these hazards as having an MPH structure, especially at short durations, and we are therefore cautious in interpreting this finding within existing theories.

An advantage of our setup is that it allows us to integrate filters for sales prices into our statistical model. The usual approach is to first drop all prices associated with sales from the data and conduct analysis with the remaining data. However, there is a concern that doing so affects estimated stochastic process for the remaining prices. Our framework allows us to model sales as part of the data. Importantly, it is not necessary to assume that sales have an MPH structure in order to get consistent estimates for the baseline hazard of regular price changes.

Given our use of a discrete time model, it is natural to wonder whether the results are sensitive to the length of the intervening time period. Figure 5 shows that in Cavallo's (2018) daily data set, there are large spikes in the hazard of price changes every seventh day, while the hazard is very low on the other days. This means that time aggregation remains an important feature of the data on the days when most prices change. Put differently, it is natural to measure price spells at the same frequencies that firms use to adjust prices.

Related Literature. Lancaster (1990) is one of the pioneers in the analysis of the continuous time MPH model. The main contributions in terms of non-parametric identification using single spell and covariates are Elbers and Ridder (1982); Heckman and Singer (1984) and Heckman and Honoré (1989) for competing risks models. The main contribution on the non-parametric identification using repeated spells is Honoré (1993). Abbring and Van Den Berg (2003) then extend this result to the competing risks model. Based on Honoré (1993)'s identification argument, Horowitz and Lee (2004) develop an estimator for two-spell data, and show how to conduct inference for the case of continuous time and continuous measurement. Their estimator, like ours, does not require specification of the frailty distribution and can be applied to censored data. The advantage of our estimator is that it is formulated for the data measured in discrete times, which is the usual format of duration datasets and that due its linearity, can use more than two spells in the estimation and is simpler to implement.

On the application of duration data to price spells we, like many others, follow the seminal work of Bils and Klenow (2004). Our findings in the IRI data give much more importance to unobserved heterogeneity than the existing literature. Perhaps the most careful analysis of the shape of the baseline hazards in the presence of unobserved heterogeneity for price changes is Fougere, Le Bihan, and Sevestre (2007) and Nakamura and Steinsson (2008). Nakamura and Steinsson (2008) conclude that the baseline hazard is steeply declining and there is not much heterogeneity across products. Our results suggest that the difference between their finding and ours is the estimation method. They use the likelihood function of a continuous time MPH model, but do not account for time aggregation in their duration data. We show in the last section that ignoring time aggregation would significantly bias our results and lead to similar conclusions.

The closest paper in terms of application is Fougere, Le Bihan, and Sevestre (2007) who estimate the baseline hazard for almost 400 product categories of similar aggregation as ours. They find very little evidence of unobserved heterogeneity within these categories. They test whether the baseline hazard is constant and fail to reject this hypothesis in more than half of the categories. These results contrast with ours. We think the main reason is the lower frequency of their data (their price data is gathered monthly while ours is gathered weekly), and more importantly, much fewer price spells per category, roughly three orders of magnitude less than us. Hence, they have less power to reject the null hypothesis of a constant baseline hazard. Fougere, Le Bihan, and Sevestre (2007) also specify and estimate a competing risks model where a price spell can end with a price increase or a price decrease. Differently than in our case, they estimate this model without unobserved heterogeneity and conclude that this extension barely affects the shape of the baseline hazard. Finally, they note that they could not find estimates for the competing risks model with unobserved heterogeneity, i.e. for the majority of the categories they fail to reach a maximum. Our GMM estimator is linear in the baseline hazard even in the competing risks extension, and hence it is ensured to converge, since its computation involves a simple matrix inversion.

2 Discrete Time MPH

2.1 Model

We consider a set of products with measure 1. Each product has a fixed type θ with cumulative distribution function $G(\theta)$, also known as the *frailty distribution*. The fixed type may be correlated with some observable individual characteristics, but we are interested in cases where the econometrician does not observe θ perfectly. For expositional simplicity, we focus on the case where the econometrician does not observe any individual characteristics.

Time is discrete and the amount of time between price changes is a random variable taking values in the positive integers $\{1, 2, \dots\}$. We call this elapsed time the spell length. The MPH model specifies that conditional on a spell length at least equal to t, the probability that the length is exactly t, i.e. the hazard at duration t, is the product of two components, the product's type θ and the *baseline hazard* b_t, which is common to all products. We assume that the frailty distribution $G(\theta)$ has a bounded support $[\theta_L, \theta_H]$ where $\theta_L \geq 0$ and $\theta_H b_t < 1$ for all $t = 1, 2, \ldots$.

The primitives of the model are the sequence of baseline hazards $\{b_1, b_2 \ldots\}$ and the frailty distribution G. Together they determine the distribution of spell lengths τ in the population. The duration distribution can be described by its cumulative distribution function, or equivalently, by its survival function $\Phi(t) \equiv \Pr[\tau > t]$, given by

$$\Phi(t) = \int \prod_{s=0}^{t} (1 - \theta b_s) dG(\theta),$$

where for notational convenience we define $b_0 = 0$.

Only the product of the baseline hazard b_t and the type θ enters the survival function. This implies that the model is homogeneous in $\{b_t\}$ and θ. That is, we can multiply the baseline hazard at all durations by a positive multiplicative constant λ and divide the type of each product by λ without affect the probability of any outcome. In what follows, we therefore identify the baseline hazard up to a multiplicative constant.

The Kaplan-Meier hazard is the probability that the spell length is exactly t conditional on it being at least t, but not otherwise conditional on the product's type:

$$H_t \equiv \frac{\Pr[\tau = t]}{\Pr[\tau \geq t]} = \frac{\Phi(t-1) - \Phi(t)}{\Phi(t-1)} = b_t \frac{\int \theta \prod_{s=0}^{t-1} (1 - \theta b_s) dG(\theta)}{\int \prod_{s=0}^{t-1} (1 - \theta b_s) dG(\theta)}. \tag{1}$$

This is the baseline hazard b_t times the *average type* among those products with spell length at least t, $\frac{\int \theta \prod_{s=0}^{t-1}(1-\theta b_s) dG(\theta)}{\int \prod_{s=0}^{t-1}(1-\theta b_s) dG(\theta)}$. This gives a clear decomposition of the evolution of the Kaplan-Meier hazard H_t into the component explained by structural duration dependence, captured through the baseline hazard b_t, and the component explained by dynamic selection of heterogeneous products, captured through changes in average type over time.

An implication of the MPH model is that the average type declines with duration:

Proposition 1 *The ratio of the Kaplan-Meier hazard to the baseline hazard, H_t/b_t, is decreasing in t.*

Proof. We let $g(\theta|t)$ be the distribution of θ among those products whose duration is at least t,

$$g(\theta|t) = \frac{\prod_{s=0}^{t-1} (1 - \theta b_s)\, g(\theta)}{\int \prod_{s=0}^{t-1} (1 - \theta' b_s)\, g(\theta') d\theta'}.$$

Consider the ratio of the densities at two values of θ for a given t:

$$\frac{g(\theta_2|t)}{g(\theta_1|t)} = \frac{g(\theta_2)\prod_{s=0}^{t-1}(1-\theta_2 b_s)}{g(\theta_1)\prod_{s=0}^{t-1}(1-\theta_1 b_s)} = \frac{g(\theta_2)}{g(\theta_1)}\prod_{s=0}^{t-1}\frac{1-\theta_2 b_s}{1-\theta_1 b_s}.$$

If $\theta_1 < \theta_2$, $1-\theta_2 b_s < 1-\theta_1 b_s$ and so this is decreasing in t. It follows that, as t increases, the distribution function $G(\theta|t)$ shifts towards lower θ in the sense of first order stochastic dominance and so its mean falls. Thus, the average type $\frac{\int \theta \prod_{s=0}^{t-1}(1-\theta b_s)dG(\theta)}{\int \prod_{s=0}^{t-1}(1-\theta b_s)dG(\theta)}$ is decreasing in t. Using equation (1), this is equal to H_t/b_t, and hence that ratio must decrease with t as well. ■

This result reflects dynamic sorting and is intuitive: products with a higher type have a higher chance of changing their price early and thus exit the pool of surviving products. As duration increases, products with lower type disproportionately remain. Lancaster (1979) discusses the same point in a related continuous time setup.

2.2 Identification with Multi-Spell Data

We now show that the model is non-parametrically identified with data on two spells. Define the survival function $\Phi(t_1, t_2) = Pr[\tau_1 > t_1, \tau_2 > t_2]$ as the probability that the first spell length is greater than t_1 and the second spell length is great than t_2.[1] The MPH model implies that

$$\Phi(t_1, t_2) = \int \left(\prod_{s=0}^{t_1}(1-\theta b_s)\right)\left(\prod_{s=0}^{t_2}(1-\theta b_s)\right) dG(\theta). \tag{2}$$

This captures the assumption that the length of the two spells is independent conditional on the product type θ.

We show how to recover the baseline hazard $\boldsymbol{b} \equiv \{b_1, b_2, \dots\}$ and the frailty distribution G from the survivor function $\Phi(t_1, t_2)$ for all $(t_1, t_2) \in \{0, 1, \dots\}^2$, up to the aforementioned multiplicative constant. Our proof is an adaptation of the identification result of Honoré (1993) to the discrete time model.

Proposition 2 *Assume $b_t > 0$ for all t. The baseline hazard $\boldsymbol{b}$ and the frailty distribution G are identified up to a multiplicative constant using the survivor function $\Phi(t_1, t_2)$.*

Proof. We first show how to identify the baseline hazard and then show how to identify the frailty distribution.

[1]If the duration distribution is defective, $\lim_{t\to\infty}\Phi(t) \geq 0$, there is a positive probability that we would not observe a second spell. Nevertheless, the model allows us to construct $\Phi(t_1, t_2)$, even if it cannot be observed with infinitely much data. Our estimator handles defective duration distributions.

Baseline Hazard. The definition of the survivor function in equation (2) implies

$$\Phi(t_1-1,t_2-1)-\Phi(t_1,t_2-1)=b_{t_1}\int\theta\left(\prod_{s=0}^{t_1-1}(1-\theta b_s)\right)\left(\prod_{s=0}^{t_2-1}(1-\theta b_s)\right)dG(\theta), \tag{3}$$

and symmetrically for $\Phi(t_1-1,t_2-1)-\Phi(t_1-1,t_2)$. Thus using $(t_1,t_2)=(t,1)$ for some $t>1$,

$$\frac{\Phi(t-1,0)-\Phi(t,0)}{\Phi(t-1,0)-\Phi(t-1,1)}=\frac{b_t\int\theta\left(\prod_{s=0}^{t-1}(1-\theta b_s)\right)dG(\theta)}{b_1\int\theta\left(\prod_{s=0}^{t-1}(1-\theta b_s)\right)dG(\theta)}=\frac{b_t}{b_1}, \tag{4}$$

where we simply cancel the common nuisance term. This equation determines b_t/b_1 for any $t>1$. Thus we have found the baseline hazard up to a multiplicative constant.

Frailty Distribution. Let $\mu_k\equiv\int\theta^k dG(\theta)$ denote the k^{th} moment of the G distribution. It exists since the distribution is bounded. Once we know the baseline hazard $\boldsymbol{b}$ up to a multiplicative constant, the model implies that the probability $\tau=t$ is a known function of μ_k with $k=1,\dots,t$,

$$Pr[\tau=t]=b_t\int\theta\prod_{s=0}^{t-1}(1-\theta b_s)dG(\theta)=b_t\sum_{k=0}^{t-1}\alpha_k(t-1;\boldsymbol{b})\mu_{k+1}, \tag{5}$$

where for all $t,k\geq 1$, the coefficients $\alpha_k(t;\boldsymbol{b})$ are defined recursively as follows:

$$\alpha_k(t;\boldsymbol{b})=\begin{cases}1 & \text{if } k=0\\ 0 & \text{if } k>t\\ \alpha_k(t-1;\boldsymbol{b})-b_t\alpha_{k-1}(t-1;\boldsymbol{b}) & \text{if } t\geq k>1\end{cases} \tag{6}$$

We know $\boldsymbol{b}$. Setting $t=1$ in equation (5) gives us an equation for μ_1. Having found $\mu_1,\dots,\mu_{k-1}$, setting $t=k$ in equation (5) gives us an equation for μ_k. Thus by induction we can find all the moments μ_k of G. Since the support of G is a bounded interval $[\theta_L,\theta_H]$, its moments uniquely determine distribution G. ■

Proposition 2 is behind our approach to estimation, where we convert this logic into moment conditions for the case where we have measures of the survivor function from a finite sample. We turn to that next.

3 Estimation and Testing

In this section, we turn the identification result in Proposition 2 into moment conditions, and use GMM to non-parametrically estimate the baseline hazard and moments of the frailty distribution. We work in a more general environment, where there is an MPH data generating process for $t \in \{\underline{T}, \dots, \bar{T}\}$ but not necessarily for durations outside this interval. More precisely, let $h_t(\theta)$ be the hazard at duration t for a product with type θ. For $t \in \{\underline{T}, \dots, \bar{T}\}$, we assume $h_t(\theta) = \theta b_t$, but we allow for arbitrary $h_t(\theta) \in [0, 1]$ outside this interval. For notational convenience, let $h_0(\theta) = 0$ for all θ.

Our estimator works with right-censored duration data. It also allows for the possibility that the duration distribution only satisfies the proportional hazard assumption at some durations, that the distribution may be defective, and that some products may have only a single spell. On the other hand, it uses the information from more than two spells when that is available. Each of these possibilities is important in our empirical application.

3.1 Measurement

We start by layering measurement assumptions on top of the duration model. If we observed a product i with type θ^i for infinitely long, we would see a vector of durations $\boldsymbol{\tau}^i = \{\tau_1^i, \tau_2^i, \dots, \tau_{\bar{K}^i}^i\}$, where $\bar{K}^i$ is either a positive integer or infinite. This allows for the possibility that the duration distribution is defective, $\prod_{t=1}^{\infty}(1 - h_t(\theta^i)) > 0$, in which case $\bar{K}^i$ is almost surely finite and $\tau_{\bar{K}^i}^i = \infty$; otherwise, $\bar{K}^i = \infty$. In any case, for $j < \bar{K}^i$, τ_j^i is a strictly positive integer. Thus in either case, $\sum_{j=1}^{\bar{K}^i} \tau_j^i = \infty$.

In real-world data, we do not observe any product for infinitely long. Instead, we assume that we observe product i for c^i periods, where the censoring time c^i is a random variable, possibly correlated with the product type. We assume that the first spell starts when we first observe the product and ends τ_1^i periods later, while the second spell ends $\tau_1^i + \tau_2^i$ periods after we first observe the product, and so on.[2]

Since the censoring time c^i is finite, we only observe a finite number of spells, $K^i \leq \bar{K}^i$ satisfying $\sum_{j=1}^{K^i-1} \tau_j^i \leq c^i$ and $\sum_{j=1}^{K^i} \tau_j^i > c^i$, with $K^i = 1$ if $\tau_1^i > c^i$. Then measured duration is a K^i-vector $\boldsymbol{\zeta}^i \equiv (\zeta_1^i, \dots, \zeta_{K^i}^i)$, where for $j = 1, \dots, K^i - 1$, $\zeta_j^i \equiv \tau_j^i$ is the completed duration of the j^{th} spell, and $\zeta_{K^i}^i \equiv c^i - \sum_{j=1}^{K^i-1} \tau_j^i + 1$ is a lower bound on the duration of the final spell. Since $\sum_{j=1}^{K^i} \tau_j^i > c^i$, $\zeta_{K^i}^i \leq \tau_{K^i}^i$.

[2] Data may also be left-censored, with an in-progress spell when we first observe the product. Without knowing the duration of that initial spell, we cannot make use of the data it generates. Instead, we redefine c^i by subtracting the measured duration of the initial left-censored spell, i.e. by acting as if we only observe the product once the next spell starts.

We let $P(c)$ denote the cumulative distribution of censoring times and let $G_{\bar{c}}(\theta)$ denote the frailty distribution conditional on $c^i \geq \bar{c}$ for any duration $\bar{c}$. We stress that θ^i captures all factors which affect the true duration of spells $\boldsymbol{\tau}^i$, with $\prod_{s=0}^{t}(1 - h_s(\theta^i))$ denoting the probability that each spell's duration strictly exceeds t given θ^i. The censoring time c^i affects the measured duration of spells $\boldsymbol{\zeta}^i$.

We assume that the baseline hazard $\boldsymbol{b}_0 = \{b_{0,\underline{T}}, \ldots, b_{0,\bar{T}}\}$ is nontrivial, $\boldsymbol{b}_0 \neq 0$, and let T_0 is the smallest $t \geq \underline{T}$ with $b_{0,t} > 0$. We then impose the following rank condition which ensures that we have variation to estimate the baseline hazard:

Assumption 1 *$\zeta_1 = T_0$ and $\zeta_2 \geq \bar{T}$ with positive probability.*

This ensures we have variation in the data to compare $b_{\bar{T}}$ to b_{T_0}. It holds if and only if

$$\big(1 - P(T_0 + \bar{T} - 2)\big) \int \prod_{t=1}^{\bar{T}-1}(1 - h_t(\theta)) dG_{T_0+\bar{T}-1}(\theta) > 0.$$

This requires that the censoring time is long enough to allow us to observe a product with two spells, one of which is completed with duration T_0 and the other has duration at least equal to $\bar{T}$. If this were violated, we would be unable to estimate baseline hazard at duration $\bar{T}$ relative to T_0. It also requires that there is a positive probability that a product with censoring time at least equal to $T_0 + \bar{T} - 1$ has a spell that lasts at least $\bar{T}$ periods. If this were violated, we would naturally be unable to estimate $b_{\bar{T}}$.

When this combination of assumptions determines the distribution of measured duration $\boldsymbol{\zeta}$ and the rank condition is satisfied, we say that $\boldsymbol{\zeta}$ is drawn from a right-censored MPH model with baseline hazard $\boldsymbol{b}_0$.

3.2 Moment Conditions for the Baseline Hazard

We now construct a consistent estimate of the baseline hazard when measured duration $\boldsymbol{\zeta}$ is drawn from a right-censored MPH model with baseline hazard $\boldsymbol{b}$, up to a multiplicative constant reflecting the standard lack-of-identification in the MPH model.

In Section 2, we argued that for any duration t_1, t_2 the model implies

$$Pr[\tau_1 = t_1, \tau_2 \geq t_2] b_{t_2} = Pr[\tau_1 \geq t_1, \tau_2 = t_2] b_{t_1}.$$

If we observed two completed spells per product, it would be straightforward to turn this result into a moment condition:

$$\mathbb{E}\Big[\mathbb{1}_{\tau_1^i = t_1, \tau_2^i \geq t_2} b_{t_2} - \mathbb{1}_{\tau_1^i \geq t_1, \tau_2^i = t_2} b_{t_1}\Big] = 0,$$

since expected value of the indicator function is the probability of the corresponding event. Censoring affects our ability to use such conditions since τ_1^i and τ_2^i are not always observed. In particular, there is no function of data that is equivalent to $\mathbb{1}_{\tau_1^i \geq t_1, \tau_2^i = t_2}$. To see why, take a product with $\tau_1^i > c^i \geq t_1$, so the first spell is right-censored. This record does not depend at all on τ_2^i and in particular does not depend on whether $\tau_2^i = t_2$.

