DISCUSSION PAPER SERIES

DP16404

Consistent Evidence on Duration Dependence of Price Changes

Fernando Alvarez, Katarína Borovičková and Robert Shimer

MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

Consistent Evidence on Duration Dependence of Price Changes

Fernando Alvarez, Katarína Borovičková and Robert Shimer

Discussion Paper DP16404
Published 28 July 2021
Submitted 27 July 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Consistent Evidence on Duration Dependence of Price Changes

Abstract

We develop an estimator and tests of a discrete time mixed proportional hazard (MPH) model of duration with unobserved heterogeneity. We allow for competing risks, observable characteristics, and censoring, and we use linear GMM, making estimation and inference straightforward. With repeated spell data, our estimator is consistent and robust to the unknown shape of the frailty distribution. We apply our estimator to the duration of price spells in weekly store data from IRI. We find substantial unobserved heterogeneity, accounting for a large fraction of the decrease in the Kaplan-Meier hazard with elapsed duration. Still, we show that the estimated baseline hazard rate is decreasing and a homogeneous firm model can accurately capture the response of the economy to a monetary policy shock even if there is significant strategic complementarity in pricing. Using competing risks and spell-specific observable characteristics, we separately estimate the model for regular and temporary price changes and find that the MPH structure describes regular price changes better than temporary ones.

JEL Classification: E31, E50, C14

Keywords: baseline hazard, Unobserved heterogeneity, Mixed proportional hazard, GMM, sticky prices, nominal rigidities

Fernando Alvarez - f-alvarez1@uchicago.edu
University of Chicago

Katarína Borovičková - katarina.borovickova@nyu.edu
New York University and CEPR

Robert Shimer - robert.shimer@gmail.com
University of Chicago and CEPR

Acknowledgements
We are grateful to David Argente, Jaroslav Borovicka, Stephane Bonhomme, Luca Dedola, Mark Gertler, Simon Gilchrist, Yuriy Gorodnichenko, Lars Hansen, Herve Le Bihan, John Leahy, Simon Mongey, Emi Nakamura, Diego Perez, Edouard Schaal, Alex Torgovitsky, Ivan 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".

Consistent Evidence on Duration Dependence of Price Changes*

Fernando Alvarez
University of Chicago

Katarína Borovičková
New York University

Robert Shimer
University of Chicago

July 27, 2021

Abstract

We develop an estimator and tests of a discrete time mixed proportional hazard (MPH) model of duration with unobserved heterogeneity. We allow for competing risks, observable characteristics, and censoring, and we use linear GMM, making estimation and inference straightforward. With repeated spell data, our estimator is consistent and robust to the unknown shape of the frailty distribution. We apply our estimator to the duration of price spells in weekly store data from IRI. We find substantial unobserved heterogeneity, accounting for a large fraction of the decrease in the Kaplan-Meier hazard with elapsed duration. Still, we show that the estimated baseline hazard rate is decreasing and a homogeneous firm model can accurately capture the response of the economy to a monetary policy shock even if there is significant strategic complementarity in pricing. Using competing risks and spell-specific observable characteristics, we separately estimate the model for regular and temporary price changes and find that the MPH structure describes regular price changes better than temporary ones.

*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".

1 Introduction

This paper makes two contributions. First, we develop a linear generalized method of moments (GMM) estimator for a discrete time mixed proportional hazard (MPH) model for duration data, including an extension to an environment with competing risks. Second, we apply our methodology to data from two large panels for the duration of price spells. We then use the estimated model to re-examine three aspects of sticky price models commonly explored in macroeconomics: the shape of the hazard of price changes as a diagnostic device for distinguishing between different structural models; how the interaction between individual firms' price setting decision affects aggregate price stickiness; and the different behavior of temporary sales versus regular price changes.

Methodological Contribution. The first contribution of this paper is to develop an estimator for a discrete time MPH model for duration data. Here we briefly describe our benchmark MPH model, for concreteness in terms of the elapsed time between price changes of products at the retailer level. 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 specific to a particular product (defined as a bar code and retailer in our empirical application), and is fixed over time. We refer to θ as the product type and assume throughout that it is unobserved. The value of b_t is the *baseline hazard* at duration t. 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 duration of any two spells is 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. We are interested in estimating b_t as well as some measure of heterogeneity as captured by G.

Our estimator builds on continuous time identification results from Honoré (1993). We first extend his identification results to discrete time, and then turn those into a set of moment conditions, linear in b_t, which allow us to estimate the baseline hazard using linear GMM. Our estimator is consistent, it is robust to the shape of the unknown frailty distribution, it allows that the MPH structure only holds on some interval of durations, and it handles both left- and right-censored data. We then extend the set-up and results to an environment with competing risks and observable characteristics, where the MPH structure may only hold for a subset of risks and observables.

We also estimate the Kaplan-Meier hazard H_t, the probability that a typical spell ends in period t conditional on not having ended earlier. We prove that $H_t = b_t\mathbb{E}[\theta|t]$ where $\mathbb{E}[\theta|t]$

is the mean frailty parameter among the spells that have had a constant price until duration t. We prove that in the model without competing risks, the average type $\mathbb{E}[\theta|t]$ must be decreasing with duration t, and use the degree to which it is decreasing as a measure of the importance of heterogeneity.

Our estimator of the baseline hazard is consistent in panel data sets that are large in the cross-sectional dimension, so long as we observe at least two (possibly right-censored) price spells for a positive fraction of products. That is, our inference relies on limits as the number of products goes to infinity, but allows for a boundedly short time dimension. Estimation and inference are simple and quick, even with large data sets, since we use linear GMM. In realistic cases, the model is over-identified, and so we use the Hansen-Sargan J test to explore the assumptions of the MPH model. We also test the prediction in the model without competing risks that the average type $\mathbb{E}[\theta|t]$ is decreasing in duration t.

Cox (1972) and Lancaster (1979) pioneered the analysis of the MPH model in continuous time; see also Lancaster (1990, Chapter 4). The main contributions in terms of non-parametric identification using single spell data and observable covariates are Elbers and Ridder (1982) and Heckman and Singer (1984). Heckman and Honoré (1989) extend this to a competing risks framework. The main contribution on non-parametric identification using repeated spells is Honoré (1993). Abbring and van den Berg (2003) extend this to handle competing risks.

The bulk of the literature estimates the continuous time MPH model using maximum likelihood, either with continuous records or with records that are time aggregated. We believe that there are several advantages to our approach. We impose no restrictions on the frailty distribution, while maximum likelihood requires specifying its family, e.g. a gamma distribution. Our estimator is linear in the baseline hazard, which makes estimation fast and inference straightfoward. In contrast, the likelihood function can be difficult to maximize, particularly when the frailty distribution is a mixture of gammas or when there are competing risks. Finally, our approach to the competing risks model allows the MPH structure to hold only for some risks, with an unspecified hazard rate for other risks. In contrast, maximizing the likelihood requires a parameterized hazard rate (typically an MPH with known frailty family) for all risks.

Horowitz and Lee (2004) build on Honoré's identification argument to construct an estimator for two-spell continuous time data with continuous records. Their estimator, like ours, does not require specification of the frailty distribution. Still, there are again several advantages to our estimator. It is linear, and hence simple to implement and conduct inference. It allows us to easily use all spells, while they only provide an explicit formula using the first two spells. It handles competing risks. It imposes no restrictions on the joint distribution of the unobserved type and the time a product is in the sample, while Horowitz and Lee (2004) impose they are independent. Finally, our estimator is formulated for data measured in discrete time, which is the usual format of duration data sets. If the data generating process were in continuous time and continuous records were available, Horowitz and Lee's approach would be consistent and our method would generally not be.

Application to Price Setting. The second contribution of our paper is applying our estimator of the MPH model to real-world data, and then using the estimates to improve our understanding of macroeconomic models with sticky prices. For this we use the IRI weekly store data, which record weekly revenue and weekly quantities for each product (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 the price as the ratio of revenue to quantity, and a price spell as the time between two price changes for a product. After cleaning, our data contains more than 21 million products. We also explore the Online Micro Price Data from Cavallo (2018) which, while much smaller in size (250,000 products), has daily frequencies and arguably much less measurement error.

We use these data sets to ask three substantive questions. First, we examine the shape of the baseline hazard rate as a diagnostic device for different structural models. As is well known, heterogeneity across products always pushes the Kaplan-Meier hazard rate down with duration due to dynamic selection. For this reason, we concentrate on the shape of baseline hazard rate. State dependent models with persistent cost shocks (Golosov and Lucas, 2007) imply that the hazard rate of price changes is increasing in the time since the last price change for any particular product. In contrast, time dependent models impose no restriction on duration dependence in the baseline hazard, although it is often assumed that the hazard of price changes is constant for each product (Calvo, 1983).

In the IRI data, we find a decreasing baseline hazard rate, despite uncovering a substantial amount of heterogeneity. Figure 1 shows that the baseline hazard rate b_t is generally decreasing between durations 2 and 60 weeks, except for a noticeable spike near one year's duration. The Kaplan-Meier hazard H_t is much steeper throughout the entire time period. As a result, the average type $\mathbb{E}[\theta|t] = H_t/b_t$, which we normalize to 1 at the start of a spell, declines sharply to 0.4 during the first 20 weeks. It then keeps declining at a slower pace, reaching 0.3 after one year. The pattern for the baseline hazard is common in most product categories, and the one for the average type holds for essentially all categories.

We find a similar pattern using the daily Online Micro Price Data, with one important exception: we uncover a sharp spike in the hazard each week, suggesting that many prices only change on a particular day of the week. This justifies our analysis of a discrete time model, where the time period of one week corresponds to the timing of price change decisions.

The second substantive question we examine is how the price setting decisions of heterogeneous firms are affected by and affect the path of prices following a monetary policy shock. We explore a time-dependent pricing model, where the distribution of the duration of price spells is exogenous and described by our estimated MPH model. When firms adjust their price, they do so in order to maximize the expected discounted deviation from a moving target, which in turn depends on both economic fundamentals and the average price set by other firms. Because of the dependence on other firms' prices, the pricing decisions exhibit strategic complementarities, which is known to increase aggregate price stickiness in this type of environment.

We first find the impulse response of the price level to a monetary policy shock using our estimated MPH model. We then compare this with two artificial economies which share the same Kaplan-Meier hazard. In one, all firms are homogeneous and the common hazard rate of each is the Kaplan-Meier hazard. In the other, all firms set prices for a fixed duration as in Taylor (1979, 1980), with a population distribution of durations that achieves the same Kaplan-Meier hazard. If there is no strategic complementarity in pricing, we show that all three economies have the same impulse response (see also Carvalho and Schwartzman, 2015). With strategic complementarity, the exact aggregation result does not hold, and so the relationship between the three impulse responses is a quantitative question. We find that the price level initially adjusts most quickly when firms are homogeneous and least quickly when prices are set for a fixed duration. The response of the price level after about two years is the reverse, largest when prices are set for a fixed duration and smallest when firms are homogeneous. The response in our estimated model lies in between, but is closer to the one in the homogeneous firm economy. We therefore conclude that researchers can safely estimate the Kaplan-Meier hazard and assume that all firms are homogeneous and have this hazard rate. On the other hand, assuming all firms are homogeneous with a constant hazard rate (Calvo, 1983) would give the wrong shape to the impulse response.

The third question we examine is how to distinguish between regular and temporary price changes. We use the extended framework with observable characteristics and competing risks to estimate separately four different baseline hazard rate functions, depending on whether the spell starts and ends with a price increase or decrease, effectively introducing a statistical filter for sales and other temporary price changes. We think, for example, of sales as being a subset of price decreases followed by price increases. Relative to the existing literature (Nakamura and Steinsson, 2008), our approach to regular price changes is consistent with the statistical models used for estimation and testing, and allows for the full generality of the MPH model presented above. When applied to the IRI data, the baseline hazard function for price increases followed by subsequent price increases is much flatter than the one for price decreases followed by price increases. That is, regular price increases have a much flatter baseline hazard rate than the one for sales, consistent with the price plan model in Eichenbaum, Jaimovich, and Rebelo (2011), whose hazard rate Alvarez and Lippi (2020) analyze. A caveat is that the data appears to be much more consistent with an MPH structure for regular price increases as compared with sales, as evidenced by the results of the J test. For this reason, it is reassuring that our econometric approach gives consistent estimates of the baseline hazard for regular price increases, even if sales do not have an MPH structure.

Related Literature. Our application of duration data to price spells builds on the seminal work of Bils and Klenow (2004). Nakamura and Steinsson (2008) and Fougere, Le Bihan, and Sevestre (2007) offer the most thorough analyses of the shape of the baseline hazard in the presence of unobserved heterogeneity for price changes. Nakamura and Steinsson (2008) use both CPI and PPI data for the US, and Fougere, Le Bihan, and Sevestre (2007) use CPI data for France. Both papers use maximum likelihood to estimate the parameters of a continuous time duration model with a parametric frailty distribution. Nakamura and Steinsson (2008) assume the monthly and bimonthly data comes from continuous time records, while Fougere, Le Bihan, and Sevestre (2007) correct for time aggregation.

The Nakamura and Steinsson (2008) approach and ours yield qualitatively similar results on our data set, although there are some significant quantitative differences. To see this, we maximize the likelihood function for a continuous time model with continuous time records using the IRI data set. This estimator recovers less heterogeneity than our GMM estimator. For instance, the average type is estimated to decrease from $\mathbb{E}[\theta|t] = 1$ during the first week to $\mathbb{E}[\theta|t] = 0.37$ at six months using the GMM estimates. Instead it decreases to only $\mathbb{E}[\theta|t] = 0.48$ using the maximum likelihood estimates. Equivalently, the baseline hazard rate estimated using GMM is flatter than the one estimated using maximum likelihood with continuous records. We also compare our results with those we obtain by maximizing the likelihood for time-aggregated records. We conclude that in our data set, time aggregation is quantitatively important but the shape of the frailty distribution and whether we assume time is continuous or discrete is less so.

The closest paper in terms of application, if not results, is Fougere, Le Bihan, and Sevestre (2007). They estimate the baseline hazard for almost 400 product categories at a similar level of aggregation to 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 hypothesize that one reason is the lower frequency of their data (their price data is gathered monthly while ours is gathered weekly). Another more important reason is that they have many fewer price spells per category, roughly three orders of magnitude fewer than ours. 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 spells can end with a price increase or a price decrease. They estimate this model without unobserved heterogeneity and conclude that this extension barely affects the shape of the baseline hazard, a result that is again very different than ours. Finally, they note that their estimator did not converge for the competing risks model with unobserved heterogeneity. Since our GMM estimator is linear in the baseline hazard even in the competing risks extension, this is not an issue for us.

Our three substantive questions are related to a variety of papers, and we turn to those next. The first is about the shape of the hazard rate in structural price setting models. The simplest model of price adjustment is Calvo (1983), where the hazard of price adjustment is a constant function of duration. If this probability differs across firms, as in Carvalho (2006), the model has an MPH structure with a constant baseline hazard and a decreasing Kaplan-Meier hazard. Another large class of models assumes state-dependent prices, e.g. Golosov and Lucas (2007). In a canonical version of this model, the desired price follows a stochastic process, exogenous to any firm, and a 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.

In most cases, adding heterogeneity across firms into a structural model of price setting will typically not lead to an MPH representation. For example, Nakamura and Steinsson (2010) use a model with heterogenous firms following a combination of time and state-dependent price rules to investigate the degree of monetary non-neutrality. In this case, the resulting statistical model will not be exactly an MPH model. Still, we have found that in quantitative versions of these models, the shape of the baseline hazard recovered using an MPH model resembles the "typical" hazard rate for the original model.

The second substantive question we analyze is the effect of monetary policy in an environment with time-dependent price setting rules. Caballero (1989), Reis (2006), and Alvarez, Lippi, and Paciello (2011) analyze models of firms' price setting based on costly information gathering. In an environment with neither strategic complementarity nor strategic substitutability in price setting, they show that optimal price setting rules are pure time-dependent. Alvarez, Lippi, and Paciello (2011) show how these decision rules apply to a once and for all monetary shock. Carvalho (2006) and Nakamura and Steinsson (2010) use models where firms follow time-dependent and state-dependent price setting rules, respectively, to evaluate real effects of monetary policy in presence of heterogeneity. Both papers find that not taking heterogeneity into account leads to underestimation of the real impact of monetary policy. Carvalho and Schwartzman (2015) obtain an analytical characterization of the cumulative impulse response of the aggregate price level to a monetary shock in the presence of heterogeneity for a general time-dependent pricing models. Our analysis further generalizes this to a case not previously considered in the literature mentioned above, where price setting exhibits strategic complementarity.

The third substantive question is related to price setting models of price plans, as introduced by Eichenbaum, Jaimovich, and Rebelo (2011) and further analyzed by Alvarez and Lippi (2020). In these models, a firm can adjust costlessly between a set of prices constituting a "price plan," but it has to either pay a fixed cost or just wait for a free adjustment opportunity to switch its price plan. A firm that follow a price plan will show frequent reversal of prices, since changes within the plan are assumed costless. As a consequence, in these models the hazard of changing the price may be decreasing; see Section F of the Appendix of Alvarez and Lippi (2020). 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.

This paper proceeds as follows. In Section 2, we describe the discrete time MPH model, prove it is nonparametrically identified using repeated spell data, and define the Kaplan-Meier hazard. In Section 3, we discuss measurement issues, especially left- and right-censoring. We then present our estimators for the baseline and Kaplan-Meier hazards. Section 4 extends the framework to allow for observable characteristics and competing risks. We then discuss our data sets in Section 5 before turning to our applied results. In Section 6 we present our estimates of the baseline hazard in both data sets, and then use the estimated model to analyze the aggregate implications of microeconomic heterogeneity for a monetary policy shock. Section 7 distinguishes temporary and regular price changes, showing that the baseline hazard is much flatter for regular price increases compared to the end of sales, and that moreover only the former fits the MPH structure. Section 8 compares our results with those we obtain using other estimation techniques, especially maximum likelihood estimation of a continuous time model with continuous records. Finally, we conclude in Section 9.

2 Discrete Time MPH

2.1 Model

We consider a continuum of products. 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, $1, 2, \ldots$. 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 the baseline hazard b_t is strictly positive for at least one value of t. We also assume that the frailty distribution $G(\theta)$ has a bounded support $[\theta_L, \theta_H]$ with $\theta_H b_t \leq 1$ and $0 \leq \theta_L b_t < 1$ for all $t = 1, 2, \ldots$. We also allow that different types of products are observed with different probability and let $\omega(\theta) > 0$ denote the weight on observations for type θ products. In our application, we will specify this to be the frequency weight (number of spells per unit of time) for a type θ product. It will be convenient to define the weighted frailty distribution

$$G(\theta|\omega) \equiv \frac{\int_{\theta_L}^{\theta} \omega(\theta') dG(\theta')}{\int_{\theta_L}^{\theta_H} \omega(\theta') dG(\theta')}. \tag{1}$$

The sequence of baseline hazards $\{b_1, b_2 \ldots\}$, the frailty distribution G, and weights ω determine the distribution of spell lengths in the population. The duration distribution can be described by its cumulative distribution function, or equivalently, by its survival function

$$\Phi_t(\omega) \equiv \int_{\theta_L}^{\theta_H} \prod_{s=0}^{t} (1 - \theta b_s) dG(\theta|\omega), \tag{2}$$

where for notational convenience we define $b_0 = 0$. This is the fraction of spells that last strictly more than t periods.

Only the product of the baseline hazard b_t and the type θ enters the survival function. This implies that we can multiply the baseline hazard at all durations by a positive multiplicative constant and divide the type of each product by the same constant without affecting the probability of any outcome. In what follows, we therefore identify the baseline hazard up to a multiplicative constant.

