Supplemental Appendix

Consistent Evidence on Duration Dependence of Price Changes

Fernando Alvarez   Katarína Borovičková   Robert Shimer

B Time-Dependent Pricing with Heterogeneous Firms

B.1 Model

Household

Using the period budget constraints (18) and the no-Ponzi scheme condition, we derive the lifetime budget constraint

$$0 \leq M_0 + A_0 + \sum_{t=0}^{\infty} Q_t \left(\Pi_t + W_t \ell_t - \int p c_t(p) d\nu_t(p) \right) - \sum_{t=1}^{\infty} Q_t (R_{t-1} - 1) M_t. \tag{O.1}$$

The household maximizes its utility (16) and (17) subject to (O.1). Note that this problem has no solution if $R_t \leq 1$ for any $t \geq 0$, since in that case the household could achieve unbounded utility through a sufficiently large choice of M_{t+1}. We thus assume that $R_t > 1$ for all t in what follows.

Denote the Lagrange multiplier on the lifetime budget constraint as λ. The first order conditions with respect to ℓ_t, $c_t(p)$ and M_{t+1} for $t \geq 0$ are

$$[\ell_t] : 0 = \alpha \beta^t - \lambda Q_t W_t \tag{O.2}$$

$$[c_t(p)] : 0 = \beta^t C_t^{\frac{1-\eta}{\eta}} c_t(p)^{-\frac{1}{\eta}} - \lambda Q_t p \tag{O.3}$$

$$[M_{t+1}] : 0 = \frac{\beta^{t+1}}{M_{t+1}} - \lambda Q_{t+1}(R_t - 1) \tag{O.4}$$

where we used

$$\frac{\partial C_t}{\partial c_t(p)} = C_t^{\frac{1}{\eta}} c_t(p)^{-\frac{1}{\eta}}.$$

Equation (O.3) implies that $c_t(p)/c_t(p') = (p'/p)^{\eta}$. Using this, together with the definition of C_t in equation (17), we get the standard demand function

$$c_t(p) = C_t P_t^{\eta} p^{-\eta},$$

where P_t is defined in equation (19).

Now we focus on an environment with constant money growth in all periods, $M_{t+1}/M_t = 1+\pi > \beta$, and look for an equilibrium in which the real interest rate converges to a constant as t gets big. To find the implied real interest rate, take ratios of equation (O.4) at t and $t+1$ and use $Q_t \equiv \prod_{s=0}^{t-1} \frac{1}{R_s}$ to get a difference equation for R_t:

$$\frac{R_t(R_{t-1}-1)}{R_t-1} = \frac{1+\pi}{\beta} \Rightarrow R_t = \frac{1+\pi}{1+\pi-\beta(R_{t-1}-1)}.$$

Define $\bar{R} \equiv (1+\pi)/\beta > 1$. This difference equation has two stationary solutions, $R_t = \bar{R}$ and $R_t = 1$. The former is globally unstable, while the latter is inconsistent with a solution to the household problem. Thus a necessary condition for the real interest rate to converge to a constant is $R_t = \bar{R}$ for all t. We impose that in what follows.

Next, we evaluate equation (O.4) at $t = 0$ and use $R_t = \bar{R}$ and $M_1 = M_0(1+\pi)$ to obtain $\lambda = 1/(M_0(\bar{R}-1))$. Equation (O.2) together with $Q_t = \bar{R}^{-t}$ imply that W_t grows at the rate π, $W_t = W_0(1+\pi)^t$. Moreover, evaluating (O.2) at $t = 0$ pins down the initial level of wages, $W_0 = \alpha M_0(\bar{R}-1)$. This proves

$$W_t = \alpha M_t(\bar{R}-1). \tag{O.5}$$

Finally, we manipulate equation (O.3). In particular, we take $(1-\eta)$ power of both sides and integrate across all p to get $\lambda Q_t C_t P_t = \beta^t$. Combining it with (O.2), we find that aggregate consumption is inversely proportional to markup,

$$C_t = \frac{W_t}{\alpha P_t}. \tag{O.6}$$

Equivalently, using equation (O.5), nominal output $P_t C_t$ is equal to a share $\bar{R}-1$ of the money supply M_t. This means that monetary policy determines nominal output.

Firms

Now we turn to the maximization problem of a firm with time-dependent pricing rule Φ. If the firm has a chance to set the price at time $t > 0$, it sets it so as to maximize the present discounted value of its profits until the next price adjustment, $\tilde{\Pi}_t(p;\Phi)$

$$\tilde{\Pi}_t(p;\Phi) \equiv \sum_{s=0}^{\infty} \frac{Q_{t+s}}{Q_t} \Phi_s c_{t+s}(p)\left(p - W_{t+s}\right),$$

where $c_{t+s}(p) = C_{t+s} P_{t+s}^{\eta} p^{-\eta}$ is the demand for a good with price p at time $t+s$, determined by the household. The household's problem implies that $Q_t W_t = \alpha \beta^t / \lambda$ (equation (O.2)) and that $P_t = W_t/(\alpha C_t)$ (equation (O.6)). We use these expressions to write firm's profit as

$$\tilde{\Pi}_t(p; \Phi) = \frac{\alpha^{1-\eta}\beta^t}{\lambda Q_t} \sum_{s=0}^{\infty} \beta^s \Phi_s C_{t+s}^{1-\eta} \left(\frac{p}{W_{t+s}}\right)^{-\eta} \left(\frac{p}{W_{t+s}} - 1\right).$$

We thus get

$$P_t^*(\Phi) = \arg\max_p \sum_{s=0}^{\infty} \beta^s \Phi_s C_{t+s}^{1-\eta} \left(\frac{p}{W_{t+s}}\right)^{-\eta} \left(\frac{p}{W_{t+s}} - 1\right).$$

We use that $W_t = W_0(1+\pi)^t$ and take the first order condition with respect to p to find that the optimal price is

$$P_t^*(\Phi) = \frac{\eta W_t}{\eta - 1} \frac{\sum_{s=0}^{\infty} \beta^s \Phi_s C_{t+s}^{1-\eta} (1+\pi)^{\eta s}}{\sum_{s=0}^{\infty} \beta^s \Phi_s C_{t+s}^{1-\eta} (1+\pi)^{(\eta-1)s}}. \tag{O.7}$$

Note that if additionally consumption C_t is constant over time, equation (O.7) reduces to

$$P_t^*(\Phi) = \frac{\eta W_t}{\eta - 1} \frac{\sum_{s=0}^{\infty} \beta^s \Phi_s (1+\pi)^{\eta s}}{\sum_{s=0}^{\infty} \beta^s \Phi_s (1+\pi)^{(\eta-1)s}} \tag{O.8}$$

for all t. This equation then implies $P_t^*(\Phi)$ grows at rate π. From the definition of the ideal price index in equation (19) and the result that nominal wages grow at rate π, we then get

$$P_t = \frac{\eta W_t}{\eta - 1} \left(\int \left(\sum_{s=1}^{\infty} \omega_s(\Phi)(1+\pi)^{(\eta-1)(s-1)} \right) \left(\frac{\sum_{s=0}^{\infty} \beta^s \Phi_s (1+\pi)^{\eta s}}{\sum_{s=0}^{\infty} \beta^s \Phi_s (1+\pi)^{(\eta-1)s}} \right)^{1-\eta} dF(\Phi) \right)^{\frac{1}{1-\eta}} \tag{O.9}$$

for all t. This ensures that P_t grows at rate π as well. Finally, since nominal output grows at rate π (equation (O.6)), real consumption must be constant. Thus we can construct a *stationary equilibrium* with constant real variables $(C_t, M_t/P_t, W_t/P_t, R_t)$ and a time-invariant distribution of relative prices $\nu_t(p) = \nu_0(pP_0/P_t)$.

B.2 Log-Linearization

The economy converges to a stationary equilibrium if money growth remains constant for a long time. Here we approximate the transitional dynamics of an economy that starts from one stationary equilibrium with zero money growth and then unexpectedly at time 1 experiences a change in the money growth rate.

First define the function $z(C, \pi, s, \Phi) \equiv \beta^s C^{1-\eta} \Phi_s (1+\pi)^{(\eta-1)s}$ and rewrite equation (O.7) as

$$\log P_t^*(\Phi) = \log\left(\frac{\eta W_t}{\eta - 1}\right) + Z(\pi, C_t, C_{t+1}, \ldots; \Phi),$$

where

$$Z(\pi, C_t, C_{t+1}, \ldots; \Phi) \equiv \log\left(\sum_{s=0}^{\infty} z(C_{t+s}, \pi, s, \Phi)(1+\pi)^s\right) - \log\left(\sum_{s=0}^{\infty} z(C_{t+s}, \pi, s, \Phi)\right).$$

We log-linearize $P_t^*(\Phi)$ around a stationary equilibrium with zero inflation where $C_t = \bar{C}$ and $\pi = 0$. A little bit of algebra yields

$$\frac{\partial Z(\pi, C_t, C_{t+1}, \ldots; \Phi)}{\partial C_{t+s}}|_{C_s=\bar{C}, \pi=0} = 0$$

$$\frac{\partial Z(\pi, C_t, C_{t+1}, \ldots; \Phi)}{\partial \pi}|_{C_s=\bar{C}, \pi=0} = \frac{\sum_{s=0}^{\infty} z(\bar{C}, 0, s, \Phi)s}{\sum_{s=0}^{\infty} z(\bar{C}, 0, s, \Phi)} = \frac{\sum_{s=0}^{\infty} \beta^s \Phi_s s}{\sum_{s=0}^{\infty} \beta^s \Phi_s}.$$

Furthermore, equation (5) states $\omega_t(\Phi) = \Phi_{t-1}\omega_1(\Phi)$, so

$$\frac{\sum_{s=0}^{\infty} \beta^s \Phi_s s}{\sum_{s=0}^{\infty} \beta^s \Phi_s} = \gamma(\Phi) - 1$$

where $\gamma(\Phi)$ is defined in (23). Denoting $p_t^*(\Phi) = \log P_t^*(\Phi)$, $w_t \equiv \log W_t$ and $p_t \equiv \log P_t$, the log-linearized dynamics of the model is thus given by

$$p_t(\Phi) = w_t + \log\frac{\eta}{\eta - 1} + (\gamma(\Phi) - 1)\,\pi \tag{O.10}$$

$$p_t = \int \sum_{s=0}^{\infty} \omega_{s+1}(\Phi) p_{t-s}(\Phi) dF(\Phi). \tag{O.11}$$

In the experiment, we assume the economy is in a stationary equilibrium with zero money growth at time $t = 0$. There is an increase of money growth to π so that $M_t = M_0(1+\pi)^t$ for $t > 0$. We now analyze the deviation of log-output from the old steady state at time t, denoted y_t, which is equal to the negative deviation of log-markup. In the equilibrium with zero money growth, equation (O.10) implies that each firm has a log-markup of $\log\frac{\eta}{\eta-1}$.

Hence, the deviation of log output from the old steady state is given by

$$\begin{aligned}
y_t &= \log\frac{\eta}{\eta-1} - (p_t - w_t) \\
&= \log\frac{\eta}{\eta-1} + w_t - \int\sum_{s=0}^{\infty}\omega_{s+1}(\Phi)p_{t-s}(\Phi)dF(\Phi) \\
&= \log\frac{\eta}{\eta-1} - \int\sum_{s=0}^{\infty}\omega_{s+1}(\Phi)\left(p_{t-s}(\Phi) - w_{t-s}\right)dF(\Phi) + \int\sum_{s=0}^{\infty}\left(w_t - w_{t-s}\right)\omega_{s+1}(\Phi)dF(\Phi).
\end{aligned}$$

Next, split the sums into two parts

$$\begin{aligned}
y_t &= \log\frac{\eta}{\eta-1} - \int\sum_{s=0}^{t-1}\omega_{s+1}(\Phi)\left(p_{t-s}(\Phi) - w_{t-s}\right)dF(\Phi) - \int\sum_{s=t}^{\infty}\omega_{s+1}(\Phi)\left(p_{t-s}(\Phi) - w_{t-s}\right)dF(\Phi) \\
&\quad + \int\sum_{s=0}^{t-1}\left(w_t - w_{t-s}\right)\omega_{s+1}(\Phi)dF(\Phi) + \int\sum_{s=t}^{\infty}\left(w_t - w_{t-s}\right)\omega_{s+1}(\Phi)dF(\Phi) \\
&= \log\frac{\eta}{\eta-1} - \int\sum_{s=0}^{t-1}\omega_{s+1}(\Phi)\left(\log\frac{\eta}{\eta-1} + \left(\gamma(\Phi) - 1\right)\pi\right)dF(\Phi) \\
&\quad - \int\sum_{s=t}^{\infty}\omega_{s+1}(\Phi)\left(\log\frac{\eta}{\eta-1}\right)dF(\Phi) \\
&\quad + \int\sum_{s=0}^{t-1}\left(s\pi\right)\omega_{s+1}(\Phi)dF(\Phi) + \int\sum_{s=t}^{\infty}\left(t\pi\right)\omega_{s+1}(\Phi)dF(\Phi)
\end{aligned}$$

where we used that for $t \leq 0$, $w_t = w_0$ and $p_t(\Phi) - w_t = \log\frac{\eta}{\eta-1}$, while for $t > 0$, it holds $w_t = w_0 + t\pi$ and the log-markup of firm Φ is given by (O.10). Finally, rearrange the terms and use the fact that $\int\sum_{s=0}^{\infty}\omega_{s+1}(\Phi)dF(\Phi) = 1$ to find the expression in the text:

$$\begin{aligned}
y_t &= -\pi\int\sum_{s=0}^{t-1}\omega_{s+1}(\Phi)\left(\gamma(\Phi) - 1\right)dF(\Phi) + \pi\int\sum_{s=0}^{t-1}\omega_{s+1}(\Phi)s\,dF(\Phi) + t\pi\int\sum_{s=t}^{\infty}\omega_{s+1}(\Phi)dF(\Phi) \\
&= -\pi\int\sum_{s=0}^{t-1}\omega_{s+1}(\Phi)\left(\gamma(\Phi) - 1\right)dF(\Phi) + \pi\int\sum_{s=0}^{t-1}\omega_{s+1}(\Phi)s\,dF(\Phi) + t\pi\left(1 - \int\sum_{s=0}^{t-1}\omega_{s+1}(\Phi)dF(\Phi)\right) \\
&= \pi\int\left(t - \sum_{s=0}^{t-1}\omega_{s+1}(\Phi)\left(\gamma(\Phi) - 1 - s + t\right)dF(\Phi)\right) \\
&= \pi\int\left(t - \sum_{s=0}^{t}\omega_{s}(\Phi)\left(\gamma(\Phi) - s + t\right)dF(\Phi)\right).
\end{aligned}$$

B.3 Calibration

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 do not use spells shorter than $\underline{T}$ in the estimation, and set $b_t = 0$ for $t < \underline{T}$.