We use two key observations to circumvent this. Consider a product where we observe at least two spells, $K^i \geq 2$. First, if $c^i \geq t_1 + t_2$, we can evaluate whether the event $\tau_1^i = t_1, \tau_2^i \geq t_2$ occurred using objects we observe. This is because we see the product for long enough to tell if the first spell lasts exactly t_1 periods; and if it does, we see it long enough to tell if the second spell lasts at least t_2 periods. The second is that the model implies probabilities are symmetric across the two spells, so the events $\tau_1^i \geq t_1, \tau_2^i = t_2$ and $\tau_1^i = t_2, \tau_2^i \geq t_1$ are equally likely. While we cannot evaluate an indicator function for the first event, we can evaluate one for the second event for all individuals with $c^i \geq t_1 + t_2$. This motivates the following moment condition:

$$\mathbb{E}\left[\mathbb{1}_{K^i \geq 2, \zeta_1^i = t_1, \zeta_2^i \geq t_2} b_{t_2} - \mathbb{1}_{K^i \geq 2, \zeta_1^i = t_2, \zeta_2^i \geq t_1} b_{t_1}\right] = 0.$$

Our main result formalizes these observations and shows how to develop a moment condition that uses information from two arbitrary spells, not just the first two spells:

Proposition 3 *Assume* $\boldsymbol{\zeta}$ *is drawn from a right-censored MPH model with baseline hazard* $\boldsymbol{b}_0$*. Define*

$$f_{t_1,t_2}^{[b]}(\boldsymbol{\zeta}; \boldsymbol{b}) \equiv \sum_{(j,k):1 \leq j < k \leq K} \left(b_{t_2} \mathbb{1}_{\zeta_j = t_1, \zeta_k \geq t_2} - b_{t_1} \mathbb{1}_{\zeta_j = t_2, \zeta_k \geq t_1}\right). \tag{7}$$

Then $\mathbb{E}\left[f_{t_1,t_2}^{[b]}(\boldsymbol{\zeta}; \boldsymbol{b})\right] = 0$ *for all* $\underline{T} \leq t_1 < t_2 \leq \bar{T}$ *if and only if* $\boldsymbol{b} = \lambda \boldsymbol{b}_0$ *for some number* λ*.*

We postpone the proof of this Proposition, since we can obtain it as a special case of Proposition 5 below. See Appendix A for the proof of that proposition and the explanation for why Proposition 3 is a special case.

We use Proposition 3 to build a GMM estimator of $\boldsymbol{b}_0$ for some strictly positive λ. Let $T = \bar{T} - \underline{T}$. We have $T(T+1)/2$ moment conditions of the form $\mathbb{E}\left[f_{t_1,t_2}^{[b]}(\boldsymbol{\zeta}; \boldsymbol{b}_0)\right] = 0$ for some $\underline{T} \leq t_1 < t_2 \leq \bar{T}$, each linear in the $T+1$ vector $\boldsymbol{b}_0$. The basic idea of GMM is to replace the expected value with the sample mean, so we have the moment condition $\frac{1}{I}\sum_{i=1}^{I} f_{t_1,t_2}^{[b]}(\boldsymbol{\zeta}^i; \boldsymbol{b}) = 0$. We estimate $\boldsymbol{b}_0$ by minimizing the quadratic form of the error in the moment conditions, weighted by a positive-definite matrix W. The "if" part of Proposition 3 gives us the necessary condition for this estimator to be consistent, while the "only if" part gives us sufficiency. We discuss further details of the GMM estimator in Appendix B. Here we highlight one important feature of our approach: since the moment conditions in equation (7) are linear in the baseline hazard, we obtain the GMM estimator of $\boldsymbol{b}_0$ in closed form.

We can also build on the proof of Proposition 2 to find moment conditions for the moments of the type distribution. Unfortunately, unless we impose that the proportional hazard structure holds at the shortest duration, $\underline{T} = 1$, and that censoring time c^i and type θ^i are independent, these conditions are difficult to interpret. We therefore do not report them.

3.3 Testing

We recognize that the multiplicative structure of the MPH model might be restrictive and hence propose two tests of the model. First, Proposition 3 gives us $T(T+1)/2$ moment conditions to estimate T parameters, where we recognize the unidentified scaling factor λ. For $T \geq 2$, the model is thus overidentified. We conduct the J-test of overidentifying restrictions for the empirical counterpart of the baseline hazard. These conditions come from the structure of the model and the fact that we have more moment conditions than parameters.

Second, Proposition 1 tells us that the ratio of the Kaplan-Meier hazard to the baseline hazard, H_t/b_t, is decreasing in t. We seek to test this prediction, but first must discuss how we estimate H_t. While the baseline hazard b_t is the same for all products, the Kaplan-Meier hazard depends on the type distribution of the products we are examining. Here we define the Kaplan-Meier hazard for all products whose censoring time is at least equal to some $\bar{c}$, i.e. for the type distribution $G_{\bar{c}}(\theta)$.[3] This is defined as

$$H_t^{\bar{c}} \equiv \frac{\Pr[\tau = t | c \geq \bar{c}]}{\Pr[\tau \geq t | c \geq \bar{c}]} = \frac{\Pr[\tau = t, c \geq \bar{c}]}{\Pr[\tau \geq t, c \geq \bar{c}]}.$$

The second equation uses the definition of a conditional probability.

Now define $\boldsymbol{H}^{\bar{c}} = (H_{\underline{T}}^{\bar{c}}, \ldots, H_{\bar{T}}^{\bar{c}})$. We can construct a consistent estimator of the Kaplan-Meier hazard $\boldsymbol{H}^{\bar{c}}$ for any $\bar{c} \geq \bar{T}$. In particular,

Proposition 4 *Assume* $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_K)$ *is drawn from a right-censored MPH model with Kaplan-Meier hazard* $\boldsymbol{H}_0^{\bar{T}}$. *Define*

$$f_{t,\bar{T}}^{[H]}\left(\boldsymbol{\zeta}; \boldsymbol{H}^{\bar{T}}\right) \equiv H_t^{\bar{T}} \mathbb{1}_{\zeta_1 \geq t, c \geq \bar{T}} - \mathbb{1}_{\zeta_1 = t, c \geq \bar{T}} \tag{8}$$

where $c = \sum_{j=1}^{K} \zeta_j - 1$. *Then* $\mathbb{E}\left[f_{t,\bar{T}}^{[H]}\left(\boldsymbol{\zeta}; \boldsymbol{H}^{\bar{T}}\right)\right] = 0$ *for all* $\underline{T} \leq t \leq \bar{T}$ *if and only if* $\boldsymbol{H}^{\bar{T}} = \boldsymbol{H}_0^{\bar{T}}$.

[3] If we assume that censoring time c^i and type θ^i are independent, then $\bar{H}^{c_1} = \bar{H}^{c_2}$ for all c_1, c_2.

We omit the proof, which is straightforward once one recognizes that the restriction $c \geq \bar{T} \geq t$ implies that $\tau_1 \geq t$ if and only if $\zeta_1 \geq t$.

Finally, we test monotonicity of the ratio $H_t^{\bar{T}}/b_t$ by looking at the inequalities

$$\left(\log H_t^{\bar{T}} - \log b_t\right) - \left(\log H_{t+1}^{\bar{T}} - \log b_{t+1}\right) \geq 0 \quad \forall t = \underline{T}, \ldots, \bar{T} - 1.$$

This gives us T inequalities, which we test jointly using Chen and Szroeder (2009).

4 Extensions

We now consider two extensions to our basic model, allowing for spell-specific observable characteristics which affect the hazard, and permitting competing risks for why a spell ends. In our empirical application, the spell-specific characteristic is whether the spell starts with a price increase or decrease; and spells end for one of those reasons as well.

4.1 Setup

We assume that each product is characterized by an unobserved type vector $\boldsymbol{\theta}$ with population distribution $G(\boldsymbol{\theta})$. In addition, we assume that each product has an observable characteristic for the j^{th} spell, $\chi_j \in \{1, \ldots, X\}$.

Both the observed and unobserved characteristics affect the joint distribution of the duration of a spell and the reason why the spell ends. We let $h_t^r(\chi_j, \boldsymbol{\theta}) \geq 0$ denote the probability that a spell with observable χ_j and unobservable $\boldsymbol{\theta}$ ends at duration $t \in \{1, 2, \ldots\}$ for reason $r \in \{1, \ldots, R\}$ conditional on not ending earlier. χ_j captures all observables that affect the hazard, and so in particular conditioning on past observables is not useful for forecasting the hazard.[4] Let $h_t(x, \boldsymbol{\theta}) \equiv \sum_{r=1}^{R} h_t^r(x, \boldsymbol{\theta})$ denote the probability of a duration t spell ending in period t. We assume $h_t(x, \boldsymbol{\theta}) < 1$ for all t, x, and $\boldsymbol{\theta}$.

The initial observable characteristic χ_1 is a random variable. Let $\pi_1(x|\boldsymbol{\theta}) \geq 0$ denote the probability that $\chi_1 = x$ given $\boldsymbol{\theta}$, with $\sum_{x=1}^{X} \pi_1(x|\boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta}$. Thereafter, the observable characteristic follows a first order Markov process. Let $\pi(x|\chi_{j-1}, \rho_{j-1}, \boldsymbol{\theta}) \geq 0$ denote the probability that $\chi_j = x$ conditional on the reason the previous spell ended $\rho_{j-1} \in \{1, \ldots, R\}$, the observable characteristic of that spell $\chi_{j-1} \in \{1, \ldots, X\}$, and the unobserved type,[5] with $\sum_{x=1}^{X} \pi(x|\chi_{j-1}, \rho_{j-1}, \boldsymbol{\theta}) = 1$ for all χ_{j-1}, ρ_{j-1}, and $\boldsymbol{\theta}$. We note that

[4] To be precise, let $\hat{h}_t^r(\chi_1, \ldots, \chi_j, \boldsymbol{\theta})$ denote the probability that a spell with current and lagged observables $\chi_1, \ldots, \chi_j$ and unobservable $\boldsymbol{\theta}$ ends at duration $t \in \{1, 2, \ldots\}$ for reason $r \in \{1, \ldots, R\}$ conditional on not ending earlier. Then we assume $h_t^r(\chi_j, \boldsymbol{\theta}) = \hat{h}_t^r(\chi_1, \ldots, \chi_j, \boldsymbol{\theta})$. One can view this as a definition of the observable state χ_j.

[5] We assume a Markovian structure for notational simplicity, but can easily relax this assumption.

we assume that the duration of one spell does not directly affect the duration of later spells, i.e. we assume that there is no lagged duration dependence. Still, we allow for the possibility that the reason one spell ends can influence the duration of the next spell. This is important in our empirical application.

For at least one observable characteristic x, reason r, and set of durations $\{\underline{T}, \ldots, \bar{T}\}$, we assume that there is a proportional hazard representation, $h_t^r(x, \boldsymbol{\theta}) = \phi(\boldsymbol{\theta}) b_t$ for all $\underline{T} \leq t \leq \bar{T}$, where $\phi(\boldsymbol{\theta})$ is a scalar function of the unobserved type vector $\boldsymbol{\theta}$ and $b_t \geq 0$ for all $t \in \{\underline{T}, \ldots, \bar{T}\}$. We focus throughout on this pair (x, r) and seek to estimate $\boldsymbol{b} = \{b_{\underline{T}}, \ldots, b_{\bar{T}}\}$ up to a multiplicative constant.

We do not impose any restrictions on $h_t^{r'}(x', \boldsymbol{\theta})$ for $(x', r') \neq (x, r)$. However, we allow for the possibility that multiple hazards have a proportional hazard representation, with potentially different scaling functions ϕ and different baseline hazards b. In this case, we can jointly estimate all the baseline hazards. We note, however, that even if all the hazards have a proportional hazard representation, the hazard of a spell with characteristic x ending for any reason, $h_t(x, \boldsymbol{\theta})$, generally does not have a proportional hazard representation. Thus we may reject the MPH model but not fail to reject this more general specification.

4.2 Identification

Consider a product with two spells that both have observable characteristic x. Take a pair of durations satisfying $\underline{T} \leq t_1 < t_2 \leq \bar{T}$. Using $h_{t_1}^r(x, \boldsymbol{\theta}) = \phi(\boldsymbol{\theta}) b_{t_1}$, the model implies that the probability that the first spell ends exactly at duration t_1 for risk r, the second spell lasts at least t_2 periods, and both spells have observable x is

$$\begin{aligned}\Pr[\tau_1 = t_1, \tau_2 \geq t_2, \rho_1 = r, \chi_1 = \chi_2 = x] \\ = b_{t_1} \int \phi(\boldsymbol{\theta}) \pi_1(x|\boldsymbol{\theta}) \pi(x|x, r, \boldsymbol{\theta}) \prod_{s=1}^{t_1-1} (1 - h_s(x, \boldsymbol{\theta})) \prod_{s=1}^{t_2-1} (1 - h_s(x, \boldsymbol{\theta}))\, dG(\boldsymbol{\theta}).\end{aligned}$$

Reversing the role of t_1 and t_2 gives

$$\begin{aligned}\Pr[\tau_1 = t_2, \tau_2 \geq t_1, \rho_1 = r, \chi_1 = \chi_2 = x] \\ = b_{t_2} \int \phi(\boldsymbol{\theta}) \pi_1(x|\boldsymbol{\theta}) \pi(x|x, r, \boldsymbol{\theta}) \prod_{s=1}^{t_2-1} (1 - h_s(x, \boldsymbol{\theta})) \prod_{s=1}^{t_1-1} (1 - h_s(x, \boldsymbol{\theta}))\, dG(\boldsymbol{\theta}).\end{aligned}$$

Combining these two we get

$$b_{t_2} \Pr[\tau_1 = t_1, \tau_2 \geq t_2, \rho_1 = r, \chi_1 = \chi_2 = x] - b_{t_1} \Pr[\tau_1 = t_2, \tau_2 \geq t_1, \rho_1 = r, \chi_1 = \chi_2 = x] = 0,$$

and so an appropriate moment condition is

$$\mathbb{E}\left[b_{t_2}\mathbb{1}_{\tau_1=t_1,\tau_2\geq t_2,\rho_1=r,\chi_1=\chi_2=x} - b_{t_1}\mathbb{1}_{\tau_1=t_2,\tau_2\geq t_1,\rho_1=r,\chi_1=\chi_2=x}\right] = 0.$$

Setting $t_1 = \underline{T}$ and normalizing $b_{\underline{T}}$, we can then vary t_2 to recover $\boldsymbol{b}$. This generalizes the identification argument in Proposition 2 to framework with observable characteristics and competing risks.

We cannot use this moment condition for GMM in our setting because we have censored data. Moreover, it does not make use of available data after the end of the second spell. Still, we show how to adapt this insight to our framework using the approach in Section 3.2.

4.3 Measurement

As in the MPH model, we assume that we observe product i for c^i periods, where the censoring time c^i may be correlated with the product's type $\boldsymbol{\theta}^i$. We still let $P(c)$ denote the cumulative distribution of censoring times and $G_{\bar{c}}(\theta)$ denote the frailty distribution conditional on $c^i \geq \bar{c}$ for any duration $\bar{c}$.

As before, we let $\boldsymbol{\zeta}^i = (\zeta_1^i, \ldots, \zeta_{K^i}^i)$ be the vector of measured durations, with the last spell right censored, so $c^i = \sum_{i=1}^{K^i} \zeta_j^i - 1$. We also let $\boldsymbol{\chi}^i = (\chi_1^i, \ldots, \chi_{K^i}^i)$ be a vector recording the observable characteristic of each spell and $\boldsymbol{\rho}^i = (\rho_1^i, \ldots, \rho_{K^i-1}^i)$ be a vector recording the risk that ended each spell. Since the last spell is right-censored, we do not observe why it ended, and hence $\boldsymbol{\rho}$ is of length $K^i - 1$.

We assume that the baseline hazard $\boldsymbol{b}_0 = \{b_{0,\underline{T}}, \ldots, b_{0,\bar{T}}\}$ is nontrivial, $\boldsymbol{b}_0 \neq 0$, and let T_0 is the smallest $t \geq \underline{T}$ with $b_{0,t} > 0$. We then generalize the rank condition in Assumption 1 to the environment with observable characteristics and competing risks:

Assumption 2 *With positive probability, there exists a* $1 \leq j < k \leq K$ *with* $\zeta_j = T_0$, $\zeta_k \geq \bar{T}$, $\rho_j = r$, *and* $\chi_j = \chi_k = x$.

This guarantees that we have variation in the data to compare $b_{\bar{T}}$ to b_{T_0}. It holds, for example, if

$$\left(1 - P(T_0 + \bar{T} - 2)\right) \int \pi_1(x|\boldsymbol{\theta})\pi(x|x, r, \boldsymbol{\theta}) \prod_{t=1}^{\bar{T}-1}(1 - h_t(x, \boldsymbol{\theta}))dG_{T_0+\bar{T}-1}(\boldsymbol{\theta}) > 0,$$

so there is a positive probability that the censoring time is at least $T_0+\bar{T}-1$ and, conditional on such a censoring time, the first two spells have observable x and a spell can last at least $\bar{T}$ periods.

We highlight a few special cases in which this reduces to the rank condition in Assumption 1. First, the observable may be degenerate, $X = 1$, so $\pi(1|\boldsymbol{\theta}) = \pi(1|1, r, \boldsymbol{\theta}) = 1$ for all r and $\boldsymbol{\theta}$. Second, the observable distribution may have full support in each of the first two spells for all r and $\boldsymbol{\theta}$. This is the case in our empirical analysis when x measures whether the spell starts with a price increase or decrease and r measures whether it ends with a price increase or decrease. Third, the observable may have full support in the first period and then be degenerate, $\pi(x|x, r, \boldsymbol{\theta}) = 1$ for all x, r, and $\boldsymbol{\theta}$. This is the case in the empirical analysis when the observable characteristic measures the product's category. Of course, combinations of these assumptions are consistent with Assumption 1 as well, e.g. we can observe both the product's category and whether the spell starts with a price increase or decrease.

When this set of assumptions determine the joint distribution of $(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho})$ and the rank condition holds, we say that $(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho})$ is drawn from a right-censored competing-risk model with baseline hazard $\boldsymbol{b}_0$ for observable characteristic x and risk r.

4.4 Moment Conditions

We now show how to estimate the baseline hazard, extending the approach in Proposition 3:

Proposition 5 *Assume $(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho})$ is drawn from a right-censored competing-risk model with baseline hazard $\boldsymbol{b}_0$ for observable characteristic x and risk r. Define*

$$f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho}; \boldsymbol{b}) \equiv \sum_{(j,k):1 \leq j < k \leq K} \left(b_{t_k} \mathbb{1}_{\zeta_j = t_1, \zeta_k \geq t_2, \rho_j = r, \chi_j = \chi_k = x} - b_{t_j} \mathbb{1}_{\zeta_j = t_2, \zeta_k \geq t_1, \rho_j = r, \chi_j = \chi_k = x} \right). \tag{9}$$

Then $\mathbb{E}\left[f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho}; \boldsymbol{b}) \right] = 0$ for all $\underline{T} \leq t_1 < t_2 \leq \bar{T}$ if and only if $\boldsymbol{b} = \lambda \boldsymbol{b}_0$ for some number λ.