2.2 Identification with Multi-Spell Data

We now show that the model is non-parametrically identified with data on two spells. To do this, we first define the two-spell survival function:

$$\Phi_{t_1,t_2}(\omega) \equiv \int_{\theta_L}^{\theta_H} \left(\prod_{s=0}^{t_1} (1-\theta b_s) \right) \left(\prod_{s=0}^{t_2} (1-\theta b_s) \right) dG(\theta|\omega). \tag{3}$$

This is the ω-weighted probability that first spell length is greater than t_1 and the second spell length is great than t_2 given our model structure. It uses the assumption that the length of the two spells is independent conditional on the product type θ.

In this section, we think of the one- and two-spell survival functions as something we can observe in the data,[1] and ask how we can use them to recover the baseline hazard $\boldsymbol{b} \equiv \{b_1, b_2, \dots\}$ (up to the aforementioned multiplicative constant) and the weighted frailty distribution $G(\cdot|\omega)$. Our proof is an adaptation of the identification result of Honoré (1993) to the discrete time model.

Proposition 1 *For an arbitrary weight ω, the baseline hazard $\boldsymbol{b}$ is identified up to a multiplicative constant using the two-spell survival function $\Phi_{t_1,t_2}(\omega)$. Given $\boldsymbol{b}$, the frailty distribution $G(\cdot|\omega)$ is identified using the one-spell survivor function $\Phi_t(\omega)$.*

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

Baseline Hazard. The definition of the two-spell survival function in equation (3) implies

$$\Phi_{t_1-1,t_2-1}(\omega) - \Phi_{t_1,t_2-1}(\omega) = b_{t_1} \int_{\theta_L}^{\theta_H} \theta \left(\prod_{s=0}^{t_1-1} (1-\theta b_s) \right) \left(\prod_{s=0}^{t_2-1} (1-\theta b_s) \right) dG(\theta|\omega),$$
$$\Phi_{t_1-1,t_2-1}(\omega) - \Phi_{t_1-1,t_2}(\omega) = b_{t_2} \int_{\theta_L}^{\theta_H} \theta \left(\prod_{s=0}^{t_1-1} (1-\theta b_s) \right) \left(\prod_{s=0}^{t_2-1} (1-\theta b_s) \right) dG(\theta|\omega).$$

It follows immediately that

$$b_{t_2}\big(\Phi_{t_1-1,t_2-1}(\omega) - \Phi_{t_1,t_2-1}(\omega)\big) = b_{t_1}\big(\Phi_{t_1-1,t_2-1}(\omega) - \Phi_{t_1-1,t_2}(\omega)\big).$$

By varying t_1 and t_2, we can recover $\boldsymbol{b}$ up to a multiplicative constant.

[1] If the duration distribution is defective, $\lim_{t\to\infty} \Phi_t(\omega) \geq 0$, there is a positive probability that we would not observe a second spell. Here we ignore that issue by assuming $\Phi_{t_1,t_2}(\omega)$ is known, but our estimator in Section 3 handles defective duration distributions.

Frailty Distribution. Let $\mu_k \equiv \int_{\theta_L}^{\theta_H} \theta^k dG(\theta|\omega)$ denote the k^{th} moment of the frailty distribution $G(\cdot|\omega)$. 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 that the completed duration of a spell is t is $\Phi_{t-1}(\omega) - \Phi_t(\omega)$, a known function of μ_k, $k = 1, \ldots, t$:

$$\Phi_{t-1}(\omega) - \Phi_t(\omega) = b_t \int_{\theta_L}^{\theta_H} \theta \prod_{s=0}^{t-1} (1 - \theta b_s) dG(\theta|\omega) = b_t \sum_{k=0}^{t-1} \alpha_k(t-1; \boldsymbol{b}) \mu_{k+1}, \tag{4}$$

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

$$\alpha_k(t; \boldsymbol{b}) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k > t \\ \alpha_k(t-1; \boldsymbol{b}) - b_t \alpha_{k-1}(t-1; \boldsymbol{b}) & \text{if } t \geq k > 1 \end{cases} \tag{5}$$

We know $\boldsymbol{b}$. Setting $t = 1$ in equation (4) gives us an equation for μ_1. Having found $\mu_1, \ldots, \mu_{k-1}$, setting $t = k$ in equation (4) gives us an equation for μ_k. Thus by induction we can find all the moments μ_k of $G(\cdot|\omega)$. Since the support of $G(\cdot|\omega)$ is a bounded interval $[\theta_L, \theta_H]$, its moments uniquely determine distribution $G(\cdot|\omega)$. ■

Proposition 1 is behind our approach to estimation, where we convert this logic into moment conditions for the case where we have measures of the survival function from a finite sample.

2.3 Kaplan-Meier Hazard

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:

$$H_t(\omega) \equiv \frac{\Phi_{t-1}(\omega) - \Phi_t(\omega)}{\Phi_{t-1}(\omega)} = b_t \frac{\int_{\theta_L}^{\theta_H} \theta \prod_{s=0}^{t-1}(1 - \theta b_s) dG(\theta|\omega)}{\int_{\theta_L}^{\theta_H} \prod_{s=0}^{t-1}(1 - \theta b_s) dG(\theta|\omega)}. \tag{6}$$

This is the baseline hazard b_t times the *average type* among those products with spell length at least t, $\frac{\int_{\theta_L}^{\theta_H} \theta \prod_{s=0}^{t-1}(1-\theta b_s) dG(\theta|\omega)}{\int_{\theta_L}^{\theta_H} \prod_{s=0}^{t-1}(1-\theta b_s) dG(\theta|\omega)}$. This gives a clear decomposition of the evolution of the Kaplan-Meier hazard $H_t(\omega)$ 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 the average type over time.

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

Proposition 2 *Assume $b_t > 0$ for all t. For any weights ω, the ratio of the Kaplan-Meier hazard to the baseline hazard, $H_t(\omega)/b_t$, is strictly decreasing in t.*

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

$$G_t(\theta|\omega) \equiv \frac{\int_{\theta_L}^{\theta} \prod_{s=0}^{t-1} (1-\theta' b_s)\, dG(\theta'|\omega)}{\int_{\theta_L}^{\theta_H} \prod_{s=0}^{t-1} (1-\theta' b_s)\, dG(\theta'|\omega)}.$$

Consider the double ratio of the densities at $\theta_1 < \theta_2$ and $t_1 < t_2$:

$$\frac{dG_{t_1}(\theta_2|\omega)}{dG_{t_2}(\theta_2|\omega)} \frac{dG_{t_2}(\theta_1|\omega)}{dG_{t_1}(\theta_1|\omega)} = \frac{\prod_{s=0}^{t_1-1} (1-\theta_2 b_s)}{\prod_{s=0}^{t_2-1} (1-\theta_2 b_s)} \frac{\prod_{s=0}^{t_2-1} (1-\theta_1 b_s)}{\prod_{s=0}^{t_1-1} (1-\theta_1 b_s)} = \prod_{s=t_1}^{t_2-1} \frac{1-\theta_1 b_s}{1-\theta_2 b_s},$$

Since $\theta_1 < \theta_2$, $1-\theta_1 b_s > 1-\theta_2 b_s$ and so $\prod_{s=t_1}^{t_2-1} \frac{1-\theta_1 b_s}{1-\theta_2 b_s} > 1$. That is,

$$\frac{dG_{t_1}(\theta_2|\omega)}{dG_{t_2}(\theta_2|\omega)} > \frac{dG_{t_1}(\theta_1|\omega)}{dG_{t_2}(\theta_1|\omega)}. \tag{7}$$

This in turn implies that G_{t_1} first order stochastically dominates G_{t_2}, $G_{t_1}(\theta|\omega) < G_{t_2}(\theta|\omega)$ for all $\theta \in (\theta_L, \theta_H)$. To prove this, suppose to the contrary that there exists a $\theta \in (\theta_L, \theta_H)$ with $G_{t_1}(\theta|\omega) \geq G_{t_2}(\theta|\omega)$. Since these are distribution functions, it follows that

$$\int_{\theta_L}^{\theta} dG_{t_1}(\theta'|\omega) \geq \int_{\theta_L}^{\theta} dG_{t_2}(\theta'|\omega) \text{ and } \int_{\theta}^{\theta_H} dG_{t_1}(\theta'|\omega) \leq \int_{\theta}^{\theta_H} dG_{t_2}(\theta'|\omega),$$

and in particular that there exists a $\theta_1 \in [\theta_L, \theta]$ and a $\theta_2 \in (\theta, \theta_H]$ such that

$$dG_{t_1}(\theta_1|\omega) \geq dG_{t_2}(\theta_1|\omega) \text{ and } dG_{t_1}(\theta_2|\omega) \leq dG_{t_2}(\theta_2|\omega).$$

That contradicts equation (7).

Since G_{t_1} first order stochastically dominates G_{t_2}, the expected value of θ is higher under the former distribution than the latter,

$$\frac{\int_{\theta_L}^{\theta_H} \theta \prod_{s=0}^{t_1-1}(1-\theta b_s) dG(\theta|\omega)}{\int_{\theta_L}^{\theta_H} \prod_{s=0}^{t_1-1}(1-\theta b_s) dG(\theta|\omega)} > \frac{\int_{\theta_L}^{\theta_H} \theta \prod_{s=0}^{t_2-1}(1-\theta b_s) dG(\theta|\omega)}{\int_{\theta_L}^{\theta_H} \prod_{s=0}^{t_2-1}(1-\theta b_s) dG(\theta|\omega)}.$$

Using equation (6), this implies $H_{t_1}(\omega)/b_{t_1} > H_{t_2}(\omega)/b_{t_2}$. ■

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.

3 Estimation and Testing

In this section, we turn the identification result in Proposition 1 into moment conditions for the baseline hazard. We work in a more general environment, where there is an MPH data generating process for durations $t \in \{\underline{T}, \ldots, \bar{T}\}$, $1 \leq \underline{T} < \bar{T} < \infty$, but not necessarily for durations outside this interval. More precisely, let $h_t(\theta)$ be the hazard at duration $t = 1, 2, \ldots$ for a product with type θ. For $t \in \{\underline{T}, \ldots, \bar{T}\}$, 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 and allows for a defective duration distribution. It recognizes that some products may have only a single spell, but uses information from all available spells. This is important in our empirical application.

After we show how to estimate the baseline hazard, we turn to estimation of the Kaplan-Meier hazard rate, the average hazard rate among all products with spells lasting t periods. Here we seek an estimator that does not rely on the MPH structure at all, but simply uses stationarity assumptions on the data generating process.

3.1 Censoring and Measurement

We start by introducing left- and right-censoring into the duration model. A product i is described by a type θ^i and a *censoring time* c^i, possibly correlated. We continue to let $G(\theta)$ denote the frailty distribution and $G(\theta|\omega)$ denote the frailty distribution weighted by ω, as in equation (1). We let $Q_c(\theta)$ be the cumulative distribution of censoring times conditional on type, with density $q_c(\theta) \equiv Q_c(\theta) - Q_{c-1}(\theta)$, allowing for an arbitrary correlation structure. The product type affects the true duration of spells through the hazard function $h_t(\theta^i)$, while the censoring time is equal to the number of (consecutive) periods during which we observe the product.

For results concerning the Kaplan-Meier hazard, we make an additional assumption on the duration of the in-progress spell when we first observe the product: it is a random draw from the stationary ergodic duration distribution for that product. More precisely, let $\tilde{\omega}_t(\theta)$ denote the probability that a type θ product last changed its price t periods ago in the stationary ergodic duration distribution. We measure this immediately after firms change their price, so $\tilde{\omega}_0(\theta)$ is the probability a firm changes its price. For arbitrary $t \geq 0$, this satisfies $\tilde{\omega}_{t+1}(\theta) = (1 - h_{t+1}(\theta))\tilde{\omega}_t(\theta)$. Using $\sum_{t=0}^{\infty} \tilde{\omega}_t(\theta) = 1$, this implies

$$\tilde{\omega}_t(\theta) = \frac{\prod_{s=0}^{t}(1 - h_s(\theta))}{\sum_{t'=0}^{\infty} \prod_{s=0}^{t'}(1 - h_s(\theta))}. \tag{8}$$

If the initial duration for a type θ product is a random draw with density $\tilde{\omega}_t(\theta)$, we call this a *stationary mixture model*.[2]

If we observed product i for infinitely long, we would see a vector of completed durations $\boldsymbol{\tau}^i = \{\tau_0^i, \tau_1^i, \ldots, \tau_{\bar{K}^i}^i\}$, where $\bar{K}^i$ is either a non-negative integer or infinite and $\sum_{j=0}^{\bar{K}^i} \tau_j^i = \infty$.[3] But we do not observe any product for infinitely long. The censoring time c^i affects the *measured* duration of spells, which we denote by $\boldsymbol{\zeta}^i = \{\zeta_0^i, \zeta_1^i, \ldots, \zeta_{K^i}^i\}$. In particular, since the censoring time c^i is finite, we only observe a finite number of spells, $K^i \leq \bar{K}^i$.

To define K^i and $\boldsymbol{\zeta}^i$, it is useful to first define the *residual* censoring time after the start of each spell, c_j^i with $c_0^i = c^i$. The initial spell $j = 0$ may be in progress when we first observe the product, and hence its measured duration may be left-censored, $\zeta_0^i \leq \tau_0^i$. Subsequently we define the residual censoring time for the j^{th} spell as $c_j^i = c_{j-1}^i - \zeta_{j-1}^i$. The j^{th} spell is uncensored if its completed duration is less than the residual censoring time, $\tau_j^i \leq c_j^i$; in this case, the measured duration is $\zeta_j^i = \tau_j^i$. If $\tau_j^i > c_j^i \geq 0$, the measured duration is right censored at $\zeta_j^i = c_j^i + 1 \leq \tau_j^i$; additionally, we set $K^i = j$.[4] We do not observe anything about spells $j > K^i$ and they are not part of the vector $\boldsymbol{\zeta}^i$.

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

Assumption 1 *$K \geq 2$, $\zeta_1 = T_0$, and $\zeta_2 \geq \bar{T}$ with G-positive probability.*

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

$$\int_{\theta_L}^{\theta_H} \left(1 - Q_{T_0+\bar{T}-1}(\theta)\right) \prod_{t=1}^{\bar{T}-1}(1 - h_t(\theta))dG(\theta) > 0.$$

This requires that there is a positive measure of product types θ with (i) a positive probability

[2]The stationary ergodic distribution is defined for a type θ product if and only if the expected duration of a spell is finite, since this is a necessary and sufficient condition for the denominator in equation (8) to be finite. We assume that this is the case when we examine the Kaplan-Meier hazard.

[3]If the duration distribution is defective, $\prod_{s=1}^{\infty}(1 - h_s(\theta^i)) > 0$, $\bar{K}^i$ will be almost surely be finite with $\tau_{\bar{K}^i}^i = \infty$. Otherwise $\bar{K}^i$ is almost surely infinite.

[4]The addition of one period to the measured duration of the right-censored spell captures the fact that we know the spell lasts strictly longer than c_j^i periods and ensures that $\zeta_{K^i}^i$ is a tight lower bound on the duration of the final spell. These definitions imply that if $c_1^i > 0$, $\sum_{j=1}^{K^i-1} \tau_j^i \leq c_1^i$ and $\sum_{j=1}^{K^i} \tau_j^i > c_1^i$, with $K^i = 1$ if $\tau_1^i > c_1^i$.

of a censoring time at least equal to $T_0+\bar{T}$ and (ii) a positive probability of a spell lasting at least $\bar{T}$ periods. If either of these assumptions were violated, no product would have $K \geq 2$ with $\zeta_1 = T_0$ and $\zeta_2 \geq \bar{T}$. In particular, the first assumption ensures that we can observe the product long enough to have $K = 2$, with spell 0 ending in the first period, spell 1 ending T_0 periods later, and spell 2 lasting at least until $\bar{T}$.

When this combination of assumptions determines the distribution of measured duration $\boldsymbol{\zeta}$ and the rank assumption 1 is satisfied, we say that $\boldsymbol{\zeta}$ is drawn from an *MPH model with baseline hazard* $\boldsymbol{b}_0$. We turn next to a consistent estimate of the baseline hazard when measured duration $\boldsymbol{\zeta}$ is drawn from an MPH model with baseline hazard $\boldsymbol{b}$.

3.2 Moment Conditions for the Baseline Hazard

In Section 2, we argued that for any durations t_1, t_2 and any weighting function ω, the model implies

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

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

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

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. For example, if there is a positive probability that $\tau_1^i > c_1^i \geq t_1$, then 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, note that the first spell is right censored for any product with $\tau_1^i > c_1^i \geq t_1$, and so 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 $c_1^i \geq t_1 + t_2$, 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 $c_1^i \geq t_1 + t_2$.

This motivates the following moment condition:

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

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 $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, \ldots, \zeta_K)$ is drawn from an MPH model with baseline hazard $\boldsymbol{b}_0$. Define*

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

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, including standard errors and clustering, in Appendix B. Here we highlight one important feature of our approach: since the moment conditions in equation (9) are linear in the baseline hazard, we obtain the GMM estimator of $\boldsymbol{b}_0$ in closed form.

The multiplicative structure of the MPH model is restrictive and can be tested using a J-test of overidentifying restrictions. In particular, Proposition 3 gives us $T(T+1)/2$ moment conditions to estimate T parameters in the vector $\boldsymbol{b}_0$. For $T \geq 2$, the model is thus overidentified.

We can also build on the proof of Proposition 1 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_1^i and type θ^i are independent, these conditions are difficult to interpret. We therefore relegate them to Appendix C.4 and do not report them in our main analysis.

3.3 Moment Conditions for the Kaplan-Meier Hazard

Proposition 2 tells us that the ratio of the Kaplan-Meier hazard to the baseline hazard, $H_t(\omega)/b_t$, is decreasing in t for any weighting function ω. Generalizing equation (6) to the case of an arbitrary type-dependent hazard function $h_t(\theta)$, the Kaplan-Meier hazard rate at duration t is

$$H_t(\omega) \equiv \frac{\int_{\theta_L}^{\theta_H} h_t(\theta) \prod_{s=0}^{t-1}(1-h_s(\theta))dG(\theta|\omega)}{\int_{\theta_L}^{\theta_H} \prod_{s=0}^{t-1}(1-h_s(\theta))dG(\theta|\omega)}. \tag{10}$$

We seek to test this prediction, but first must describe our choice of ω and how we estimate $H_t(\omega)$.

Anticipating our theoretical model of price setting described in Section 6.3, we would like to measure the hazard rate for a typical spell, rather than a typical product. In a stationary mixture model, defined in Section 3.1, a product changes its price with a fixed probability per unit of time. We thus want weights ω^* equal to the expected number of times that a type θ product changes its price during the observation window:

$$\omega^*(\theta) \equiv \frac{\sum_{c=1}^{\infty} cq_c(\theta)}{\sum_{t=0}^{\infty}\prod_{s=0}^{t}(1-h_s(\theta))}. \tag{11}$$

The numerator is the expected censoring time for a type θ product. We multiply this by the expected number of price changes per unit of time for that product type, $\tilde{\omega}_0(\theta)$ defined in equation (8).[5]

Unfortunately, censoring limits our ability to estimate $H_t(\omega^*)$ for $t \in \{\underline{T}, \dots, \bar{T}\}$. In particular, if $c \leq \bar{T}$, we cannot observe a spell $j \geq 1$ which lasts $\bar{T}$ periods.[6] This means that we cannot estimate the Kaplan-Meier hazard at duration $\bar{T}$ for such products. Since we want to estimate the Kaplan-Meier hazard at all durations $\{\underline{T}, \dots, \bar{T}\}$ across a fixed set of products, we must focus attention on the subset of productst that we observe for more than $\bar{T}$ periods. That is, we thus seek to estimate $H_t(\omega^f)$, where the feasible frequency weight ω^f satisfies

$$\omega^f(\theta) \equiv \frac{\sum_{c=\bar{T}+1}^{\infty} cq_c(\theta)}{\sum_{t=0}^{\infty}\prod_{s=0}^{t}(1-h_s(\theta))}. \tag{12}$$

This gives zero weight to any product that we observe for less than $\bar{T}$ periods, but is otherwise the same as $\omega^*(\theta)$.