We also want to estimate moments of the frailty distribution. Since the feasible mixture hazard gives zero weight to any product that we observe for less than $\bar{T}$ periods (see Section 4.4), we wish to measure a comparable frailty distribution, and so weight products by the product of the expected censoring time and frequency of price adjustment,

$$w^f_{\underline{T}}(\theta) \equiv \left(\sum_{c=\bar{T}+1}^{\infty} c q_c(\theta) \right) \omega_1(\Phi(\theta)).$$

Denote the weighted distribution $F^f_{\underline{T}}$:

$$dF^f_{\underline{T}}(\theta) = \frac{w^f_{\underline{T}}(\theta) dG(\theta)}{\int w^f_{\underline{T}}(\theta') dG(\theta')},$$

and let $\mu = (\mu_1, \mu_2, \mu_3)$ denote its first three moments. We can use purely cross-sectional data to estimate μ. In particular, we look at the fraction of spells with duration $\underline{T}$

$$\Psi_{\underline{T}} - \Psi_{\underline{T}-1} = \int \theta b_{\underline{T}} dF^f_{\underline{T}}(\theta).$$

Rearranging the terms, we can equivalently write

$$\Psi_{\underline{T}} - \Psi_{\underline{T}-1} - \mu_1 b_{\underline{T}} = 0. \tag{O.12}$$

This is an equation for μ_1. To obtain an equation for μ_2, we look at the fraction of spells with duration $\underline{T} + 1$

$$\Psi_{\underline{T}+1} - \Psi_{\underline{T}} = \int \theta b_{\underline{T}+1}(1 - \theta b_{\underline{T}}) dF^f_{\underline{T}}(\theta) = b_{\underline{T}+1}(\mu_1 - \mu_2 b_{\underline{T}}). \tag{O.13}$$

Finally, we use the share of spells with duration $\underline{T} + 2$ to find an equation for μ_3.

Assuming that $\zeta = (\zeta_0, \zeta_1, \ldots, \zeta_K)$ follows a stationary mixture model, we translate equations (O.12), (O.13) and equation for μ_3 into moment conditions for μ_1, μ_2, μ_3,

$$f_{\bar{T}}^{[\mu_1]}(\zeta; \mu, b) \equiv \frac{c}{c-\bar{T}} \sum_{j=1}^{K} \left(\mathbb{1}_{\zeta_j = \underline{T}, c_j \geq \bar{T}} - b_{\underline{T}} \mu_1 \mathbb{1}_{\zeta_j \geq \underline{T}, c_j \geq \bar{T}} \right) \tag{O.14}$$

$$f_{\bar{T}}^{[\mu_2]}(\zeta; \mu, b) \equiv \frac{c}{c-\bar{T}} \sum_{j=1}^{K} \left(\mathbb{1}_{\zeta_j = \underline{T}+1, c_j \geq \bar{T}} - b_{\underline{T}+1} \left(\mu_1 - b_{\underline{T}} \mu_2 \right) \mathbb{1}_{\zeta_j \geq \underline{T}, c_j \geq \bar{T}} \right) \tag{O.15}$$

$$f_{\bar{T}}^{[\mu_3]}(\zeta; \mu, b) \equiv \frac{c}{c-\bar{T}} \sum_{j=1}^{K} \left(\mathbb{1}_{\zeta_j = \underline{T}+2, c_j \geq \bar{T}} - b_{\underline{T}+2} \left(\mu_1 - (b_{\underline{T}} + b_{\underline{T}+1}) \mu_2 + b_{\underline{T}} b_{\underline{T}+1} \mu_3 \right) \mathbb{1}_{\zeta_j \geq \underline{T}, c_j \geq \bar{T}} \right), \tag{O.16}$$

where we use weights $\frac{c}{c-\bar{T}}$ following the same logic as in equation (12). In Online Appendix E.3, we prove that $\mathbb{E}\left[f_{\bar{T}}^{[\mu_i]}(\zeta; \mu, b)\right] = 0$, for $i = 1, 2, 3$.

We find that $\hat{\mu} = (1, 1.331, 2.137)$. Let $G_{\underline{T}}^{f}(\theta)$ be the distribution of θ implied by the stationary MPH model,

$$G_{\underline{T}}^{f}(\theta) = \int_{\Phi_{\underline{T}}/\Phi_{\underline{T}-1} \geq 1-\theta b_{\underline{T}}} dF_{\underline{T}}^{f}(\Phi).$$

We assume that $G_{\underline{T}}^{f}$ 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 mixture 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 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 7 shows that we fit the estimated mixture hazard and average type for $2 \leq t \leq 60$ very well. Using the estimated baseline hazard and the frailty distribution with the above parameters, we use equation (9) to compute the implied mixture 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 7.

We use the mixture hazard estimated in Section 6.1 for $2 \leq t \leq 60$ and assume that it is given by $H_t = \gamma_0 + \gamma_1/t$ for $60 < t \leq 500$, and 1 for $t = 501$. We estimate γ_0 and γ_1 by fitting this function to the estimated mixture hazard for weeks $10 \leq t \leq 60$. 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 $H_t = \bar{\theta}_t b_t$. For any initial distribution $G(\theta)$, we can compute distribution of types among products surviving to t, $G(\theta|t)$, using the distribution $G(\theta|t-1)$ and baseline hazard at t, b_t. We use this relationship together with $H_t = \bar{\theta}_t b_t$, where H_t is known, to recover b_t for $t > 60$.

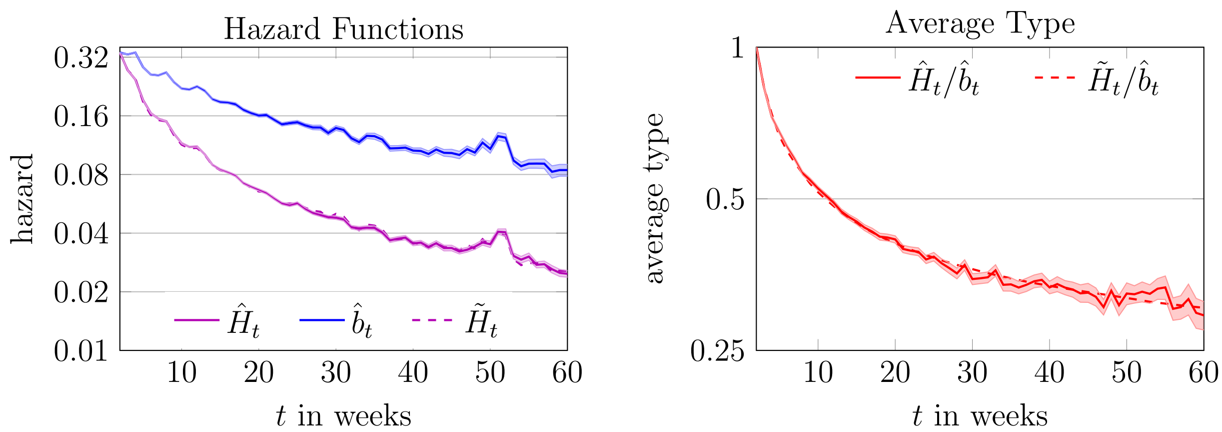

Figure 7: Fit of the mixture hazard by estimated frailty distribution and $\hat{b}_t$. The left panel shows the baseline hazard $\hat{b}_t$ as well as two estimates of the mixture hazard H_t. $\hat{H}_t$ uses GMM on purely cross-sectional data, following equation (4) in Proposition 4. $\tilde{H}_t$ uses the estimated type distribution and $\hat{b}_t$. The right panel compares the implied measures of the "average type," calculated as the ratio of mixture and baseline hazards, as in Figure 1.

C Price Plans

C.1 Shape of Hazards

In this section, we consider a trinomial discrete-time process $\tilde{x}_t$ with exit at two boundaries $\underline{x}, \bar{x}$, and derive formulae for the hazard of exiting through one or both of the boundaries. We then explain how to apply these results to the conditional and unconditional hazard of changing price plans.

We consider the following model. Time is discrete. The process $\tilde{x}_t$ attains one of $N \geq 3$ values $\{\underline{x}, \underline{x}+\Delta, \ldots, \bar{x}-\Delta, \bar{x}\}$. We denote $x_n \equiv \underline{x}+(n-1)\Delta$ for $n \in \{1, \ldots, N\}$, with $\bar{x} = \underline{x}+(N-1)\Delta$.

We assume $\tilde{x}_0 = x_{n_0}$ for some $n_0 \in \{1, \ldots, N\}$. Subsequently, for $\underline{x} < \tilde{x}_t < \bar{x}$, $\tilde{x}_{t+1}$ takes on one of three possible values: $\tilde{x}_{t+1} = \tilde{x}_t + \Delta$ with probability 1/4, $\tilde{x}_{t+1} = \tilde{x}_t - \Delta$ with probability 1/4, or $\tilde{x}_{t+1} = \tilde{x}_t$ otherwise. There is exit at the boundary, so if $\tilde{x}_t = \underline{x}$, there is a 1/4 chance of exit, a 1/4 chance of $\tilde{x}_t = \underline{x}+\Delta$, and a 1/2 chance of $\tilde{x}_t = \underline{x}$. The outcomes at the upper boundary are symmetric. All shocks are independent over time.

Let $m_t(x)$ be the fraction of realizations with $\tilde{x}_t = x$ at time t. The law of motion of

$m_t(x)$ is given by the Kolmogorov forward equation

$$m_{t+1}(x) = \begin{cases} \frac{1}{4}m_t(x-\Delta) + \frac{1}{2}m_t(x) + \frac{1}{4}m_t(x+\Delta) \text{ if } \underline{x} < x < \bar{x} \\ \frac{1}{4}m_t(x-\Delta) + \frac{1}{2}m_t(x) \text{ if } x = \bar{x} \\ \frac{1}{4}m_t(x+\Delta) + \frac{1}{2}m_t(x) \text{ if } x = \underline{x} \end{cases} \tag{O.17}$$

with initial condition $m_0(x_{n_0}) = 1$ and $m_0(x_n) = 0$ for $n \neq n_0$.

Proposition 6 *The solution of the equation* (O.17) *with the initial condition* $m_0(x_n) = 1$ *at* $n = n_0$ *and zero otherwise is given by*

$$m_t(x_n) = \sum_{j=1}^{N} \lambda_j^t \sin\left(\frac{n}{N+1} j\pi\right) \gamma_j, \tag{O.18}$$

where

$$\lambda_j = \frac{1}{2}\left(1 + \cos\left(\frac{1}{N+1} j\pi\right)\right), \quad \gamma_j = \frac{2}{N+1} \sin\left(\frac{n_0}{N+1} j\pi\right).$$

See Online Appendix G for the proof.

We now turn to duration analysis. The survival function at duration t is $S_t = \sum_{n=1}^{N} m_t(x_n)$. The hazard of exiting at boundary $\bar{x}$ ($\underline{x}$), denoted $h_t^{\bar{x}}$ ($h_t^{\underline{x}}$), is given by the ratio of the outflow through $\bar{x}$ ($\underline{x}$) and the survival function. The hazard of exiting at either boundary, h_t, is given by the sum of the two hazards:

$$h_t^{\bar{x}} = \frac{1}{4}\frac{m_t(\bar{x})}{S_t}, \quad h_t^{\underline{x}} = \frac{1}{4}\frac{m_t(\underline{x})}{S_t}, \quad h_t = h_t^{\bar{x}} + h_t^{\underline{x}}.$$

The following proposition establishes the shape of h_t for large t.

Proposition 7 *For* $1 < n_0 < N$, $h_1 = 0$. *For all* n_0, $\lim_{t\to\infty} h_t = 1 - \lambda_1$. *If* $\frac{N+1}{3} \leq n_0 \leq \frac{2(N+1)}{3}$, *then the hazard* h_t *converges to its asymptote from below. If* $n_0 < \frac{N+1}{3}$ *or* $n_0 > \frac{2(N+1)}{3}$, *the hazard* h_t *converges to its asymptote from above. The expected time to hit either* $\underline{x}$ *or* $\bar{x}$ *is* $2n_0(N+1-n_0)$.

See Online Appendix G for the proof.

We next turn to analyze shapes of $h_t^{\underline{x}}$ and $h_t^{\bar{x}}$ for large t.