The proof is in Appendix A.

We turn next to the Kaplan-Meier hazard. Define the Kaplan-Meier hazard for observable characteristic x and risk r, $\boldsymbol{H}^{\bar{c}} = (H_{\underline{T}}^{\bar{c}}, \ldots, H_{\bar{T}}^{\bar{c}})$ where

$$\bar{H}_t^{\bar{c}} \equiv \frac{Pr(\tau_1 = t, \chi_1 = x, \rho_1 = r, c \geq \bar{c})}{Pr(\tau_1 \geq t, \chi_1 = x, c \geq \bar{c})}.$$

This is the probability that the first spell ends at duration t for risk r conditional on lasting at least t periods, having observable x, and the product's time in the sample exceeding $\bar{c}$.

Proposition 6 *Assume $(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho})$ is drawn from a right-censored competing-risk model with*

Kaplan-Meier hazard $\boldsymbol{H}_0^{\bar{T}}$ *for observable characteristic* x *and risk* r. *Define*

$$f_t^{[H,x,r]}\left(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{H}\right) \equiv H_t \mathbb{1}_{\zeta_1 \geq t, \chi_1 = x, c \geq \bar{T}} - \mathbb{1}_{\zeta_1 = t, \chi_1 = x, \rho_1 = r, c \geq \bar{T}} \tag{10}$$

Then $\mathbb{E}\left[f_t^{[H,x,r]}\left(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{H}\right)\right] = 0$ *for* $\underline{T} \leq t \leq \bar{T}$ *if and only if* $\boldsymbol{H} = \boldsymbol{H}_0^{\bar{T}}$.

We omit the straight-forward proof.

5 Data

In the empirical application, we use IRI weekly store data,[6] described in Bronnenberg, Kruger, and Mela (2008).

5.1 Construction of Price Spells

The IRI data set contains weekly revenue and quantity sold for a large number of products in chain grocery and drug stores for years 2001–2011. The data cover 30 large product categories[7] (coffee, carbonated beverages, detergents, for example) and include approximately 2.6 million distinct items defined by its barcode (Universal Product Code, UPC).

We define a product as a particular UPC in a particular store. We use revenue and quantity sold to compute the average weekly price for each product. We turn data on price levels into data on price spells by first computing the price changes and then defining the price spell as the time elapsed between two price changes. In particular, suppose that price changes occur at times $t_0, t_1, \ldots, t_{K-1}$ and that the last price observation is at time t_K. Then we construct $\zeta_j = t_j - t_{j-1}$ for $j = 1, \ldots, K-1$, $\zeta_K = t_K - t_{K-1} + 1$, where the latter reflects the fact that the earliest possible date when the last price can change is $t_K + 1$ and the hence the price will be at least $t_K - t_{K-1} + 1$ periods long. The censoring time is $c = t_K - t_0$.

We use price levels to determine whether the price spell follows a price decrease or increase; we can determine this for every price spell. We further use prices to determine whether a price spell ends with a price increase or decrease; we can construct such indicator for every complete spell. We use this information to estimate a richer model with observable characteristics and competing risks.

Missing observations are prevalent. For example, if the product has not been sold in a given week, the store does not report quantity or revenue. We address this problem by

[6]All estimates and analyses in this paper based on Information Resources Inc. data are by the authors and not by Information Resources Inc.

[7]There are 31 product categories in IRI but we exclude cigarettes from our analysis because their price is regulated.

selecting the longest period with no missing observations for a given product and use only this interval to construct price spells.

We work with average weekly prices which brings in two issues for the spell construction. First, some changes in average weekly price are due to the fact that some customers shop with coupons, which we cannot directly observe. We impose a lower bound on the price change of 0.1 percent to exclude such price changes. Second, if the product's price changes in the middle of a week, it generates a spurious spell of duration one week.[8] We therefore set $\underline{T} = 2$ and do not estimate the baseline hazard in week 1.

Any price spell occurring before the first observed price change is left-censored. We exclude such spell but we use the price level of this censored spell to determine whether the first (not left-censored) price spell occurs after price increase or decrease.

Table 1 shows summary statistics of the price spells by product category, focusing on price spells longer than $\underline{T} = 2$ weeks. The pooled sample contains 21,717,549 products, yielding 684,919,778 pairs of durations where both durations exceed $\underline{T}$.

5.2 Choice of $\underline{T}$ and $\bar{T}$

Since we observe average weekly prices, price changes occurring in the middle of the week generate spurious price spells with duration one week. The one-week spells are thus not generated by an MPH model and including them into estimation would bias estimates of the baseline hazard at all durations. At the same time, we want to choose the lowest possible value for $\underline{T}$. Therefore, we set $\underline{T} = 2$.

We provide further justification for this choice. An implication of any mixture model where each product has two independent spell durations from its type-specific distribution, and of the MPH model in particular, is that the autocorrelation of the duration of two completed spells is non-negative, and strictly positive when there is heterogeneity in mean duration. To understand why, note that conditional on a product's type, the autocorrelation of duration is zero by assumption. But with heterogeneity, the autocorrelation captures differences in the type-specific means and is necessarily positive.

Inspired by this, we measure the autocorrelation of the duration of price spells in the data. If we use all price spells, including one-week spells, we find correlation of 0.029 when spells are measured in levels, and -0.042 when duration is measured in logs. Once we exclude spells lasting one week, the correlation increases to 0.235 in levels and 0.233 when measured in logs. These correlations increase further to 0.248 in levels and 0.256 in logs when we consider

[8] For example, suppose that the price of a product increases from \$1 to \$2 in the middle of a week. Then we would measure average price of \$1 in week 1, \$1.5 in week 2 and \$2 in week 3, which looks like as if there were two price changes.

	number of products with $\tilde{K}^i \geq 1$	$\tilde{K}^i \geq 2$	number of pairs	percentiles of c^i 50^{th}	90^{th}	percentiles of ζ_j^i 50^{th}	90^{th}
Yoghurt	1,402,766	1,155,766	98,999,368	30	130	3	10
Carb. Beverage	1,819,607	1,321,762	90,836,025	24	129	3	8
Salty Snack	2,481,250	1,670,539	72,485,278	16	85	3	9
Frozen Dinner	2,272,888	1,693,017	70,495,598	15	74	3	8
Cold Cereal	1,429,028	1,038,096	56,080,465	21	117	4	12
Beer	701,604	470,815	37,454,496	17	114	3	11
Milk	549,261	426,316	34,036,391	36	165	4	14
Soup	1,286,921	897,080	33,873,770	17	92	4	14
Spaghetti Sauce	501,088	353,379	25,015,292	17	105	3	11
Frozen Pizza	711,065	519,293	24,984,150	14	74	3	8
Margarine	244,844	204,293	23,833,374	45	188	4	13
Hot Dog	213,598	172,031	19,603,427	27	143	3	9
Coffee	793,004	455,555	13,969,362	8	60	3	10
Toilet Tissue	412,746	312,604	10,791,034	25	99	3	11
Laundry Det.	804,837	489,482	9,993,575	9	50	3	9
Facial Tissue	250,134	185,450	9,557,189	24	108	3	11
Peanut Butter	203,380	150,692	9,255,148	25	130	4	13
Mayonnaise	186,392	136,585	7,992,048	25	138	4	14
Mus & Ketchup	217,559	143,485	7,659,886	19	124	4	16
Paper Towel	340,032	252,339	6,939,886	24	88	3	13
HH Cleaners	413,061	232,276	5,959,387	9	57	4	11
Toothpaste	716,457	322,194	4,615,305	5	30	3	8
Shampoo	1,134,428	352,570	2,483,449	3	14	3	7
Diapers	602,164	247,864	1,918,554	4	22	3	7
Sugar Sub.	94,528	56,644	1,818,682	12	95	4	17
Deodorant	972,970	291,558	1,633,620	3	13	3	6
Toothbrush	512,729	178,488	1,097,352	4	18	3	7
Blades	297,314	114,407	1,076,134	5	25	3	10
Photo	65,503	28,187	358,959	5	30	3	8
Razors	86,391	26,001	102,574	3	12	2	6

Table 1: Descriptive statistics by product category. For this table, we consider only spells $\zeta_j^i \geq \underline{T} = 2$ and use $\tilde{K}^i$ to denote the number of such spells for the product i. The first column shows the number of products with at least one spell longer than $\underline{T}$. The second column reports the number of products with at least two such spells. The third column reports the number of pairs of spells where both are longer than $\underline{T}$. Columns 4 and 5 show the median and 90^{th} percentile value of the censoring time c^i. The last two columns show the median and 90^{th} percentile value of the spell length.

spells at least 3 weeks long. With all spells, both correlations very close to zero, and one is even negative, suggesting that the data are not likely to be coming from a mixture model. When we exclude one-week spells, the correlation is different from zero, which does not change even after exluding two-week spells. Thus, there is a reason for excluding one-week spells but not two-week spells, and hence we set $\underline{T} = 2$.

The choice of $\bar{T}$ is guided by the nature of the data and our need to balance two forces. On the one hand, we want to choose a large value for $\bar{T}$ to learn about the baseline hazard at long durations. At the same time, the number of spells longer than $\bar{T}$ decreases quickly with $\bar{T}$. Indeed, Table 1 shows that depending on the product category, the median spell duration is 2–4 weeks and the 90^{th} percentile varies between 6 and 17 weeks and so data are thin at durations longer than half a year. While this does not constitute a problem for estimating the baseline hazard, smaller sample size will be reflected in larger standard errors, the choice of $\bar{T}$ affects our estimates of the Kaplan-Meier hazard at all durations because we condition on $c \geq \bar{T}$. Balancing these forces, we choose $\bar{T} = 60$ weeks, a little over a year, because there is an interesting pattern in the hazard at 52 weeks. Figure 4 shows estimates beyond 60 weeks. The estimates are noisy but follow the same trend from before $\bar{T} = 60$ so our main results are for $\bar{T} = 60$.

6 Results

We start this section with a brief discussion of models of price adjustments and their implications for the shape of the hazard. The shape of the baseline hazard, that is the hazard adjusted for heterogeneity, allows us to distinguish different models which have different predictions for the monetary policy. We then present our estimates of the baseline hazard for both the MPH model and the extension with competing risks of price increases and decreases.

6.1 Models of Price Adjustment

The simplest model of price adjustment is Calvo (1983), where a firm has a constant probability of being able to adjust its price. Thus, for an individual firm, the hazard of price adjustment as a function of duration is constant. If this probability differs across firms, then the Kaplan-Meier hazard is decreasing.

With time-dependent pricing (Taylor, 1979, 1980), a firm changes its price after a fixed number of periods. For an individual firm, the hazard of adjusting the price is zero prior to that duration, one at that duration, and then undefined at longer durations.

A large class of models assumes state-dependent prices, e.g. Golosov and Lucas (2007). In their simplest form, the desired price follows a stochastic process and the firm can adjust its price at any time by paying a fixed cost. Under some regularity conditions, a firm optimally adjusts its price when the difference between the current and desired price is too high, and the hazard of price changes is increasing in duration: the longer the elapsed time since the last price adjustment, the more likely the firm is to adjust the price in the next period.

Eichenbaum, Jaimovich, and Rebelo (2011) and Alvarez and Lippi (2020) feature two types of adjustments. A firm can adjust costlessly between a set of prices constituting a "price plan," but it has to pay a fixed cost to switch its price plan. For example, the price plan may include a regular price and a sale price. In these models, the hazard of changing the price may be decreasing. In particular, when a price plan containing two prices can be modified with probability λ in each period, the hazard of changing the price is $1/(2t) + \lambda$ at duration t.

We note that, with the exception of Calvo, these models do not have the MPH representation. However, we have found that in quantitative versions of these models, the MPH structure is a close approximation to the true amount of heterogeneity.

Many papers use sticky price models to explore the real effects of monetary policy. In Appendix C we solve a simple model where firms follow heterogeneous time-dependent pricing rules and show how heterogeneity affects the path of average prices. To be concrete, suppose that prices are strategic complements, so firms want to set a lower price when other firms have low prices. We can then compare two economies with the same Kaplan-Meier hazard but different individual hazards. We show that if firms have heterogeneous hazards, the response of average prices to a monetary policy shock is dampened compared to an economy where all firms have the same hazard. Conversely, given a particular Kaplan-Meier hazard, the real effects of a monetary policy shock are larger when firms are heterogeneous.

Nakamura and Steinsson (2008) show that distinguishing between sales prices and regular prices has important implications both for the frequency and hazard of price changes. In particular, sales are more transient than regular price changes and are not typically related to macroeconomic conditions. Following their pioneering work, most researchers have dropped all price changes associated with sales from the data set before estimating the hazard of regular price changes. We are concerned that doing so may affect the estimated stochastic process for the regular price changes. In our case, this problem is particularly acute, since we do not observe sales directly, but instead must infer them from the nature of the price change, e.g. a short-lived low price between two higher prices. Even if one could directly observe sales prices, dropping a subset of price changes can bias estimates of the hazard for the remaining price changes, a standard issue in competing risk models.

Our approach instead allows us to view sales as part of the data, albeit a part that does not necessarily fit the MPH structure, e.g. because sales have a fixed duration that varies across products. We propose circumventing sales by focusing on outcomes—that is, competing risks—that represent regular price changes, and measure duration dependence for only those risks. In this paper we focus on price increases following price increases and on price decreases following price decreases, which we call *price trends*. Our approach can be used to look at other risks, e.g. setting a price that has not been observed in the recent past. In a data set with a reliable sales flag, one could use our competing risks framework to look at price spells that neither start nor end with sales.

6.2 Baseline Hazard and Heterogeneity

We start with presenting the results for the baseline model with $X = 1, R = 1$ on the pooled sample; see Figure 1. Appendix G shows results for each product category separately.

The left panel of Figure 1 shows Kaplan-Meier and baseline hazards. The Kaplan-Meier hazard is steeply declining between weeks 2 and 45, especially so between weeks 2 and 12. There is a pronounced peak at 52 weeks. In contrast to the Kaplan-Meier hazard, the baseline hazard is constant until week 4, after which it declines modestly until week 45. We observe a small increase between weeks 45 and 52, and then a subsequent drop. The difference between the Kaplan-Meier and baseline hazard points to substantial unobserved heterogeneity. Recall from equation (1) that

$$H_t = b_t \mathbb{E}[\theta|t] \Rightarrow \mathbb{E}[\theta|t] = \frac{H_t}{b_t}.$$

The average type at duration t, $\mathbb{E}[\theta|t]$, reflects the extent of dynamic sorting in the economy. A flat average type suggests that there is little dynamic sorting and hence little heterogeneity, while a steeply declining average type suggests a lot of heterogeneity. The right panel of Figure 1 shows that the average type is steeply declining, especially between weeks 2 and 12 when the average type drops by 70 percent. It then continues to decline until week 47, to the trough of 0.23, but at a lower pace. The average type then increases at 52 weeks, reflecting that the sharp increase in the Kaplan-Meier is not captured by the baseline hazard.

We now turn to the two tests of the model. The J-statistic is $J = 10{,}498$, while the critical value of the χ^2 distribution with $M - T = 1{,}712$ degrees of freedom is 1,749, implying that we reject the model at any conventional significant level. However, recall the dimensionality of our problem: we have $M = 1{,}770$ moment conditions and more than 21 million products. It is common to fail the J-test in such a situation. In what follows, we investigate the source of this rejection in more detail.

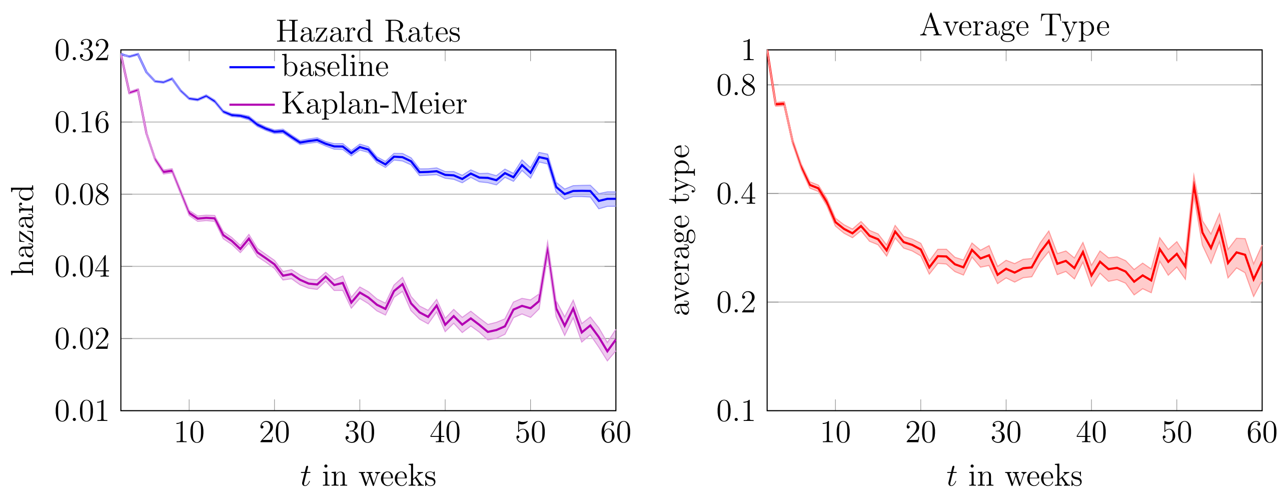

Figure 1: Kaplan-Meier and baseline hazard for the pooled sample. The purple line shows the Kaplan-Meier hazard, the blue line is the estimated baseline hazard. The red line shows the "average type" at given duration, calculated as the ratio of Kaplan-Meier and baseline hazards. Shaded regions show two standard error bands. Standard errors are clustered at the store $\times$ product category level. The baseline hazard is normalized to equal the Kaplan-Meier hazard at duration 2 weeks.

Our second test is whether the average type is decreasing. The right panel of Figure 1 shows a declining trend through durations 2 to 47 weeks, but a formal test rejects the null hypothesis due to the very tight standard errors.

Our conclusion is that the baseline hazard is declining, although much less so than the Kaplan-Meier hazard, suggesting the presence of substantial unobserved heterogeneity. We find evidence for Taylor-type price setting, with a mild spike in the baseline hazard at week 52. Still, all of these results are tempered by the fact that the model fails the overidentifying test as well as the test for dynamic sorting.