We now give a consistent estimator of the Kaplan-Meier hazard $\boldsymbol{H}(\omega^f) = (H_{\underline{T}}(\omega^f), \dots, H_{\bar{T}}(\omega^f))$:

[5]The denominator is also equal to the expected duration of a price spell for a type θ product.

[6]Recall all observations start with an ongoing spell $j=0$ of unknown duration. Thus $c^i \leq \bar{T}$ implies the residual censoring time at the start of spell 1 satisfies $c_1^i \leq c^i - 1$ and hence $c_1^i < \bar{T}$.

Proposition 4 *Assume $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, \ldots, \zeta_K)$ is drawn from a stationary mixture model with Kaplan-Meier hazard $\boldsymbol{H}_0(\omega^f)$. Define*

$$f_{t,\bar{T}}^{[H]}(\boldsymbol{\zeta}; \boldsymbol{H}) \equiv \begin{cases} \frac{c}{c-\bar{T}} \sum_{j=1}^{K} \left(H_t \mathbb{1}_{\zeta_j \geq t, c_j \geq \bar{T}} - \mathbb{1}_{\zeta_j = t, c_j \geq \bar{T}} \right) & \text{if } c > \bar{T} \\ 0 & \text{if } c \leq \bar{T} \end{cases} \tag{13}$$

where $c = \sum_{j=0}^{K} \zeta_j - 1$ and $c_j = \sum_{j'=j}^{K} \zeta_{j'} - 1$. Then $\mathbb{E}\left[f_{t,\bar{T}}^{[H]}(\boldsymbol{\zeta}; \boldsymbol{H}) \right] = 0$ for all $\underline{T} \leq t \leq \bar{T}$ if and only if $\boldsymbol{H} = \boldsymbol{H}_0(\omega^f)$.

We prove this result in Appendix A. The basic idea is that $\frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{\bar{K}^i} \mathbb{1}_{\tau_j^i \geq t}$ and $\frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{\bar{K}^i} \mathbb{1}_{\tau_j^i = t}$ are consistent (but infeasible) estimates of the survivor function and density at duration t. For $t \leq \bar{T}$ and $c_j^i \geq \bar{T}$, the indicator functions are equivalent to ones using measured duration ζ_j^i instead of completed duration τ_j^i, suggesting a path towards a feasible estimator. The only difficulty is that this underweights products with short censoring times c^i. We prove that when durations are drawn from a stationary mixture model, the multiplicative factor $\frac{c^i}{c^i - \bar{T}}$ is the correct reweighting.

We test monotonicity of the ratio $H_t(\omega^f)/b_t$ by looking at the inequalities

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

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

4 Observables and Competing Risks

We now consider two extensions to our basic model, allowing for (possibly spell-specific) observable characteristics which affect the hazard, and permitting competing risks for why a spell ends. In our empirical application, the observable characteristics include the product's category and whether the spell starts with a price increase or decrease; and the competing risk is a spell ending with a price increase or decrease.

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 and spell has an observable characteristic, say $\chi_j \in \{1, \ldots, X\}$ for the j^{th} spell. We allow for correlation between observables and unobservables, as discussed below.

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 $h_t^r(\chi,\boldsymbol{\theta}) \geq 0$ denote the probability that a spell with observable χ and unobservable $\boldsymbol{\theta}$ ends at duration $t \in \{1,2,\dots\}$ for reason $r \in \{1,\dots,R\}$ conditional on not ending earlier. χ captures all observables that affect the hazard, and so in particular conditioning on past observables is not useful for forecasting the hazard.[7] Let $h_t(\chi,\boldsymbol{\theta}) \equiv \sum_{r=1}^{R} h_t^r(\chi,\boldsymbol{\theta})$ denote the probability of a duration t spell with observable χ and unobservable $\boldsymbol{\theta}$ ending in period t. We assume that $h_t(\chi,\boldsymbol{\theta}) \leq 1$ for all $\boldsymbol{\theta}$, so this is a proper probability, and $\int h_t(\chi,\boldsymbol{\theta})dG(\boldsymbol{\theta}) < 1$, so there is a chance of observing spells with observable χ at any duration.

The initial (left-censored) observable characteristic χ_0 is a random variable. Let $\pi_0(x|\boldsymbol{\theta}) \geq 0$ denote the probability that $\chi_0 = x$ given $\boldsymbol{\theta}$, with $\sum_{x=1}^{X} \pi_0(x|\boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta}$. Thereafter, the observable characteristic follows a first order Markov process. For $j \geq 1$, let $\pi(x|\chi_{j-1},\rho_{j-1},\boldsymbol{\theta}) \geq 0$ denote the probability that $\chi_j = x$ conditional on the observable characteristic of the previous spell $\chi_{j-1} \in \{1,\dots,X\}$, the reason the previous spell ended $\rho_{j-1} \in \{1,\dots,R\}$, and the unobserved type, 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 $\boldsymbol{\theta}$.[8] This structure is rich enough to allow for an arbitrary relationship between observable and unobservable characteristics.

For at least one observable characteristic x, reason r, and set of durations $\{\underline{T},\dots,\bar{T}\}$, 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 $t \in \{\underline{T},\dots,\bar{T}\}$. We focus throughout on this pair (x,r) and seek to estimate $\boldsymbol{b} = \{b_{\underline{T}},\dots,b_{\bar{T}}\}$ 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.

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

[8] We assume a Markovian structure for notational simplicity, but can easily relax this assumption. A substantive assumption is 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. We do not know how to relax this assumption without imposing additional structure elsewhere in the model. 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.

4.2 Identification

To understand identification, it is useful to consider an environment with left-censoring but no right-censoring. Consider a product where spell 0 has observable characteristic x_0 and ends for reason r_0. We compute the conditional probability that the first uncensored spell has observable characteristic x and ends for reason r at duration t_1, while the second uncensored spell also has characteristic x and lasts at least t_2 periods. We assume $\underline{T} \leq t_1 < t_2 \leq \bar{T}$ and so use $h^r_{t_1}(x, \boldsymbol{\theta}) = \phi(\boldsymbol{\theta}) b_{t_1}$:

$$\begin{aligned}
&\Pr[\tau_1 = t_1, \tau_2 \geq t_2, \rho_1 = r, \chi_1 = \chi_2 = x | \chi_0 = x_0, \rho_0 = r_0] \\
&\qquad = b_{t_1} \int_{\theta_L}^{\theta_H} \phi(\boldsymbol{\theta}) \pi(x|x_0, r_0, \boldsymbol{\theta}) \pi(x|x, r, \boldsymbol{\theta}) \prod_{s=1}^{t_1 - 1} (1 - h_s(x, \boldsymbol{\theta})) \prod_{s=1}^{t_2 - 1} (1 - h_s(x, \boldsymbol{\theta}))\, dG(\boldsymbol{\theta}).
\end{aligned}$$

Reversing the role of t_1 and t_2 gives

$$\begin{aligned}
&\Pr[\tau_1 = t_2, \tau_2 \geq t_1, \rho_1 = r, \chi_1 = \chi_2 = x | \chi_0 = x_0, \rho_0 = r_0] \\
&\qquad = b_{t_2} \int_{\theta_L}^{\theta_H} \phi(\boldsymbol{\theta}) \pi(x|x_0, r_0, \boldsymbol{\theta}) \pi(x|x, r, \boldsymbol{\theta}) \prod_{s=1}^{t_2 - 1} (1 - h_s(x, \boldsymbol{\theta})) \prod_{s=1}^{t_1 - 1} (1 - h_s(x, \boldsymbol{\theta}))\, dG(\boldsymbol{\theta}).
\end{aligned}$$

Combining these two we get

$$\begin{aligned}
&b_{t_2} \Pr[\tau_1 = t_1, \tau_2 \geq t_2, \rho_1 = r, \chi_1 = \chi_2 = x | \chi_0 = x_0, \rho_0 = r_0] \\
&\qquad\qquad - b_{t_1} \Pr[\tau_1 = t_2, \tau_2 \geq t_1, \rho_1 = r, \chi_1 = \chi_2 = x | \chi_0 = x_0, \rho_0 = r_0] = 0.
\end{aligned}$$

Since this equation holds for any (x_0, r_0), it is true when we integrate across that left-censored spell distribution. This gives us a moment condition:

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

We can then vary t_1 and t_2 to recover $\boldsymbol{b}$ up to a multiplicative constant. This generalizes the identification argument in Proposition 1 to a 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. The remainder of this section shows how to adapt this insight to our framework, building on the approach in Section 3.

4.3 Censoring and 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 $\boldsymbol{\theta}$. We still let $Q_c(\boldsymbol{\theta})$ denote the cumulative distribution of censoring times conditional on type $\boldsymbol{\theta}$ and $G(\boldsymbol{\theta})$ denote the frailty distribution.

As before, we let $\boldsymbol{\zeta}^i = (\zeta_0^i, \zeta_1^i, \dots, \zeta_{K^i}^i)$ be the vector of measured durations, with the first spell left censored and the last spell right censored, so $c^i = \sum_{j=0}^{K^i} \zeta_j^i - 1$. We also let $\boldsymbol{\chi}^i = (\chi_0^i, \chi_1^i, \dots, \chi_{K^i}^i)$ be a vector recording the observable characteristic of each spell and $\boldsymbol{\rho}^i = (\rho_0^i, \rho_1^i, \dots, \rho_{K^i-1}^i)$ 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 rather than $K^i + 1$.

We assume that the baseline hazard $\boldsymbol{b}_0 = \{b_{0,\underline{T}}, \dots, b_{0,\bar{T}}\}$ is nontrivial, $\boldsymbol{b}_0 \neq 0$, and let T_0 denote 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 there is a positive probability that the left-censored spell has observable x_0 and ends for reason r_0 for some (x_0, r_0), and

$$\int_{\theta_L}^{\theta_H} \left(1 - Q_{T_0+\bar{T}-1}(\boldsymbol{\theta})\right)\pi(x|x_0, r_0, \boldsymbol{\theta})\pi(x|x, r, \boldsymbol{\theta}) \prod_{t=1}^{\bar{T}-1}(1 - h_t(x, \boldsymbol{\theta}))dG(\boldsymbol{\theta}) > 0,$$

so there is a positive probability that the censoring time is at least $T_0 + \bar{T}$, spells 1 and 2 both have observable x, and the second spell lasts at least $\bar{T}$ periods.

We highlight two special cases in which this reduces to the rank condition in Assumption 1. First, the observable distribution may have full support for 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. Second, the observable may be fully persistent and $\pi_0(x|\boldsymbol{\theta}) > 0$. 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 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*

$$f^{[b,x,r]}_{t_1,t_2}(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho}; \boldsymbol{b}) \equiv \sum_{(j,k):1\leq j<k\leq K} \left(b_{t_k} \mathbb{1}_{\zeta_j=t_1,\zeta_k\geq t_2,\rho_j=r,\chi_j=\chi_k=x} - b_{t_j} \mathbb{1}_{\zeta_j=t_2,\zeta_k\geq t_1,\rho_j=r,\chi_j=\chi_k=x} \right). \tag{14}$$

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

5 Data

For most of our empirical applications, we use IRI weekly store data,[9] described in Bronnenberg, Kruger, and Mela (2008). We also use Online Micro Price Data

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, [10] such as coffee, carbonated beverages, and detergents, and include approximately 2.6 million distinct items defined by its store and barcode (Universal Product Code, UPC).

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 change from the prior week and then defining a price spell as the time elapsed between two price changes. In particular, suppose that the first price observation occurs at time t_0, the last one at time t_{K+1} and that price changes occur at times $t_1, \ldots, t_K$. Then we construct $\zeta_j = t_{j+1} - t_j$ for $j = 0, \ldots, K-1$, and $\zeta_K = t_{K+1} - t_K + 1$, reflecting the fact that the earliest possible date

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

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

when the price set at t_K can change is $t_{K+1}+1$ and the hence the price spell will be at least ζ_K periods long. Spell 0 has measured duration ζ_0 and is left-censored. The censoring time is $c = t_{K+1} - t_0$.

We use price levels to determine whether the price spell follows a price decrease or increase; we can determine this for every price spell except the left-censored ones. We further use prices to determine whether a price spell ends with a price increase or decrease; we can construct such indicator for every spell which is not right-censored. We use this information to estimate the 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 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 the 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.[11] We therefore set $\underline{T} = 2$ and do not estimate the baseline hazard in week 1.

Table 1 shows summary statistics of the price spells by product category, focusing on price spells longer than $\underline{T} = 2$ weeks (which are not left-censored). The pooled sample contains 21,717,549 products, yielding 684,919,778 pairs of durations where both durations exceed $\underline{T}$. It is the total number of pairs of durations that enters into the sums in Propositions 3 and 5.

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. We think of these as coming from a different model, possibly without an MPH structure, and so drop them to avoid biasing 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

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

Table 1: Descriptive statistics by product category, IRI data. 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 where both are longer than $\underline{T}$. The last two columns show the median and 90^{th} percentile value of the censored spell length.

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 a correlation of 0.029 when duration is measured in levels, and -0.042 when duration is measured in logs. This suggests that the data are unlikely to come from a mixture model. But 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 a bit further to 0.248 in levels and 0.256 in logs when we consider spells at least 3 weeks long. That is, once we exclude one-week spells, the correlation is different from zero, and it is fairly stable if we exclude two-week spells. This is behind our decision to set $\underline{T} = 2$. We examine the robustness of our results to this assumption in Appendix D.3.

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. This means that 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 9 in Appendix D.3 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 $\bar{T} = 60$.

5.3 Online Micro Price Data

We are also interested in looking at higher frequencies than are available in the IRI data. To do this, we use the daily Online Micro Price data, the open access data from the Billion Prices Project presented by Cavallo (2018).[12] In this data set, we observe daily posted prices for many products, which we use to construct price spells. We focus on the U.S. stores. We impute some missing observations. If the last non-missing price before missing observations is the same as the first non-missing price after the missing observations, we impute this price to the missing observations. If these two prices are not the same, we do not impute any price. After this imputation, we consider the longest window without missing data for each

[12] http://www.thebillionpricesproject.com/datasets/

product in the sample, as we did in the IRI data. The resulting sample contains 71,925 products with at least one spell, $K^i \geq 1$, out of which 48,550 products have at least two spells, $K^i \geq 2$. Since we observe the posted price directly, we do not need to exclude price spells of length 1. For this data, we therefore choose $\underline{T} = 1$ and $\bar{T} = 70$, that is, 10 weeks.

6 Results

We start this section by presenting our estimates of the baseline and Kaplan-Meier hazards for the MPH model. We first analyze our main data set with weekly price observations before turning to a data set with daily price data to explore whether time aggregation affects our results. Finally, we develop a simple theoretical framework in which we explore how heterogeneity and duration dependence interact to create real effects from monetary policy shocks.

6.1 Baseline Hazard and Heterogeneity

Here we present estimates for the basic model with $X = 1, R = 1$ on the pooled sample; see Figure 1. Figures 14 and 15 in Appendix F show results for each product category separately.

The left panel of Figure 1 shows the Kaplan-Meier and baseline hazards. The Kaplan-Meier hazard[13] is steeply declining throughout the whole investigated range, except for the peak at 52 weeks. In contrast, the baseline hazard is constant until week 4, after which it modestly declines, also showing a peak at 52 weeks.

The distinction between the Kaplan-Meier and baseline hazard points to substantial unobserved heterogeneity. Recall from equation (6) that the ratio of the Kaplan-Meier hazard to the baseline hazard is the average type, which captures the extent of dynamic sorting. 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. Within 20 weeks, the average type drops by 60 percent and it continues to decline thereafter, albeit more slowly. This means that heterogeneity plays an important role in shaping the Kaplan-Meier hazard.

We now turn to the two tests of the model. Recall there are $T = \bar{T} - \underline{T} = 58$ independent moments and $M = 1{,}770$ moment conditions. The J-statistic is $J = 10{,}498$, while the critical value of the χ^2 distribution with $M - T$ degrees of freedom is 1,749, implying that we reject the model at any conventional significance level. In what follows, we investigate the source

[13]To estimate the Kaplan-Meier hazard, we need an assumption that the data come from a stationary mixture model. In Appendix D.1, we propose a test of the stationarity assumption and find that the data look close to stationary.

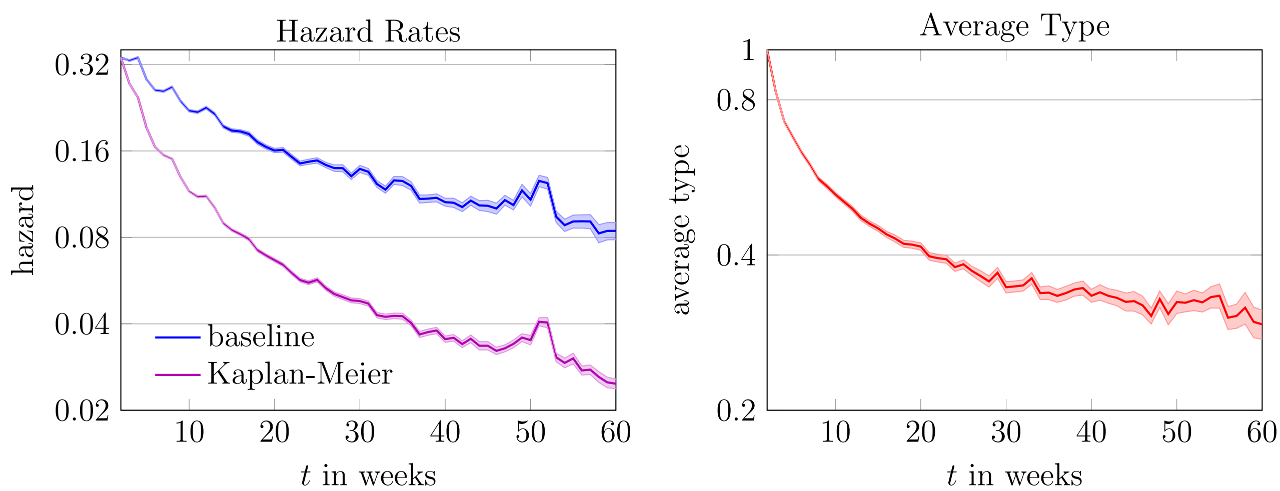

Figure 1: Kaplan-Meier and baseline hazard for pooled IRI data, log scale. The purple line shows the Kaplan-Meier hazard, the blue line is the estimated baseline hazard. The red line shows the "average type" at given duration, calculated as the ratio of Kaplan-Meier and baseline hazards. Shaded regions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

of this rejection in more detail, but note here that we have 21 million products and a model that is significantly over-identified. Perhaps it is not surprising that we fail the J-test in such a situation.

Our second test is whether the average type is decreasing. The right panel of Figure 1 shows a declining trend through durations 2 to 60 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 of a mild spike in the baseline hazard at week 52, suggestive of certain time-dependent pricing rules (Taylor, 1979, 1980). 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. In Section 7, we investigate whether this failure can be driven by the fact that sales follow different dynamics than regular price changes, which we do not distinguish in this section.

6.2 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.