Proposition 8 *For* $1 < n_0 < N$, *hazard* $h_1^{\underline{x}} = h_1^{\bar{x}} = 0$ *and* $\lim_{t\to\infty} h_t^{\underline{x}} = \lim_{t\to\infty} h_t^{\bar{x}} = \frac{1}{2}(1-\lambda_1)$. *For* $n_0 < \frac{N+1}{2}$, *hazard* $h_t^{\underline{x}}$ *converges to its asymptote from above and hazard* $h_t^{\bar{x}}$ *converges from below. For* $n_0 > \frac{N+1}{2}$, *hazard* $h_t^{\underline{x}}$ *converges to its asymptote from below and hazard* $h_t^{\bar{x}}$ *converges from above. For* $n_0 = \frac{N+1}{2}$, *both converge from below. .*

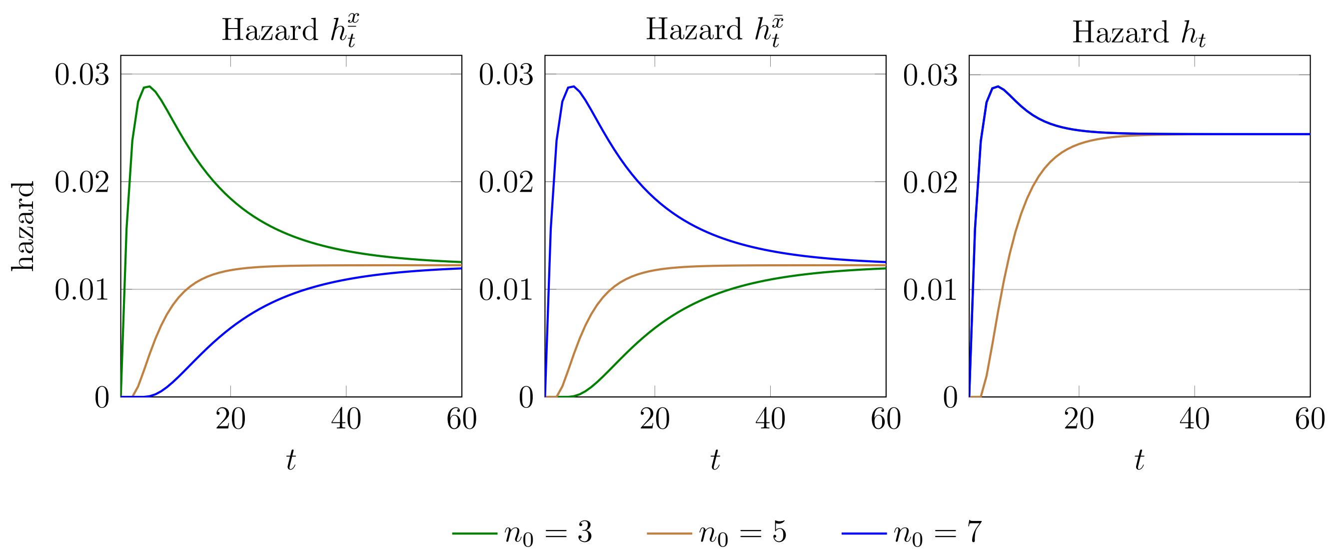

Figure 8: Hazards $h_t^{\underline{x}}$, $h_t^{\bar{x}}$ and h_t for $N = 9$ and different choices of $n_0 \in \{3, 5, 7\}$. Note the lines for $n_0 = 3$ and $n_0 = 7$ are identical in the right panel.

See Online Appendix G for the proof.

Figure 8 shows the shapes for the three hazards for $N = 9$ and different initial condition n_0. The expected time until hitting a barrier when $n_0 = 5$ and $N = 9$ is 32 periods.

To apply these results to the model described in Section 8, it is enough to appropriately choose the barriers $\underline{x}$, $\bar{x}$ and thus $N = (\bar{x} - \underline{x})/\Delta + 1$, and initial point n_0. For the unconditional hazard of changing plans after a price increase, we choose $\underline{x} = \underline{\underline{p}}$, $\bar{x} = \bar{\bar{p}}$, and $n_0 = (\bar{p} - \underline{\underline{p}})/(\bar{\bar{p}} - \underline{\underline{p}})$ since we know the initial condition for the process is $\bar{p}$.

If, starting with a price increase, we want to study the conditional hazard of changing plans conditional on not changing the price within a plan, then $\bar{x} = \bar{\bar{p}}$, $\underline{x} = \underline{p}$ and the initial point is $\bar{p}$ which corresponds to $n_0 = (\bar{p} - \underline{p})/(\bar{\bar{p}} - \underline{p})$.

C.2 Constructing Price Plans Data

We describe in detail how we construct spells for price plans.

In the Online Micro Price data, where we observe posted prices, construction is straightforward. We take any three consecutive prices of a product p_1, p_2, p_3. If $p_1 = p_3$, then prices $\{p_1, p_2\}$ constitute a price plan and we keep it. We measure duration of the spell which started with p_3 and consider the next price p_4. If $p_4 = p_2$, the spell ends with a price change within the plan, if $p_4 \neq p_2$, the spell ends with a price outside the plan. If p_3 is right censored, we keep the plan and treat the spell as right-censored.

In the IRI data, one-week spells might be spurious and hence we want to avoid them.

We proceed according to this algorithm:

1. We start from a price p_1 observed for at least 2 periods, say $t_1, \dots t_2 - 1$. This is potentially one element of the price plan.

2. Let $p_2 = p(t_2 + 1)$ (not $p(t_2)$). This price lasts from $t_2, ..., t_3 - 1$ or $t_2 + 1, ...t_3 - 1$.

3. If $p_2 = p(t_2)$ or $p_2 = p(t_2 + 2)$, i.e. price lasts at least two periods, we have another element of the price plan. Otherwise stop.

4. Let $p_3 = p(t_3 + 1)$. This price lasts from $t_3, ..., t_4 - 1$ or $t_3 + 1, ..., t_4 - 1$.

5. If $p_1 = p_3$, we have a price plan $\{p_1, p_2\}$. We do not require p_3 to last as least two periods.[21] Otherwise stop.

6. Let $p_4 = p(t_4 + 1)$.

7. If $p_4 = p(t_4)$ or $p_4 = p(t_4 + 2)$, that is, price lasts at least two periods, we look at whether $p_4 = p_2$. If so, this is a within-plan price change. If not, this is a between-plan price change.

8. If $p_4 \neq p(t_4)$ and $p_4 \neq p(t_4 + 2)$, we cannot establish the nature of the price change, and hence we drop the plan.

9. If p_3 is right-censored, we keep the plan.

We use these price plans to estimate the model with competing risks, where one risk represents within-plan price change, and another a between-plan price change.

We also use these price plan spells to construct data which we use to estimate unconditional hazard of changing the price between plans. In Online Micro Price, starting from any p_3 which is part of a plan, i.e. $p_1 = p_3$, we measure duration until the first time the price changes to a new plan. That is, we look at the elapsed time between switch to the price p_3 and a new price p' with $p' \neq p_1$, $p' \neq p_2$. This is the duration of the price plan. If we never observe such a price, the duration is right censored with duration given by the time elapsed between p_3 and the last time the product is in a dataset.

In the IRI data, we start from any p_3 which is part of the plan and following our notation above, lasted from $t_3, \dots t_4 - 1$. If duration of p_3 is right censored or the price $p_4 = p(t_4 + 1)$ is such that $p_2 \neq p_4$, we record duration of the spell starting with p_3. If $p_2 = p_4$ and p_4 lasted till $t_5 - 1$, we look at price $p_5 = p(t_5 + 1)$. If $p_5 = p(t_5)$ or $p_5 = p(t_5 + 2)$, that is, p_5 lasts

[21] This is because $p_3 = p_1$, where p_1 has already been established to be a potential element of a price plan.

	new plan (uncond.)	within plan	new plan (cond.)
starting with p_h			
number of pairs	5,069,174	2,595,883	2,747,753
number of products	988,593	532,911	701,286
J-statistic	8,257	3,174	2,767
starting with p_l			
number of pairs	2,107,146	1,734,287	565,776
number of products	433,304	345,617	178,390
J-statistic	7,090	2,738	2,295

Table 1: Descriptive statistics for price trends and price reversals, IRI data. For this table, we consider only pairs of spells which are used in estimation. The first row reports the number of pairs, the second row reports how many products have at least one such pair. The third row reports the J-statistic. The critical value for the J-statistic 1,207.

at least 2 periods, we compare p_3 and p_5. If $p_5 \neq p_3$, the duration of spell is given by time elapsed between t_3 and the end of price p_5. If $p_5 = p_3$, we repeat this step.

Table 1 shows the number of observations (pairs of products as well as products with $n \geq 2$) for six different data sets, the ones used to estimate the conditional and unconditional hazard of a new plan, as well as the hazard of a within-plan price change, starting from either the high or low price within the plan. We also show the J-statistic from a test of the overidentifying restrictions implied by the proportional hazard structure.

D Price Trends and Price Reversals

We analyze the hazard of adjusting the price allowing for different dynamics for sales and regular prices, without relying on any particular model of sales. To do this, we use the MPH model extended to allow observable spell-specific characteristic and competing risks, as introduced in Section 3.5 and characterized in Appendix A. In particular, we distinguish spells based on whether they started with a price increase or price decrease, and so we have two observable characteristics ($X = 2$ in Appendix A). Let χ^i_j denote the observable characteristic of the j^{th} spell of product i. For mnemonic convenience let $\chi^i_j =$"+" if the j^{th} spell of product i follows a price increase and $\chi^i_j =$"−" if it follows a price decrease. We also distinguish whether a spell ends with a price increase or decrease, and so we have two competing risks ($R = 2$ in Appendix A). We use ρ^i_j to denote the reason for why the j^{th} spell of product i ends, and let $\rho^i_j =$"+" if it ends with a price increase and $\rho^i_j =$"−" if it ends with a price decrease. Spells with $\chi^i_j = \rho^i_j$ represent price trends, while spells with $\chi^i_j \neq \rho^i_j$

	b^{++}	b^{+-}	b^{-+}	b^{--}
number of pairs	42,077,438	275,322,989	72,940,882	17,876,354
number of products	4,313,488	9,416,198	5,637,378	2,911,651
J-statistic	3,920	8,737	7,910	3,401

Table 2: Descriptive statistics for price trends and price reversals, IRI data. For this table, we consider only pairs of spells which are used in estimation. For b^{+-}, the first row reports the number of pairs (j,k) such that $\underline{T} \leq \zeta_j \leq \bar{T}$, $\underline{T} \leq \zeta_k$, $\chi_j = \chi_k =$"+", $\rho_j =$"−". The second row reports how many products have at least one such pair. Columns for b^{++}, b^{-+} and b^{--} are analogous. The third row shows the J-statistic for the estimated model. The critical value for the J-statistic 1,749.

are price reversals.

We separately estimate four different baseline hazards, one for each possible combination of observable characteristic and risk. We use b_t^{++} (b_t^{+-}) to denote the baseline hazard that a spell which starts with a price increase subsequently ends with a price increase (decrease) at duration t. Similarly, b_t^{-+} (b_t^{--}) denotes the baseline hazard that a spell which starts with a price decrease subsequently ends with a price increase (decrease) at duration t. We do not restrict the unobserved type vector for the different observables and risks, and so some products may have a relatively high risk of certain outcomes and a relatively low risk for other outcomes, for example. Additionally, the validity of the proportional hazard assumption for one combination of observable characteristics and risks does not impact the consistency of the estimates for other combinations.

This richer model allows for the possibility that price trends have different dynamics than price reversals. We estimate the baseline hazard using the moment conditions specified in Proposition 5 in Appendix A, with the level of the baseline hazard at duration $\underline{T}$ set equal to the mixture hazard for that observable-risk pair from equation (O.21) in Online Appendix E.2. Table 2 shows descriptive statistics of the data used, while Figure 9 shows the estimated baseline hazards.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 reversals are declining, with b_t^{-+}, i.e. temporary sales, showing the sharpest decline. We conclude from these findings that the shape of the baseline hazard we recovered in Figure 1 is primarily driven by price reversals, especially those associated with sales. Price reversals are common in the data: the hazard for a price reversal is higher than the hazard of a price trend at all durations, regardless of whether the spell started with a price increase or a price decrease.

The model is over-identified and so we can again apply the J-test. We run a separate J-test for each hazard, since each baseline hazard can be estimated without assuming a MPH

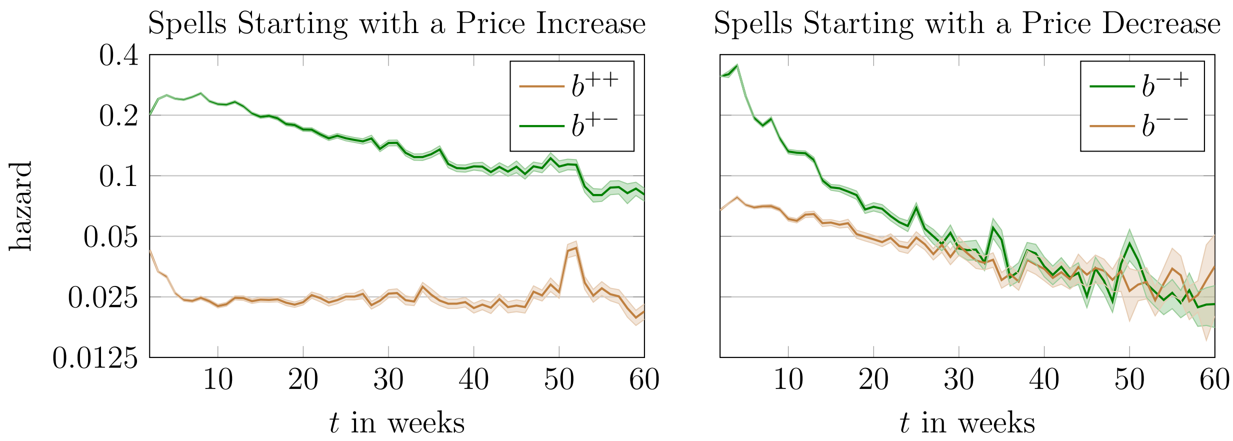

Figure 9: Baseline hazards in the competing risks model for pooled IRI data, log scale. b_t^{++} is the baseline hazard for spell which begin and end with a price increase; b_t^{--} for spells which begin and end with a price decrease; b_t^{+-} for spells which begin with a price increase and end with 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. The baseline hazard at duration $\underline{T} = 2$ is set to be equal to the corresponding mixture hazard at duration $\underline{T}$.

structure for the other competing hazards. The five percent critical value is 1,749 for each observable and 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.