6.3 Hazard Rates for Price Increases and Price Decreases

We next estimate a richer model of price changes. Inspired by our discussion of theoretical models, we distinguish spells based on whether they started with a price increase or price decrease. Thus we set $X = 2$ and for mnemonic convenience let $\chi^i_j = +$ if the j^{th} spell of product i follows a price increase and $\chi^i_j = -$ if it follows a price decrease. We also distinguish whether a spell ends with a price increase or decrease, $R = 2$, and let $\rho^i_j = +$ if the j^{th} spell of product i ends with a price increase and $\rho^i_j = -$ if it ends with a price decrease. Spells with $\chi^i_j = \rho^i_j$ represent price trends, while other spells are *price reversals.*

We separately estimate four different baseline hazards, one for each possible combination of x and r. We use b_t^{++} to denote the baseline hazard that a spell after a price increase ends with a price increase at duration t; b_t^{+-} the baseline hazard that a spell following a price increase ends with a price decrease at duration t. Similarly, b_t^{-+} denotes the baseline hazard that a spell following a price decrease ends with a price increase at duration t and b_t^{--} denotes the baseline hazard that a spell following a price decrease ends with a price decrease at duration t. We allow for four different functions determining unobserved heterogeneity, $\phi^{++}(\boldsymbol{\theta}), \phi^{+-}(\boldsymbol{\theta})$ and $\phi^{-+}(\boldsymbol{\theta}), \phi^{--}(\boldsymbol{\theta})$.

This richer model allows for the possibility that price trends have different dynamics than price reversals. We estimate this richer model using the moment conditions specified in Section 4. Figure 2 shows the results. The figure reveals interesting patterns. The baseline hazards for price trends, b_t^{++} and b_t^{--}, are rather flat, especially the hazard for two consecutive price increases. The baseline hazards for the price reversal are declining, with b_t^{-+} showing the sharpest decline. The Kaplan-Meier hazard shows a steeper decline in all four cases, pointing to the presence of unobserved heterogeneity.

The right panel of Figure 2 shows the average type, the ratio of the Kaplan-Meier and baseline hazard for the four cases. We note that in the competing risks model, the average type does not have to be decreasing even if the model is correctly specified. Still, the average type is informative about the degree of dynamic sorting. We recover substantial heterogeneity in all four cases, but more so for price reversals than price trends. Different shapes for the average type in the four panels suggest that functions $\phi^{++}(\boldsymbol{\theta})$, $\phi^{+-}(\boldsymbol{\theta})$, $\phi^{-+}(\boldsymbol{\theta})$, and $\phi^{--}(\boldsymbol{\theta})$ are different, and so we cannot reduce dimensionality of unobserved heterogeneity.

The model is over-identified and so we can again apply the J-test. We run a separate J-test for each hazard. This is conceptually correct since each baseline hazard can be estimated without assuming a MPH structure for the other competing hazards. The 5% critical value is 1,749 for each risk, and the test statistics are $J^{++} = 3{,}920$, $J^{+-} = 8{,}737$, $J^{-+} = 7{,}910$, and $J^{--} =3{,}401$. Even though we still reject the model at the five percent level, the rejection is "milder" than in the case of one baseline hazard, where we have $J =10{,}498$, and especially so for the for price trends. Rejection is strongest for the case of sales, price decreases followed by price increases.

We investigate the nature of the failure of proportional hazard assumption more systematically in the next subsection. We conclude that the dynamics of price trends is well described by the MPH model and that the baseline hazard is fairly flat. On the other hand, we conclude that MPH model is not a good description of the dynamics of price reversals.

A consequence of these findings is that the shape of the baseline hazard we recovered in Figure 1 is primarily driven by price reversal, especially those associated with sales, where the hazard is steeply declining. Price reversal are common in the data: 72.3 percent of spells

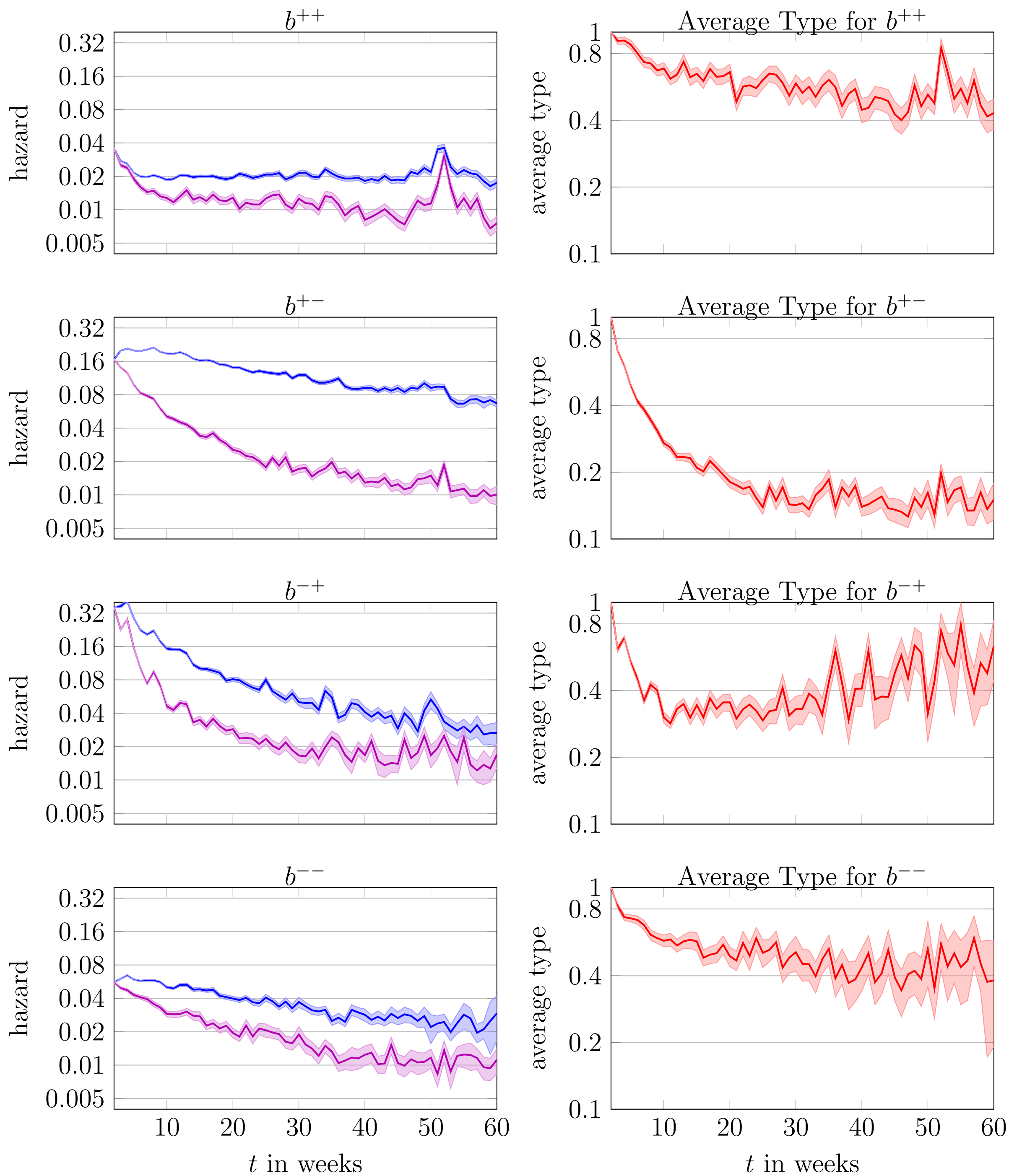

Figure 2: Competing risks model. Blue lines show the baseline hazard, purple lines the corresponding Kaplan-Meier hazard. b_t^{++} is the baseline hazard for spell which begin and end with a price increase; b_t^{--} for spells which begin and end with a price decrease; b_t^{+-} for spells which begin with a price increase and end with a price decrease; and b_t^{-+} for spells which begin with a price decrease and end with a price increase. The bottom row shows the average type for each case. The shaded regions show two standard error bands. Standard errors are clustered at the store $\times$ product category level.

starting with a price increase end with a price decrease, while 72.4 percent of spells starting witha a price decrease end with a price increase.

6.4 Sensitivity of Results to the Choice of $\underline{T}$ and $\bar{T}$

We examine the sensitivity of our results to the choice of $\underline{T}$ and $\bar{T}$. This allows us to see if there is a systematic failure of the MPH assumption. The idea is the following. Suppose we want to learn about the relative baseline hazards at duration 10 and 20, b_{10}/b_{20}. The MPH model admits several ways of recovering the ratio. We can directly recover the ratio b_{10}/b_{20} from equation (9) by choosing $t_1 = 10$ and $t_2 = 20$. But there are other options which use information on spells at other durations. Specifically, we can use this moment condition to recover b_{10}/b_t and b_{20}/h_t for some $t \neq 10, 20$, and combine them to find b_{10}/b_{20}. Our estimator uses all such conditions. If it is the case that the MPH model is not correctly specified at t, then including t into estimation will affect the relative hazards b_{10}/b_{20}.

Let $b_t(\underline{T}, \bar{T})$ denote the GMM estimate of the baseline hazard at duration $t \in \{\underline{T}, \dots, \bar{T}\}$ using some values $\underline{T}$ and $\bar{T}$. We first fix $\bar{T} = 60$ and estimate the model for different values of $\underline{T} = 2, 3, \dots, 10$. To help visualize the impact of $\underline{T}$ on the shape of the baseline hazard, we normalize $b_2(2, 60) = 1$ and then recursively set $b_{\underline{T}}(\underline{T}, 60) = b_{\underline{T}}(\underline{T} - 1, 60)$ for $\underline{T} > 2$. If the model is correctly specified for $t \in \{\underline{T}, \dots, \bar{T}\}$, we should find that $b_t(\underline{T}, \bar{T}) = b_t(\underline{T}', \bar{T})$ for all $\underline{T} < \underline{T}' < t \leq \bar{T}$. Substantial deviations from this indicate systematic violations of the MPH assumption.

Figure 3 shows the results. The choice of $\underline{T}$ has little effect on the hazard of price trend, b^{++} and b^{--}, consistent with a correctly-specified model, but it substantially affects the hazard of price reversals, especially so b^{-+}.

To analyze the role of $\bar{T}$, we fix $\underline{T} = 2$ and estimate the model for $\bar{T} \in \{10, 20, \dots, 90\}$. We now normalize $b_2(2, \bar{T}) = 1$ for each value of $\bar{T}$. Figure 4 shows that the choice of $\bar{T}$ does not affect the estimates.

This exercise does not reveal systematic violation of the MPH structure for b^{++} and b^{--}. However, it brings up the concern that the hazards b^{+-} and b^{-+} are not well described by the MPH, at least at short durations. One hypothesis for the failure of the MPH model is that the product type $\phi(\boldsymbol{\theta})$ is not fixed over time. We investigate this by restricting the censoring time to at most 80 weeks for every product. Figure 7 in Appendix D depicts the results. With the shorter censoring time, the choice of $\underline{T}$ matters less for all four hazards. The baseline hazards b^{++} or b^{--} are insensitive to the choice of $\underline{T}$, supporting our conclusion that these are well described by the MPH model. The estimates of b^{+-} or b^{-+} still depend on the choice of $\underline{T}$, but much less so than in the case of unrestricted censoring time. This is

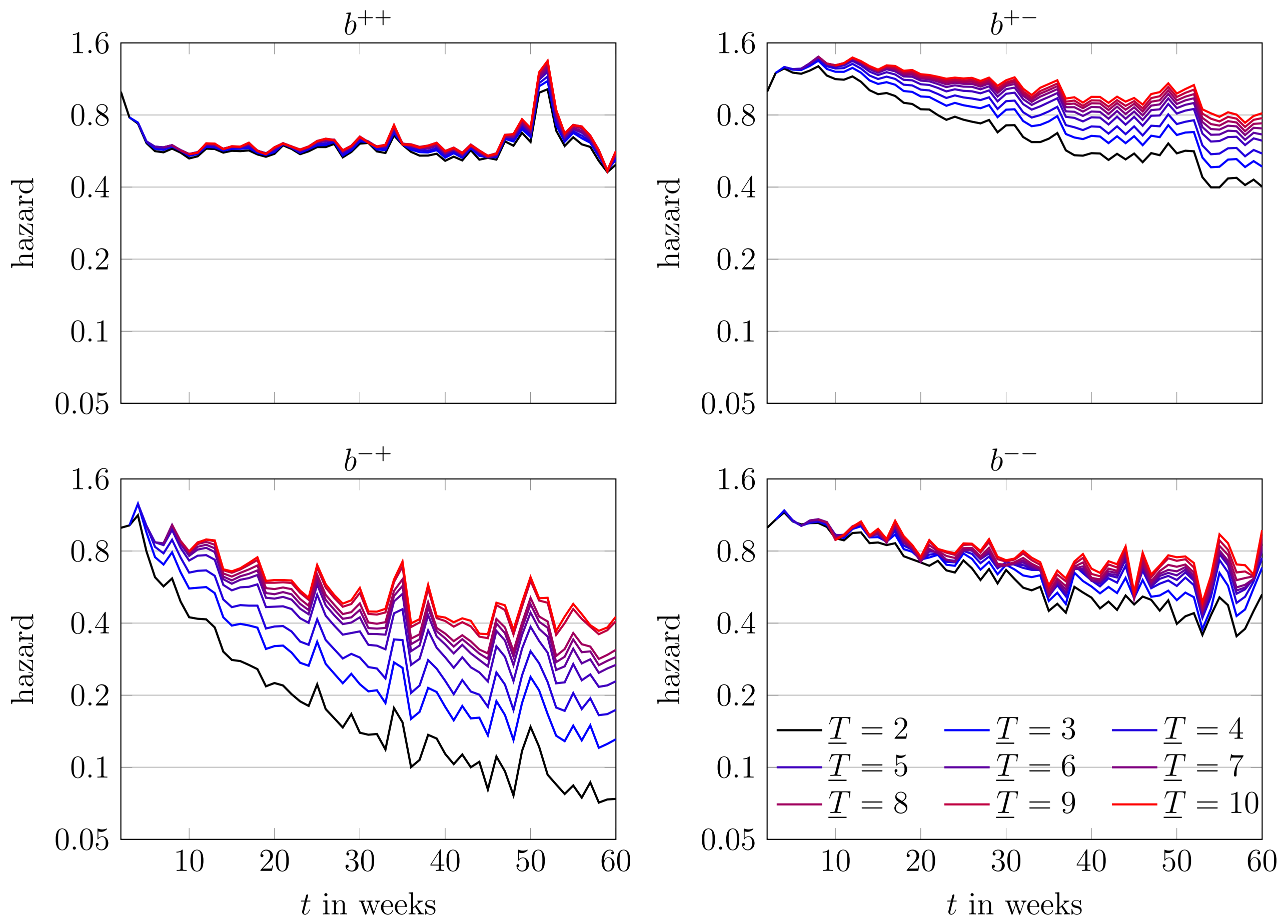

Figure 3: Baseline hazard for the competing risks model, the pooled sample, estimated using different values of $\underline{T} \in \{2, \ldots, 10\}$ and $\bar{T} = 60$.

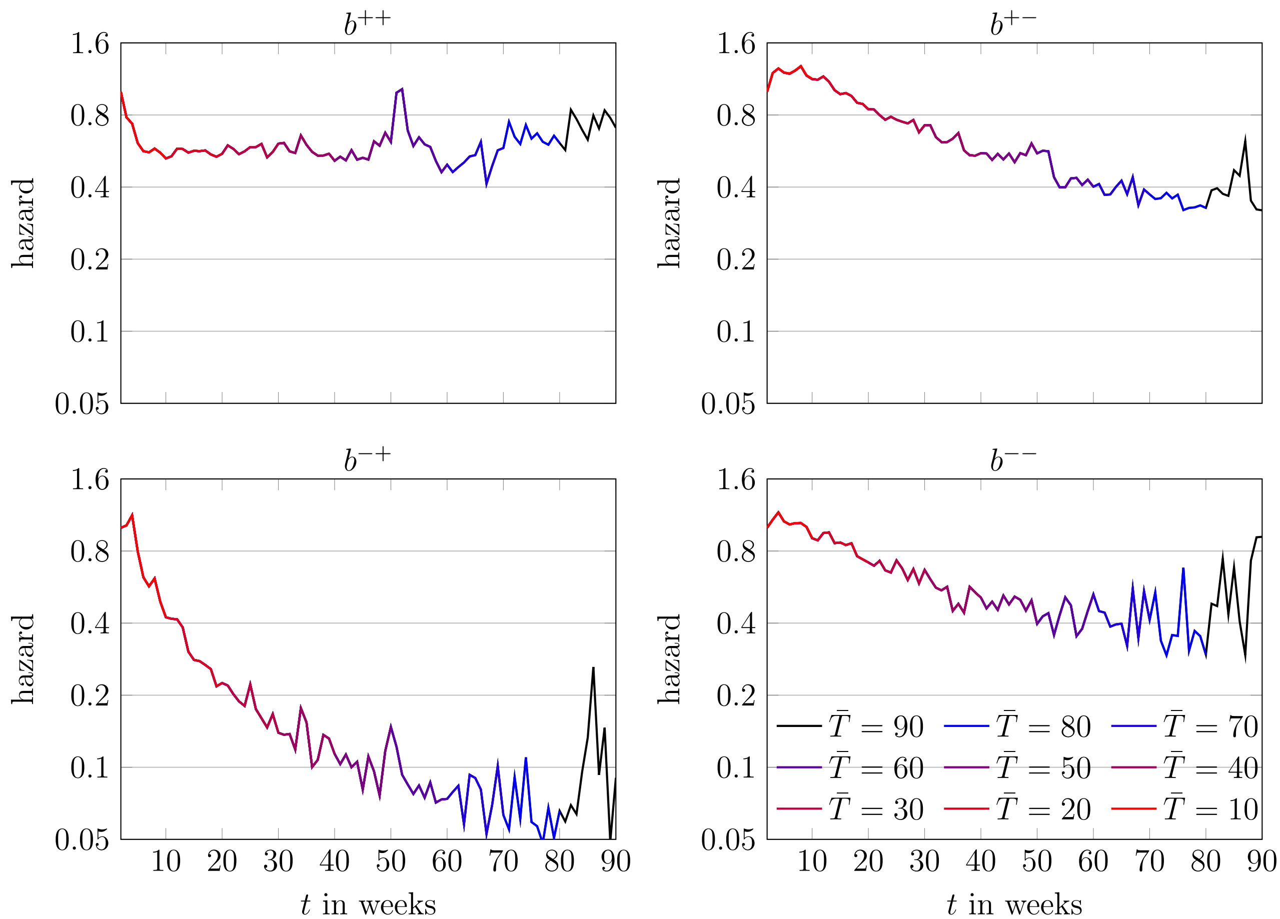

Figure 4: Baseline hazard for the competing risks model, the pooled sample, estimated using different values of $\bar{T} \in \{10, 20, \ldots, 90\}$ and $\underline{T} = 2$.

consistent with time-varying types.

Based on these test results we believe that the richer model with competing risks and observable characteristics is a useful framework to analyze the data. The baseline hazard for two consecutive price increases is decreasing until week 6 which covers a substantial amount of price changes: 76.8% of complete spells which start after a price increase last at most 6 weeks (among complete spells which start and end with a price increase, 76.7% last at most 6 weeks). During first 6 weeks, the baseline hazard drops by almost 50%. The hazard is then flat until week 45. This shape of the hazard is consistent with price plan models with Calvo-type switching between plans. There is a pronounced spike at around one year, consistent with Taylor-type pricing. The baseline hazard for two consecutive price decreases is mildly decreasing over the examined range.

Our analysis suggests that price reversals are not well described by the MPH model. One possible reason is that temporary changes might have fixed duration which does not fit into the MPH framework.