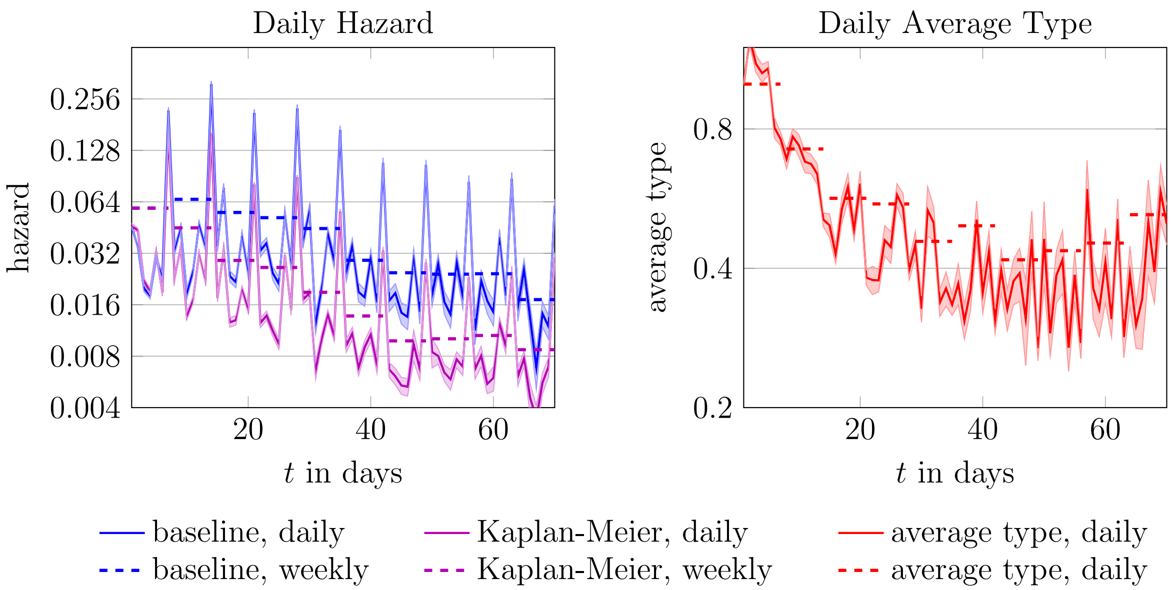

Figure 2: Kaplan-Meier and baseline hazard for Online Micro Price Data using daily and weekly data, log scale. Solid lines use daily data, dashed lines weekly data. The purple line shows the Kaplan-Meier hazard, the blue line is the estimated baseline hazard. The red line shows the "average type" at 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. Kaplan-Meier hazard and baseline hazard are in daily units throughout.

Figure 2 shows the estimates using daily data for $\underline{T} = 1$ day and $\bar{T} = 70$ days (solid lines). We observe that the price-change hazard 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.

To see how much information we gain from using daily data, we aggregate the 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. We then estimate the model again. The dashed lines in Figure 2 show the results, with hazards adjusted to have the same (daily) units. We find that the estimates are similar, even though weekly data recover somewhat less heterogeneity that the daily data.

In Appendix D.2, we conduct a similar exercise with the IRI data, where we aggregate weekly data up to a monthly frequency. We find that the monthly data understate the extent of heterogeneity, even more so than in Online Micro Price Data (Figure 8). We think that this is because in Online Micro Price Data, large spikes in the hazard every seven days indicate that the decision to change prices is made at the weekly frequency, and hence using weekly instead of daily data does not make significant difference. With the IRI data, the week is the relevant unit for measurement, and so monthly data miss selection which happens within a month, leading to an understatement of the extent of heterogeneity.

6.3 Aggregate Implication of Firm-Level Heterogeneity

Finally, we discuss the macroeconomic implications of our estimates using a model of heterogeneous firms with time-dependent pricing rules. We are interested in understanding the dynamic response of the price level to a monetary policy shock. In a richer model with price rigidity, this would translate into the impact on output.

We consider a discrete time economy populated by heterogeneous firms which use time-dependent pricing rules. We assume firms are described by their type θ with population distribution G and support $[\theta_L, \theta_H]$. $\Phi_t(\theta)$ is the survival function for type θ firms, the probability that a type θ firm keeps the same price for more than t periods, with $\Phi_0(\theta) = 1$. We assume the expected duration of a completed spell is finite for all θ, which is equivalent to assuming that $\sum_{t=0}^{\infty} \Phi_t(\theta)$ is finite. We also assume that if a type θ firm adjusts its price at t, it sets it to a new log price denoted $\nu_t(\theta)$.

We assume all products are in the stationary ergodic duration distribution. Let $\tilde{\omega}_t(\theta)$ denote the fraction of type θ firms that have kept the same price for t periods, measured immediately after the price adjustment, so $\tilde{\omega}_0(\theta)$ is the fraction of the time that a type θ firm adjusts its price. Generalizing equation (8) to the case of arbitrary survival functions gives

$$\tilde{\omega}_t(\theta) = \frac{\Phi_t(\theta)}{\sum_{s=0}^{\infty} \Phi_s(\theta)}.$$

This means that the average log price charged by type θ firms at time t is a weighted average of past values of the new prices $\nu_t(\theta)$:

$$p_t(\theta) \equiv \sum_{s=0}^{\infty} \tilde{\omega}_s(\theta)\nu_{t-s}(\theta) = \frac{\sum_{s=0}^{\infty} \Phi_s(\theta)\nu_{t-s}(\theta)}{\sum_{s=0}^{\infty} \Phi_s(\theta)}. \tag{15}$$

This varies over time if $\nu_t(\theta)$ is not constant. We turn to its determinants next.

For all $t \leq 0$, we assume that all firms set a log price normalized to 0. Thereafter, we assume that a type θ firm, if it is able to adjust its price at t, sets a price to minimize the discounted squared difference between the actual log price and a frictionless target log price P_{t+s}^* during the time until the next price change:

$$\nu_t(\theta) \equiv \arg\min_{\nu} \sum_{s=0}^{\infty} \beta^s \Phi_s(\theta)(\nu - P_{t+s}^*)^2,$$

where $\beta \in [0,1]$ is the discount factor. The first order condition implies

$$\nu_t(\theta) = \frac{\sum_{s=0}^{\infty} \beta^s \Phi_s(\theta) P^*_{t+s}}{\sum_{s=0}^{\infty} \beta^s \Phi_s(\theta)}.$$

We also assume that for $t \geq 1$, the frictionless target log price P^*_t is a weighted average of the new long-run steady state log price $\delta \neq 0$ and the average log price charged by other firms P_t, $P^*_t = (1-\alpha)\delta + \alpha P_t$, where

$$P_t \equiv \int_{\theta_L}^{\theta_H} p_t(\theta) dG(\theta). \tag{16}$$

To the extent that the target price is increasing in the average log price set by other firms, we say there is strategic complementarity in pricing decisions. The parameter α with $0 \leq \alpha < 1$ captures this strategic complementarity.

In summary, the log price set by firms at t satisfies

$$\nu_t(\theta) = \begin{cases} (1-\alpha)\delta + \alpha \dfrac{\sum_{s=0}^{\infty} \beta^s \Phi_s(\theta) P_{t+s}}{\sum_{s=0}^{\infty} \beta^s \Phi_s(\theta)} & \text{if } t > 0 \\ 0 & \text{if } t \leq 0. \end{cases} \tag{17}$$

We are interested in solving equations (15)–(17) to characterize the evolution of the average log price P_t under the restriction that P_t is bounded. In Appendix C.1, we prove that there is a unique such price sequence and find a contraction with modulus α which characterizes the entire sequence P_t.

When $\alpha = 0$, we also obtain a useful aggregation result (see also Carvalho and Schwartzman, 2015, Proposition 2′). We prove in Appendix C.2 that the dynamics of P_t are identical in a heterogeneous firm economy and in an economy where there is only a single type of firm with a survival function

$$\bar{\Phi}_t \equiv \int_{\theta_L}^{\theta_H} \Phi_t(\theta) dG(\theta|\tilde{\omega}_0). \tag{18}$$

This is a weighted average of the heterogeneous firms' survival functions, with weights given by the frequency of price changes. Put differently, it is the survival function associated with the Kaplan-Meier hazard $H_t(\omega^*)$, as defined in equations (10) and (11). Thus if we can measure the Kaplan-Meier hazard, we do not need to understand the extent of heterogeneity in the economy without strategic complementarity. This is a useful benchmark.

For other values of the complementarity parameter α, we do not have a closed-form solution and the exact aggregation result fails. We therefore solve the model numerically to investigate the difference between the path of the average log price in a heterogeneous economy and in other economies with the same Kaplan-Meier hazard rate.

More precisely, we calibrate the model using the estimated baseline hazard presented in Section 6.1. See Appendix C.4 for details. We set the weekly discount factor to $\beta = 0.999$ (five percent annual discounting) and consider a monetary shock which increases the steady (log) price from 0 to $\delta = 1$. We examine three different values of the complementarity parameter, $\alpha = 0$, $\alpha = 0.5$ and $\alpha = 0.95$, with the aggregation result not holding in the latter two economies. We calibrate the distribution of survival functions using our estimated moments of the frailty distribution. The estimated model does a good job of matching the Kaplan-Meier hazard (Figure 6 in the Appendix).

We then consider three different economies with the same Kaplan-Meier hazard. The first is the estimated MPH model. The second is a homogeneous firm economy, where each firm's hazard is the Kaplan-Meier hazard. The third is a heterogeneous firm economy where every firm adjusts its price at fixed intervals (Taylor, 1979, 1980), but different firms have different intervals so as to match the Kaplan-Meier hazard.[14]

Figure 3 displays these results. The left panels show the evolution of the average log price P_t, while the right panel shows the fraction of the gap between the current average log price and the asymptotic log price δ that is closed in period t, $l_t \equiv \ln(\delta - P_{t-1}) - \ln(\delta - P_t)$. We observe that average log price dynamics in the MPH economy is similar to that in the homogeneous firm economy.[15] The average log price is somewhat less similar in the Taylor economy, especially when strategic complementarities are strong ($\alpha = 0.95$). For example, the initial speed of price convergence in the MPH and homogeneous economies is significantly faster than in the Taylor economy, while after half a year it is noticeably slower.

At least for this question and this data set, our analysis shows that a researcher can usefully approximate the true heterogeneous firm economy with a representative firm that has the empirical Kaplan-Meier hazard. This approximation is useful for two reasons. First, it may be easier to model a homogeneous firm economy. And second, as we have shown in Proposition 4, it is possible to estimate the Kaplan-Meier hazard without first estimating a mixture model.

This approximate aggregation result, that only the Kaplan-Meier hazard affects aggregate price dynamics, extends Carvalho and Schwartzman (2015), who analyze the role of heterogeneity for the cumulative impulse response of the aggregate price level to a monetary

[14]There is a sense in which the latter two economies represent polar cases of heterogeneity. In any mixture model, the correlation between the duration of any two spells is non-negative. In a homogeneous firm economy, this correlation is zero, while in the Taylor economy it is 1.

[15]We use the model-implied Kaplan-Meier hazard for this exercise. An alternative would be to use the estimated Kaplan-Meier hazard without first imposing the MPH structure. Since the calibrated model hits the empirical the Kaplan-Meier hazard almost exactly, the implied paths for the log price are indistinguishable.

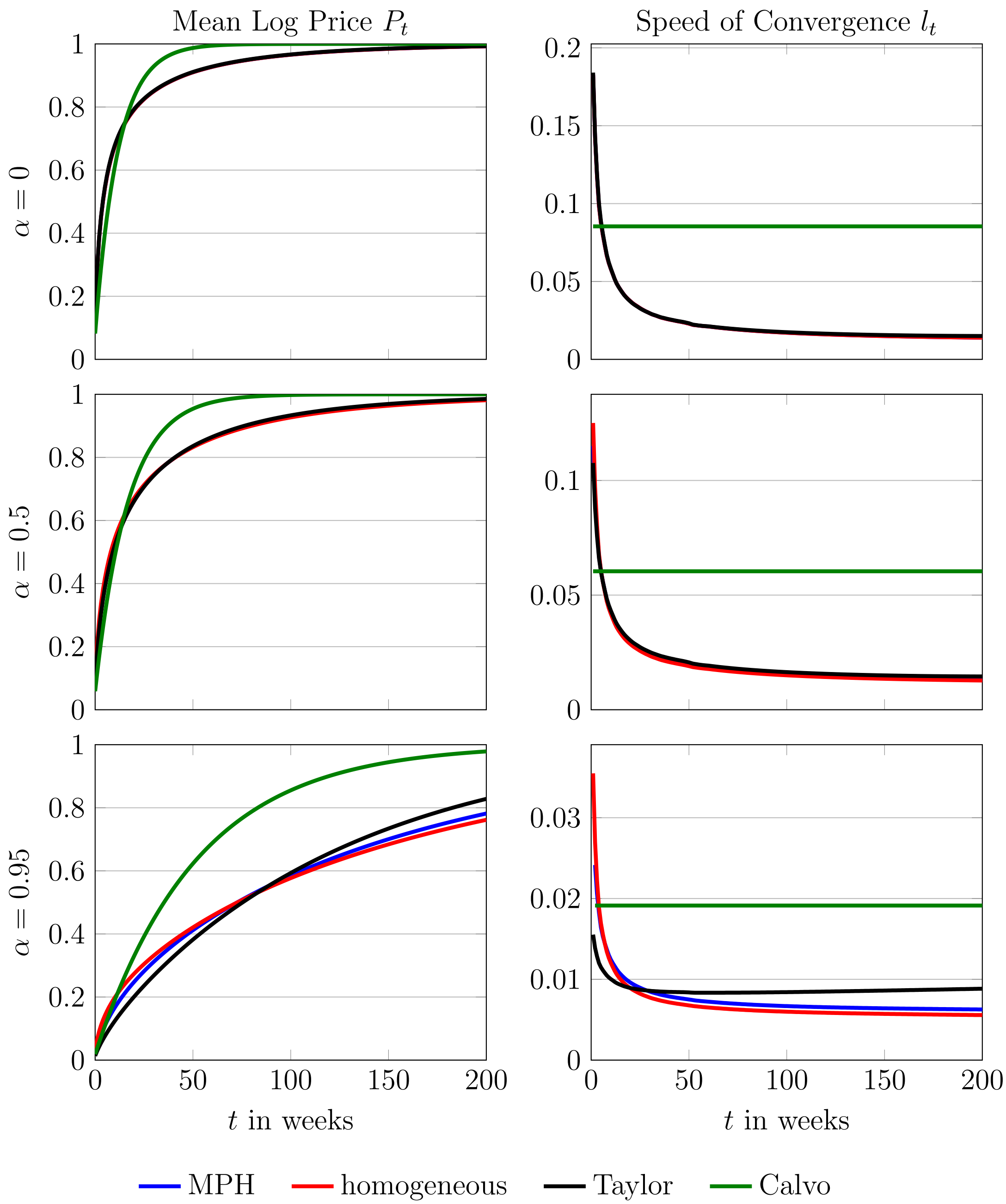

Figure 3: Mean log price P_t and the fraction of the gap closed in period t, l_t, in a heterogeneous and single-firm economy. The top panels have $\alpha = 0$ (no strategic complementarity in pricing), middle have $\alpha = 0.5$ (weak strategic complementarity), and the bottom panels have $\alpha = 0.95$ (strong strategic complementarity). Throughout we assume $\beta = 0.999$ per week. The blue lines show the estimated MPH model. The red dashed line shows the economy with a single firm with survival function equal to the Kaplan-Meier survival function. The purple line shows the economy with heterogeneous firms, each with a fixed price duration (Taylor, 1979, 1980), and the same Kaplan-Meier survival function. The green line shows a single firm with a constant probability of changing its price (Calvo, 1983), and hence a different Kaplan-Meier survival function.

shock in the case without strategic complementarity, $\alpha = 0$. Their analysis implies that the impulse response of aggregate prices to a one-time monetary shock in an economy with heterogeneous sectors is the same as the impulse response of a one sector economy with survival function given by equation (18). We use our estimated model to turn this into an approximation result in an environment with a strategic complementarity. To be clear, for arbitrary distributions of hazard rates, this approximation result would not hold.

In addition to this exercise, we examine whether the path for P_t can be well approximated by a single firm with a Calvo price setting rule, a common assumption in the literature. We show in Appendix C.3 that in this case, the average price equals $\delta(1-x^t)$ where x depends on the firm's Calvo parameter as well as the value of the complementarity parameter α and the discount factor β. This implies a constant speed of convergence l_t in the Calvo model. The time path for the average price is quite far from this exponential structure in our estimated model, and hence cannot be well-approximated by a single Calvo firm, depicted by the green lines in Figure 3.

7 Disentangling Regular Price Changes from Sales

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, for example if 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 measuring duration dependence for only those risks. 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.

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 $\chi_j^i = +$ if the j^{th} spell of product i follows a price increase and $\chi_j^i = -$ if it follows a price decrease. We also distinguish whether a spell ends with a price increase or decrease, $R = 2$, and let $\rho_j^i = +$ if the j^{th} spell of product i ends with a price increase and $\rho_j^i = -$ if it ends with a price decrease. Spells with $\chi_j^i = \rho_j^i$ 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, $\phi^{++}(\boldsymbol{\theta})$, $\phi^{+-}(\boldsymbol{\theta})$, $\phi^{-+}(\boldsymbol{\theta})$, and $\phi^{--}(\boldsymbol{\theta})$.

This richer model allows for the possibility that price trends have different dynamics than price reversals. We estimate it using the moment conditions specified in Proposition 5. Figure 4 shows the results and some 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.

We conclude from these findings is that the shape of the baseline hazard we recovered in Figure 1 is primarily driven by price reversals, especially those associated with sales (price increases following price decreases), where the hazard is steeply declining. Price reversal are common in the data: 72.3 percent of spells starting with a price increase end with a price decrease, while 72.4 percent of spells starting with a price decrease end with a price increase.

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 five percent 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" for price trends than price reversals, and is especially mild compared to the estimates of the MPH model, where we had $J = 10{,}498$.

Figures 16 and 17 in Appendix F show estimated b^{++} and b^{--} for individual product categories. The results are in line with those for the pooled sample. The hazard b^{++} declines for about 6 weeks and then is flat, and the baseline hazard b^{--} is declining in most categories. The value of the J-test for individual categories is lower than the value on the pooled sample.

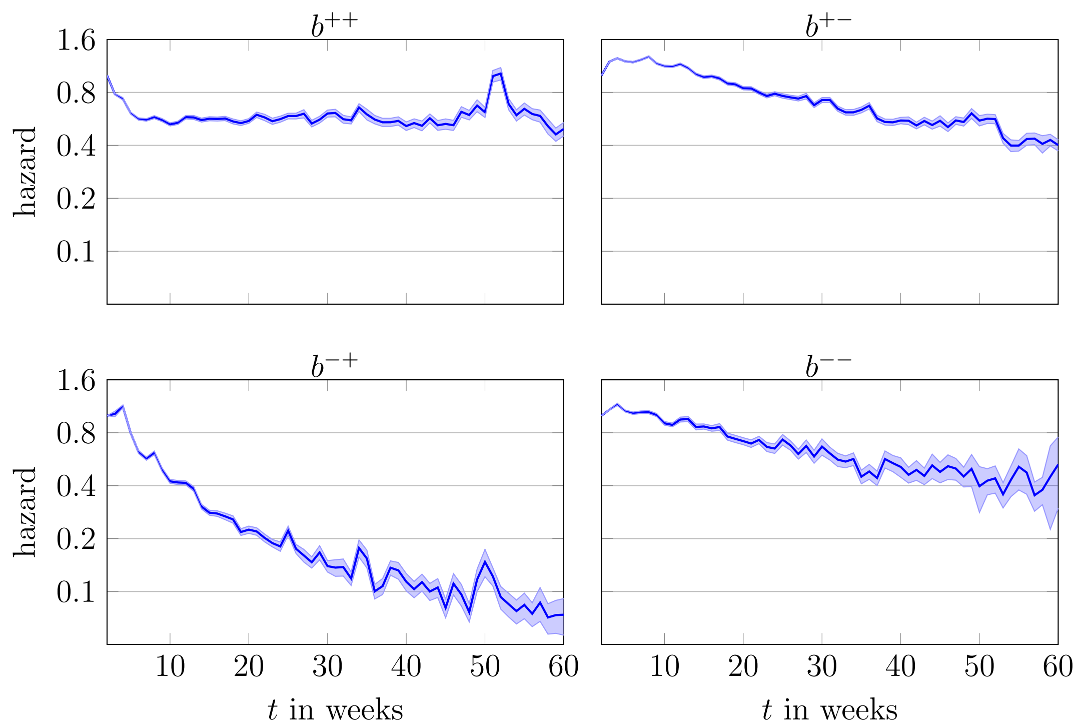

Figure 4: Baseline hazards in the competing risks model for pooled IRI data, log scale. b_t^{++} is the baseline hazard for spell which begin and end with a price increase; b_t^{--} for spells which begin and end with a price decrease; b_t^{+-} for spells which begin with a price increase and end with a price decrease; and b_t^{-+} for spells which begin with a price decrease and end with a price increase. The shaded regions show two standard error bands. Standard errors are clustered at the store $\times$ product category level. We normalize the baseline hazard to be 1 at duration $= \underline{T} = 2$.