Figures 16 and 17 in Online Appendix J 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 because of the smaller sample. 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 the proportional hazard assumption more systematically in Online Appendix H.4.

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 findings, we believe distinguishing price trends from price reversals is useful in this data. The baseline hazard for two consecutive price increases is decreasing until week 6, which covers a substantial amount of price changes: 76.8 percent 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 percent last at most 6 weeks). During first 6 weeks, the baseline hazard drops by almost 50 percent and then is then flat until week 45. 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.

E Mixture Hazard

We provide proof of Proposition 4 and show how we compute mixture hazard at $\underline{T}$ in the model with competing risks.

E.1 Mixture Hazard Proofs

Proof of Proposition 4. Take an entity i with measured durations $\zeta^i = (\zeta_0^i, \zeta_1^i, \ldots, \zeta_{K^i}^i)$ and hence censoring time $c^i = \sum_{j=0}^{K^i} \zeta_j^i - 1$. Let $z^i = \{z_1^i, \ldots, z_{c^i}^i\}$ 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 entity, with zeros in any period when a new spell does not start.

We first claim 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 Φ^i. Here we use the assumption that initial duration is drawn from the stationary distribution. This implies that with probability $\omega_1(\Phi^i)$, a spell ends 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 spell does not end, the probability that the measured duration of the spell is at least t is given by the type-specific survival function Φ_{t-1}^i. This uses 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 $s \in \{1, \dots, c^i - \bar{T}\}$, the expected value of $\mathbb{1}_{z^i_s \geq t}$ conditional on c^i and Φ^i is $\Phi^i_{t-1}\omega_1(\Phi^i)$. It follows that

$$\mathbb{E}\left[\sum_{j=1}^{K^i} \mathbb{1}_{\zeta^i_j \geq t, c^i_j \geq \bar{T}} \middle| c^i, \Phi^i\right] = \mathbb{E}\left[\sum_{s=1}^{c^i - \bar{T}} \mathbb{1}_{z^i_s \geq t} \middle| c^i, \Phi^i\right]$$
$$= \begin{cases} (c^i - \bar{T})\Phi^i_{t-1}\omega_1(\Phi^i) & \text{if } c^i > \bar{T} \\ 0 & \text{if } c^i \leq \bar{T}. \end{cases}$$

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

$$\mathbb{E}\left[\frac{c}{c - \bar{T}} \sum_{j=1}^{K} \mathbb{1}_{\zeta_j \geq t, c_j \geq \bar{T}} \middle| \Phi^i\right] = \left(\sum_{c=\bar{T}+1}^{\infty} c q_c(\Phi^i)\right) \Phi^i_{t-1}\omega_1(\Phi^i).$$

Finally, integrating across Φ using the distribution F, we get

$$\mathbb{E}\left[\frac{c}{c - \bar{T}} \sum_{j=1}^{K} \mathbb{1}_{\zeta_j \geq t, c_j \geq \bar{T}}\right] = \int \left(\sum_{c=\bar{T}+1}^{\infty} c q_c(\Phi)\right) \Phi_{t-1}\omega_1(\Phi) dF(\Phi) \tag{O.19}$$

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

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

$$\mathbb{E}\left[\frac{c}{c - \bar{T}} \sum_{j=1}^{K} \mathbb{1}_{\zeta_j \geq t+1, c_j \geq \bar{T}}\right] = \int \left(\sum_{c=\bar{T}+1}^{\infty} c q_c(\Phi)\right) \Phi_t\omega_1(\Phi) dF(\Phi).$$

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

$$\mathbb{E}\left[\frac{c}{c - \bar{T}} \sum_{j=1}^{K} \mathbb{1}_{\zeta_j = t, c_j \geq \bar{T}}\right] = \int \left(\sum_{c=\bar{T}+1}^{\infty} c q_c(\Phi)\right) (\Phi_{t-1} - \Phi_t)\omega_1(\Phi) dF(\Phi). \tag{O.20}$$

Then using equations (12), (O.19) and (O.20), it follows immediately that $\mathbb{E}\left[f^{[H]}_{t,\bar{T}}(\zeta; H)\right] = 0$ for $t = 1, \dots, \bar{T}$ if and only if $H_t = H^f_t$ defined in equation (11), proving the first result.

Now assume Φ and c are independent, so $q_c(\Phi) = q_c$ for all c and Φ. Then a similar logic to the derivation of equation (O.19) implies

$$\mathbb{E}\left[\sum_{j=1}^{K}\mathbb{1}_{\zeta_j\geq t, c_j\geq \bar{T}}\right]=\left(\sum_{c=\bar{T}+1}^{\infty}(c-\bar{T})q_c\right)\int\Phi_{t-1}\omega_1(\Phi)dF(\Phi).$$

And a similar logic to the derivation of equation (O.20) implies

$$\mathbb{E}\left[\sum_{j=1}^{K}\mathbb{1}_{\zeta_j=t, c_j\geq \bar{T}}\right]=\left(\sum_{c=\bar{T}+1}^{\infty}(c-\bar{T})q_c\right)\int(\Phi_{t-1}-\Phi_t)\omega_1(\Phi)dF(\Phi).$$

Combining these with equation (13), we get that $\mathbb{E}\left[f_{t,\bar{T}}^{[HI]}(\zeta;H)\right]=0$ for $t=1,\ldots,\bar{T}$ if and only if

$$H_t=\frac{\int(\Phi_{t-1}-\Phi_t)\omega_1(\Phi)dF(\Phi)}{\int\Phi_{t-1}\omega_1(\Phi)dF(\Phi)}.$$

Since $q_c(\Phi)=q_c$ implies $\bar{c}(\Phi)=\bar{c}$, this is equal to H_t^* in equation (10), proving the second result. ■

E.2 Mixture Hazard with Competing Risks and Observables

In several applications, it is useful to estimate the mixture hazard associated with risk r and observable characteristic x at duration $\underline{T}$, $H_{\underline{T}}^{x,r}$. We use the following formula

$$0=\sum_{i=1}^{I}\sum_{j=1}^{K}\left(\hat{H}_{\underline{T}}^{x,r}\mathbb{1}_{\zeta_j^i\geq t,\chi_j^i=x}-\mathbb{1}_{\zeta_j^i=t,\rho_j^i=r,\chi_j^i=x}\right), \tag{O.21}$$

where $i=1,\ldots I$ indexes the products. In words, the mixture hazard at duration $\underline{T}$ associated with risk r and observable characteristic x is given by the ratio of the spells with characteristic x which ended at duration $\underline{T}$ due to risk r, and the number of all spells with characteristic x lasting at least $\underline{T}$. This is analogous to the formula (12) for the mixture hazard when the type and censoring time are independent, with the exception that we select spells which ended due to risk r. Note that since we are interested in estimating mixture hazard at just one duration, $\underline{T}$, there is not need to condition on the censoring time exceeding a certain number.

E.3 Moments of the Frailty Distribution

We use insights from Section E.1 to prove that the moment conditions for the first three moments of the frailty distribution, which we use in Section B.3, have the expected value of zero.

Consider first equation (O.14). To see that $\mathbb{E}\left[f_{\bar{T}}^{[\mu_1]}(\zeta;\mu,b)\right]=0$, we use equations (O.19) and (O.20). We use equation (O.19) and set $t=\underline{T}$ to find an expression for the expected value of the second term in $f_{\bar{T}}^{[\mu_1]}$:

$$\begin{aligned}\mathbb{E}\left[\frac{c}{c-\bar{T}}\sum_{j=1}^{K}\mathbb{1}_{\zeta_j\geq\underline{T},c_j\geq\bar{T}}\right]&=\int\left(\sum_{c=\bar{T}+1}^{\infty}cq_c(\Phi)\right)\Phi_{\underline{T}-1}\omega_1(\Phi)dF(\Phi)\\&=\int w_{\underline{T}}^{f}(\Phi)dF(\Phi)\end{aligned}$$

where we used that $\Phi_{\underline{T}-1}=1$ because $b_t=0$ for $t<\underline{T}$.

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

$$\begin{aligned}\mathbb{E}\left[\frac{c}{c-\bar{T}}\sum_{j=1}^{K}\mathbb{1}_{\zeta_j=\underline{T},c_j\geq\bar{T}}\right]&=\int\left(\sum_{c=\bar{T}+1}^{\infty}cq_c(\Phi)\right)(\Phi_{\underline{T}-1}-\Phi_{\underline{T}})\omega_1(\Phi)dF(\Phi)\\&=\int\left(\sum_{c=\bar{T}+1}^{\infty}cq_c(\Phi)\right)\theta(\Phi)b_{\underline{T}}\Phi_{\underline{T}-1}\omega_1(\Phi)dF(\Phi)\\&=b_{\underline{T}}\int\theta(\Phi)w_{\underline{T}}^{f}(\Phi)dF(\Phi),\end{aligned}$$

where $\theta(\Phi)$ is the type, as defined in Section 4.5.

Since

$$\mu_1=\int\theta(\Phi)dF_{\underline{T}}^{f}=\frac{\int\theta(\Phi)w_{\underline{T}}^{f}(\Phi)dF(\Phi)}{\int w_{\underline{T}}^{f}(\Phi)dF(\Phi)},$$

the result follows. The proof that $\mathbb{E}\left[f_{\bar{T}}^{[\mu_2]}(\zeta;\mu,b)\right]=0$ in equation (O.15) and $\mathbb{E}\left[f_{\bar{T}}^{[\mu_3]}(\zeta;\mu,b)\right]=0$ in equation (O.16) is similar and so we omit it.

F GMM Estimation

We provide additional details on the GMM estimation.

F.1 GMM Estimator

We start with the GMM estimator for the baseline hazard. Proposition 2 gives us one moment condition for the choice t_1, t_2 such that $\underline{T} \le t_1 < t_2 \le \bar{T}$:

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

Let $Y(\underline{T}, \bar{T}) = \{(t_1, t_2) : \underline{T} \le t_1 < t_2 \le \bar{T}\}$. Denoting $T = \bar{T} - \underline{T}$, 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 $f^{[b]}(\zeta; b)$ be a vector function with m^{th} element corresponding to the choice $y_m \in Y(\underline{T}, \bar{T})$, given by $f^{[b]}_{y_{m_1}, y_{m_2}}(\zeta; b)$.

Since the baseline hazard is identified up to scale, we choose one normalization. We choose $T_0 \in \{\underline{T}, \ldots \bar{T}\}$ to be the shortest duration such that there exists a product i with at least two spells and measured complete duration of one of its spells equal to T_0, $K^i \ge 2$, and $1 \le j < k \le K^i$ such that $\zeta^i_j = T_0, \zeta^i_k = t$ for any $t \in \{T_0, \ldots \bar{T}\}$.[22] Without loss of generality, we normalize $b_{T_0} = 1$.

Let $b_{./T_0}$ be the vector b without its component b_{T_0}, that is, $b_{./T_0} = (b_{\underline{T}}, \ldots b_{T_0 - 1}, b_{T_0+1}, \ldots b_{\bar{T}})$. Linearity of $f^{[b]}_{t_1,t_2}(\zeta; b)$ and normalization of b_{T_0} implies that we can write

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

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

$$\mathbb{E}\left[U^{[b]}(\zeta)\right] b_{./T_0} - \mathbb{E}\left[V^{[b]}(\zeta)\right] = 0. \tag{O.22}$$

We now discuss GMM estimator for the mixture hazard. Proposition 4 gives us one moment condition for each $\underline{T} \le t \le \bar{T}$. Define $f^{[H]}_{\bar{T}}$ as a vector function, with m^{th} element given by $f^{[H]}_{m+\underline{T}-1,\bar{T}}(\zeta; H^{\bar{T}})$ for $m = 1, \ldots, T+1$. Since equation (12) is linear in $H^{\bar{T}}$, we can write $f^{[H]}_{m+\underline{T}-1,\bar{T}}(\zeta; H^{\bar{T}}) = U^{[H]} H^{\bar{T}} - V^{[H]}$, 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

$$\mathbb{E}\left[U^{[H]}(\zeta)\right] H^{\bar{T}} - \mathbb{E}\left[V^{[H]}(\zeta)\right] = 0. \tag{O.23}$$

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

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

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

Then the moment conditions are

$$\mathbb{E}\left[U(\zeta)\right]\beta - \mathbb{E}\left[V(\zeta)\right] = 0. \tag{O.24}$$

To estimate the model, we replace expected values with sample means. In particular, indexing the products with $i = 1, \dots I$, we have

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

The sample analog of (O.24) is $U_I\beta - V_I = 0$. For a given positive-definite $(M+T+1)\times(M+T+1)$ weighting matrix W, the estimator $\hat{\beta} \in \mathbb{R}_+^{2T+1}$ solves

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

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

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

In practice, we use the identity matrix as a weighting matrix to estimate the parameters. To construct the J-statistic, we estimate the variance-covariance matrix as discussed in the next section.