6.5 Higher Frequency Data

We study the price data through the lens of a discrete time model and naturally wonder if the frequency of the data affects our results. To explore this, we repeat our analysis using daily Online Micro Price data, the open access data from the Billion Prices Project presented by Cavallo (2018).[9] In this data set, we observe daily posted prices for many products, which we use to construct price spells.[10]

The top row of Figure 5 shows the estimates using daily data for $\underline{T} = 1$ day and $\bar{T} = 70$ days (ten weeks). We observe that the hazard of changing the price spikes every seventh day. This suggests that even though the data are daily, the decision to change prices is taken at the weekly frequency and hence a week might be a natural time unit.

We then aggregate data to weekly frequency, that is, any spell lasting 1–7 days is recorded as duration of 1 week, spells lasting 8–14 days as duration of 2 weeks, etc, and estimate the model again. The bottom row of Figure 5 shows the results, with the solid lines depicting estimates using weekly data and dashed line showing weekly averages of daily estimates. The results are very similar, even though weekly data recover a little bit less heterogeneity that the daily data.

[9] http://www.thebillionpricesproject.com/datasets/. We use the US store number 1.

[10] We observed posted price directly and hence we do not need to exclude price spells of length 1.

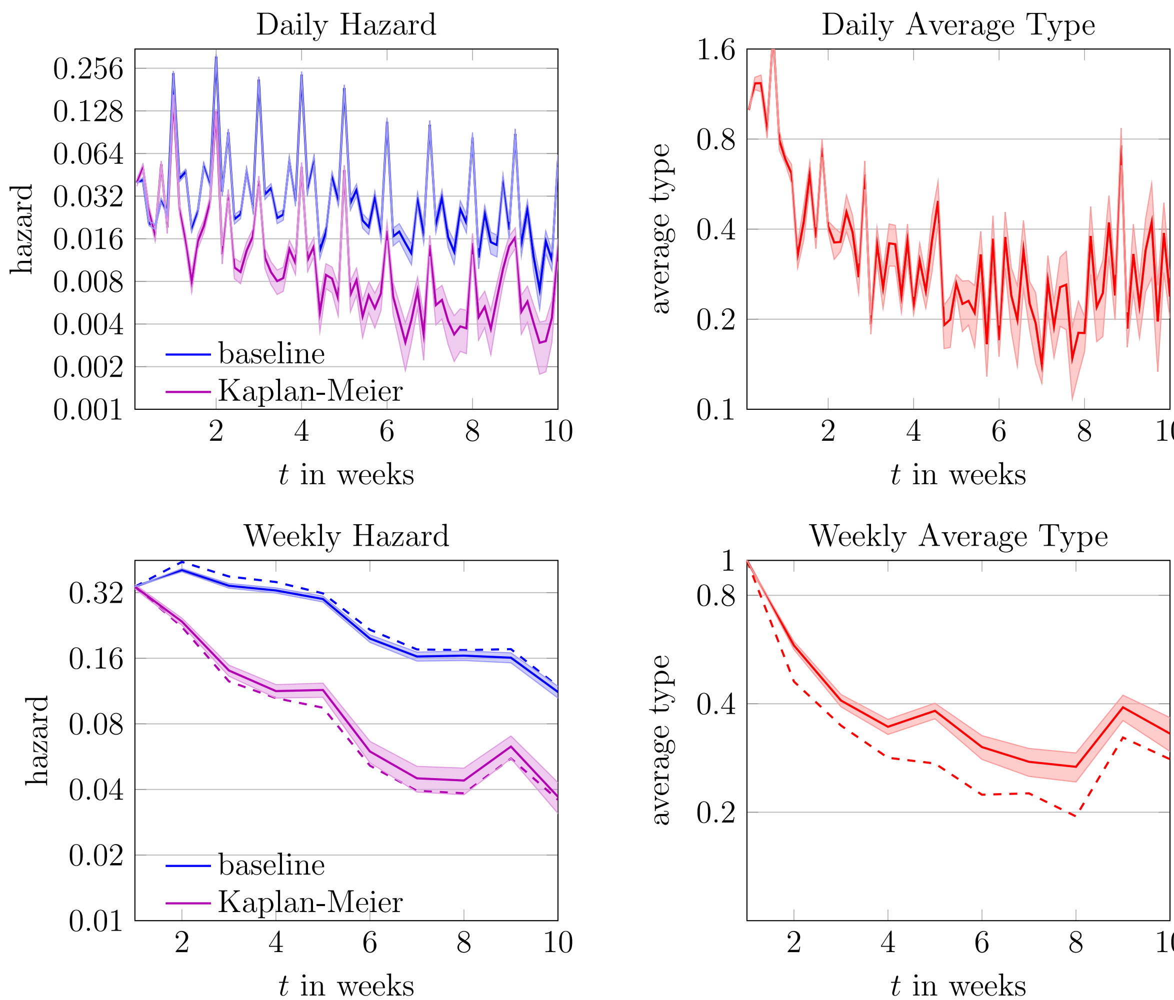

Figure 5: Kaplan-Meier and baseline hazard for Online Micro Price Data using daily and weekly data. The top row uses daily data, the bottom row daily data aggregated to weekly frequency. The purple line shows the Kaplan-Meier hazard, the blue line is the estimated baseline hazard. The red line shows the “average type” at given duration, calculated as the ratio of Kaplan-Meier and baseline hazard. Shaded regions show two standard error bands. The baseline hazard is normalized to be equal to the Kaplan-Meier hazard at duration 1 day in the daily data, or 1 week in the weekly data. The dashed lines in the bottom row show weekly averages of daily estimates, normalized to be equal to the weekly values at week 1.

7 Comparison to Other Estimation Methods

The usual approach to estimating the MPH model is via maximum likelihood (ML) for the continuous time model. Formulating the likelihood requires an assumption on the frailty distribution. It is convenient to choose a gamma distribution with mean 1 and variance v, and then estimate v together with the baseline hazard. There are two approaches how to deal with the fact that the model is continuous but the data are measured discretely. One, as in for example in Fougere, Le Bihan, and Sevestre (2007), is specify the likelihood so that the probability of observing a discrete duration t means that the realization of the corresponding continuous time variable τ was in the interval $(t-1, t]$ and that the vector of baseline hazards that can be estimated corresponds to $b_t \equiv \int_{t-1}^{t} b(s)ds$. We formulate likelihood for this model in Appendix E, and refer to it as continuous-time discrete-measurement (CT-DM). Another approach, as in Nakamura and Steinsson (2008), is to assume that the baseline hazard is piece-wise constant, $b(t+\gamma) = b_t$ for all $\gamma \in [0, 1)$, and observing discrete time duration t means that the corresponding continuous time variable τ also reached this value, $\tau = t$. We call this continuous-time continuous-measurement (CT-CM) and develop the model together with its likelihood in Appendix F. We note that this is the procedure which is coded in Stata, although implementing this in Stata requires some tricks to allow for a flexible baseline hazard. For comparison, we refer to the discrete time model formulated in this paper as DT-DM (discrete-time discrete-measurement).

We make two simplifying assumptions when formulating likelihoods for CT-DM and CT-CM models. First, in line with the literature, we assume that censoring time c is independent of product's types θ. Second, we use at most two spells per product which allows us to represent the data in a simple way. For each combination of durations (t_1, t_2), with $t_1 \geq 1$ and $t_2 \geq 0$, it is enough to store the number of products with these measured durations and the share of these with the right-censored first and/or second spell. Due to this simplification, maximizing the likelihood is very fast but we are aware of the fact that usefulness of this trick disappears in a general setup where different products have a different number of spells.

Figure 6 shows the results from such estimation using at most two spells per product. In both CT-DM and CT-CM we assume that the baseline hazard after duration $\bar{T}$ is constant. The purple line shows the Kaplan-Meier hazard. Since we assume that censoring time and product's type are independent, we do not condition on $c \geq \bar{T}$ when estimating Kaplan-Meier hazard and use all first spells of all products to estimate it. The blue line shows the baseline hazard estimated from the discrete time model (DT-DM) using GMM. The other solid lines show ML estimates for the continuous time model, either with discrete measurement CT-DM(1) (black line) or continuous time measurement CT-CM(1) (green line). The CT-CM(1) baseline hazard is very close to the Kaplan-Meier hazard, implying it recovers almost no heterogeneity. The CT-DM(1) model which properly takes into account time aggregation, gives an estimate basically identical to our DT-DM model. In general, CT-DM and DT-DM models are not the same and so we should not expect them to deliver the same estimates. There is, however, an important special case when they are, which is when the baseline hazard is constant.

Heckman and Singer (1984) pointed out that imposing a specific distribution for the ML estimation can bias the estimates of the baseline hazard. We investigate whether misspecification of the frailty distribution can explain the difference between CT-CM(1) and DT-DM. We cannot formulate the likelihood without choosing a frailty distribution but we can choose a more flexible distribution than a single gamma, for example a mixture of several gamma distributions. In the CT-CM model, we could not find the second gamma distribution and hence the estimates of CT-CM(1) and CT-CM(2) are identical. In the CT-DM model, modeling the frailty as a mixture of distributions does not affect the baseline hazard as CT-DM(1) and CT-DM(2) are very close. We therefore conclude that in this case, imposing a specific functional form on the frailty distribution does not affect results.

The bottom line is that the estimates from CT-DM and DT-DM model are similar. Our conclusion from this exercise is that the most important factor explaining the difference between the CT-CM and DT-DM model is the failure of CT-CM to deal with discrete data.

In closing, we note that there are several advantages to using the DT-DM model and the GMM estimator we developed over CT-DM. First, the estimator does not require us to specify the frailty distribution. Second, it is linear in $\boldsymbol{b}$ and hence is very simple and fast to solve. Third, we prove in Proposition 3 that we find a global maximum. In contrast, the estimator for the CT-DM model is based on maximizing the likelihood and requires choosing a frailty distribution. The log-likelihood is non-linear $\boldsymbol{b}$ and finding its maximum can be slow. Importantly, there is no guarantee that we find a global maximum. Finally, we showed that our method is easily extended to handle competing risks, something that is extremely hard to handle in the CT-DM framework. Fougere, Le Bihan, and Sevestre (2007) try to estimate competing risks model with unobserved heterogeneity but say on page 260 that "…convergence of the maximum likelihood procedure is very difficult to reach."

8 Conclusion

We develop a new consistent estimator of the baseline hazard in the MPH model using duration data with repeated observations. Our estimator has many desirable features: it is linear in the baseline hazard and hence easy to implement; it does not require specifying

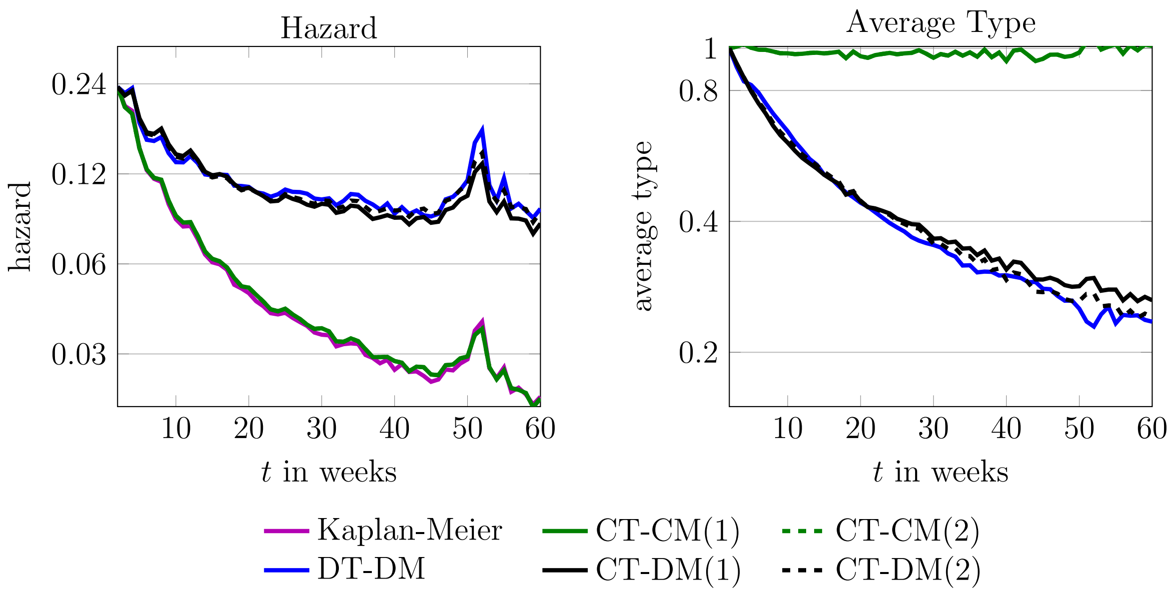

Figure 6: Baseline hazard estimated using different methods (left panel) and the corresponding average type (right panel), using two-spell data. The purple line is the Kaplan-Meier hazard, the blue line is the discrete time model. The green lines correspond to continuous time with continuous time measurement (CT-CM), where the frailty distribution is a single gamma distribution (green solid line) or a mixture of 2 gamma distributions (green dashed line). The black lines correspond to the continuous time, discrete measurement (CT-DM) model, where the frailty distribution is a single gamma distribution (black solid line) or a mixture of 2 gamma distributions (black dashed line). The right panel shows the "average type" corresponding to the estimates, using the same color coding as in the left panel. The average type is computed as a ratio of the Kaplan-Meier and baseline hazard.

the frailty distribution; and it handles right-censored data, competing risks, and discrete observable characteristics. Importantly, it works in an environment where duration takes on one of a finite number of possible values, which is the format or real-world data. We further propose and implement two tests of the MPH specification.

We treat the frailty distribution as a nuisance parameter. However, under two additional assumptions, namely that $\underline{T} = 1$ and that the censoring time c and type θ are independent, it is straightforward to construct an estimator of the moments of the frailty distribution using the logic in our identification proposition.

We also estimate the baseline hazard for price changes, distinguishing between price trends, which we interpret as regular price changes, and price reversals, which include sales. Our framework is general enough to handle different notions of sales. For example, we could have defined a sale as a temporary cut in price from a "normal" price p to a sale price $p' < p$, followed by a reversal back to p. We could also include other variable into the vector of observable characteristics, such as bins of marginal cost or of the average price of competitors. All these options can be handled through appropriately defining observables x and risks r in our framework.

The model and its estimator can also be applied in other fields. In the labor market, it can be used to study duration dependence in transitions between employment, unemployment and out of labor force. Worker's current labor market status is an observable characteristic and the next labor market status can be treated as a competing risk.

References

Abbring, Jaap H., and Gerard J. Van Den Berg, 2003. "The Identifiability of the Mixed Proportional Hazards Competing Risks Model." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 65 (3): 701–710.

Alvarez, Fernando, and Francesco Lippi, 2020. "Temporary Price Changes, Inflation Regimes, and the Propagation of Monetary Shocks." *American Economic Journal: Macroeconomics*. 12 (1): 104–52.

Bils, Mark, and Pete Klenow, 2004. "Some Evidence on the Importance of Sticky Prices." *Journal of Political Economy*. 112 (5): 947–985.

Bronnenberg, Bart J., Michael W. Kruger, and Carl F. Mela, 2008. "Database paper: The IRI marketing data set." *Marketing Science*. 27 (4): 745–748.

Calvo, Guillermo A, 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics*. 12 (3): 383–398.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, 2011. "Robust Inference With Multiway Clustering." *Journal of Business and Economic Statistics*. 29 (2): 238–249.

Cavallo, Alberto, 2018. "Scraped Data and Sticky Prices." *The Review of Economics and Statistics*. 100 (1): 105–119.

Chen, Le-Yu, and Jerzy Szroeter, 2014. "Testing multiple inequality hypotheses: A smoothed indicator approach." *Journal of Econometrics*. 178: 678 – 693.

Chen, Le-Yu Chen, and Jerzy Szroeder, 2009. "Hypothesis Testing of Multiple Inequalities: The Method of Constraint Chaining.." Cemmap Working Paper CWP13/09.

Cox, David R., 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society. Series B (Methodological)*. 34 (2): 187–220.

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo, 2011. "Reference Prices, Costs, and Nominal Rigidities." *The American Economic Review*. 101 (1): 234–262.

Elbers, Chris, and Geert Ridder, 1982. "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *The Review of Economic Studies*. 49 (3): 403–409.

Fougere, Denis, Hervé Le Bihan, and Patrick Sevestre, 2007. "Heterogeneity in Consumer Price Stickiness: A Microeconometric Investigation." *Journal of Business and Economic Statistics*. 25 (3): 247–264.

Golosov, Mikhail, and Robert Lucas, 2007. "Menu Costs and Phillips Curves." *Journal of Political Economy*. 115: 171–199.

Heckman, James J., and Bo E. Honoré, 1989. "The Identifiability of the Competing Risks Model." *Biometrika*. 76 (2): 325–330.

Heckman, James J, and B Singer, 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica*. 52 (2): 271–320.

Honoré, Bo E., 1993. "Identification Results for Duration Models with Multiple Spells." *Review of Economic Studies*. 60 (1): 241–246.

Horowitz, Joel L., and Sokbae Lee, 2004. "Semiparametric Estimation of a Panel Data Proportional Hazards Model with Fixed Effects." *Journal of Econometrics*. 119 (1): 155–198.

Lancaster, Tony, 1979. "Econometric Methods for the Duration of Unemployment." *Econometrica*. 47 (4): 939–956.

Lancaster, Tony, 1990. *Mixture Models*p. 58–82, . Econometric Society Monographs, Cambridge University Press.

Nakamura, Emi, and Jón Steinsson, 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models." *Quarterly Journal of Economics*. 123 (4): 1415–1464.

Newey, Whitney K., and Daniel McFadden, 1994. "Large Sample Estimation and Hypothesis Testing." *Handbook of Econometrics*. 4: 2111–2245.

Politis, Dimitris N., 2011. "Higher-Order Accurate, Positive Semidefinite Estimation of Large-Sample Covariance and Spectral Density Matrices." *Econometric Theory*. 27 (4): 703–744.

Taylor, John B., 1979. "Staggered Wage Setting in a Macro Model." *The American Economic Review*. 69 (2): 108–113.

Taylor, John B., 1980. "Aggregate Dynamics and Staggered Contracts." *Journal of Political Economy*. 88 (1): 1–23.

A Omitted Proofs

Since Proposition 3 is a special case of Proposition 5, we prove the latter proposition first and then turn to the special case.

To simplify the exposition in this appendix, we introduce the following notation. For any K vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_K)$ and $k \leq K$, we define $\boldsymbol{\zeta}_k$ to be a vector consisting of the first k elements of $\boldsymbol{\zeta}$, that is, $\boldsymbol{\zeta}_k = (\zeta_1, \ldots, \zeta_k)$. For $k > K$, we construct $\boldsymbol{\zeta}_k$ by adding $k - K$ zeros to the end of $\boldsymbol{\zeta}$ to construct a k vector, $\boldsymbol{\zeta}_k = (\zeta_1, \ldots, \zeta_K, 0, \ldots, 0\}$. Next, for $j < k$ we let $\boldsymbol{\zeta}_{k/j}$ denote the vector $\boldsymbol{\zeta}_k$ without the j^{th} element, that is, $\boldsymbol{\zeta}_{k/j} \equiv (\zeta_1, \ldots, \zeta_{j-1}, \zeta_{j+1}, \ldots, \zeta_k)$.