In particular, we cannot reject the specification for 8 categories for b^{++} and 21 categories for b^{--}.

We investigate the nature of the failure of proportional hazard assumption more systematically in Appendix D.3. We conclude that the dynamics of price trends is well described by the MPH assumption, and that the baseline hazard is fairly flat. On the other hand, we conclude that MPH assumption is not a good description of the dynamics of price reversals. One possible reason is that temporary changes might have fixed duration which does not fit into the MPH framework.

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.

8 Comparison to Other Estimation Methods

The usual approach to estimating the MPH model is via maximum likelihood for the continuous time model. Formulating the likelihood requires an assumption on the frailty distribution. It is convenient to assume a gamma distribution, since this makes it possible to integrate out the frailty distribution analytically and obtain a simpler expression for the likelihood function. Since the gamma distribution is described by its mean and variance, and the mean can be normalized to 1, one then finds the baseline hazard and variance to maximize the likelihood.

One issue which arises is that time in the model is continuous but the data are typically measured discretely. One 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 nonnegative integer t, and to assume that observing an integer duration t means that the exact (continuous time) duration was t. Maximum likelihood estimation can be done using a built-in procedure in Stata with some tricks, as we explain in Appendix E.3.

Figure 5 compares estimates from Stata with ours. In both cases, the average type is normalized to 1, so the baseline hazards are equal to the Kaplan-Meier hazard at duration 2 weeks. The baseline hazard estimated using maximum likelihood is somewhat steeper than the one estimated using GMM. We think there are two potential biases in the maximum likelihood estimates. First, maximum likelihood requires choosing a family for the frailty distribution. Heckman and Singer (1984) pointed out that imposing a specific distribution can bias the estimates of the baseline hazard. Another possibility is time aggregation. The likelihood is formulated in continuous time while data are discrete, and failure to properly account for time aggregation can bias the estimates. We investigate these reasons in detail in Appendix E and conclude that in our data set, time aggregation plays a more important role.

In closing, we note that there are several advantages to using the GMM estimator we developed over maximum likelihood. First, our 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 proved in Proposition 3 that we find a global maximum. In contrast, the log-likelihood is non-linear $\boldsymbol{b}$ and finding its maximum can be slow.[16] Importantly, there is no guarantee that maximum likelihood finds a global maximum. Finally, we showed that our method is easily extended to a competing risks framework with spell-specific observable characteristics. We can handle these even if the proportional hazard assumption only holds for some risks and some observables. This set of assumptions has proved to be extremely hard to handle in the maximum likelihood framework. For example, Fougere, Le Bihan, and Sevestre (2007) try to estimate a competing risks model with unobserved heterogeneity but say on page 260 that "...convergence of the maximum likelihood procedure is very difficult to reach."

9 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 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 of real-world data. We further propose and implement two tests of the MPH specification.

We treat the frailty distribution as a nuisance parameter in most of our analysis. However, in the Appendix, we present estimators of the first three moments of the frequency-weighted

[16] Using our pooled IRI sample, it took 15 hours to estimate the baseline hazard using the ML method in Stata on a computer with 256GB memory. It took 70 minutes to estimate it (including standard errors) using GMM. A computer with 60GB memory failed to deliver ML estimates but produced GMM estimates.

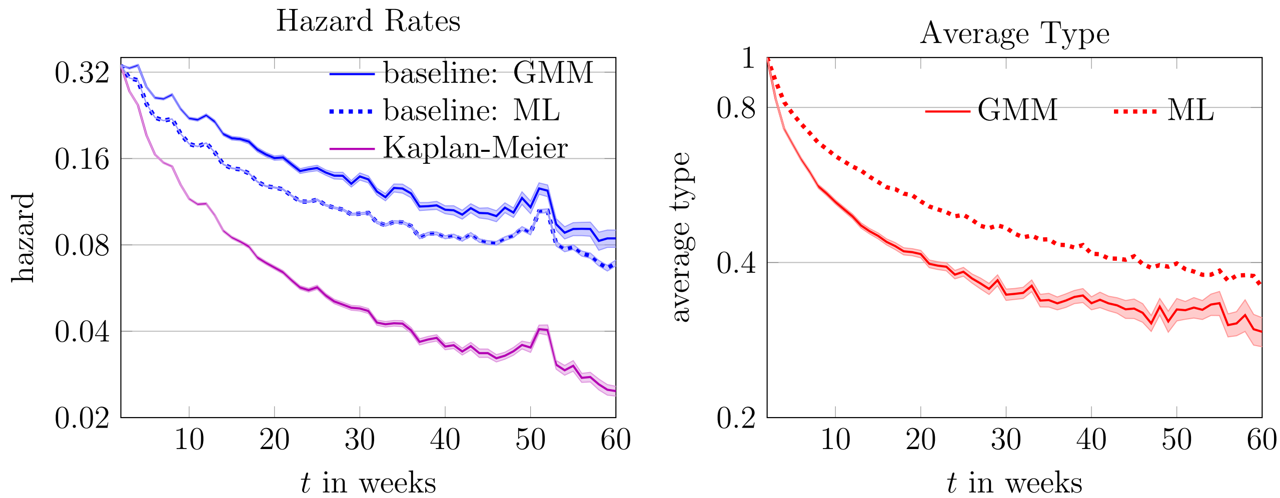

Figure 5: The left hand figure shows the Kaplan-Meier and baseline hazard for pooled IRI data, log scale. The purple line shows the Kaplan-Meier hazard, the blue solid line is our GMM estimate of the baseline hazard, and the blue dotted line is the maximum likelihood estimate of the baseline model under the assumption that the data generating process is in continuous time and the frailty distribution is gamma. The right hand figure shows the average type, with our GMM estimate the solid line and the maximum likelihood estimate the dotted line.

frailty distribution following the identification proof and insights from the construction of the estimator of Kaplan-Meier hazard.

We also estimate a competing risk model for the direction of 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.

Alvarez, Fernando, Francesco Lippi, and Luigi Paciello, 2011. "Optimal Price Setting with Observation and Menu Costs." *Quarterly Journal of Economics*. 126 (4): 1909–1960.

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.

Caballero, Ricardo J., 1989. "Time Dependent Rules, Aggregate Stickiness and Information Externalities." Discussion Papers 198911, Columbia University.

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.

Carvalho, Carlos, 2006. "Heterogeneity in Price Stickiness and the Real Effects of Monetary Policy." *Frontiers of Macroeconomics*. 2 (1).

Carvalho, Carlos, and Felipe Schwartzman, 2015. "Selection and Monetary Non-Neutrality in Time-Dependent Pricing Models." *Journal of Monetary Economics*. 76: 141–156.

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

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." *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." *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 Burton 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. *The Econometric Analysis of Transition Data* . No. 17 in 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.

Nakamura, Emi, and Jón Steinsson, 2010. "Monetary Non-neutrality in a Multisector Menu Cost Model." *Quarterly Journal of Economics*. 125 (3): 961–1013.

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.

Reis, Ricardo, 2006. “Inattentive Producers.” *Review of Economic Studies*. 73 (3): 793–821.

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

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

Appendix

A Omitted Proofs

A.1 Kaplan-Meier Hazard Moments

Proof of Proposition 4. Take a product i with measured durations $\boldsymbol{\zeta}^i = (\zeta_0^i, \zeta_1^i, \ldots, \zeta_{K^i}^i)$ and hence censoring time $c^i = \sum_{j=0}^{K^i} \zeta_j^i - 1$. Let $\boldsymbol{z}^i = \{z_1^i, \ldots, z_{c^i}^i\}$ be a vector of length c^i with the following elements:

$$z_s^i = \begin{cases} \zeta_k^i & \text{for } s = \sum_{j=0}^{k-1} \zeta_j^i \text{ and } k = 1, \ldots, K^i \\ 0 & \text{otherwise.} \end{cases}$$

z_s^i encodes the measured duration of any spell that starts s periods into the observation window for the product, with zeros in any period when a new spell does not start.

We first claim is that for any product i and any duration $t = 1, \ldots, \bar{T} + 1$,

$$\sum_{j=1}^{K^i} \mathbb{1}_{\zeta_j^i \geq t, c_j^i \geq \bar{T}} = \sum_{s=1}^{c^i - \bar{T}} \mathbb{1}_{z_s^i \geq t},$$

where we understand that the left hand side evaluates to 0 when $c^i \leq \bar{T}$. The left-hand sum counts the number of spells (except the initial left-censored one) with duration at least t and residual censoring time at least $\bar{T}$. The right-hand sum counts the same spells by dropping all those that start after $c^i - \bar{T}$, when the residual censoring time would be less than $\bar{T}$.

Next, we compute the expected value of $\sum_{s=1}^{c^i - \bar{T}} \mathbb{1}_{z_s^i \geq t}$ for any $t = 1, \ldots, \bar{T} + 1$ conditional on c^i and θ^i. Here we use the assumption that initial duration is drawn from the stationary ergodic distribution. This implies that with probability $\tilde{\omega}_0(\theta^i) = 1 / \sum_{t'=0}^{\infty} \prod_{s=0}^{t'} (1 - h_s(\theta^i))$ (see equation 8), the firm changes its price in any period $s \geq 1$, in which case $z_s^i > 0$, while otherwise $z_s^i = \mathbb{1}_{z_s^i \geq t} = 0$. If the firm does change its price, the probability that the measured duration of the price spell is at least t is given by the type-specific survivor function $\prod_{s=0}^{t-1} (1 - h_s(\theta^i))$. This use the fact that right censoring is not an issue for $t \leq \bar{T} + 1$ and $s \leq c^i - \bar{T}$.

Putting this together, in any period $s \in \{1, \ldots, c^i - \bar{T}\}$, the expected value of $\mathbb{1}_{z_s^i \geq t}$ conditional on c^i and θ^i is

$$\frac{\prod_{s=0}^{t-1} (1 - h_s(\theta^i))}{\sum_{t'=0}^{\infty} \prod_{s=0}^{t'} (1 - h_s(\theta^i))}.$$

It follows that

$$\mathbb{E}\left[\sum_{j=1}^{K^i} \mathbb{1}_{\zeta_j^i \geq t, c_j^i \geq \bar{T}} \middle| c^i, \theta^i\right] = \mathbb{E}\left[\sum_{s=1}^{c^i-\bar{T}} \mathbb{1}_{z_s^i \geq t} \middle| c^i, \theta^i\right] = \begin{cases} \frac{(c^i-\bar{T})\prod_{s=0}^{t-1}(1-h_s(\theta^i))}{\sum_{t'=0}^{\infty}\prod_{s=0}^{t'}(1-h_s(\theta^i))} & \text{if } c^i > \bar{T} \\ 0 & \text{if } c^i \leq \bar{T}. \end{cases}$$

Now condition only on θ^i. Using the conditional distribution of c given θ we get

$$\begin{aligned} \mathbb{E}\left[\frac{c}{c-\bar{T}}\sum_{j=1}^{K} \mathbb{1}_{\zeta_j \geq t, c_j \geq \bar{T}} \middle| \theta^i\right] &= \frac{\left(\sum_{c=\bar{T}+1}^{\infty} c q_c(\theta^i)\right)\left(\prod_{s=0}^{t-1}(1-h_s(\theta^i))\right)}{\sum_{t'=0}^{\infty}\prod_{s=0}^{t'}(1-h_s(\theta^i))} \\ &= \omega^f(\theta^i)\prod_{s=0}^{t-1}(1-h_s(\theta^i)), \end{aligned}$$

where the feasible weight ω^f is defined in equation (12). Finally, integrating across θ using the frailty distribution G, we get

$$\mathbb{E}\left[\frac{c}{c-\bar{T}}\sum_{j=1}^{K} \mathbb{1}_{\zeta_j \geq t, c_j \geq \bar{T}}\right] = \int_{\theta_L}^{\theta_H} \omega^f(\theta)\prod_{s=0}^{t-1}(1-h_s(\theta))dG(\theta) \tag{19}$$

for all $t = 1, \ldots, \bar{T}+1$.

For any $t = 1, \ldots, \bar{T}$, this implies

$$\mathbb{E}\left[\frac{c}{c-\bar{T}}\sum_{j=1}^{K} \mathbb{1}_{\zeta_j \geq t+1, c_j \geq \bar{T}}\right] = \int_{\theta_L}^{\theta_H} \omega^f(\theta)\prod_{s=0}^{t}(1-h_s(\theta))dG(\theta).$$

We can then take first differences for any such t to get

$$\mathbb{E}\left[\frac{c}{c-\bar{T}}\sum_{j=1}^{K} \mathbb{1}_{\zeta_j = t, c_j \geq \bar{T}}\right] = \int_{\theta_L}^{\theta_H} \omega^f(\theta)h_t(\theta)\prod_{s=0}^{t-1}(1-h_s(\theta))dG(\theta). \tag{20}$$

Then using equations (13), (19) and (20), it follows immediately that $\mathbb{E}\left[f_{t,\bar{T}}^{[H]}(\boldsymbol{\zeta}; \boldsymbol{H})\right] = 0$ for $t = 1, \ldots, \bar{T}$ if and only if

$$H_t = \frac{\int_{\theta_L}^{\theta_H} \omega^f(\theta)h_t(\theta)\prod_{s=0}^{t-1}(1-h_s(\theta))dG(\theta)}{\int_{\theta_L}^{\theta_H} \omega^f(\theta)\prod_{s=0}^{t-1}(1-h_s(\theta))dG(\theta)} = \frac{\int_{\theta_L}^{\theta_H} h_t(\theta)\prod_{s=0}^{t-1}(1-h_s(\theta))dG(\theta|\omega^f)}{\int_{\theta_L}^{\theta_H} \prod_{s=0}^{t-1}(1-h_s(\theta))dG(\theta|\omega^f)},$$

where the last equation uses equation (1). This is equal to $H_t(\omega^f)$, proving the result. ■

A.2 Baseline Hazard Moments

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*

$$f_{j,k,\boldsymbol{t},\boldsymbol{x},\boldsymbol{r}}(\boldsymbol{\zeta}, \boldsymbol{\chi}, \boldsymbol{\rho}; \boldsymbol{b}) \equiv \\ b_{t_k} \mathbb{1}_{K \geq k, \boldsymbol{\chi}_k = \boldsymbol{x}, \boldsymbol{\rho}_{k-1} = \boldsymbol{r}, \boldsymbol{\zeta}_{k-1} = \boldsymbol{t}_{k-1}, \zeta_k \geq t_k} - b_{t_j} \mathbb{1}_{K \geq k, \boldsymbol{\chi}_k = \boldsymbol{x}, \boldsymbol{\rho}_{k-1} = \boldsymbol{r}, \boldsymbol{\zeta}_{k-1/j} = \boldsymbol{t}_{k-1/j}, \zeta_j = t_k, \zeta_k \geq t_j}. \tag{21}$$

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 (21) 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 (21) 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

$$\mathbb{E}\left[\mathbb{1}_{K\geq k, \boldsymbol{\chi}_k=\boldsymbol{x}, \boldsymbol{\rho}_{k-1}=\boldsymbol{r}, \boldsymbol{\zeta}_{k-1}=\boldsymbol{t}_{k-1}, \zeta_k\geq t_k}\right] = \psi(\boldsymbol{t}_{k-1/j}, t_j, t_k, \boldsymbol{x}, \boldsymbol{r}; j, k) b_{0,t_j}, \tag{22}$$

where

$$\begin{aligned}
&\psi(\boldsymbol{t}_{k-1/j}, t_j, k_j, \boldsymbol{x}, \boldsymbol{r}; j, k) \equiv \left(1 - P\left(\sum_{l=1}^{k} t_l\right)\right) \times \\
&\int_{\theta_L}^{\theta_H} \phi(\boldsymbol{\theta})\pi_1(x_1|\boldsymbol{\theta}) \prod_{l=1}^{k} \left(\pi(x_l|x_{l-1}, r_{l-1}, \boldsymbol{\theta})^{\mathbb{1}_{l\neq 1}} h_{t_l}^{r_l}(x_l, \boldsymbol{\theta})^{\mathbb{1}_{l\neq j, l\neq k}} \prod_{s=1}^{t_l-1} (1 - h_s(x_l, \boldsymbol{\theta})) \right) dG_{\sum_{l=1}^{k} t_l - 1}(\boldsymbol{\theta}).
\end{aligned} \tag{23}$$

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

$$\mathbb{E}\left[\mathbb{1}_{K\geq k, \boldsymbol{\chi}_k=\boldsymbol{x}, \boldsymbol{\rho}_{k-1}=\boldsymbol{r}, \boldsymbol{\zeta}_{k-1/j}=\boldsymbol{t}_{k-1/j}, \zeta_j=t_k, \zeta_k\geq t_j}\right] = \psi(\boldsymbol{t}_{k-1/j}, t_k, t_j, \boldsymbol{x}, \boldsymbol{r}; j, k) b_{0,t_k}. \tag{24}$$

Moreover, equation (23) and the commutative property of multiplication implies

$$\psi(\boldsymbol{t}_{k-1/j}, t_k, t_j, \boldsymbol{x}, \boldsymbol{r}; j, k) = \psi(\boldsymbol{t}_{k-1/j}, t_j, t_k, \boldsymbol{x}, \boldsymbol{r}; j, k). \tag{25}$$

The result then follows from equations (22), (24), and (25). ■

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

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

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

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 (21). 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 $\underline{T}\leq t_1<t_2\leq\bar{T}$, 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 $1\leq j<k$. By equation (14), this gives us $f^{[b,x,r]}_{t_1,t_2}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})$. Since the expected value of each component of this sum is zero, this implies $\mathbb{E}\left[f^{[b,x,r]}_{t_1,t_2}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b}_0)\right]=0$.

Finally, note that the function $f^{[b,x,r]}_{t_1,t_2}$ defined in equation (14) is linear in the baseline hazard, $f^{[b,x,r]}_{t_1,t_2}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\lambda\boldsymbol{b})=\lambda f^{[b,x,r]}_{t_1,t_2}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\boldsymbol{b})$ for all t_1, t_2, $\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho}$, $\boldsymbol{b}$, and λ. Thus $\mathbb{E}\left[f^{[b,x,r]}_{t_1,t_2}(\boldsymbol{\zeta},\boldsymbol{\chi},\boldsymbol{\rho};\lambda\boldsymbol{b})\right]=0$ as well.

Sufficiency: Recall that T_0 is the smallest $t\in\{\underline{T},\ldots,\bar{T}\}$ 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 (14) implies that

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

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 $T_0 \in \{\underline{T}, \bar{T}\}$ 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 $t \in \{T_0, \bar{T}\}$.[17] 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, $\boldsymbol{b}_{./T_0} = (b_{\underline{T}}, \ldots b_{T_0-1}, b_{T_0+1}, \ldots b_{\bar{T}})$.