Results in Newey and McFadden (1994) imply that this GMM estimator using the moment conditions in Proposition 2 and 4 is consistent without any additional assumptions.[23] In particular, we do not need to impose that the space of possible parameters β is compact since our estimator is linear.

F.2 Clustered Standard Errors and J-test

Recall that the GMM formula for the variance-covariance matrix of the parameter vector β is

$$VAR \equiv \frac{1}{I}(F'WF)^{-1}F'W\Omega W'F(F'W'F)^{-1}, \tag{O.25}$$

[23]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.

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

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

where $\hat{\beta}$ is a GMM estimate of β.

To implement one-way clustering, we follow Cameron, Gelbach, and Miller (2011). Formula (O.25) 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 $\bar{f}_q$ as the sum of the moment conditions across products in cluster q,

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

Then

$$\Omega_I^{[cluster]} = \frac{\mathcal{Q}}{\mathcal{Q}-1} \frac{I-1}{I-(2T+1)} \frac{1}{I} \sum_{q=1}^{\mathcal{Q}} \bar{f}_q \bar{f}_q', \tag{O.26}$$

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 (O.25).

Finally, we use the variance-covariance matrix to compute the J-statistic

$$J = I \left(\frac{1}{I} \sum_{i=1}^{I} f(\zeta^i; \widehat{\beta}) \right)' \Omega_I^{[cluster]^{-1}} \left(\frac{1}{I} \sum_{i=1}^{I} f(\zeta^i; \widehat{\beta}) \right), \tag{O.27}$$

which, under the null hypothesis that the model is correctly specified, is distributed χ^2_{M-T}.

F.3 Practical Consideration

It is 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 eigenvalues which are either negative or 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, have 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.

G Proofs for the Price Plans Model

It is useful to state the following two lemmas.

Lemma 2 *For all integers $N \geq 1$ and numbers $n, n' \in \{1, 2, \ldots, N\}$,*

$$\sum_{j=1}^{N} \sin\left(\frac{n}{N+1} j\pi\right) \sin\left(\frac{n'}{N+1} j\pi\right) = \begin{cases} \frac{N+1}{2} & \text{if } n = n' \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 3 *For integers $N \geq 2$ and $1 \leq k \leq N$,*

$$\sum_{n=1}^{N} \sin\left(\frac{n}{N+1} k\pi\right) = \begin{cases} 0 & \text{if } k \text{ is even} \\ \cot\left(\frac{k\pi}{2(N+1)}\right) & \text{if } k \text{ is odd.} \end{cases}$$

The proofs are in Supplemental Appendix L.

Proof of Proposition 6. We proceed in several steps. First, we can easily verify that a function of the form $\lambda^t \varphi(x)$ with $\varphi(x) = \sin(\frac{x}{\alpha} + \beta)$, for some constants α, β solves equation (O.17) in the interior if $\lambda = \frac{1}{2}\left(1 + \cos(\frac{\Delta}{\alpha})\right)$ by just plugging it into equation (O.17) and simplifying using $\sin(\frac{x \pm \Delta}{\alpha} + \beta) = \sin(\frac{x}{\alpha} + \beta)\cos(\frac{\Delta}{\alpha}) \pm \cos(\frac{x}{\alpha} + \beta)\sin(\frac{\Delta}{\alpha})$. Second, we impose that the solution satisfies the boundary conditions, which gives us two conditions

$$\sin\left(\frac{\bar{x}}{\alpha} + \beta + \frac{\Delta}{\alpha}\right) = 0, \qquad \sin\left(\frac{\underline{x}}{\alpha} + \beta - \frac{\Delta}{\alpha}\right) = 0.$$

These conditions hold for any integer $j = 1, 2, \dots$ for which

$$\frac{\bar{x}}{\alpha} + \beta + \frac{\Delta}{\alpha} = j\pi, \qquad \frac{\underline{x}}{\alpha} + \beta - \frac{\Delta}{\alpha} = 0.$$

For given j, we thus obtain (α_j, β_j) as a solution this system, and plug it into $\varphi(x)$ using $x_n = \underline{x} + (n-1)\Delta$ and $\bar{x} = \underline{x} + (N-1)\Delta$ to obtain $\varphi_j(x_n) = \sin(\frac{n}{N+1} j\pi)$. Next, we plug the solution for α into the expression for λ to obtain $\lambda_j = \frac{1}{2}\left(1 + \cos(\frac{1}{N+1} j\pi)\right)$.

We note that for each n we can regard $\varphi_n = \{\varphi_j(x_n)\}_{j=1}^{N}$ as an N dimensional vector. Lemma 2 implies that the set of vectors $\{\varphi_n\}$ for $n = 1, 2, \dots, N$ are orthogonal with each other. Since the law of motion is linear, the solution $m_t(x)$ can be expressed as a linear combination of these solutions, with weights γ_j. We choose γ_j such that $m_t(x)$ satisfies the initial condition $m_0(x)$. Again using Lemma 2, this means

$$\gamma_j = \frac{\langle \varphi_j, m_0 \rangle}{\langle \varphi_j, \varphi_j \rangle} = \frac{2}{N+1} \sum_{n=1}^{N} \sin\left(\frac{n}{N+1} j\pi\right) m_0(x_n) = \frac{2}{N+1} \sin\left(\frac{n_0}{N+1} j\pi\right).$$

■

Proof of Proposition 7. Since $m_0(\bar{x}) = 0$ if $n_0 < N$ and $m_0(\underline{x}) = 0$ if $n > 1$, we obtain immediately from its definition that $h_0 = 0$ if $1 < n_0 < N$.

We next analyze the limit of h_t. It is useful to write h_t using the survival function as

$$h_t = \frac{S_t - S_{t+1}}{S_t} = \sum_{j=1}^{N} (1 - \lambda_j) w_{j,t}$$
$$w_{j,t} = \frac{\lambda_j^t \gamma_j \sum_{n=1}^{N} \sin\left(\frac{n}{N+1} j\pi\right)}{\sum_{k=1}^{N} \lambda_k^t \gamma_k \sum_{n=1}^{N} \sin\left(\frac{n}{N+1} k\pi\right)}.$$

Since $1 > \lambda_1 > \lambda_2 > \dots \lambda_N > 0$, it follows that $\lim_{t\to\infty} w_{1,t} = 1$ while $\lim_{t\to\infty} w_{j,t} = 0$ for $j > 1$, and hence $\lim_{t\to\infty} h_t = 1 - \lambda_1$.

We next analyze the shape of h_t for large t. We first observe that $w_{1,t} > 0$ for all t, and, following Lemma 3, $w_{j,t} = 0$ for j even. This implies that the behavior of h_t for large t is governed by λ_1 and λ_3, with $h_t \approx (1 - \lambda_1) w_{1,t} + (1 - \lambda_3)(1 - w_{1,t})$. We are interested in whether this is bigger or smaller than the asymptotic hazard $1 - \lambda_1$.

To answer this, we look at $w_{1,t}$, which for large t is also determined by λ_1 and λ_3:

$$
\begin{aligned}
w_{1,t} &\approx \frac{\lambda_1^t \gamma_1 \sum_{n=1}^N \sin\left(\frac{n}{N+1}\pi\right)}{\lambda_1^t \gamma_1 \sum_{n=1}^N \sin\left(\frac{n}{N+1}\pi\right) + \lambda_3^t \gamma_3 \sum_{n=1}^N \sin\left(\frac{n}{N+1}3\pi\right)} \\
&= \frac{\lambda_1^t \gamma_1 \cot\left(\frac{\pi}{2(N+1)}\right)}{\lambda_1^t \gamma_1 \cot\left(\frac{\pi}{2(N+1)}\right) + \lambda_3^t \gamma_3 \cot\left(\frac{3\pi}{2(N+1)}\right)},
\end{aligned}
$$

where the equation uses Lemma 3. Note that for $N \geq 3$, $\cot\left(\frac{\pi}{2(N+1)}\right) > \cot\left(\frac{3\pi}{2(N+1)}\right) > 0$. Additionally, we know from Proposition 6 that $\gamma_1 = \frac{2}{N+1} \sin\left(\frac{n_0}{N+1}\pi\right)$, which is positive for all $n_0 \in \{1, 2, \ldots, N\}$. Thus the asymptotic behavior of $w_{1,t}$ depends on $\gamma_3 = \frac{2}{N+1} \sin\left(\frac{3n_0}{N+1}\pi\right)$:

- if $n_0 < \frac{N+1}{3}$ or $n_0 > \frac{2(N+1)}{3}$, $\gamma_3 > 0$, and so $w_{1,t} < 1$ for all large t. Since $h_t \approx (1-\lambda_1)w_{1,t} + (1-\lambda_3)(1-w_{1,t})$ for large t, it follows that $h_t > 1-\lambda_1$, and so the hazard converges to its asymptote from above.

- if $\frac{N+1}{3} < n_0 < \frac{2(N+1)}{3}$, $\gamma_3 < 0$, and so $w_{1,t} > 1$ for all large t. Again, $h_t \approx (1-\lambda_1)w_{1,t} + (1-\lambda_3)(1-w_{1,t})$ implies $h_t < 1-\lambda_1$ for large t, and so the hazard converges to its asymptote from below.

Finally, we can have $n_0 = \frac{N+1}{3}$ or $n_0 = \frac{2(N+1)}{3}$, in which case $\gamma_3 = w_{3,t} = 0$ for all t. Note that since n_0 is an integer, these cases require $N \geq 5$. In this case, we have $\cot\left(\frac{5\pi}{2(N+1)}\right) > 0$ as well, and so the asymptotic slope of $w_{1,t}$ depends on the sign of γ_5. It is straightforward to verify that $\gamma_5 < 0$ whenever $\gamma_3 = 0$, and so again $w_{1,t} > 1$. Now $h_t \approx (1-\lambda_1)w_{1,t} + (1-\lambda_5)(1-w_{1,t})$ implies $h_t < 1-\lambda_1$ for large t, which means that in this borderline case, the hazard is again converging to its asymptote from below.

Finally, denote $T(x_n)$ the expected time to hitting either barrier if the current state is x_n. Then for $x_n \in \{\underline{x} + \Delta, \ldots, \bar{x} - \Delta\}$ it holds

$$
T(x_n) = 1 + \frac{1}{4}T(x_{n+1}) + \frac{1}{4}T(x_{n-1}) + \frac{1}{2}T(x_n),
$$

with $T(\underline{x}) = 1 + \frac{1}{4}T(\underline{x}+\Delta) + \frac{1}{2}T(\underline{x})$ and $T(\bar{x}) = 1 + \frac{1}{4}T(\bar{x}-\Delta) + \frac{1}{2}T(\bar{x})$. This is a second-order difference equation and it can be verified that its solution is given by $T(x_n) = 2n(N+1-n)$. ■

Proof of Proposition 8. We focus on analyzing $h_t^{\underline{x}}$. The analysis of $h_t^{\bar{x}}$ is symmetric. Using the derived formulae, we can write

$$
h_t^{\underline{x}} = \frac{1}{4}\frac{m_t(\underline{x})}{S_t} = \frac{1}{4}\frac{\sum_{k=1}^N \lambda_k^t \gamma_k \sin\left(\frac{1}{N+1}k\pi\right)}{\sum_{n=1}^N \sum_{k=1}^N \lambda_k^t \gamma_k \sin\left(\frac{n}{N+1}k\pi\right)} = \frac{1}{4}\frac{\sum_{k=1}^N \lambda_k^t y_k}{\sum_{k=1}^N \lambda_k^t z_k}
$$

where weights y_k and z_k are defined as

$$y_k \equiv \gamma_k \sin\left(\frac{1}{N+1}k\pi\right), \qquad z_k \equiv \gamma_k \sum_{n=1}^{N} \sin\left(\frac{n}{N+1}k\pi\right).$$

Since $1 > \lambda_1 > \lambda_2 > \dots \lambda_N > 0$, it follows that $\lim_{t\to\infty} h_t^x = \frac{1}{4}\frac{y_1}{z_1}$. Moreover, using Lemma 3, we get $z_1 = \gamma_1 \cot(\pi/(2(N+1)))$, which implies $\frac{1}{4}\frac{y_1}{z_1} = \frac{1}{2}(1-\lambda_1)$. Thus half the asymptotic hazard comes at the lower bound.

At large t, h_t^x is shaped by the largest roots. Note that Lemma 3 implies that $z_k = 0$ for k even, and so we can write

$$h_t^x \approx \frac{1}{4}\frac{\lambda_1^t y_1 + \lambda_2^t y_2}{\lambda_1^t z_1} = \frac{1}{4}\left(\frac{y_1}{z_1} + \frac{\lambda_2^t}{\lambda_1^t}\frac{y_2}{z_1}\right).$$

Observe that $z_1 = \gamma_1 \cot(\pi/(2(N+1))) > 0$. Hence, if $y_2 < 0$, then h_t^x converges to its asymptote from below. Since by definition $y_2 = \gamma_2 \sin\left(\frac{2\pi}{N+1}\right)$, $y_2 < 0$ if and only if $\gamma_2 \equiv \frac{2}{N+1}\sin\left(\frac{n_0}{N+1}2\pi\right) < 0$. Equivalently, if $n_0 > \frac{N+1}{2}$, the hazard converges to its limit from below. Conversely, if that $n_0 < \frac{N+1}{2}$, the hazard h_t^x converges to its limit from above.