The key step in proving Proposition 5 is the statement and proof of Lemma 1.

Lemma 1 *Assume $\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho}$ are drawn from a right-censored competing-risk model with baseline hazard $\boldsymbol{b}_0$ for observable characteristic x and risk r. Take any $k > j \geq 1$ and vector $\boldsymbol{t} = (t_1, \ldots, t_k) \in \{1, 2, \ldots\}^k$ with $t_j, t_k \in \{\underline{T}, \ldots, \bar{T}\}$. Also take any $\boldsymbol{x} \in \{1, \ldots, X\}^k$ with $x_j = x_k = x$ and $\boldsymbol{r} \in \{1, \ldots, R\}^{k-1}$ with $r_j = r$. Define*

$$f_{j,k,\boldsymbol{t},\boldsymbol{x},\boldsymbol{r}}(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho}; \boldsymbol{b}) \equiv \\ b_{t_k} \mathbb{1}_{K \geq k, \boldsymbol{\chi}_k = \boldsymbol{x}, \boldsymbol{\rho}_{k-1} = \boldsymbol{r}, \boldsymbol{\zeta}_{k-1} = \boldsymbol{t}_{k-1}, \zeta_k \geq t_k} - b_{t_j} \mathbb{1}_{K \geq k, \boldsymbol{\chi}_k = \boldsymbol{x}, \boldsymbol{\rho}_{k-1} = \boldsymbol{r}, \boldsymbol{\zeta}_{k-1/j} = \boldsymbol{t}_{k-1/j}, \zeta_j = t_k, \zeta_k \geq t_j}. \quad (11)$$

Then $\mathbb{E}\left[f_{j,k,\boldsymbol{t},\boldsymbol{x},\boldsymbol{r}}(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho}; \boldsymbol{b}_0)\right] = 0$.

Proof of Lemma 1. We first claim that the first indicator function in equation (11) evaluates to 1 if and only if these conditions hold:

1. without censoring, the product has sufficiently many spells, $\bar{K} \geq k$;
2. the observable characteristics for the first k spells is $\boldsymbol{x}$, $\boldsymbol{\chi}_k = \boldsymbol{x}$;
3. the risk for the first $k-1$ spells is $\boldsymbol{r}$, $\boldsymbol{\rho}_{k-1} = \boldsymbol{r}$;
4. we observe the product for sufficiently long, $\sum_{l=1}^{K} \zeta_l \geq \sum_{l=1}^{k} t_l$;
5. the uncensored durations satisfy $\boldsymbol{\tau}_{k-1} = \boldsymbol{t}_{k-1}$ and $\tau_k \geq t_k$.

If the first condition failed, we could never observe k spells. The second and third conditions ensure we observe the desired pattern of observable characteristics and risks. The fourth condition ensures we observe the product sufficiently long to see $\boldsymbol{\zeta}_{k-1} = \boldsymbol{t}_{k-1}$ and $\zeta_k \geq t_k$. Finally, if the last condition failed, we might observe k spells, but they would not satisfy $\boldsymbol{\zeta}_{k-1} = \boldsymbol{t}_{k-1}$ and $\zeta_k \geq t_k$. On the other hand, if all five conditions are satisfied, we measure $K \geq k$, $\boldsymbol{\zeta}_{k-1} = \boldsymbol{\tau}_{k-1}$, $\zeta_k \geq t_k$, $\boldsymbol{\rho}_{k-1} = \boldsymbol{r}$, and $\boldsymbol{\chi}_k = \boldsymbol{x}$.

Analogously, the second indicator function in equation (11) evaluates to 1 if and only if the first four conditions hold and the uncensored durations satisfy $\boldsymbol{\tau}_{k-1/j} = \boldsymbol{t}_{k-1/j}$, $\tau_j = t_k$ and $\tau_k \geq t_j$.

Next, we use the MPH model to compute the probability of a realization of the event in the first indicator function, conditional on $\boldsymbol{\theta}$. This is

$$
\begin{aligned}
&\Pr[\boldsymbol{\chi}_k = \boldsymbol{x}, \boldsymbol{\rho}_{k-1} = \boldsymbol{r}, \boldsymbol{\zeta}_{k-1} = \boldsymbol{t}_{k-1}, \zeta_k \geq t_k | \boldsymbol{\theta}] \\
&= \pi_1(x_1|\boldsymbol{\theta}) \prod_{l=1}^{k} \left(\pi(x_l|x_{l-1}, r_{l-1}, \boldsymbol{\theta})^{\mathbb{1}_{l\neq 1}} h_{t_l}^{r_l}(x_l, \boldsymbol{\theta})^{\mathbb{1}_{l\neq k}} \prod_{s=1}^{t_l-1} (1 - h_s(x_l, \boldsymbol{\theta})) \right) \\
&= b_{0,t_j} \phi(\boldsymbol{\theta}) \pi_1(x_1|\boldsymbol{\theta}) \prod_{l=1}^{k} \left(\pi(x_l|x_{l-1}, r_{l-1}, \boldsymbol{\theta})^{\mathbb{1}_{l\neq 1}} h_{t_l}^{r_l}(x_l, \boldsymbol{\theta})^{\mathbb{1}_{l\neq j, l\neq k}} \prod_{s=1}^{t_l-1} (1 - h_s(x_l, \boldsymbol{\theta})) \right).
\end{aligned}
$$

The first equation uses the structure of the model, in particular the fact that we are computing the probability of a particular sequence of observable characteristics and spell durations. The second equation uses the fact that $r_j = r$, $x_j = x$, and $h_{t_j}^{r}(x, \boldsymbol{\theta}) = \phi(\boldsymbol{\theta}) b_{0,t_j}$ since $t_j \in \{\underline{T}, \ldots, \bar{T}\}$. Integrating across the distribution of $\boldsymbol{\theta}$ conditional on censoring time equal to at least $\sum_{l=1}^{k} t_l - 1$ gives us

$$
\mathbb{E}\left[\mathbb{1}_{K\geq k, \boldsymbol{\chi}_k = \boldsymbol{x}, \boldsymbol{\rho}_{k-1} = \boldsymbol{r}, \boldsymbol{\zeta}_{k-1} = \boldsymbol{t}_{k-1}, \zeta_k \geq t_k}\right] = \psi(\boldsymbol{t}_{k-1/j}, t_j, t_k, \boldsymbol{x}, \boldsymbol{r}; j, k) b_{0,t_j}, \tag{12}
$$

where

$$
\begin{aligned}
&\psi(\boldsymbol{t}_{k-1/j}, t_j, k_j, \boldsymbol{x}, \boldsymbol{r}; j, k) \equiv \left(1 - P\left(\sum_{l=1}^{k} t_l\right)\right) \times \\
&\int \phi(\boldsymbol{\theta}) \pi_1(x_1|\boldsymbol{\theta}) \prod_{l=1}^{k} \left(\pi(x_l|x_{l-1}, r_{l-1}, \boldsymbol{\theta})^{\mathbb{1}_{l\neq 1}} h_{t_l}^{r_l}(x_l, \boldsymbol{\theta})^{\mathbb{1}_{l\neq j, l\neq k}} \prod_{s=1}^{t_l-1} (1 - h_s(x_l, \boldsymbol{\theta})) \right) dG_{\sum_{l=1}^{k} t_l - 1}(\boldsymbol{\theta}).
\end{aligned} \tag{13}
$$

Now swap the role of t_j and t_k but leave $\boldsymbol{t}_{k-1/j}$, $\boldsymbol{r}$, and $\boldsymbol{x}$ unchanged. The same logic implies

$$
\mathbb{E}\left[\mathbb{1}_{K\geq k, \boldsymbol{\chi}_k = \boldsymbol{x}, \boldsymbol{\rho}_{k-1} = \boldsymbol{r}, \boldsymbol{\zeta}_{k-1/j} = \boldsymbol{t}_{k-1/j}, \zeta_j = t_k, \zeta_k \geq t_j}\right] = \psi(\boldsymbol{t}_{k-1/j}, t_k, t_j, \boldsymbol{x}, \boldsymbol{r}; j, k) b_{0,t_k}. \tag{14}
$$

Moreover, equation (13) and the commutative property of multiplication implies

$$
\psi(\boldsymbol{t}_{k-1/j}, t_k, t_j, \boldsymbol{x}, \boldsymbol{r}; j, k) = \psi(\boldsymbol{t}_{k-1/j}, t_j, t_k, \boldsymbol{x}, \boldsymbol{r}; j, k). \tag{15}
$$

The result then follows from equations (12), (14), and (15). ■

Proof of Proposition 5. We first prove that $\mathbb{E}\left[f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\lambda\boldsymbol{b}_0)\right]=0$ for all $\underline{T}\le t_1<t_2\le\bar{T}$ and λ (necessity). Then we prove $\mathbb{E}\left[f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})\right]=0$ for all $\underline{T}\le t_1<t_2\le\bar{T}$ only if $\boldsymbol{b}=\lambda\boldsymbol{b}_0$ (sufficiency) for some λ.

Necessity: We show in two steps that function $f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})$ is the sum of functions defined in Lemma 1, each of which have expected value zero. First, take $1\le j<k$, a pair (t_j,t_k) with $t_j,t_k\in\{\underline{T},\dots,\bar{T}\}$, an observable characteristic x, and a risk r. Define the following function

$$f_{j,k,t_j,t_k,x,r}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})\equiv b_{t_k}\mathbb{1}_{K\ge k,\zeta_j=t_j,\zeta_k\ge t_k,\rho_j=r,\chi_j=\chi_k=x}-b_{t_j}\mathbb{1}_{K\ge k,\zeta_j=t_k,\zeta_k\ge t_j,\rho_j=r,\chi_j=\chi_k=x}.$$

Then let $\boldsymbol{t}$ be an arbitrary k vector of durations with j^{th} element t_j and k^{th} element t_k, $\boldsymbol{x}$ be an arbitrary k vector of observables with j^{th} and k^{th} element x, and $\boldsymbol{r}$ be an arbitrary $k-1$ vector of risks with j^{th} element r. Summing across all such vectors, we get

$$f_{j,k,t_j,t_k,x,r}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})=\sum_{\boldsymbol{t}_{k-1/j},\boldsymbol{x}_{k-1/j},\boldsymbol{r}_{k-1/j}}f_{j,k,\boldsymbol{t},\boldsymbol{x},\boldsymbol{r}}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b}),$$

where this follows directly from the definition of $f_{j,k,\boldsymbol{t},\boldsymbol{x},\boldsymbol{r}}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})$ in equation (11). Lemma 1 states that the expected value of each component of the sum is zero for $\boldsymbol{b}=\boldsymbol{b}_0$. Thus the expected value of $f_{j,k,t_j,t_k,x,r}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b}_0)$ is zero.

Second, fix a pair of durations (t_1,t_2) with $\underline{T}\le t_1<t_2\le\bar{T}$, an observable characteristic x, and a risk r. Sum $f_{j,k,t_1,t_2,x,r}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})$ across all pairs of spells (j,k) with $1\le j<k$. By equation (9), this gives us $f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})$. Since the expected value of each component of this sum is zero, this implies $\mathbb{E}\left[f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b}_0)\right]=0$.

Finally, note that the function $f_{t_1,t_2}^{[b,x,r]}$ defined in equation (9) is linear in the baseline hazard, $f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\lambda\boldsymbol{b})=\lambda f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})$ for all t_1, t_2, $\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho}$, $\boldsymbol{b}$, and λ. Thus $\mathbb{E}\left[f_{t_1,t_2}^{[b,x,r]}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\lambda\boldsymbol{b})\right]=0$ as well.

Sufficiency: Recall that T_0 is the smallest $t\in\{\underline{T},\dots,\bar{T}\}$ with $b_{0,t}>0$. We prove that any solution must take the form $\boldsymbol{b}=\lambda\boldsymbol{b}_0$ where $\lambda=b_{T_0}/b_{0,T_0}$.

Equation (9) implies that

$$b_{T_0}\sum_{(j,k):1\le j<k\le K}\mathbb{E}\left[\mathbb{1}_{\zeta_j=t,\zeta_k\ge T_0,\chi_j=\chi_k=x,\rho_j=r}\right]=b_t\sum_{(j,k):1\le j<k\le K}\mathbb{E}\left[\mathbb{1}_{\zeta_j=T_0,\zeta_k\ge t,\chi_j=\chi_k=x,\rho_j=r}\right].$$

Assumption 2 states that $\mathbb{E}\left[\mathbb{1}_{\zeta_{j'}=T_0,\zeta_{k'}\geq t,\chi_j=\chi_k=x,\rho_j=r}\right] > 0$ for some $1 \leq j' < k' < K$ and any $t \leq \bar{T}$. Therefore the sum on the right hand side of this equation is strictly positive, allowing us to pin down the ratio b_t/b_{T_0}:

$$\frac{b_t}{b_{T_0}} = \frac{\sum_{(j,k):1\leq j<k\leq K} \mathbb{E}\left[\mathbb{1}_{\zeta_j=t,\zeta_k\geq T_0,\rho_j=r,\chi_j=\chi_k=x}\right]}{\sum_{(j,k):1\leq j<k\leq K} \mathbb{E}\left[\mathbb{1}_{\zeta_j=T_0,\zeta_k\geq t,\rho_j=r,\chi_j=\chi_k=x}\right]}.$$

From the 'necessity' part of the proof, we know $b_t/b_{T_0} = b_{0,t}/b_{0,T_0}$ solves this equation, so this must be the only solution. ∎

Proof of Proposition 3. Set $X = R = 1$. This implies $\pi_1(1|\boldsymbol{\theta}) = \pi(1|1,1,\boldsymbol{\theta}) = 1$, so Assumption 1 is equivalent to Assumption 2 in this case. Then

$$f^{[b,1,1]}_{t_1,t_2}(\boldsymbol{\zeta},\mathbf{1},\mathbf{1};\boldsymbol{b}) = f^{[b]}_{t_1,t_2}(\boldsymbol{\zeta};\boldsymbol{b}),$$

where $\mathbf{1}$ is a vector of 1's, and so the results in Proposition 5 imply the proof of this proposition. ∎

B GMM Estimation

B.1 GMM Estimator

Proposition 3 gives us one moment condition for the choice t_1, t_2 such that $\underline{T} \leq t_1 < t_2 \leq \bar{T}$:

$$\mathbb{E}\left[f^{[b]}_{t_1,t_2}(\boldsymbol{\zeta};\boldsymbol{b})\right] = 0.$$

Let $Y(\underline{T},\bar{T}) = \{(t_1,t_2) : \underline{T} \leq t_1 < t_2 \leq \bar{T}\}$. This set has $M = T(T+1)/2$ elements which we index with m and refer to it as $y_m = (y_{m_1}, y_{m_2})$. Let $\boldsymbol{f}^{[b]}(\boldsymbol{\zeta};\boldsymbol{b})$ be a vector function with m^{th} element corresponding to the choice $y_m \in Y(\underline{T},\bar{T})$, given by $f^{[b]}_{y_{m_1},y_{m_2}}(\boldsymbol{\zeta};\boldsymbol{b})$.

Since the baseline hazard is identified up to scale, we choose our normalization. Choose $T^* \in \{\underline{T},\bar{T}\}$ to be the shortest for which there exists product i with at least two spells, $K^i \geq 2$, and $1 \leq j < k \leq K^i$ such that $\zeta^i_j = T_0, \zeta^i_k = t$ for any $t \in \{T_0,\bar{T}\}$.[11] Without loss of generality, we normalize $b_{T_0} = 1$.

Let $\boldsymbol{b}_{./T_0}$ be the vector $\boldsymbol{b}$ without its component b_{T_0}, that is, $\boldsymbol{b}_{./T_0} = (b_{\underline{T}}, \ldots b_{T_0-1}, b_{T_0+1}, \ldots b_{\bar{T}})$.

[11] If no product with at least two spells has a complete spell of duration t, then we estimate $\hat{b}_t = 0$ and so we cannot use it for normalization.

Linearity of $f^{[b]}_{t_1,t_2}(\boldsymbol{\zeta};\boldsymbol{b})$ and normalization of b_{T_0} implies that we can write

$$\boldsymbol{f}^{[b]}(\boldsymbol{\zeta};\boldsymbol{b}) = U^{[b]}(\boldsymbol{\zeta})\boldsymbol{b}_{\cdot/T_0} - V^{[b]}(\boldsymbol{\zeta}),$$

where $U^{[b]}$ is $M \times T$ matrix, and $V^{[b]}(\boldsymbol{\zeta})$ is a vector of length M. With this notation, we can write

$$\mathbb{E}\left[U^{[b]}(\boldsymbol{\zeta})\right]\boldsymbol{b}_{\cdot/T_0} - \mathbb{E}\left[V^{[b]}(\boldsymbol{\zeta})\right] = 0. \tag{16}$$

Proposition 4 gives us one moment condition for each $\underline{T} \leq t \leq \bar{T}$. Define $\boldsymbol{f}^{[H]}_{\bar{T}}$ as a vector function, with m^{th} element given by $f^{[H]}_{m+\underline{T}-1,\bar{T}}(\boldsymbol{\zeta};\boldsymbol{H}^{\bar{T}})$ for $m = 1,\ldots,T+1$. Since equation (8) is linear in $\boldsymbol{H}^{\bar{T}}$, we can write $f^{[H]}_{m+\underline{T}-1,\bar{T}}(\boldsymbol{\zeta};\boldsymbol{H}^{\bar{T}}) = U^{[H]}\boldsymbol{H}^{\bar{T}} - V^{[H]}$, where $U^{[H]}$ is a $(T+1)\times(T+1)$ matrix and $V^{[H]}$ is a $(T+1)\times 1$ vector. With this notation, the moment condition from Proposition 4 becomes

$$\mathbb{E}\left[U^{[H]}(\boldsymbol{\zeta})\right]\boldsymbol{H}^{\bar{T}} - \mathbb{E}\left[V^{[H]}(\boldsymbol{\zeta})\right] = 0. \tag{17}$$

We stack these moment conditions for $\boldsymbol{b}$ and $\boldsymbol{H}^{\bar{T}}$. Define

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{b}_{\cdot/T^*} \\ \boldsymbol{H}^{\bar{T}} \end{pmatrix}, \boldsymbol{f}(\boldsymbol{\zeta};\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{f}^{[b]}(\boldsymbol{\zeta};\boldsymbol{b}) \\ \boldsymbol{f}^{[H]}_{\bar{T}}(\boldsymbol{\zeta};\boldsymbol{H}^{\bar{T}}) \end{pmatrix}, U = \begin{pmatrix} U^{[b]} & 0 \\ 0 & U^{[H]} \end{pmatrix}, V = \begin{pmatrix} V^{[b]} \\ V^{[H]} \end{pmatrix}.$$

Then the moment conditions are

$$\mathbb{E}\left[U(\boldsymbol{\zeta})\right]\boldsymbol{\beta} - \mathbb{E}\left[V(\boldsymbol{\zeta})\right] = 0. \tag{18}$$

To estimate the model, we replace expected values with sample means:

$$U_I \equiv \frac{1}{I}\sum_{i=1}^{I} U(\boldsymbol{\zeta}^i), \quad V_I \equiv \frac{1}{I}\sum_{i=1}^{I} V(\boldsymbol{\zeta}^i).$$

The sample analog of (18) is $U_I\boldsymbol{\beta} - V_I = 0$. For a given positive-definite $(M+T+1)\times(M+T+1)$ weighting matrix W, the estimator $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{2T+1}_{+}$ solves

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{2T+1}_{+}} \left(U_I\boldsymbol{\beta} - V_I\right)' W \left(U_I\boldsymbol{\beta} - V_I\right).$$

This is a linear-quadratic maximization problem and its solution is known in a closed form:

$$\hat{\boldsymbol{\beta}} = \left(U_I'\left(W + W'\right)U_I\right)^{-1} U_I'\left(W + W'\right)V_I.$$

In practice, we choose the identity matrix as a weighting matrix.