[17]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

$$\mathbb{E}\left[U^{[b]}(\boldsymbol{\zeta})\right]\boldsymbol{b}_{\cdot/T_0} - \mathbb{E}\left[V^{[b]}(\boldsymbol{\zeta})\right] = 0. \tag{26}$$

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 (13) 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

$$\mathbb{E}\left[U^{[H]}(\boldsymbol{\zeta})\right]\boldsymbol{H}^{\bar{T}} - \mathbb{E}\left[V^{[H]}(\boldsymbol{\zeta})\right] = 0. \tag{27}$$

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

$$\boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{b}_{\cdot/T_0}\\ \boldsymbol{H}^{\bar{T}}\end{pmatrix}, \boldsymbol{f}(\boldsymbol{\zeta};\boldsymbol{\beta}) = \begin{pmatrix}\boldsymbol{f}^{[b]}(\boldsymbol{\zeta};\boldsymbol{b})\\ \boldsymbol{f}^{[H]}_{\bar{T}}(\boldsymbol{\zeta};\boldsymbol{H}^{\bar{T}})\end{pmatrix}, U = \begin{pmatrix}U^{[b]} & 0\\ 0 & U^{[H]}\end{pmatrix}, V = \begin{pmatrix}V^{[b]}\\ V^{[H]}\end{pmatrix}.$$

Then the moment conditions are

$$\mathbb{E}\left[U(\boldsymbol{\zeta})\right]\boldsymbol{\beta} - \mathbb{E}\left[V(\boldsymbol{\zeta})\right] = 0. \tag{28}$$

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 (28) 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).[18]

B.2 Clustered Standard Errors

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

$$VAR \equiv \frac{1}{I}(F'WF)^{-1}F'W\Omega W'F(F'W'F)^{-1}, \tag{29}$$

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 (29) still applies but with cluster-robust sample analog of Ω. Let $\mathcal{Q}$ denote the number of clusters indexed by $q = 1, \ldots, \mathcal{Q}$. 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

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

where $2T+1$ is the number of parameters. The term $\frac{\mathcal{Q}}{\mathcal{Q}-1}\frac{I-1}{I-(2T+1)}$ 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 (29).

[18] 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 percent, and that they are small in absolute value, of the order of 10^{-13}. 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

C.1 General Model

It is useful to define

$$\Omega_t(\theta) = \frac{\beta^t \Phi_t(\theta)}{\sum_{s=0}^{\infty} \beta^s \Phi_s(\theta)} \quad \text{and} \quad \Psi_t(\theta) = \frac{\Phi_t(\theta)}{\sum_{s=0}^{\infty} \Phi_s(\theta)},$$

so that we can write equations (15) and (17) as

$$p_t(\theta) = \sum_{s=0}^{t} \Psi_s(\theta)\nu_{t-s}(\theta),$$

$$\nu_t(\theta) = (1-\alpha)\delta + \alpha \sum_{s=0}^{\infty} \Omega_s(\theta) P_{t+s}$$

for $t \geq 0$. We have simplified the first equation using the assumption that $\nu_t(\theta) = 0$ for $t \leq 0$.

Substitute the second equation into the first to get

$$p_t(\theta) = (1-\alpha)\delta p_t^0(\theta) + \alpha \sum_{s=0}^{\infty} k_{t,s}(\theta) P_s, \tag{30}$$

where

$$p_t^0(\theta) = \sum_{s=0}^{t} \Psi_s(\theta) \quad \text{and} \quad k_{t,s}(\theta) = \sum_{x=\max(0,t-s)}^{t} \Psi_x(\theta)\Omega_{x+s-t}(\theta).$$

Here $p_t^0(\theta)$ is the average price among θ-type firms which would prevail in an economy with no strategic complementarity, and $k_{t,s}(\theta)$ is a θ-type kernel given by a convolution of $\Omega_t(\theta)$ and $\Psi_t(\theta)$. One can verify that

$$\sum_{s=0}^{\infty} k_{t,s}(\theta) = \sum_{x=0}^{t} \Psi_x(\theta) \leq 1.$$

Average both sides of equation (30) across the type distribution $G(\theta)$ to get

$$P_t = (1-\alpha)\delta P_t^0 + \alpha \sum_{s=0}^{\infty} K_{t,s} P_s, \tag{31}$$

where

$$P_t^0 \equiv \int_{\theta_L}^{\theta_H} p_t^0(\theta) dG(\theta) \quad \text{and} \quad K_{t,s} \equiv \int_{\theta_L}^{\theta_H} k_{t,s}(\theta) dG(\theta).$$

Since $\sum_{s=0}^{\infty} K_{t,s} \leq 1$ for all t, the mapping (31) is a contraction with modulus $\alpha < 1$, and so has a unique solution in the space of bounded functions.

C.2 Special Case: No Strategic Complementarity

We say there is no strategic complementarity when $\alpha = 0$. In this case, the optimal price for a firm that adjusts its price is simply δ, and so the average price in the economy is

$$P_t = \delta \int_{\theta_L}^{\theta_H} \frac{\sum_{s=0}^{t} \Phi_s(\theta)}{\sum_{s=0}^{\infty} \Phi_s(\theta)} dG(\theta).$$

Changing the order of summation and integration, we get that

$$\begin{aligned}
P_t &= \delta \sum_{s=0}^{t} \int_{\theta_L}^{\theta_H} \frac{\Phi_s(\theta)}{\sum_{s'=0}^{\infty} \Phi_{s'}(\theta)} dG(\theta) = \delta \sum_{s=0}^{t} \int_{\theta_L}^{\theta_H} \Phi_s(\theta) \tilde{\omega}_0(\theta) dG(\theta) \\
&= \left(\delta \sum_{s=0}^{t} \bar{\Phi}_s \right) \left(\int_{\theta_L}^{\theta_H} \tilde{\omega}_0(\theta) dG(\theta) \right) \\
&= \delta \frac{\sum_{s=0}^{t} \bar{\Phi}_s}{\sum_{s'=0}^{\infty} \bar{\Phi}_{s'}}
\end{aligned}$$

where $\bar{\Phi}_s$ is the frequency-weighted Kaplan-Meier survival function defined in equation (18). In the last equality, we used that

$$\sum_{s=0}^{\infty} \bar{\Phi}_s = \left(\int_{\theta_L}^{\theta_H} \tilde{\omega}_0(\theta) dG(\theta) \right)^{-1}.$$

Thus, the price level P_t is the same as in an economy with a single firm with the frequency-weighted Kaplan-Meier survival function.

C.3 Special Case: Calvo

It is useful to analyze this problem for a single Calvo firm with parameter θ. In this case, $\Omega_t = (1 - \beta(1-\theta))\beta^t(1-\theta)^t$ and $\Psi_t = \theta(1-\theta)^t$, and so suppressing dependence on the parameter θ, equations (15) and (17) then are

$$\begin{aligned}
p_t &= \theta \sum_{s=0}^{t} (1-\theta)^s \nu_{t-s}, \\
\nu_t &= (1-\alpha)\delta + \alpha(1 - \beta(1-\theta)) \sum_{s=0}^{\infty} \beta^s (1-\theta)^s p_{t+s}.
\end{aligned}$$

These can in turn be reduced to a pair of linear first-order difference equations,

$$\begin{aligned} p_{t+1} &= \theta\nu_{t+1} + (1-\theta)p_t, \\ \nu_t &= (1-\alpha)\delta(1-\beta(1-\theta)) + \alpha(1-\beta(1-\theta))p_t + \beta(1-\theta)\nu_{t+1}. \end{aligned}$$

The solution is of the form $p_t = \delta(1-c_1x_1^t - c_2x_2^t)$ where $x_1 < x_2$ are roots of the quadratic equation

$$x^2\beta(1-\theta) - x\big(1-\alpha\theta+\beta(1-\theta)(1-(1-\alpha)\theta)\big) + 1 - \theta = 0$$

and c_1 and c_2 are constants to be determined. It holds that $0 < x_1 < 1 < x_2$, and so we set $c_2 = 0$ to have a non-explosive path for the average price. The initial condition $p_0 = 0$ then pins down $c_1 = 1$.

C.4 Calibration of the Model

For the numerical exercise, we use estimates from our baseline model. We assume that the baseline hazard is given by our estimates presented in Section 6.1. We also want to estimate moments of the frailty distribution. Since $\underline{T} > 1$, it is not feasible to estimate moments distribution of $dG(\theta|\tilde{\omega}^f)$ because we do not know what $h_s(\theta)$ is for $s < \bar{T}$. However, it is feasible to estimate moments of the distribution $G(\cdot|\tilde{\omega}^f_{\underline{T}})$, where

$$\tilde{\omega}^f_{\underline{T}}(\theta) \equiv \left(\sum_{c=1}^{\infty} cq_c(\theta)\right)\tilde{\omega}_{\bar{T}-1}(\theta),$$

is the weight given by the expected censoring time for a type θ product, multiplied by the probability that a spell of this product reaches $\bar{T}$. It is useful to note that manipulating this expression using equations (8) and (12) leads to

$$\tilde{\omega}^f_{\underline{T}}(\theta) = \omega^f(\theta)\prod_{s=0}^{\underline{T}-1}(1-h_s(\theta)).$$

Denote $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ the first three moments of the distribution $G(\cdot|\tilde{\omega}^f_{\underline{T}})$. Assuming that $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, \ldots, \zeta_K)$ is drawn from a stationary mixture model, we can translate equation

(5) in the identification proof into moment conditions for μ_1, μ_2, μ_3:

$$
\begin{aligned}
f_{\bar{T}}^{[\mu_1]}(\boldsymbol{\zeta}; \boldsymbol{\mu}, \boldsymbol{b}) &\equiv \frac{c}{c-\bar{T}} \sum_{j=1}^{K} \left(\mathbb{1}_{\zeta_j=\underline{T}, c_j \geq \bar{T}} - b_{\underline{T}} \mu_1 \mathbb{1}_{\zeta_j \geq \underline{T}, c_j \geq \bar{T}} \right) \\
f_{\bar{T}}^{[\mu_2]}(\boldsymbol{\zeta}; \boldsymbol{\mu}, \boldsymbol{b}) &\equiv \frac{c}{c-\bar{T}} \sum_{j=1}^{K} \left(\mathbb{1}_{\zeta_j=\underline{T}+1, c_j \geq \bar{T}} - b_{\underline{T}+1} \left(\mu_1 - b_{\underline{T}} \mu_2 \right) \mathbb{1}_{\zeta_j \geq \underline{T}, c_j \geq \bar{T}} \right) \\
f_{\bar{T}}^{[\mu_3]}(\boldsymbol{\zeta}; \boldsymbol{\mu}, \boldsymbol{b}) &\equiv \frac{c}{c-\bar{T}} \sum_{j=1}^{K} \left(\mathbb{1}_{\zeta_j=\underline{T}+2, c_j \geq \bar{T}} - b_{\underline{T}+2} \left(\mu_1 - (b_{\underline{T}} + b_{\underline{T}+1}) \mu_2 + b_{\underline{T}} b_{\underline{T}+1} \mu_3 \right) \mathbb{1}_{\zeta_j \geq \underline{T}, c_j \geq \bar{T}} \right),
\end{aligned}
$$

where we use weights $\frac{c}{c-\bar{T}}$ following the same logic as in equation (13).

To see that $\mathbb{E}\left[f_{\bar{T}}^{[\mu_1]}(\boldsymbol{\zeta}; \boldsymbol{\mu}, \boldsymbol{b}) \right] = 0$, we use equations (19) and (20). We use equation (19) and set $t = \underline{T}$ to find an expression for the expected value of the second term in $f_{\bar{T}}^{[\mu_1]}$:

$$
\begin{aligned}
\mathbb{E}\left[\frac{c}{c-\bar{T}} \sum_{j=1}^{K} \mathbb{1}_{\zeta_j \geq \underline{T}, c_j \geq \bar{T}} \right] &= \int_{\theta_L}^{\theta_H} \omega^f(\theta) \prod_{s=0}^{\underline{T}-1} (1 - h_s(\theta)) dG(\theta) \\
&= \int_{\theta_L}^{\theta_H} \tilde{\omega}_{\underline{T}}^f(\theta) dG(\theta).
\end{aligned}
$$

In the next step, use equation (20) with $t = \underline{T}$ to find the expected value of the first term of $f_{\bar{T}}^{[\mu_1]}$:

$$
\begin{aligned}
\mathbb{E}\left[\frac{c}{c-\bar{T}} \sum_{j=1}^{K} \mathbb{1}_{\zeta_j = \underline{T}, c_j \geq \bar{T}} \right] &= \int_{\theta_L}^{\theta_H} \omega^f(\theta) h_t(\theta) \prod_{s=0}^{t-1} (1 - h_s(\theta)) dG(\theta) \\
&= \int_{\theta_L}^{\theta_H} \omega^f(\theta) \theta b_{\underline{T}} \prod_{s=0}^{\underline{T}-1} (1 - h_s(\theta)) dG(\theta) \\
&= b_{\underline{T}} \int_{\theta_L}^{\theta_H} \theta \tilde{\omega}_{\underline{T}}^f(\theta) dG(\theta).
\end{aligned}
$$

Since $\mu_1 = \frac{\int_{\theta_L}^{\theta_H} \theta \tilde{\omega}_{\underline{T}}^f(\theta) dG(\theta)}{\int_{\theta_L}^{\theta_H} \tilde{\omega}_{\underline{T}}^f(\theta) dG(\theta)}$, the result follows. The proof that $\mathbb{E}\left[f_{\bar{T}}^{[\mu_2]}(\boldsymbol{\zeta}; \boldsymbol{\mu}, \boldsymbol{b}) \right] = 0$ and $\mathbb{E}\left[f_{\bar{T}}^{[\mu_3]}(\boldsymbol{\zeta}; \boldsymbol{\mu}, \boldsymbol{b}) \right] = 0$ is similar and so we omit it.

We find that $\hat{\boldsymbol{\mu}} = (1, 1.331, 2.137)$. We assume that $G(\cdot|\tilde{\omega}_{\underline{T}}^f)$ has a beta distribution over the interval $[\theta_L, \theta_H]$, and choose its two parameters $\tilde{\alpha}, \tilde{\beta}$ together with θ_L, θ_H to match the first two estimated moments of the distribution and to minimize the mean squared error between the model-implied and estimated Kaplan-Meier hazard. We find $\theta_L = 0.156$, $\theta_H = 6.071$, $\tilde{\alpha} = 1.668$, $\tilde{\beta} = 10$. This distribution has a mass point at $\theta_{\max} = 1/\hat{b}_2$, with

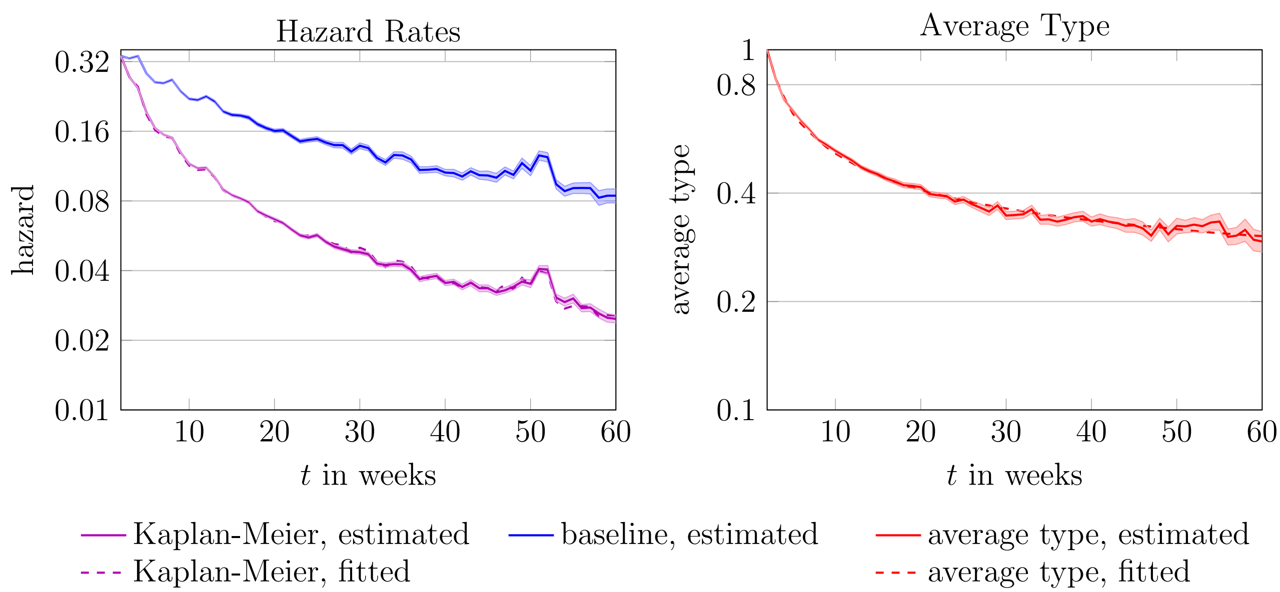

Figure 6: Kaplan-Meier and average type implied by the calibrated model and its comparison to estimates for pooled IRI data, log scale. The purple line shows the Kaplan-Meier hazard, the blue line is the estimated baseline hazard, and the red line shows the "average type" at given duration, calculated as the ratio of Kaplan-Meier and baseline hazards, as in Figure 1. The dashed lines show the fitted Kaplan-Meier hazard and the average type.

mass 0.0057, which ensures that $1-\theta b_t$ is always positive for all t. The third moment of this distribution, which is not targeted in the calibration, is 2.178, very close to the estimated $\hat{\mu}_3 = 2.137$.

Figure 6 shows that we fit the estimated Kaplan-Meier hazard and average type for $t \leq 60$ very well. Using the estimate baseline hazard and the frailty distribution with the above parameters, we use equation (6) to compute the implied Kaplan-Meier hazard, call it $\tilde{H}_t$, and then compute the average type as $\tilde{H}_t/\hat{b}_t$. These are depicted with dashed lines in Figure 6.

We use the Kaplan-Meier hazard estimated in Section 6.1 for $t <= 60$ and assume that it is given by $H_t = \gamma_0 + \gamma_1/t$ for $60 < t \leq 500$, and zero for any $t > 500$. We estimate γ_0 and γ_1 by fitting this function using the estimated baseline hazard for weeks 10–60. We find $\gamma_0 = 0.009$ and $\gamma_1 = 1.142$. We then use the model structure to recover the baseline hazard for $t > 60$ using the decomposition $H_t = b_t E[\theta|t]$. For any initial distribution $dG(\theta)$, we can compute distribution of types among products surviving to t, $dG(\theta|t)$, using the distribution $dG(\theta|t-1)$ and baseline hazard at t, b_t. We use this relationship together with $H_t = b_t E[\theta|t]$, where H_t is known, to recover b_t.

Online Appendix

D Additional Empirical Results

We report additional empirical results in this section.

D.1 Ergodic Distribution

To estimate the Kaplan-Meier hazard, we assume that when we first observe a product, the duration of the in-progress spell is a random draw from the stationary ergodic duration distribution for that product. A testable implication of that assumption is that, conditional on censoring time, the share of products changing its price in any week is constant. We implement a test in the following way. For all products with censoring time c, we compute the fraction of price changes that occur by week t since the start of the in-progress spell; we call it $F_c(\frac{t}{c})$. We then average the cumulative distribution function F_c across all $c > \bar{T}$, using the number of products with the corresponding value of c as weights. Figure 7 shows that the corresponding empirical density lies within five percent of a uniform density. It is close enough to uniform that we think the stationary mixture assumption is an empirically useful starting point.

D.2 Aggregation to Monthly to Frequency

We aggregate weekly data to monthly frequency in the following way: spells with duration 2–5 weeks are coded as duration of one month, spells of with duration 6–9 as duration of two months, and so on. We then estimate the MPH model using setting $\underline{T} = 1$ and $\bar{T} = 15$ months. To display the results, we convert the baseline hazard and Kaplan-Meier hazard into weekly units by

$$h_t^w = 1 - (1 - h_t^m)^{1/4},$$

where h_t^m is monthly hazard, either Kaplan-Meier or baseline, in month t and h_t^w is a weekly hazard in month t.

Figure 8 compares the estimates using weekly and monthly data.