Finally, it remains to analyze the case of $n_0 = \frac{N+1}{2}$, which implies $y_2 = 0$. For large t,

$$h_t^x \approx \frac{1}{4}\frac{\lambda_1^t y_1 + \lambda_3^t y_3}{\lambda_1^t z_1 + \lambda_3^t z_3}.$$

Some algebra yields that, for $z_1 > 0$, $z_3 < 0$, $z_1 + z_3 > 0$,

$$\frac{\lambda_1^t y_1 + \lambda_3^t y_3}{\lambda_1^t z_1 + \lambda_3^t z_3} < \frac{y_1}{z_1} \Leftrightarrow \frac{y_1}{z_1} < \frac{y_3}{z_3},$$

and so h_t^x converges to its limit $\frac{1}{4}\frac{y_1}{z_1}$ from below.

We next verify that these conditions hold in our case. For $n_0 = \frac{N+1}{2}$, we have $\gamma_1 =$

$2/(N+1)$ and $\gamma_3 = -2/(N+1)$. Using Lemma 3, we have

$$z_1 = \frac{2}{N+1} \cot\left(\frac{\pi}{2(N+1)}\right) > 0$$
$$z_3 = -\frac{2}{N+1} \cot\left(\frac{3\pi}{2(N+1)}\right) < 0$$
$$z_1 + z_3 = \frac{2}{N+1} \left(\cot\left(\frac{\pi}{2(N+1)}\right) - \cot\left(\frac{3\pi}{2(N+1)}\right)\right) > 0$$
$$\frac{y_1}{z_1} = \frac{\sin\left(\frac{\pi}{(N+1)}\right)}{\cot\left(\frac{\pi}{2(N+1)}\right)} = \frac{\sin\left(\frac{\pi}{(N+1)}\right)\sin\left(\frac{\pi}{2(N+1)}\right)}{\cos\left(\frac{\pi}{2(N+1)}\right)} = 2\sin^2\left(\frac{\pi}{2(N+1)}\right).$$

where in the last row we used that $\sin(2x) = 2\sin(x)\cos(x)$. Similar steps lead to

$$\frac{y_3}{z_3} = 2\sin^2\left(\frac{3\pi}{2(N+1)}\right),$$

and hence indeed

$$\frac{y_1}{z_1} = 2\sin^2\left(\frac{\pi}{2(N+1)}\right) < 2\sin^2\left(\frac{3\pi}{2(N+1)}\right) = \frac{y_3}{z_3}.$$

■

H Additional Empirical Results

H.1 IRI Data by Category

We provide summary statistics for the IRI sample used in estimation. Table 3 summarizes the number of price spells by product category in the IRI data.

H.2 Test of Ergodic Distribution of In-Progress Duration

To estimate the mixture hazard, we assume that when we first observe a product, the duration of the in-progress spell is a random draw from the stationary duration distribution for that product. This assumption has a testable implication: conditional on censoring time, the share of products changing its price in any week should be 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		number of	percentiles of ζ_j^i	
	≥ 1 spell	≥ 2 spells	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
Total	21,717,549	13,898,768	684,919,778	3	10

Table 3: Descriptive statistics by product category, IRI data. For this table, we consider only spells $j \geq 1$ with $\zeta_j^i \geq \underline{T} = 2$. 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.

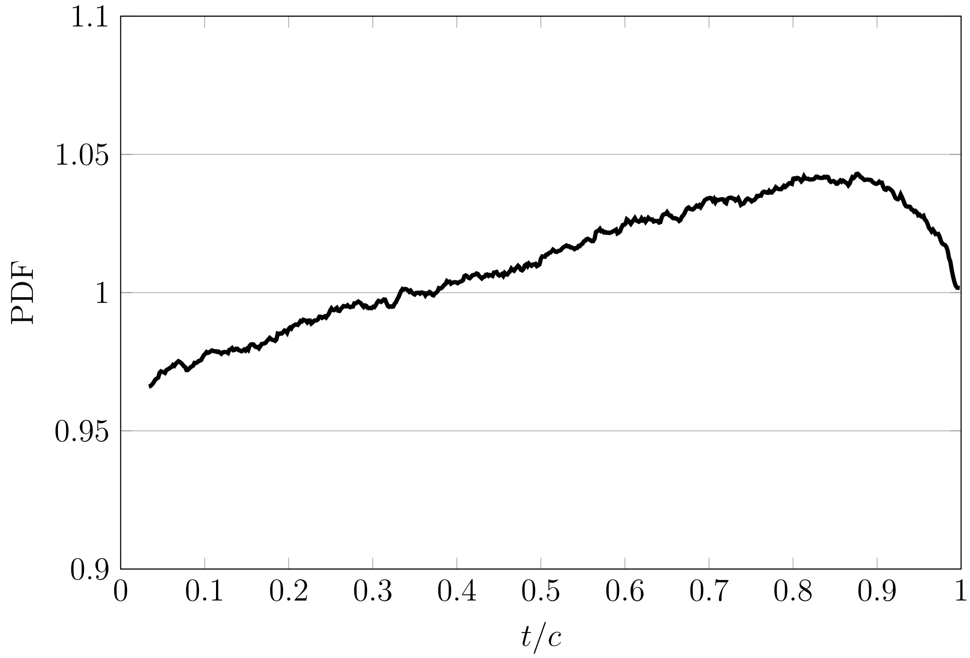

Figure 10: Empirical density of times when products change prices, measured from the start of an in-progress spell, pooled IRI data.

number of products with the corresponding value of c as weights. Figure 10 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.

H.3 Correlation between Spells

We compute correlation between two spells within narrowly defined categories to understand the extent of unobserved heterogeneity in such categories. As in our main sample, we select the longest cycle for each product. We drop spells lasting one week, and then keep first two spells of each product. Out of these products, we keep only those with two completed spells shorter than 30 weeks, and censoring time (minus the sum of one-week spells) at least 60 weeks. This guarantees that we can observe two completed spells of less than 30 weeks for each selected product.

We then compute correlation between the two spells within narrowly defined cells of products, which is a cross between good category (30 categories) and a retail chain (163 chains). This gives us 4,890 potential cells. There are 2,041 cells with at least 100 products in it. We find that the median (mean) correlation within all cells is 0.13 (0.13).

H.4 Sensitivity of Results to the Choice of $\underline{T}$ and $\bar{T}$

The choice of $\underline{T}$ and $\bar{T}$ is guided by the nature of our data. We choose $\underline{T} = 2$ to exclude spurious price spells lasting 1 week from our analysis, as explained in the main text. The choice of $\bar{T}$ has 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. On the other hand, the number of spells longer than $\bar{T}$ decreases quickly with $\bar{T}$. Indeed, Table 3 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 mixture 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 11 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 $\bar{T} = 60$.

We next examine the sensitivity of our results to the choice of $\underline{T}$ and $\bar{T}$. This allows us to see if there is a systematic failure of the MPH assumption. The idea is the following. Suppose we want to learn about the relative baseline hazards at duration 10 and 20, b_{10}/b_{20}. The MPH model admits several ways of recovering the ratio. We can directly recover the ratio b_{10}/b_{20} from equation (4) 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}/b_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 11 shows the results for the benchmark 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.

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 panel of Figure 11 shows that the choice of $\bar{T}$ does not affect the estimates.

Finally, we offer one additional justification for our choice of $\underline{T}$. An implication of any

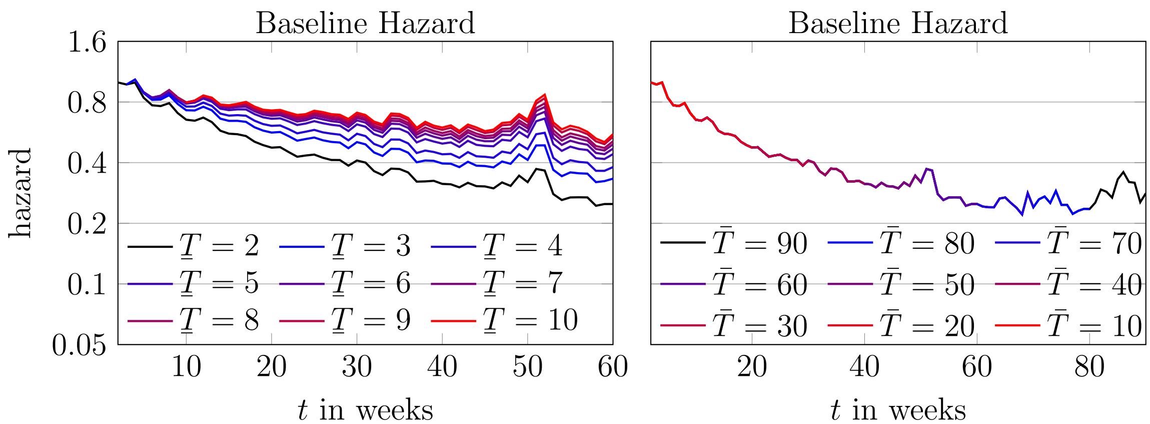

Figure 11: Baseline hazard for pooled IRI data, log scale, estimated using different values of $\underline{T} \in \{2, \ldots, 10\}$ and $\bar{T} = 60$ in the left panel, and using different values for $\bar{T} \in \{10, 20, \ldots 90\}$ and $\underline{T} = 2$ in the right panel.

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 correlation of the completed duration of two spells for a given product is non-negative, and strictly positive when there is heterogeneity in mean duration. To understand why, note that conditional on a product's type, the cross-spell correlation of duration is zero by assumption. But with heterogeneity, the correlation captures differences in the type-specific means and is generally positive.

Inspired by this, we measure the autocorrelation of the duration of price spells in the data. To avoid introducing bias due to censoring, we select products with at least 60 weeks of censoring time and first two completed spells shorter than 30 weeks. This guarantees that a product is in the dataset for long enough to us to observe two completed spells shorter than 30 weeks. If we include all spells, including one-week spells, we find a correlation of 0.0032 when duration is measured in levels, and -0.0321 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 0.176 in levels and 0.176 when measured in logs. That is, once we exclude one-week spells, the correlation is positive, which further justifies our decision to exclude one week spells.

H.5 Robustness to Iterated GMM

Our model is over-identified, with many more moment conditions than parameters. In such situations, iterated GMM might perform better than 1-step or 2-step GMM. We implement iterated GMM as follows. We estimate baseline hazard b using identity matrix as a weighting matrix in our GMM estimation and denote this estimate $\hat{b}_{(1)}$. Using $\hat{b}_{(1)}$, we estimate

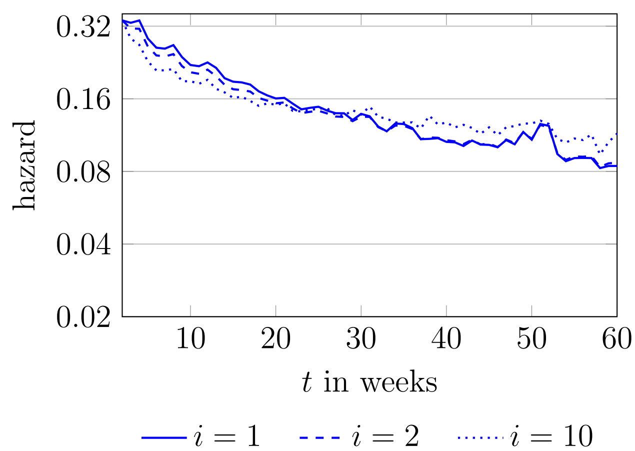

Figure 12: Robustness results for the baseline hazard for pooled IRI data, log scale. Solid line represents estimates after $i = 1$ iteration, dashed line after $i = 2$ iterations and dotted after $i = 10$ iterations. The baseline hazard is normalized to equal the mixture hazard at duration 2 weeks.

variance-covariance matrix Ω using (O.26), denoting it $\hat{\Omega}_{(1)}$. In the next step, we estimate b using $\hat{\Omega}_{(1)}^{-1}$ as a weighting matrix to obtain $\hat{b}_{(2)}$. That is, in the i^{th} iteration, we $\hat{\Omega}_{(i-1)}^{-1}$ to obtain $\hat{b}_{(i)}$, and then $\hat{b}_{(i)}$ to estimate $\hat{\Omega}_{(i)}$.

Figure 12 shows the results for the baseline hazard for the IRI data. We depict estimates for $i = 1$, $i = 2$ and $i = 10$. We conclude that the estimates are robust to the method we use.

I Maximum Likelihood Estimators

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 and taking a stance on whether the data are measured in continuous or discrete time. In this section, we investigate the role of frailty distribution and timing assumption on estimates of the baseline hazard.

We formulate the MPH model in continuous 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 durations are measured exactly in continuous time; we call this model Continuous Time with Continuous Measurement (CT-CM).

Initially, we assume that the frailty distribution is gamma with mean m and variance v, a convenient assumption as it allows us to integrate out the frailty distribution analytically and obtain a simpler expression for the likelihood function. We later relax this assumption by assuming that the frailty distribution is a mixture of several gamma distributions.

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.

Finally, we estimate the CT-CM model using all spells using Stata's built-in procedure streg, and compare it to the baseline hazard estimated using our GMM estimator.

I.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 up to the next integer values. We assume each product has a type θ drawn from a Gamma distribution with mean m and variance v, though we later consider an extension to the case where the frailty distribution is a mixture of Gamma distributions. A censoring time $c \in \mathbb{R}_+$ is drawn from a continuous distribution, with $\bar{Q}_c^1$ denoting the probability that censoring time following the first price change is at least $c-1$ periods. In contrast to our GMM estimates of the discrete time model, we impose that c and θ are independent random variables, and so $\bar{Q}_c^1$ does not depend on θ.