Proposition 3 and 4 imply consistency of GMM without any other assumptions. In particular, we do not need to impose that the space of possible parameters $\boldsymbol{\beta}$ is compact since our estimator is linear; see Newey and McFadden (1994).[12]

B.2 Clustered Standard Errors

Recall that the GMM formula for the variance-covariance matrix of the parameter vector $\boldsymbol{\beta}$ is

$$VAR \equiv \frac{1}{I}(F'WF)^{-1}F'W\Omega W'F(F'W'F)^{-1}, \tag{19}$$

where F is the score matrix $F \equiv E[\nabla_{\boldsymbol{\beta}} \boldsymbol{f}]$ and $\Omega = E[\boldsymbol{f}\boldsymbol{f}']$. To get an estimate of the variance-covariance matrix, we replace F and Ω with its sample analogs F_I and Ω_I:

$$F_I \equiv \frac{1}{I}\sum_{i=1}^{I} \nabla_{\boldsymbol{\beta}} \boldsymbol{f}(\boldsymbol{\zeta}^i; \widehat{\boldsymbol{\beta}}) = U_I, \qquad \Omega_I \equiv \frac{1}{I}\sum_{i=1}^{I} \boldsymbol{f}(\boldsymbol{\zeta}^i; \widehat{\boldsymbol{\beta}})\boldsymbol{f}(\boldsymbol{\zeta}^i; \widehat{\boldsymbol{\beta}})',$$

where $\widehat{\boldsymbol{\beta}}$ is a GMM estimate of $\boldsymbol{\beta}$.

To implement one-way clustering, we follow Cameron, Gelbach, and Miller (2011). Formula (19) still applies but with cluster-robust sample analog of Ω. Let Q denote the number of clusters indexed by $q = 1, \ldots, Q$. If a product i belongs to cluster q, we say $\mathbb{1}_{i \in q} = 1$. Define $\bar{\boldsymbol{f}}_q$ as the sum of the moment conditions across products in cluster q,

$$\bar{\boldsymbol{f}}_q = \sum_{i=1}^{I} \boldsymbol{f}(\boldsymbol{\zeta}^i; \widehat{\boldsymbol{\beta}})\mathbb{1}_{i \in q}.$$

Then

$$\Omega_I^{[cluster]} = \frac{Q}{Q-1}\frac{I-1}{I-(2T+1)}\frac{1}{I}\sum_{q=1}^{Q} \bar{\boldsymbol{f}}_q \bar{\boldsymbol{f}}_q',$$

where $2T+1$ is the number of parameters. The term $\frac{Q}{Q-1}\frac{I-1}{I-(2T+1)}$ is adjustment for the degrees of freedom; without this adjustment, the clustered standard errors are biased downwards. We obtain the variance-covariance matrix by substituting $\Omega_I^{[cluster]}$ into equation (19).

[12]Theorem 2.7 states conditions for consistency of estimators without compactness. Example 1.2 on page 2134 then shows that these conditions are satisfied for the linear GMM estimators.

B.3 Practical Consideration

It is a known that in practice matrix Ω_I (or $\Omega_I^{[cluster]}$) can be badly scaled, especially with a large number of moments as we have. This is not necessarily an issue for estimating of the variance-covariance matrix VAR but is for the $J-$test which requires inverting the matrix Ω_I (or $\Omega_I^{[cluster]}$).

Moreover, in our application, Ω_I has some negative eigenvalues. This is a result of numerical imprecisions; matrix Ω_I as well $\Omega_I^{[cluster]}$ is positive semidefinite in any sample by construction.

We address both of these issues in one step, following Cameron, Gelbach, and Miller (2011) and Politis (2011). We construct matrix Ω_I, compute its eigenvalues and replace all negative one and those close to zero in absolute term, with a small positive number ε to construct Ω_I^+, a positive definite matrix. Specifically, we write $\Omega_I = A\Lambda A'$, where $\Lambda = Diag(\lambda_1, \ldots, \lambda_K)$ are the eigenvalues of Ω_I, and A is a matrix of eigenvectors. We define $\lambda_j^+ = \max(\varepsilon, \lambda_j)$ and $\Lambda^+ = Diag(\lambda_1^+, \ldots, \lambda_K^+)$. We then construct $\Omega_I^+ = A\Lambda^+ A'$.

We need to balance two forces when choosing ε. It has to be small enough so that it does not affect results as the sample size grows, and at the same time, it has to be big enough to address the problem of ill-conditioned matrix. Politis (2011) suggests to choose $\varepsilon = I^{-a}$ for $a \in [1, 2]$; we follow this suggestion and choose $a = 1.5$.

We find that Ω_I with no clustering and $\Omega_I^{[cluster]}$ with one-way clustering has a small share of negative eigenvalues, less than 2.5%, and that they are small in absolute value, of the order of 10^{-18}. This gives us confidence that these are indeed numerical imprecisions which we correct with the above described procedure.

C Time-Dependent Pricing with Heterogeneous Firms

Consider a discrete time economy populated by heterogeneous firms indexed by a parameter θ. All firms use time-dependent pricing rules. Let $\Phi_t(\theta)$ be the probability that a type θ firm does not adjust price within t periods of its last price change, with $\Phi_0(\theta) = 1$. In the MPH model, $\Phi_t(\theta) = \prod_{s=0}^{t}(1 - \theta b_s)$ where $b_0 = 0$ and b_s, $s \geq 1$, is the baseline hazard. Let $D(\theta) \equiv \sum_{t=0}^{\infty} \Phi_t(\theta) = \sum_{t=1}^{\infty} t(\Phi_{t-1}(\theta) - \Phi_t(\theta))$ denote the expected time between price changes, so $D(\theta)$ is decreasing in θ in the MPH model.

A firm with type θ that adjusts its price at time t chooses a price $p_t(\theta)$ to minimize $\sum_{s=0}^{\infty} \Phi_s(\theta)(p_t(\theta) - p_{t+s}^*)^2$, where p_{t+s}^* is the "target" price at time $t + s$. This implies

$$p_t(\theta) \equiv \frac{\sum_{s=0}^{\infty} p_{t+s}^* \Phi_s(\theta)}{\sum_{s=0}^{\infty} \Phi_s(\theta)}$$

is the price set by a type θ firm at t. In particular, if p^*_{t+s} is increasing in s, then $p_t(\theta)$ is decreasing in θ in the MPH model.

Now let $G(\theta)$ denote the distribution of θ among the firms that change their price at an arbitrary date t; for notational simplicity, we assume that this is constant over time. The average price among firms that change their price at t is

$$\bar{p}_t \equiv \int p_t(\theta) dG(\theta). \tag{20}$$

Similarly, the average duration of a price is

$$\bar{D} \equiv \int D(\theta) dG(\theta). \tag{21}$$

We could also compute the Kaplan-Meier survivor function at duration t among the price-changing firms:

$$\Phi_t^{KM} = \int \Phi_t(\theta) dG(\theta).$$

If there were a single firm with a time-dependent pricing rule with survivor function Φ_t^{KM}, its optimal price would be

$$\begin{aligned} \bar{p}_t^{KM} &\equiv \frac{\sum_{s=0}^\infty p^*_{t+s} \Phi_s^{KM}(\theta)}{\sum_{s=0}^\infty \Phi_s^{KM}(\theta)} \\ &= \frac{\sum_{s=0}^\infty p^*_{t+s} \int \Phi_s(\theta) dG(\theta)}{\sum_{s=0}^\infty \int \Phi_s(\theta) dG(\theta)} \\ &= \frac{\int \left(\sum_{s=0}^\infty p^*_{t+s} \Phi_s(\theta)\right) dG(\theta)}{\int \sum_{s=0}^\infty \Phi_s(\theta) dG(\theta)} \\ &= \frac{\int p_t(\theta) D(\theta) dG(\theta)}{\int D(\theta) dG(\theta)}, \end{aligned} \tag{22}$$

where the first equation uses the definition of $\Phi_s^{KM}(\theta)$, the second flips the order of summation and integration, and the third uses the definition of $D(\theta)$ and $p_t(\theta)$.

Now compute the ratio of the covariance of a product's price to the duration of the price, divided by the mean duration of prices:

$$\begin{aligned} \frac{\text{cov}(p_t(\theta), D(\theta))}{\bar{D}} &= \frac{\int (p_t(\theta) - \bar{p}_t)(D(\theta) - \bar{D}) dG(\theta)}{\bar{D}} \\ &= \frac{\int p_t(\theta) D(\theta) dG(\theta)}{\bar{D}} - \int p_t(\theta) dG(\theta) = \bar{p}_t^{KM} - \bar{p}_t. \end{aligned}$$

The first equation is the definition of the covaraince, while the second expands the integrand in the numerator and simplifies. The last equation follows immediately from equations (20), (21), and (22). In the MPH model with an increasing sequence for target prices p_t^*, $p_t(\theta)$ and $D(\theta)$ are both decreasing in θ. Thus the covariance is positive whenever the type distribution is nondegenerate. This implies $\bar{p}_t^{KM} > \bar{p}_t$, so ignoring heterogeneity across firms would lead us to overstate average prices. The opposite happens when the sequence of target prices is decreasing.

To see how this might matter in an equilibrium economy, suppose that there are strategic complementarities in price setting, meaning that the target price is an increasing function of the average price prevailing among all firms. Consider the impact of a one-time expansionary (contractionary) monetary policy shock, which will lead to a gradual increase (decrease) in the target price due to strategic complementarities and price stickiness. Given any such sequence for the target price, our argument so far states that the average price among price changers is higher (lower) when all firms have a common survivor function Φ_t^{KM} than when they are heterogeneous with individual survivor functions $\Phi_t(\theta)$. Strategic complementarity implies a feedback from this to the target price, with a higher (lower) target when firms are homogeneous. That is, prices are less sticky when firms are homogeneous. That is, ignoring heterogeneity in price stickiness minimizes price stickiness and hence minimizes the real effects of the monetary policy shock.

D Additional Empirical Results

We report additional empirical results in this section. Figure 7 shows sensitivity of the baseline hazard in the competing risks model using data with censoring time is bounded from above by 80 weeks for each product.

E Continuous Time with Discrete Measurement

In this appendix we formulate a continuous time MPH model with discrete time measurement (CT-DM), which is correctly specified in real-world data where durations are rounded to integer values. We assume each product has a censoring time $c \in \mathbb{R}_+$ with continuous cumulative distribution P and a type θ drawn from a Gamma distribution with mean m and variance v. We later consider an extension to the case where the frailty distribution is a mixture of Gamma distributions. In contrast to our GMM estimates of the discrete time model, we impose that c and θ are independent random variables.

In the continuous time mixed proportional hazard model, we assume that for any $t \in \mathbb{R}_+$, the probability that the true duration of a spell is at least t for a product with type θ is

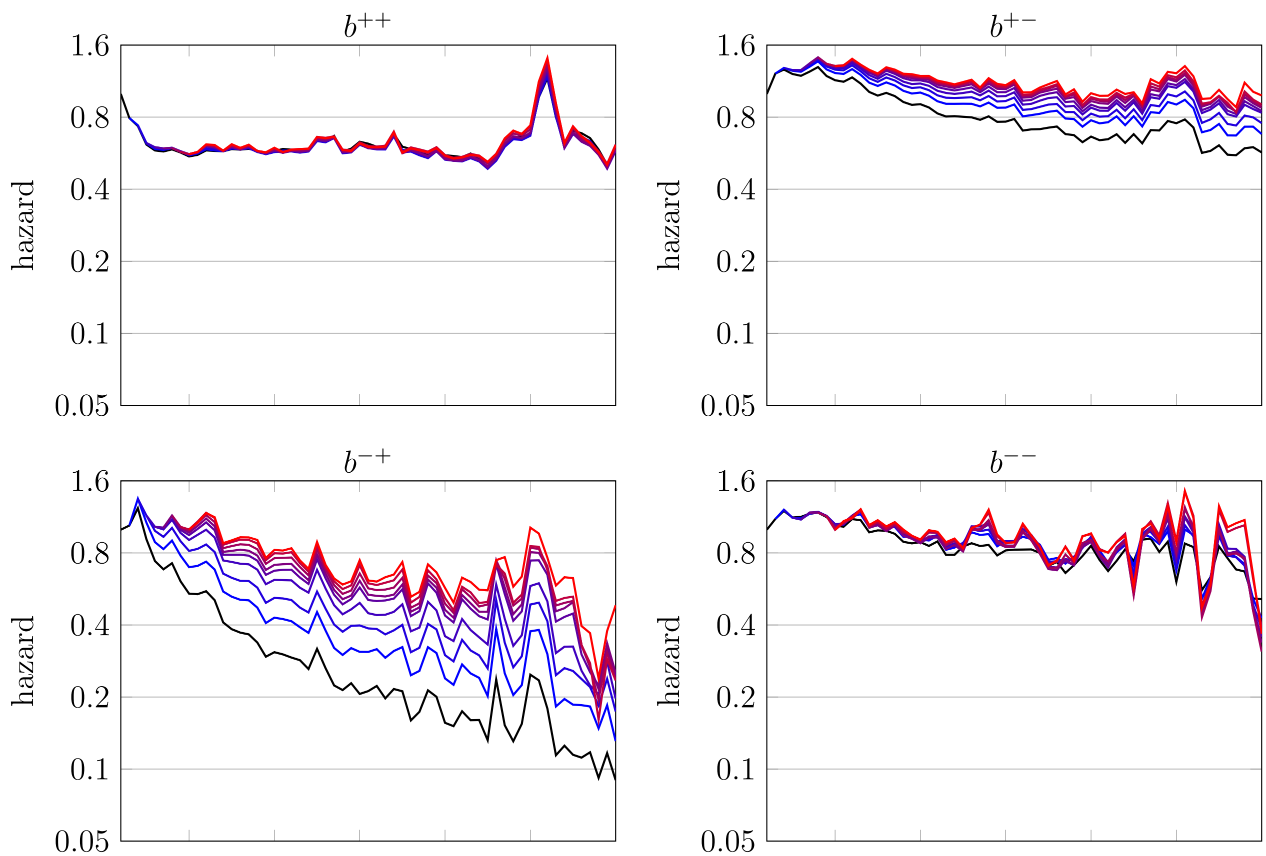

Figure 7: Baseline hazard for the competing risks model, pooled sample, estimated using different values of $\underline{T} \in \{2, \ldots, 10\}$, $\bar{T} = 60$ and censoring time restricted to be at most 80 weeks.

$e^{-\theta \int_0^t b(s)ds}$ for all $t \geq 0$. With discrete measurement, we assume that the measured duration is always rounded up to the next integer. That is, for $t = 1, 2, \dots$, the probability that measured duration is at least t is $e^{-\int_0^{t-1} \theta b(s)ds}$.

In the CT-DM model, there is no hope of recovering the baseline hazard at all real durations, since we only observe integer outcomes. Instead, for any $t = 1, 2, \dots$, define $b_t \equiv \int_{t-1}^t b(s)ds$. Additionally, for notational convenience continue to assume $b_0 = 0$. Our objective is to recover $\boldsymbol{b} \equiv \{b_1, \dots, b_T, b_{T+1}\}$, where sparsity of data lead us to impose $b_t = b_{T+1}$ for all $t \geq T+1$. It is also useful to define the integrated hazard $z_t \equiv \sum_{s=0}^t b_s = \int_0^t b(s)ds$, so the probability that measured duration of a spell is at least $t = 1, 2, \dots$ for a type θ product is $e^{-\theta z_{t-1}}$.

We formulate the likelihood function for case where we observe two spells per product. The data we observe is censored, $(c^i, d_1^i, d_2^i, \zeta_1^i, \zeta_2^i)$ for a typical individual i, where ζ_j^i is the measured duration of j^{th} spell and d_j^i equals one if j^{th} spell is censored. If the first spell right-censored (and hence the second spell is not observed), we code the duration of the second spell as $\zeta_2^i = 0$ and $d_2^i = 1$. Under our assumptions we can write down the likelihood of different outcomes. First, we may observe two completed spells, $\zeta_1^i = t_1 \in \{1, 2, \dots\}$, $\zeta_2^i = t_2 \in \{1, 2, \dots\}$, and $d_1^i = d_2^i = 0$. The probability of this event is

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i = t_1, \zeta_2^i = t_2, d_1^i = d_2^i = 0}\right] = \\ \big(1 - P(t_1 + t_2 - 1)\big) \int_0^\infty e^{-\theta(z_{t_1-1} + z_{t_2-1})}(1 - e^{-\theta b_{t_1}})(1 - e^{-\theta b_{t_2}}) \frac{e^{-\frac{m\theta}{v}} \left(\frac{m\theta}{v}\right)^{\frac{m^2}{v}}}{\theta \Gamma(m^2/v)} d\theta.$$

The integrand is equal to the probability that the censoring time exceeds $t_1 + t_2$, $c^i \geq t_1 + t_2$, multiplied by the probability that the uncensored durations (τ_1^i, τ_2^i) are exactly (t_1, t_2) given θ, multiplied by the density of a Gamma distribution with mean m and variance v. Here Γ is the gamma function. Solve the integral to get

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i = t_1, \zeta_2^i = t_2, d_1^i = d_2^i = 0}\right] = \big(1 - P(t_1 + t_2 - 1)\big) f_0^{CT-DM}(t_1, t_2; \boldsymbol{z}, m, v)$$

where

$$f_0^{CT-DM}(t_1, t_2; \boldsymbol{z}, m, v) \equiv \left(1 + \frac{v}{m}(z_{t_1-1} + z_{t_2-1})\right)^{-\frac{m^2}{v}} - \left(1 + \frac{v}{m}(z_{t_1} + z_{t_2-1})\right)^{-\frac{m^2}{v}} \\ - \left(1 + \frac{v}{m}(z_{t_1-1} + z_{t_2})\right)^{-\frac{m^2}{v}} + \left(1 + \frac{v}{m}(z_{t_1} + z_{t_2})\right)^{-\frac{m^2}{v}}.$$

We note the explicit dependence of this function on the integrated hazard $\boldsymbol{z} = \{z_1, z_2, \dots\}$, as well as the mean and variance of the frailty distribution.