D.3 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

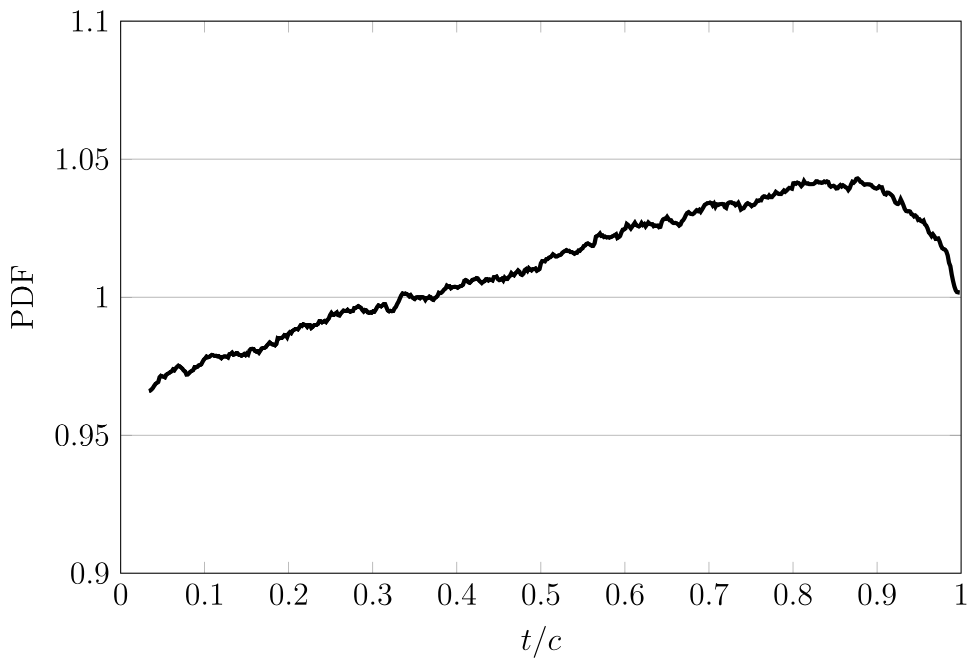

Figure 7: Empirical density of times when products change prices, measured from the start of an in-progress spell, pooled IRI data.

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 $t \in \{\underline{T}, \ldots, \bar{T}\}$ using some values $\underline{T}$ and $\bar{T}$. We first fix $\bar{T} = 60$ and estimate the model for different values of $\underline{T} = 2, 3, \ldots, 10$. 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 $t \in \{\underline{T}, \ldots, \bar{T}\}$, 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.

The left panel of Figure 9 shows the results for the benchmark model and Figure 10 for the competing risks model. The choice of $\underline{T}$ affects the estimate of the baseline hazard in the benchmark model. This is in line with the fact that we reject the model using the J-test. 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

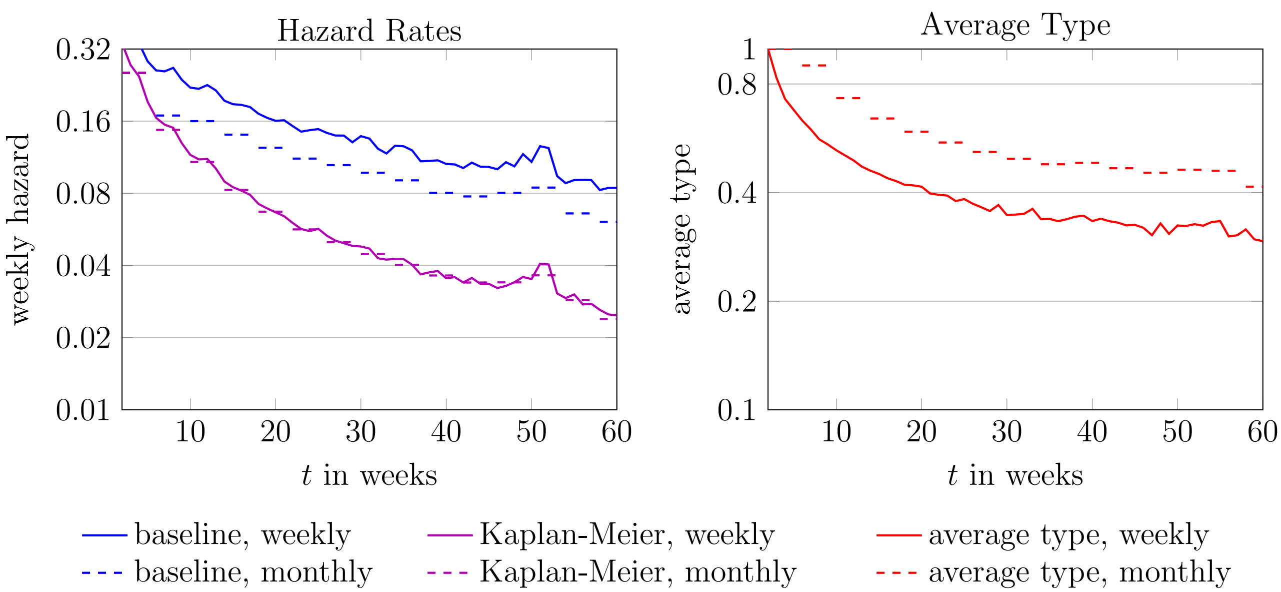

Figure 8: Kaplan-Meier and baseline hazard using weekly and monthly pooled IRI data, log scale. The solid lines uses weekly data, the dashed lines are data aggregated to monthly frequency. The purple line shows the Kaplan-Meier hazard, the blue line is the estimated baseline hazard. The red line shows the "average type" at given duration, calculated as the ratio of Kaplan-Meier and baseline hazard. The baseline hazard is normalized to be equal to the Kaplan-Meier hazard at duration 2 weeks in the weekly data, or 1 month in the monthly data. Kaplan-Meier and baseline hazard are in weekly units.

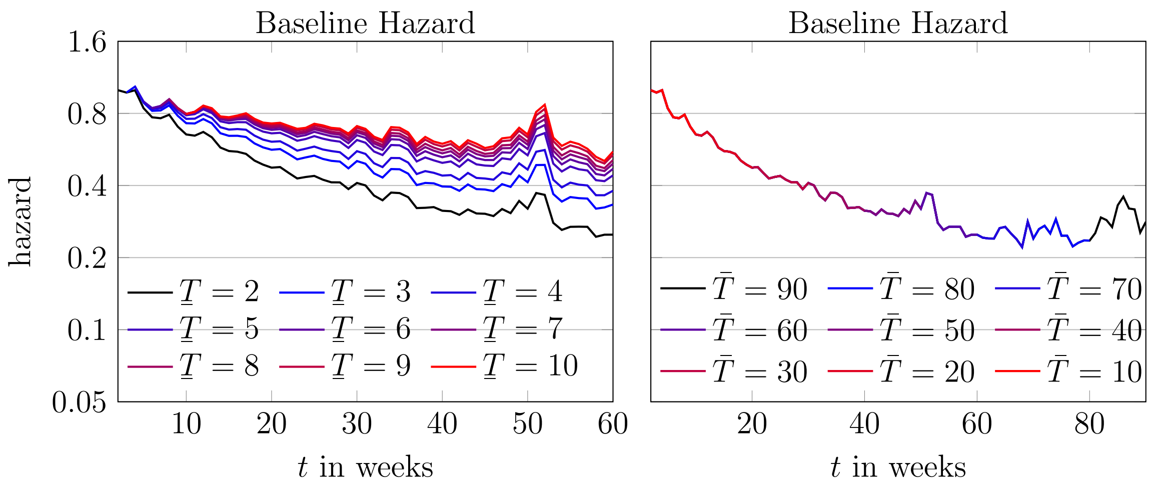

Figure 9: Baseline hazard for pooled IRI data, log scale, estimated using different values of $\underline{T} \in \{2, \ldots, 10\}$ and $\bar{T} = 60$ in the left panel, and using different values for $\bar{T} \in \{10, 20, \ldots 90\}$ and $\underline{T} = 2$ in the right panel.

so b^{-+}.

To analyze the role of $\bar{T}$, we fix $\underline{T} = 2$ and estimate the model for $\bar{T} \in \{10, 20, \ldots, 90\}$. We now normalize $b_2(2, \bar{T}) = 1$ for each value of $\bar{T}$. The right pane of Figure 9 and Figure 11 show 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 c^i to at most 80 weeks for every product. With the shorter censoring time, the choice of $\underline{T}$ matters less for all four hazards, see Figure 12. 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 consistent with time-varying types.

E Maximum Likelihood Estimators

We investigate reasons for differences between GMM and ML estimates presented in Figure 5. We formulate the MPH model in continous time and write down the likelihood function under two different timing assumptions. First, we assume that the data are generating by a continuous time model but durations are measured only in discrete times; we call this model Continuous Time with Discrete Measurement (CT-DM). Second, we assume that the baseline

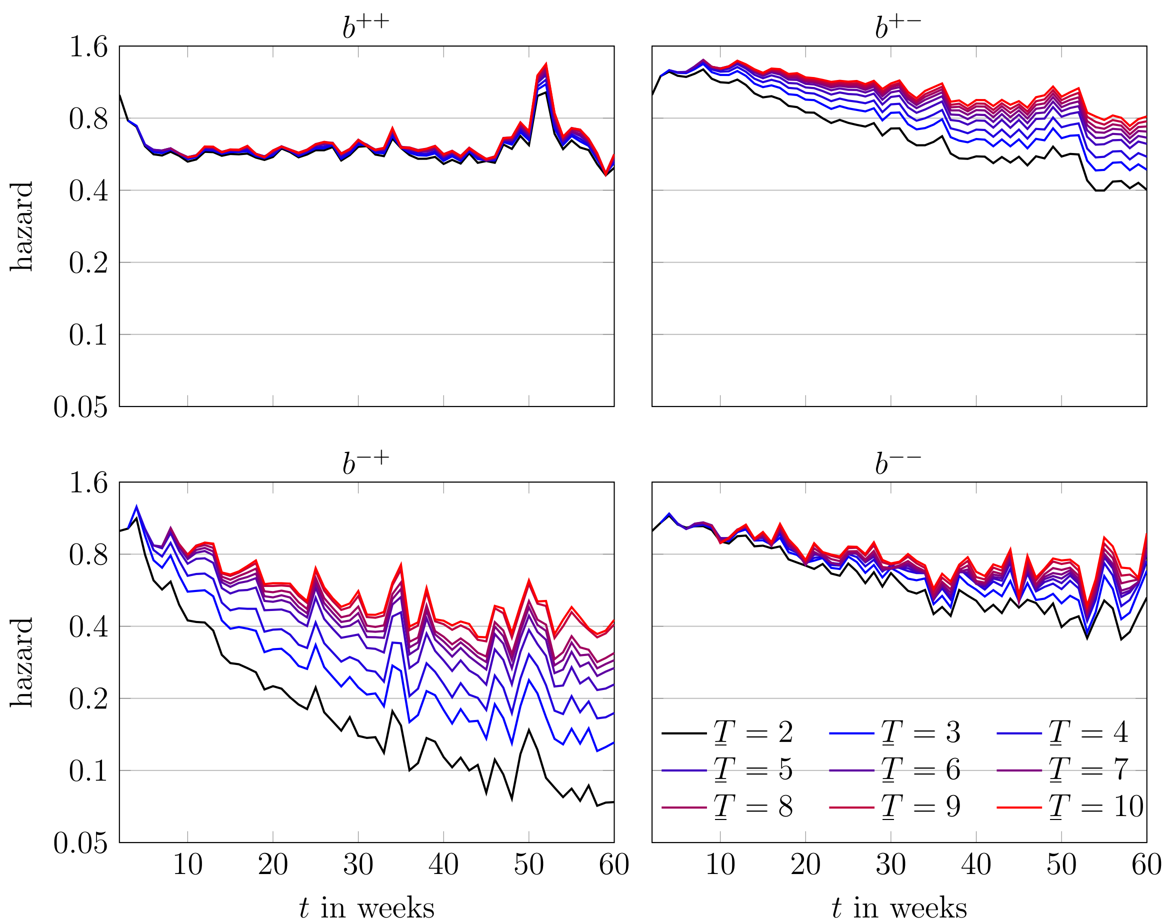

Figure 10: Baseline hazard for the competing risks model, pooled IRI data, log scale, estimated using different values of $\underline{T} \in \{2, \dots, 10\}$ and $\bar{T} = 60$.

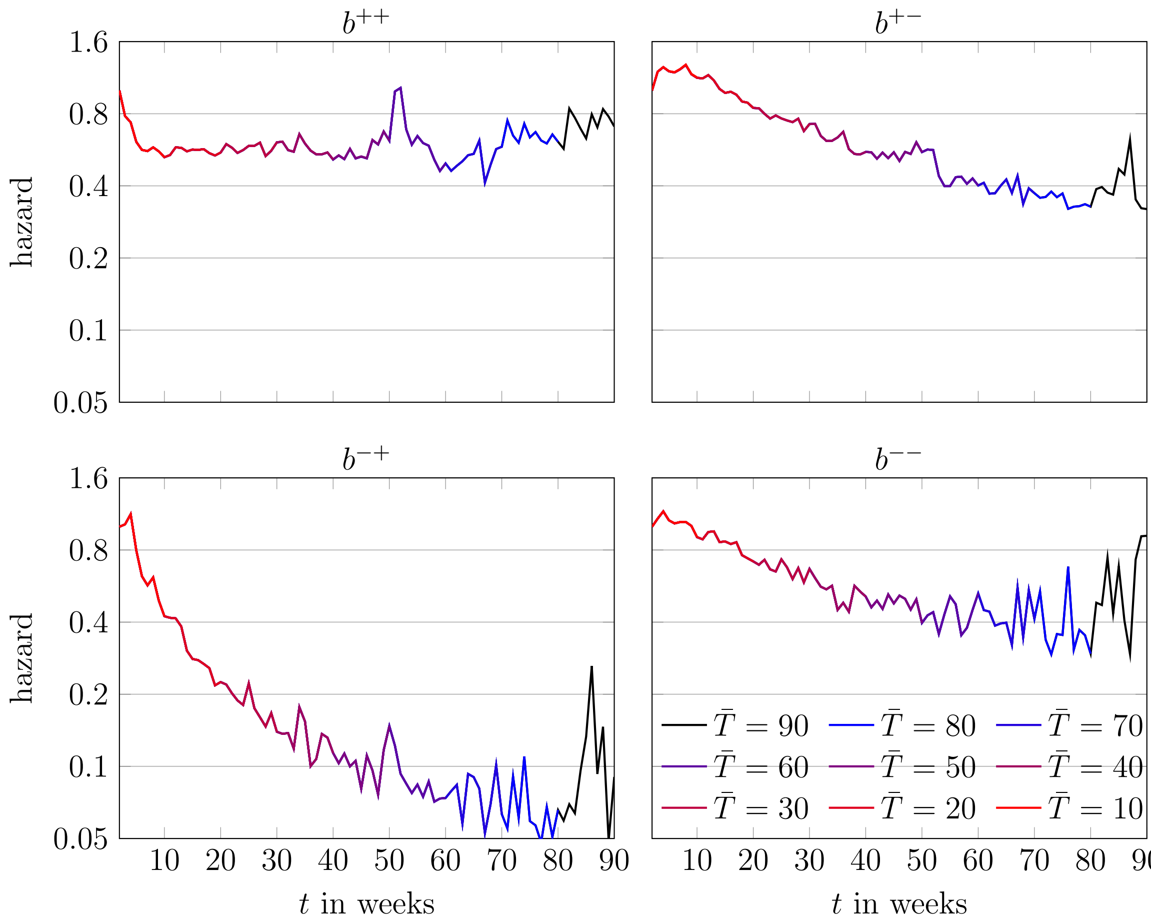

Figure 11: Baseline hazard for the competing risks model, pooled IRI data, log scale, estimated using different values of $\bar{T} \in \{10, 20, \ldots, 90\}$ and $\underline{T} = 2$.

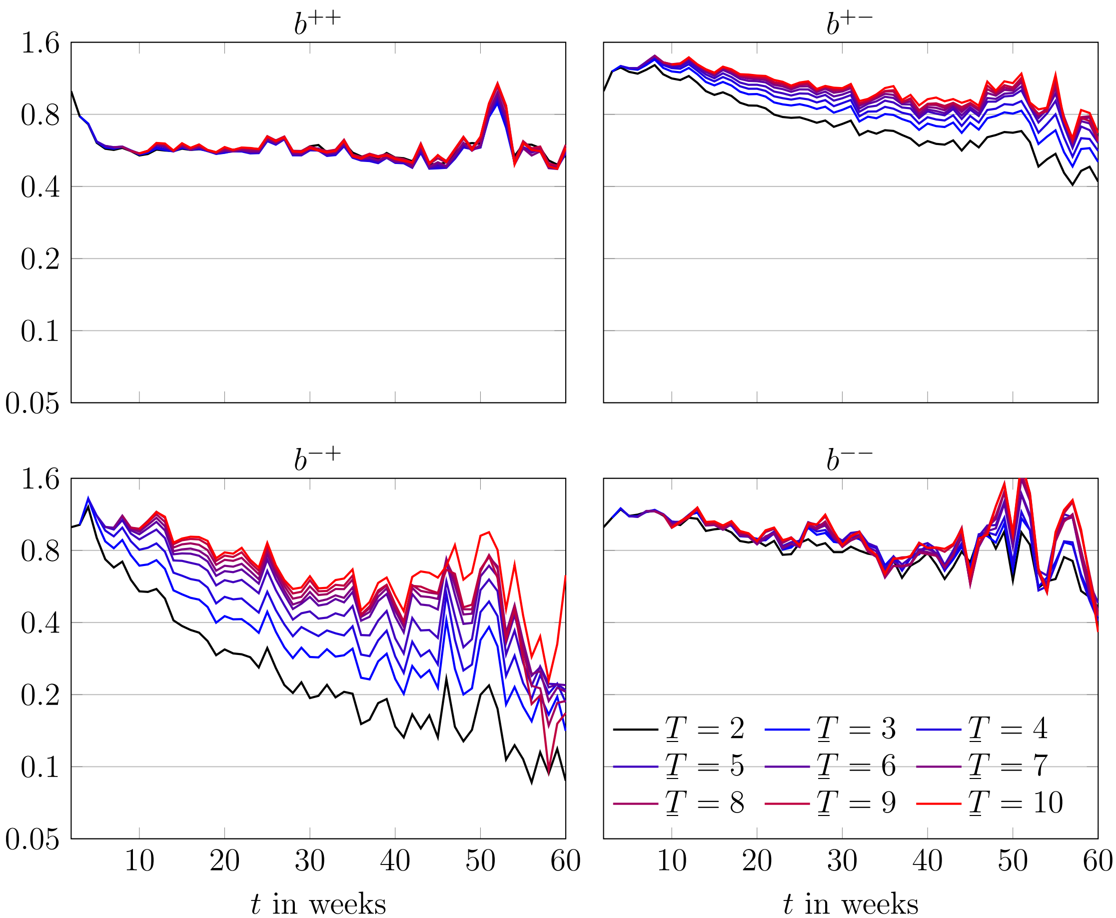

Figure 12: Baseline hazard for the competing risks model, pooled IRI data, log scale, estimated using different values of $\underline{T} \in \{2, \ldots, 10\}$, $\bar{T} = 60$ and censoring time restricted to be at most 80 weeks.

hazard is piece-wise constant and that observed discrete duration corresponds to continuous time duration; we call this model Continuous Time with Continuous Measurement (CT-CM).

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.

E.1 Continuous Time with Discrete Measurement

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 $e^{-\theta \int_0^t b(s)ds}$ 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 $\boldsymbol{b} \equiv \{b_1, \dots, b_{\bar{T}}, b_{\bar{T}+1}\}$, where sparsity of data lead us to impose $b_t = b_{\bar{T}+1}$ for all $t \geq \bar{T} + 1$. 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

$$\mathcal{L}^{CT-DM} = \frac{1}{N} \sum_{i=1}^{N} \log f_{d_1^i + d_2^i}^{CT-DM}(\zeta_1^i, \zeta_2^i; \boldsymbol{z}, m, v). \tag{32}$$

We impose $z_0 = 0$, which holds by definition. We also normalize $m = 1$.[19] 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

$$\mathcal{L}^{CT-DM} = \frac{1}{N} \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} w_k f_{d_1^i + d_2^i}^{CT-DM}(\zeta_1^i, \zeta_2^i; \boldsymbol{z}, m_k, v_k) \right). \tag{33}$$

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

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

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.