In the continuous time MPH model, the probability that a spell lasts 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 $t = 1, 2, \ldots$, 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 $t = 1, 2, \ldots$, define $b_t \equiv \int_{t-1}^t b(s)ds$. Additionally, for notational convenience, we continue to assume $b_0 = 0$. Our objective is to recover $b \equiv \{b_1, \ldots, b_{\bar{T}}, b_{\bar{T}+1}\}$, 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 $t = 1, 2, \ldots$ for a type θ product is $e^{-\theta z_{t-1}}$.

We formulate the likelihood function for the case where we observe two spells per product. For a typical product i, we observe $(c^i, d_1^i, d_2^i, \zeta_1^i, \zeta_2^i)$ where c^i is residual censoring time for the first spell after a left-censored spell, 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 $d_2^i = 1$.

Following the analysis of single spell data in Meyer (1990), we write 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

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

The integrand is equal to the probability that the censoring time exceeds t_1+t_2, 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, where we use Γ to denote the gamma function. We integrate this expression to get

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

where

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

We note the explicit dependence of this function on the integrated hazard $z = \{z_1, z_2, \dots\}$, 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

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

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

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

where

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

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

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

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

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

where

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

We can use the probability of these three events to compute the log-likelihood. We treat $\bar{Q}^1_c$ as a nuisance parameter and take advantage of the fact that each of the probabilities is multiplicatively separable in the terms involving $\bar{Q}^1_c$ to get

$$\mathcal{L}^{CT-DM} = \frac{1}{N} \sum_{i=1}^N \log f^{CT-DM}_{d_1^i+d_2^i}(\zeta_1^i, \zeta_2^i; z, m, v). \tag{O.28}$$

By definition, $z_0 = 0$ and we normalize $m = 1$.[24] Given data, we can search for values of 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

[24]The likelihood is unaffected by doubling m, quadrupling v, and halving z.

distribution. Then the likelihood is

$$\mathcal{L}^{CT-DM} = \frac{1}{N}\sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} w_k f^{CT-DM}_{d_1^i+d_2^i}(\zeta_1^i, \zeta_2^i; z, m_k, v_k)\right). \tag{O.29}$$

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

An interesting and open question is whether this model is identified. We are unaware of any existing results on identification of continuous time models with discrete measurement and repeat spells.[25] Conversely, we are unaware of any examples that illustrate a failure of identification. Here we give one such example when data are censored, finding two different CT-DM models with non-trivially different baseline hazards that generate the same data.

Model I has a Gamma-distributed frailty with mean $m = 1$, variance $v = 1$ and baseline hazards $b_1 = \int_0^1 b(s)ds = 1$, $b_2 = \int_1^2 b(s)ds = 0.5$, implying $b_2/b_1 = 0.5$. Model II has a two-type distribution, with $\theta_1 = 0.2819$, $\theta_2 = 1.7950$, shares $dG(\theta_1) = 0.5254$ and $dG(\theta_2) = 1 - dG(\theta_1)$, and baseline hazards $b_1 = \int_0^1 b(s)ds = 0.9071$ and $b_2 = \int_1^2 b(s)ds = 0.4495$, implying $b_2/b_1 = 0.4955$. We assume our dataset is censored so that we observe all products for exactly three periods.

Let Φ_{t_1,t_2} denote the survival function as in equation (3), and let z_t be the integrated hazard defined before as $z_t = \sum_{s=0}^{t} b_s$. For model I, the survival function is

$$\Phi^I_{t_1,t_2} = \int_0^\infty e^{-\theta(z_{t_1}+z_{t_2})}\frac{e^{-\theta}}{\Gamma(1)}d\theta = \frac{1}{1+z_{t_1}+z_{t_2}},$$

where the second term in the integral is the density of the Gamma distribution with $m = 1, v = 1$. For model II, we have

$$\Phi^{II}_{t_1,t_2} = e^{-\theta_1(z_{t_1}+z_{t_2})}dG(\theta_1) + e^{-\theta_2(z_{t_1}+z_{t_2})}dG(\theta_2).$$

With censoring time equal to three periods, we can measure $\Phi_{1,0}$, $\Phi_{2,0}$, $\Phi_{1,1}$, and $\Phi_{2,1}$ (but not $\Phi_{2,2}$). It is easy to verify that for both processes $\Phi_{1,0} = 0.5$, $\Phi_{2,0} = 0.4$, $\Phi_{1,1} = 0.3333$, and $\Phi_{2,1} = 0.2857$.

To summarize, we have found two different CT-DM models which explain the same data. The two models have non-trivially different baseline hazards, in the sense that b_2/b_1 is not

[25] Ridder (1990), Brinch (2011), Abbring and Ridder (2015) study identification of the continuous time MPH model, or the less restrictive Generalized Accelerated Failure-Time model, with time-aggregated records and single-spell data. As in Elbers and Ridder (1982) and Heckman and Singer (1984), identification comes from observable characteristics which affect the hazard function.

the same, which shows lack of identification of both the frailty distribution and the baseline hazard. This raises a concern that maximum likelihood estimates of the CT-DM model may depend on the assumed functional form of the frailty distribution. Perhaps for this reason, we are unaware of any attempts to estimate the CT-DM model using repeated spell data.

I.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 counter-CDF $\bar{Q}_c^1$ and density $\bar{q}_c^1$, 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.

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.

For a typical product i, we observe the vector $(c^i, d_1^i, d_2^i, \zeta_1^i, \zeta_2^i)$. Under the assumption of the Gamma frailty distribution with mean m and variance v, 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

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

where Γ is the gamma function. The integrand is equal to the probability that the censoring time exceeds $t_1 + t_2$, 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. We then solve the integral to get

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

where

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

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

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

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

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

where

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

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

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

This is the density 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

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

where

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

As in the CT-DM model, we use the density 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 $\bar{Q}_c^1$ and $\bar{q}_c^1$, allowing us to treat them as nuisance parameters. The part of the likelihood that we are interested in is

$$\mathcal{L}^{CT-CM} = \frac{1}{N} \sum_{i=1}^{N} \log f^{CT-CM}_{d_1^i+d_2^i}(\zeta_1^i, \zeta_2^i; z, m, v). \tag{O.30}$$

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

$$\mathcal{L}^{CT-CM} = \frac{1}{N} \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} w_k f_{d_1^i + d_2^i}^{CT-CM}(\zeta_1^i, \zeta_2^i; z, m_k, v_k) \right). \tag{O.31}$$

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 dataset, we need to impose some restrictions on the baseline hazard in order to maximize either likelihood (O.30) or (O.31). We assume that the baseline hazard is piecewise constant between integer values and so z is piecewise linear.

I.3 Results

We use IRI pooled sample data where we use the first two spells per product. We emphasize that all durations in this data set are coded to equal an integer. We then estimate three models. The first is the CT-DM model. We are unaware of previous attempts to estimate this model using repeated spell data. The second is the CT-CM model, which ignores the fact that durations are coded to integer values. Gagliarducci (2005) and Nakamura and Steinsson (2008) use this model to examine labor market data and price stickiness, respectively. The third is our 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.

The right panel of 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 a more flexible distribution than a single gamma, for example a mixture of several gamma

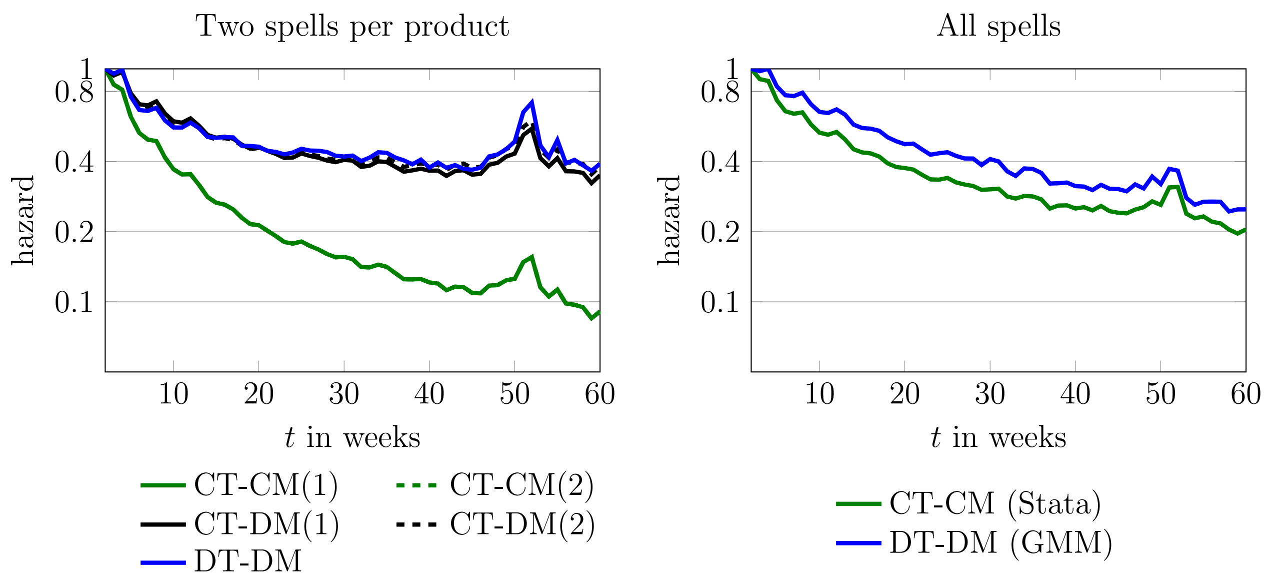

Figure 13: Baseline hazard estimated using different methods with IRI data, log scale. The green lines correspond to continuous time with continuous time measurement (CT-CM), the black lines correspond to the continuous time, discrete measurement (CT-DM) model. The blue line is the discrete time model with discrete measurement (DT-DM). The left panel uses two spells per product, the right panel uses all spells. For the CT estimates in the left panel, we assume that the frailty distribution is either a single gamma (solid line) or a mixture of two gammas (dashed line). For the CT model in the right panel we assume that the frailty distribution is gamma, but the estimator uses a full set of duration dummy variables to overcome this assumption.

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.

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. This is the estimation method employed by Nakamura and Steinsson (2008). 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. We use Stata to estimate the baseline hazard using all spells (not only first two spells per product).

The left panel of Figure 13 compares estimates from Stata with ours for the baseline MPH model from Section 6. The baseline hazard estimated using maximum likelihood is somewhat steeper than the one estimated using GMM, which is the result of the failure of CT-CM to account for time-aggregated data. Using more spells per product helps to overcome this misspecfication, since the difference between CT-CM and DT-DM is much smaller.

I.4 Advantages of GMM

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 b and hence is very simple and fast to solve. Third, Proposition 2 establishes that we find a global optimum. In contrast, the log-likelihood is non-linear in b and finding its maximum can be slow.[26] Importantly, there is no guarantee that maximum likelihood finds a global maximum. Fourth, our model formulated in discrete time is identified even with censored data, while an example in Supplemental Appendix I.1 illustrates that the time-aggregated continuous-time model may not be identified. 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 was central to the explorations of price plans in section 8.. This set of assumptions has proven to be extremely hard to handle in the maximum likelihood framework. For example, Fougere, Le Bihan, and Sevestre (2007) try to estimate a CT-DM competing risks model without unobserved heterogeneity but say on page 260 that "…convergence of the maximum likelihood procedure is very difficult to reach."

[26]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.

J Baseline Hazards for Product Categories

Here we present our results by product category. Figure 14 shows the baseline and mixture 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 of price trends and price reversals with observable characteristics, developed in Section D.

K Additional Figures for Price Plan Hazards

We estimate the unconditional and conditional hazards of changing plans, and the conditional hazard of changing a price within the plan, using the Online Micro Price Data. Figure 18 shows estimates of the three baseline hazards. We observe that the unconditional hazard of switching to a new plan is mildly increasing, while the conditional hazards are flat or declining.

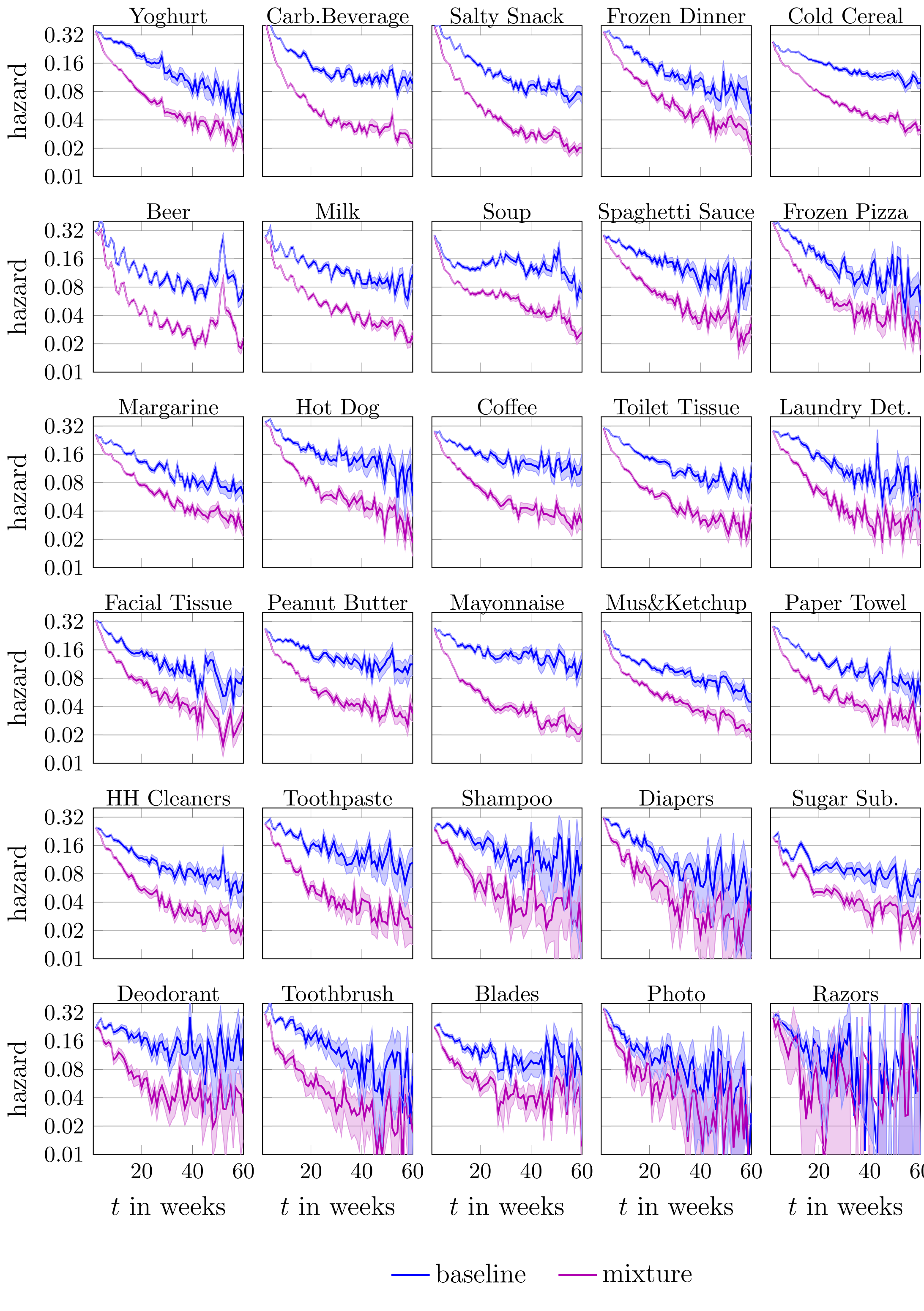

Figure 14: Mixture and baseline hazards for individual product categories, IRI data, log scale. Product categories are sorted by the number of spell pairs.