Second, we may observe a completed spell followed by a censored spell, $\zeta_1^i = t_1 \in \{1, 2, \dots\}$, $\zeta_2^i = t_2 \in \{0, 1, \dots\}$, $d_1^i = 0$, $d_2^i = 1$. The probability of this event is

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i = t_1, \zeta_2^i = t_2, d_1^i = 0, d_2^i = 1}\right] = \\ \big(P(t_1 + t_2) - P(t_1 + t_2 - 1)\big) \int_0^\infty e^{-\theta(z_{t_1 - 1} + z_{t_2})}(1 - e^{-\theta b_{t_1}}) \frac{e^{-\frac{m\theta}{v}} \left(\frac{m\theta}{v}\right)^{\frac{m^2}{v}}}{\theta \Gamma(m^2/v)} d\theta.$$

This is the probability that the censoring time is exactly $t_1 + t_2$, $c^i = t_1 + t_2$ multiplied by the probability that $\tau_1^i = t_1$ and $\tau_2^i > t_2$. Again, solve the integral to get

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i = t_1, \zeta_2^i = t_2, d_1^i = 0, d_2^i = 1}\right] = \big(P(t_1 + t_2) - P(t_1 + t_2 - 1)\big) f_1^{CT-DM}(t_1, t_2; \boldsymbol{z}, m, v)$$

where

$$f_1^{CT-DM}(t_1, t_2; \boldsymbol{z}, m, v) \equiv \left(1 + \frac{v}{m}(z_{t_1 - 1} + z_{t_2})\right)^{-\frac{m^2}{v}} - \left(1 + \frac{v}{m}(z_{t_1} + z_{t_2})\right)^{-\frac{m^2}{v}}.$$

Finally, we may observe a single censored spell, $\zeta_1^i = t_1 \in \{1, 2, \dots\}$ and $d_1^i = d_2^i = 1$. The probability of this event is

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i = t_1, d_1^i = 1}\right] = \big(P(t_1) - P(t_1 - 1)\big) \int_0^\infty e^{-\theta z_{t_1}} \frac{e^{-\frac{m\theta}{v}} \left(\frac{m\theta}{v}\right)^{\frac{m^2}{v}}}{\theta \Gamma(m^2/v)} d\theta.$$

This is the probability that the censoring time is t_1, $c^i = t_1$, multiplied by the probability that $\tau_1^i > t_1$. Solve the integral to get

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i = t_1, d_1^i = 1}\right] = \big(P(t_1) - P(t_1 - 1)\big) f_2^{CT-DM}(t_1, 0; \boldsymbol{z}, m, v)$$

where

$$f_2^{CT-DM}(t_1, 0; \boldsymbol{z}, m, v) \equiv \left(1 + \frac{v}{m} z_{t_1}\right)^{-\frac{m^2}{v}}.$$

We can use the probability of these three events to compute the log-likelihood. We treat P as a nuisance parameter and take advantage of the fact that each of the probabilities is multiplicatively separable in the terms involving P to get

$$\mathcal{L}^{CT-DM} = \frac{1}{N} \sum_{i=1}^N \log f_{d_1^i + d_2^i}^{CT-DM}(\zeta_1^i, \zeta_2^i; \boldsymbol{z}, m, v). \tag{23}$$

We impose $z_0 = 0$, which holds by definition. We also normalize $m = 1$.[13] Given a data set, we can search for values of $\boldsymbol{z}$ and v to maximize this likelihood, subject to the constraint $z_{t+1} - z_t = b_{T+1}$ for $t \geq T$. We then first difference the integrated hazard z_t to recover the baseline hazard, $b_t = z_t - z_{t-1}$.

It is straightforward to extend this analysis to the case where the frailty is a mixture of K gamma distributions. Let $\{m_k, v_k, w_k\}$ denote the mean, variance, and weight on each distribution. Then the likelihood is

$$\mathcal{L}^{CT-DM} = \frac{1}{N} \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} w_k f_{d_1^i + d_2^i}^{CT-DM} (\zeta_1^i, \zeta_2^i; \boldsymbol{z}, m_k, v_k) \right). \tag{24}$$

We again impose $z_0 = 0$ and fix $\sum_{k=1}^{K} w_k = 1$ and m_k, v_k, and w_k all nonnegative to have a mixture model. We also normalize $\sum_{k=1}^{K} w_k m_k = 1$. We then search for values of $\boldsymbol{z}$ and distributional parameters which maximize the likelihood for fixed K.

F Continuous Time with Continuous Measurement

F.1 Likelihood Function

We next turn to the continuous time model with continuous time measurement (CT-CM). As in CT-DM, we assume each product has a censoring time $c \in \mathbb{R}_+$ with continuous cumulative distribution P and a type θ drawn from a Gamma distribution with mean m and variance v. We later consider an extension to the case where the frailty distribution is a mixture of Gamma distributions. We again impose that c and θ are independent random variables.

We also assume that for any $t \in \mathbb{R}_+$, the probability that the true duration of a spell is at least t for a product with type θ is $e^{-\theta z(t)}$ for all $t \geq 0$, where $z(t) \equiv \int_0^t b(s)ds$. As usual, measured durations may be censored, but here we assume that we can measure the exact duration or censoring time for each spell

The data we observe is $(c^i, d_1^i, d_2^i, \zeta_1^i, \zeta_2^i)$ for a typical individual i. Under the assumption of a Gamma frailty distribution with mean m and variance v, independent of c^i, we can write down the likelihood of different outcomes. First, we may observe two completed spells, $\zeta_1^i = t_1 \geq 0$, $\zeta_2^i = t_2 \geq 0$, and $d_1^i = d_2^i = 0$. The density of this event is

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i = t_1, \zeta_2^i = t_2, d_1^i = d_2^i = 0}\right] = \big(1 - P(t_1 + t_2)\big) b(t_1) b(t_2) \int_0^\infty \theta^2 e^{-\theta(z_{t_1} + z_{t_2})} \frac{e^{-\frac{m\theta}{v}} \left(\frac{m\theta}{v}\right)^{\frac{m^2}{v}}}{\theta \Gamma(m^2/v)} d\theta.$$

[13] The likelihood is unaffected by doubling m, quadrupling v, and halving $\boldsymbol{z}$.

The integrand is equal to the probability that the censoring time exceeds t_1+t_2, $c^i \geq t_1+t_2$, multiplied by the density that the uncensored durations (τ_1^i, τ_2^i) are exactly (t_1, t_2) given θ, multiplied by the density of a Gamma distribution with mean m and variance v. Again, Γ is the gamma function. Solve the integral to get

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i=t_1,\zeta_2^i=t_2,d_1^i=d_2^i=0}\right] = \big(1-P(t_1+t_2)\big) f_0^{CT-CM}(t_1,t_2;\boldsymbol{z},m,v)$$

where

$$f_0^{CT-CM}(t_1,t_2;\boldsymbol{z},m,v) \equiv b(t_1)b(t_2)\big(m^2+v\big)\left(1+\frac{v}{m}(z(t_1)+z(t_2))\right)^{-2-\frac{m^2}{v}}.$$

Second, we may observe a completed spell followed by a censored spell, $\zeta_1^i = t_1 \geq 0$, $\zeta_2^i = t_2 \geq 0$, $d_1^i = 0$, $d_2^i = 1$. The density of this event is

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i=t_1,\zeta_2^i=t_2,d_1^i=0,d_2^i=1}\right] = h(t_1+t_2)b(t_1)\int_0^\infty \theta e^{-\theta(z_{t_1}+z_{t_2})}\frac{e^{-\frac{m\theta}{v}}\left(\frac{m\theta}{v}\right)^{\frac{m^2}{v}}}{\theta\Gamma(m^2/v)}d\theta.$$

This is the probability that the censoring time is exactly t_1+t_2, $c^i = t_1+t_2$ multiplied by the probability that $\tau_1^i = t_1$ and $\tau_2^i > t_2$. Again, solve the integral to get

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i=t_1,\zeta_2^i=t_2,d_1^i=0,d_2^i=1}\right] = h(t_1,t_2)f_1^{CT-CM}(t_1,t_2;\boldsymbol{z},m,v)$$

where

$$f_1^{CT-CM}(t_1,t_2;\boldsymbol{z},m,v) \equiv b(t_1)m\left(1+\frac{v}{m}(z(t_1)+z(t_2))\right)^{-1-\frac{m^2}{v}}.$$

Finally, we may observe a single censored spell, $\zeta_1^i = t_1 \geq 0$ and $d_1^i = d_2^i = 1$. The probability of this event is

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i=t_1,d_1^i=1}\right] = h(t_1)\int_0^\infty e^{-\theta z_{t_1}}\frac{e^{-\frac{m\theta}{v}}\left(\frac{m\theta}{v}\right)^{\frac{m^2}{v}}}{\theta\Gamma(m^2/v)}d\theta$$

This is the probability that the censoring time is t_1, $c^i = t_1$, multiplied by the probability that $\tau_1^i > t_1$. Solve the integral to get

$$\mathbb{E}\left[\mathbb{1}_{\zeta_1^i=t_1,d_1^i=1}\right] = h(t_1)f_2^{CT-CM}(t_1,0;\boldsymbol{z},m,v)$$

where

$$f_2^{CT-CM}(t_1,0;\boldsymbol{z},m,v) \equiv \left(1+\frac{v}{m}z_{t_1}\right)^{-\frac{m^2}{v}}.$$

As in the CT-DM model, we use the probability of these three events to compute the log-likelihood, taking advantage of the fact that each of the probabilities is multiplicatively separable in the terms involving P to treat P as a nuisance parameter. This gives us the portion of the likelihood that we are interested in:

$$\mathcal{L}^{CT-CM} = \frac{1}{N}\sum_{i=1}^{N} \log f^{CT-CM}_{d_1^i+d_2^i}(\zeta_1^i, \zeta_2^i; \boldsymbol{z}, m, v). \tag{25}$$

As usual, we normalize $m = 1$.

It is again straightforward to extend this analysis to the case where the frailty is a mixture of K gamma distributions. Let $\{m_k, v_k, w_k\}$ denote the mean, variance, and weight on each distribution. Then the likelihood is

$$\mathcal{L}^{CT-CM} = \frac{1}{N}\sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} w_k f^{CT-CM}_{d_1^i+d_2^i}(\zeta_1^i, \zeta_2^i; \boldsymbol{z}, m_k, v_k)\right). \tag{26}$$

We again impose $\sum_{k=1}^{K} w_k = 1$ and m_k, v_k, and w_k all nonnegative to have a mixture model. We also normalize $\sum_{k=1}^{K} w_k m_k = 1$.

Given any finite data set, we need to impose some restrictions on the baseline hazard in order to maximize either likelihood (25) or (26). We assume that the baseline hazard is piecewise constant and so z is piecewise linear.

F.2 Estimation of CT-CM Model in Stata

Stata has a built-in command for parametric estimation of the MPH model with multiple spells (streg) and observable characteristics. Even though it is necessary to specify frailty distribution as well as the function form of the baseline hazard, one can use a full set of dummy variables for duration to "over-ride" the parametric form of the baseline hazard and estimate it flexibly. Since we are interested in estimating hazards up to duration T, we have only one dummy variable for spells longer than T. This dummy is equal to 1 if the measured duration exceeds $T + 1$ and zero otherwise. We find that the maximum likelihood estimates in Stata following this procedure coincide with the CT-CM(1), the version that uncovers very little evidence of heterogeneity.

In IRI data, the root-mean-square difference between baseline hazards estimated in Stata and CT-CM(1) is 1.4×10^{-3}; the absolute value of the difference in variances is 2.3×10^{-4}. We thus conclude that existing estimates of the continuous time MPH model using Stata may be biased towards finding little heterogeneity.

G Baseline Hazards for Product Categories

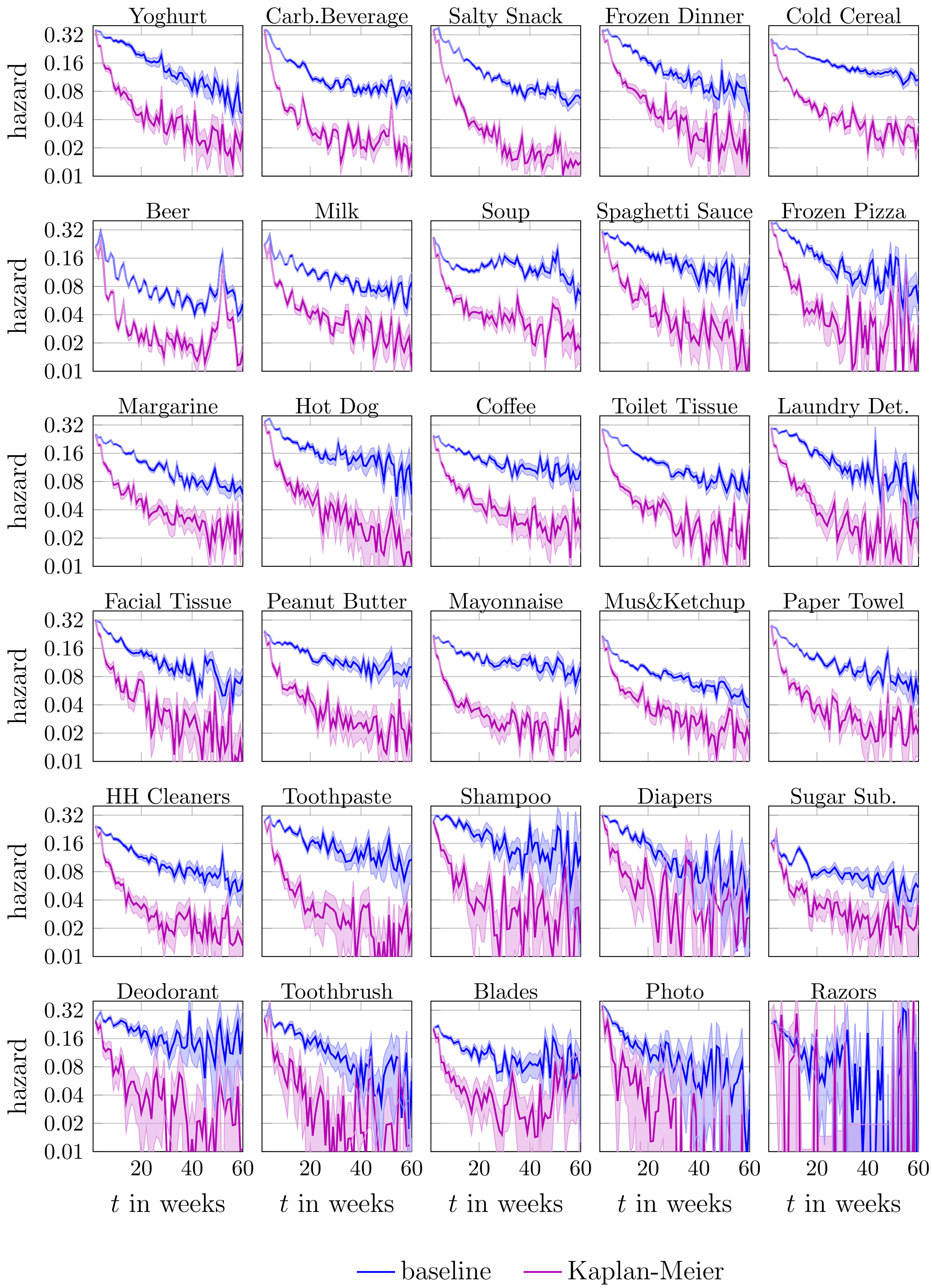

Figure 8: Kaplan-Meier and baseline hazards for individual product categories. Product categoriess are sorted by the number of spell pairs.

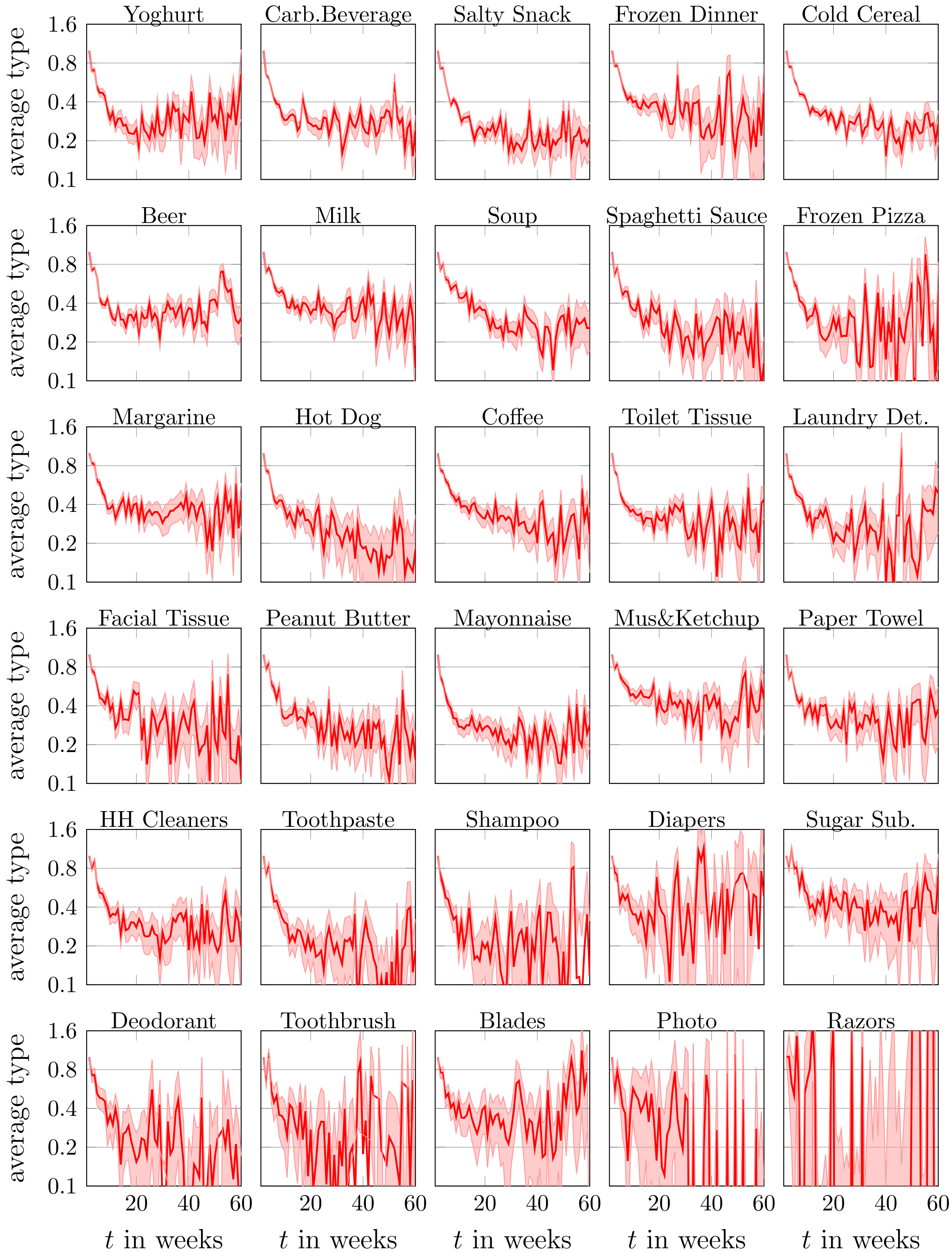

Figure 9: Average type for individual product categories. Product categories are sorted by the number of spell pairs.