E.2 Continuous Time with Continuous Measurement

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

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

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:

$$\mathcal{L}^{CT-CM} = \frac{1}{N} \sum_{i=1}^{N} \log f_{d_1^i + d_2^i}^{CT-CM}(\zeta_1^i, \zeta_2^i; \boldsymbol{z}, m, v). \tag{34}$$

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

$$\mathcal{L}^{CT-CM} = \frac{1}{N} \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} w_k f_{d_1^i + d_2^i}^{CT-CM}(\zeta_1^i, \zeta_2^i; \boldsymbol{z}, m_k, v_k) \right). \tag{35}$$

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 (34) or (35). We assume that the baseline hazard is piecewise constant and so z is piecewise linear.

E.3 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 functional 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 $\bar{T}$, we have only one dummy variable for spells longer than $\bar{T}$. This dummy is equal to 1 if the measured duration exceeds $\bar{T}+1$ and zero otherwise. We find that when we use two spells per product, the maximum likelihood estimates in Stata coincide with the CT-CM model estimates with one gamma distribution.

E.4 Results

We use IRI pooled sample data where we use first two spells per product. On this sample, we estimate the baseline hazard using CT-CM, CT-DM as well as the discrete time model with discrete measurement (DT-DM) using our GMM estimator. For the CT-CM and CT-DM models we assume that the frailty distribution is either gamma or a mixture of gammas.

Figure 13 shows the results. The hazards are normalized to be equal 1 at duration of 2 weeks. The blue line shows the baseline hazard estimated from the discrete time model with discrete measurement (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-DM(1) model, which properly takes into account time aggregation, gives an estimate basically identical to our DT-DM model. The CT-CM(1) baseline hazard is much lower, recovering little heterogeneity. 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

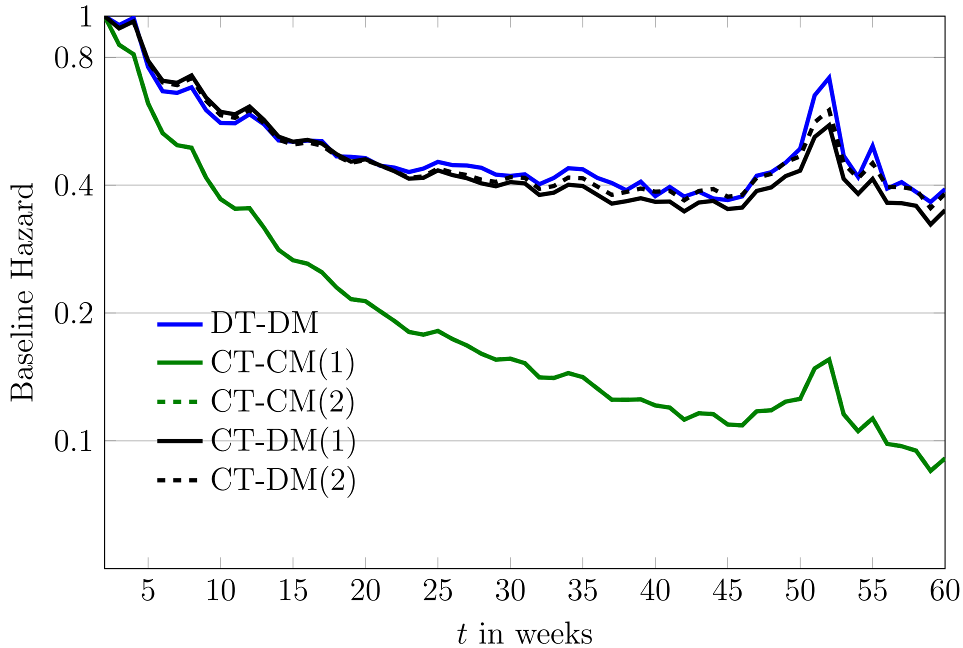

Figure 13: Baseline hazard estimated using different methods with two-spell IRI data, log scale. The blue line is the discrete time model with discrete measurement (DT-DM). The green lines correspond to continuous time with continuous time measurement (CT-CM), where the frailty distribution is a single 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).

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 and 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.

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.

F Baseline Hazards for Product Categories

Here we present our results by product category. Figure 14 shows the baseline and KM hazards and Figure 15 shows the average type estimated using the GMM conditions for the MPH model. Figures 16 and 17 show the baseline hazard for price trends, b^{++} and b^{--} respectively, estimated using the GMM conditions for the competing risks model with observable characteristics.

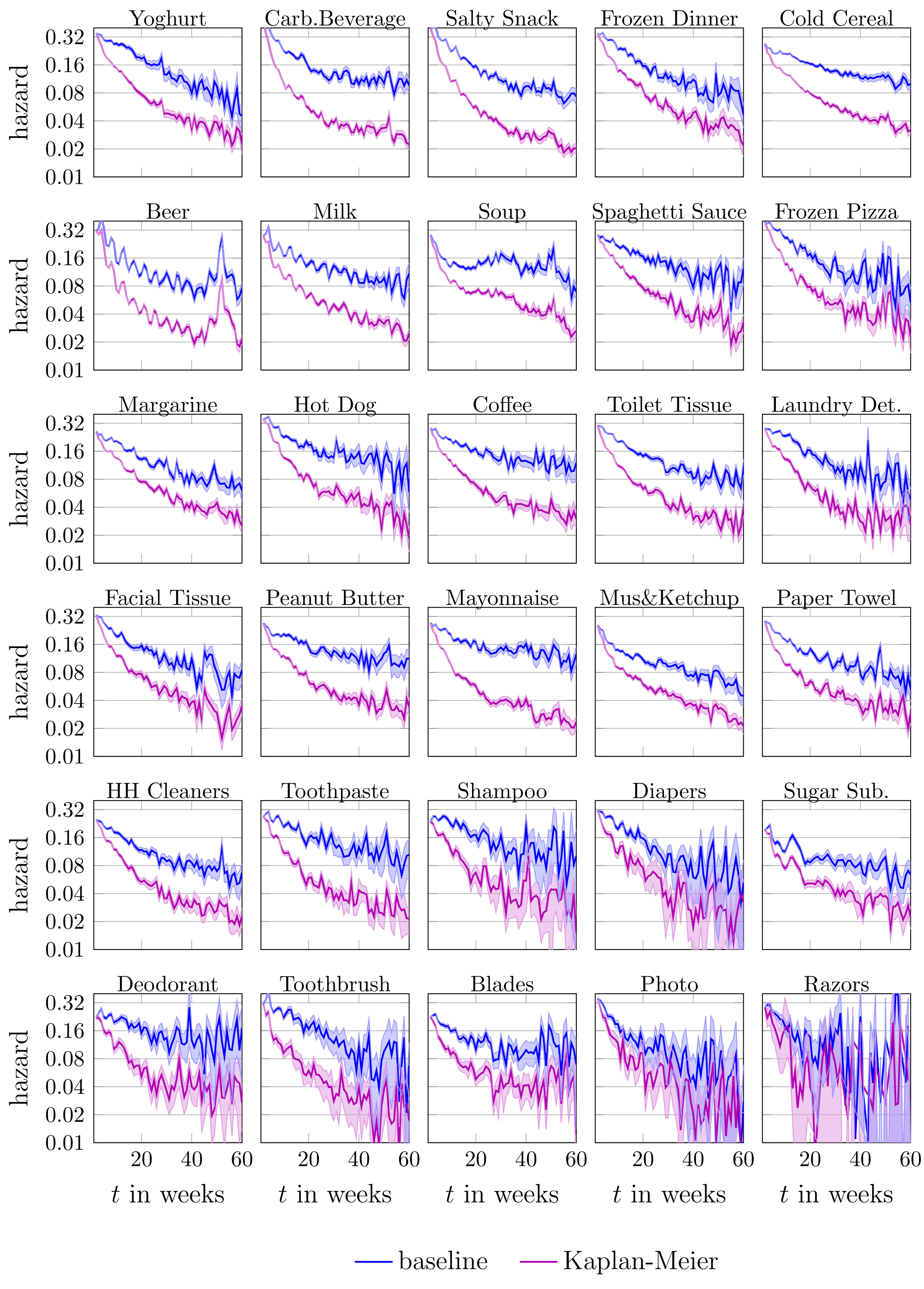

Figure 14: Kaplan-Meier and baseline hazards for individual product categories, IRI data, log scale. Product categories are sorted by the number of spell pairs.

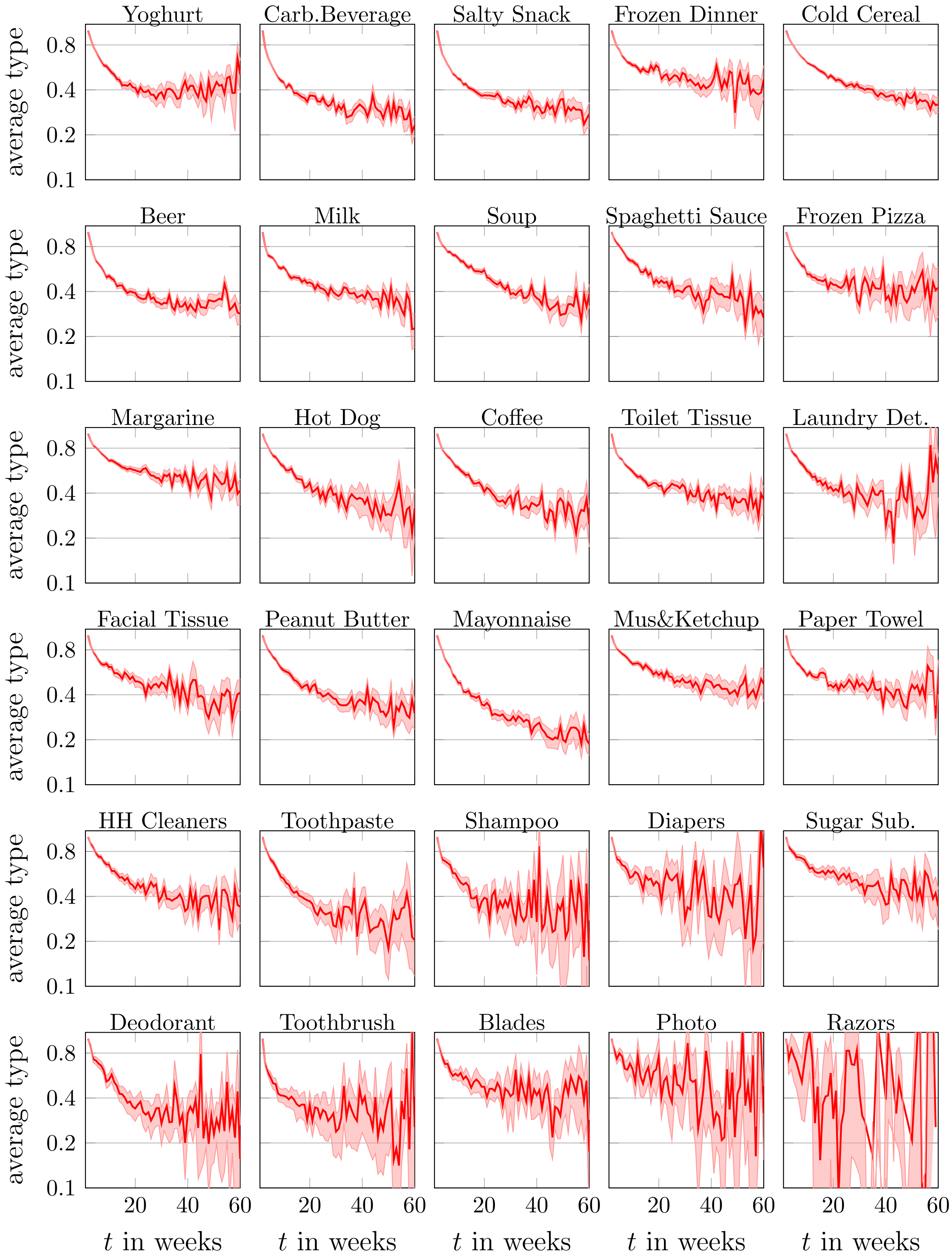

Figure 15: Average type for individual product categories, IRI data, log scale. Product categories are sorted by the number of spell pairs.

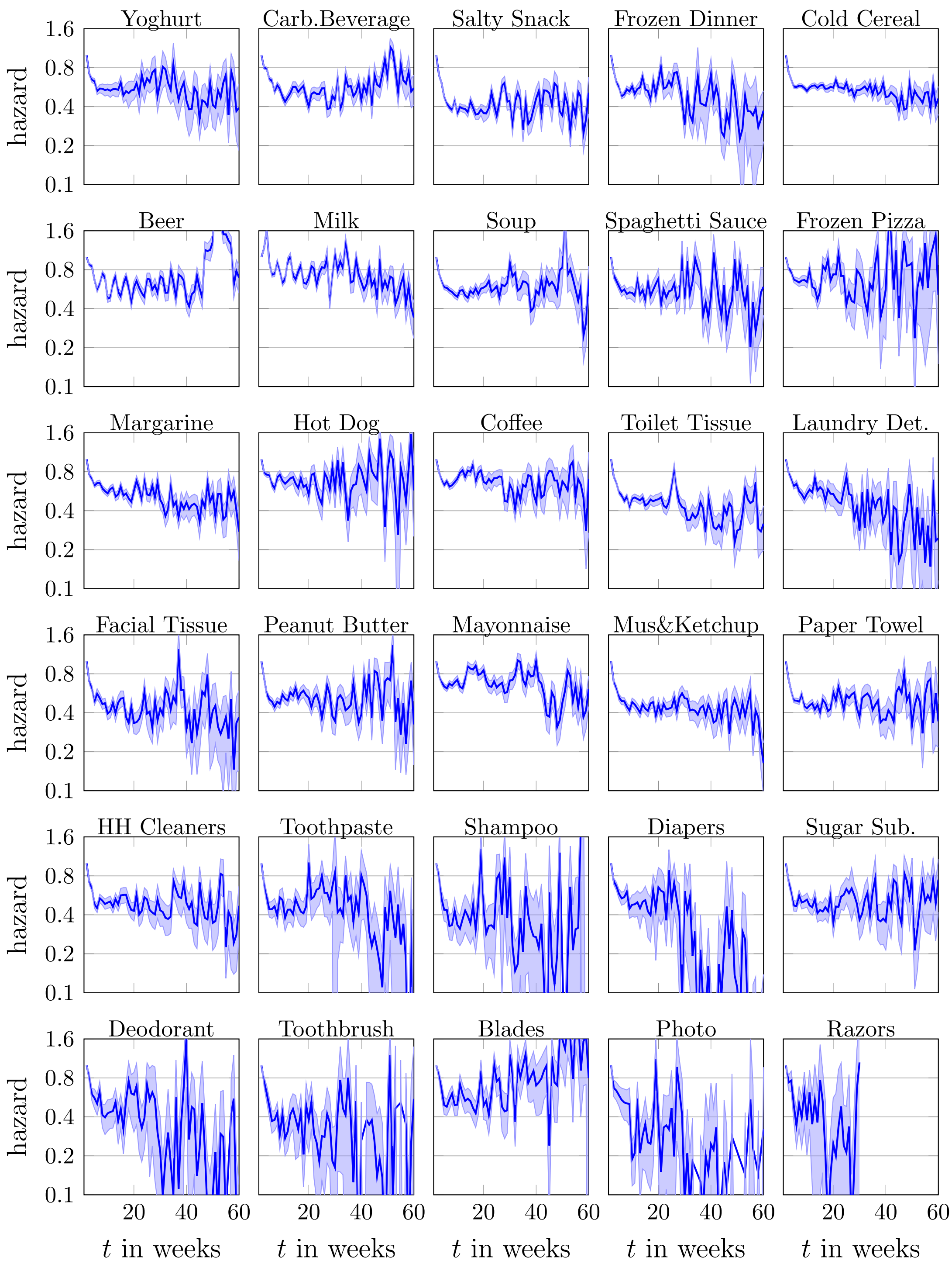

Figure 16: Baseline hazard b^{++} in the competing risks model for individual product categories, IRI data, log scale. Product categories are sorted by the number of spell pairs.

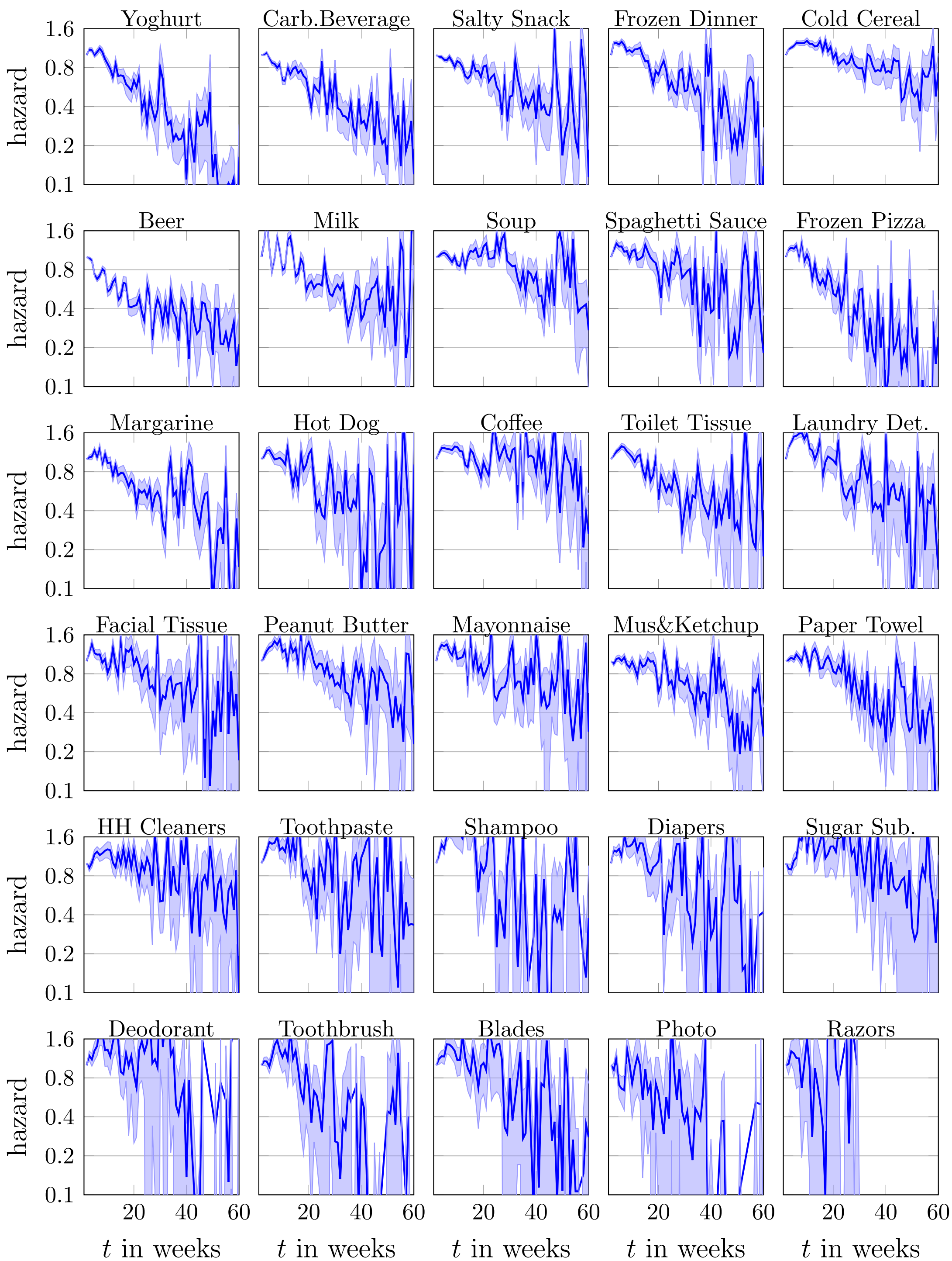

Figure 17: Baseline hazard b^{--} in the competing risks model for individual product categories, IRI data, log scale. Product categories are sorted by the number of spell pairs.