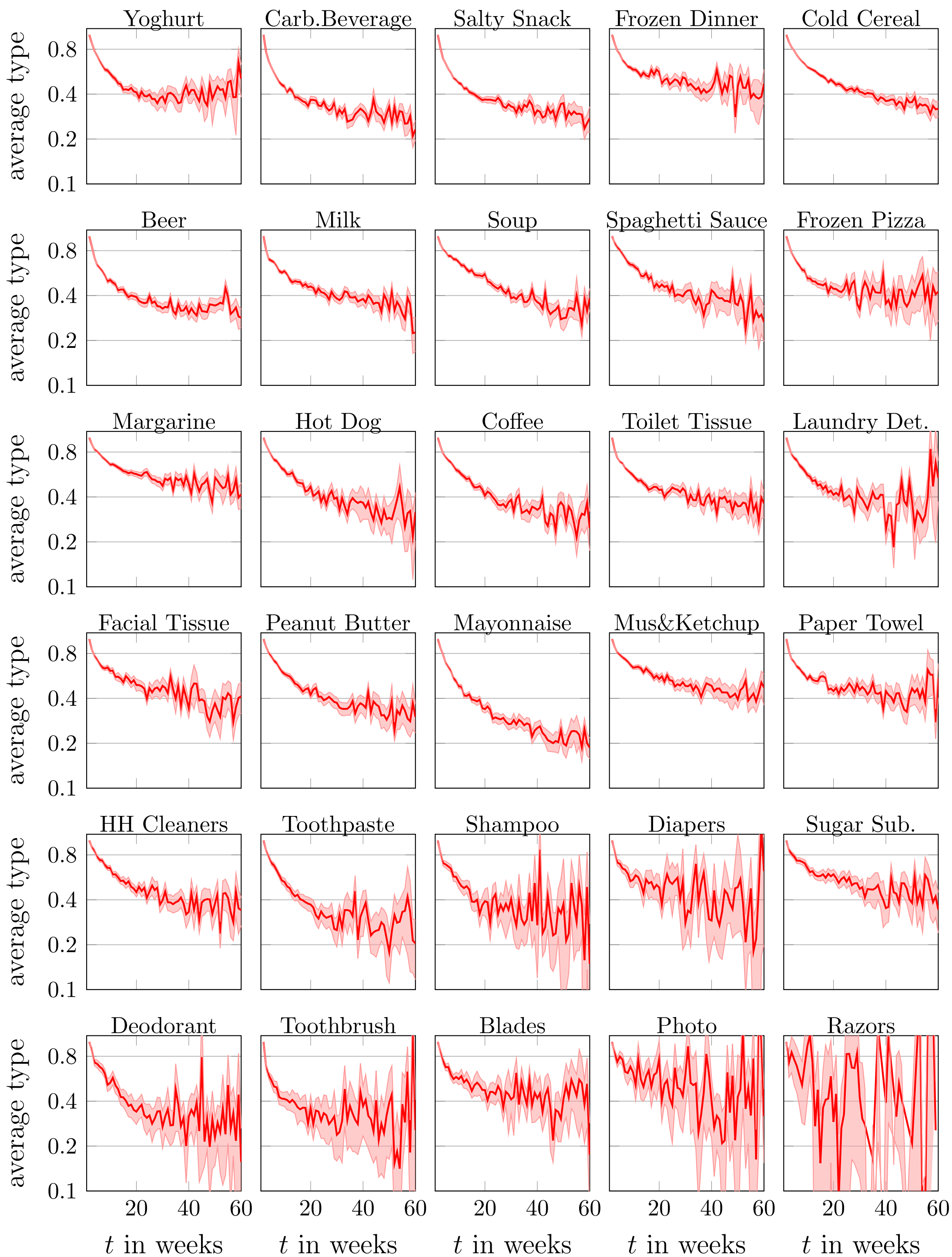

Figure 15: Average type for individual product categories, IRI data, log scale. Product categories are sorted by the number of spell pairs.

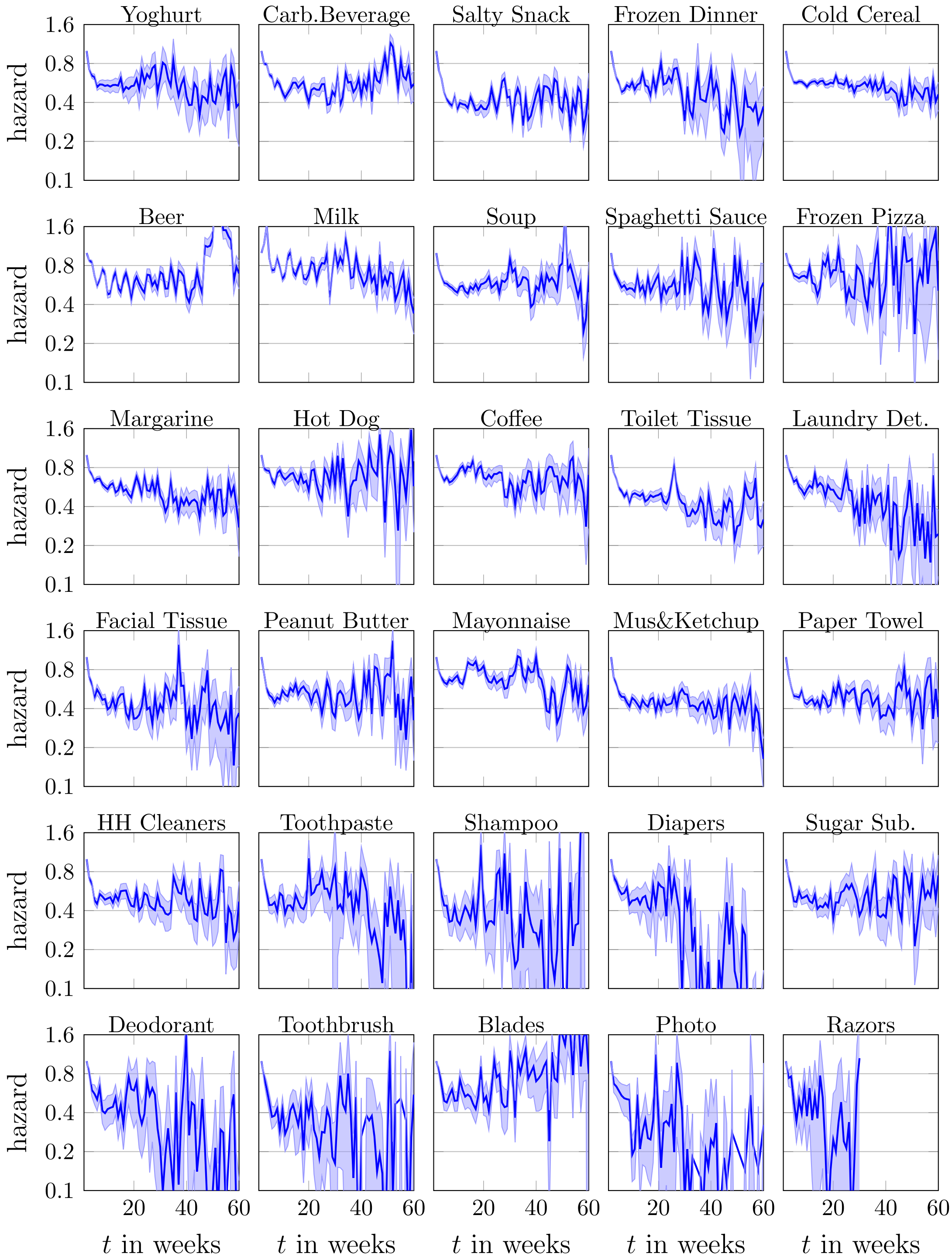

Figure 16: Baseline hazard b^{++} in the model with price trends and price reversals for individual product categories, IRI data, log scale. Product categories are sorted by the number of spell pairs.

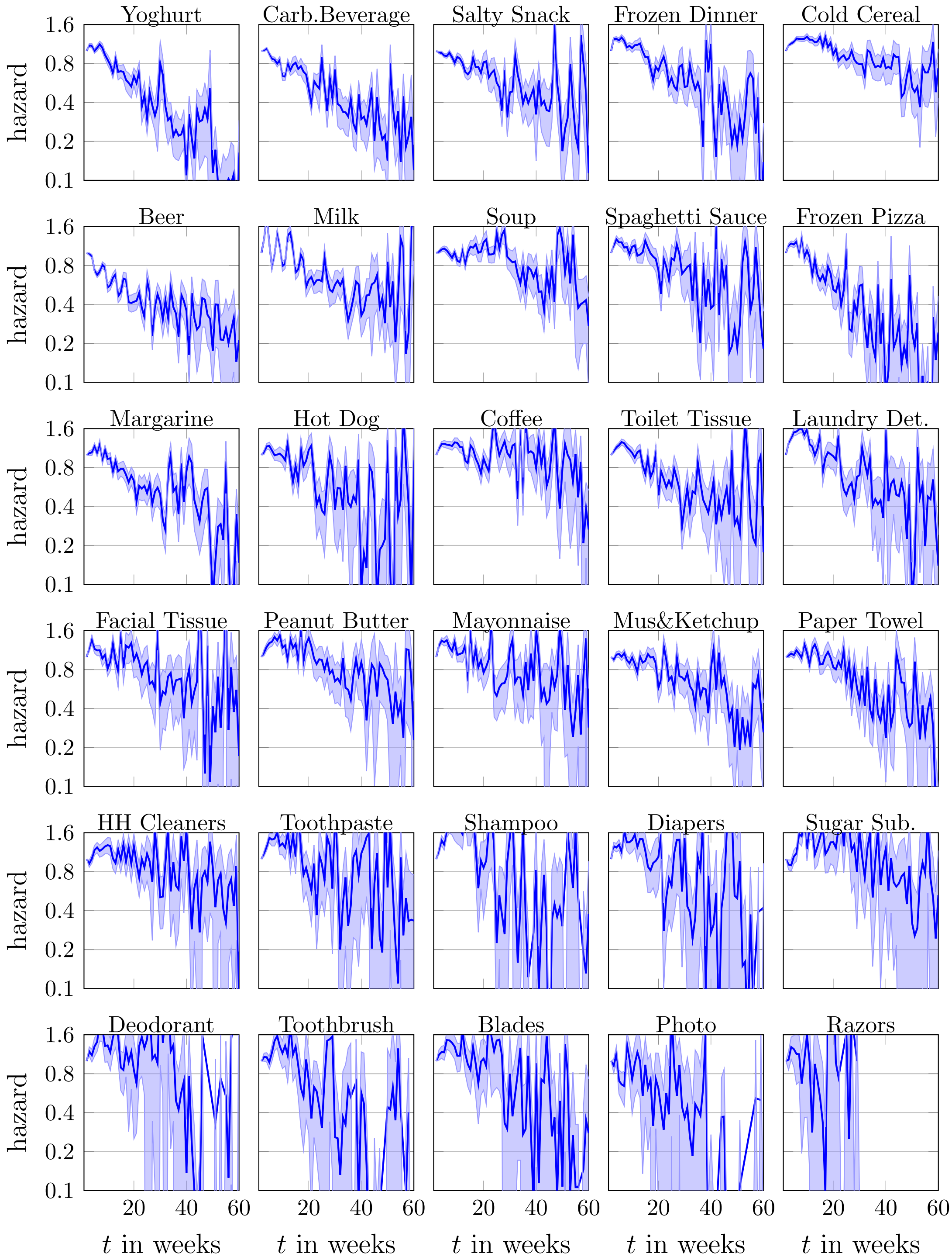

Figure 17: Baseline hazard b^{--} in the model with price trends and price reversals for individual product categories, IRI data, log scale. Product categories are sorted by the number of spell pairs.

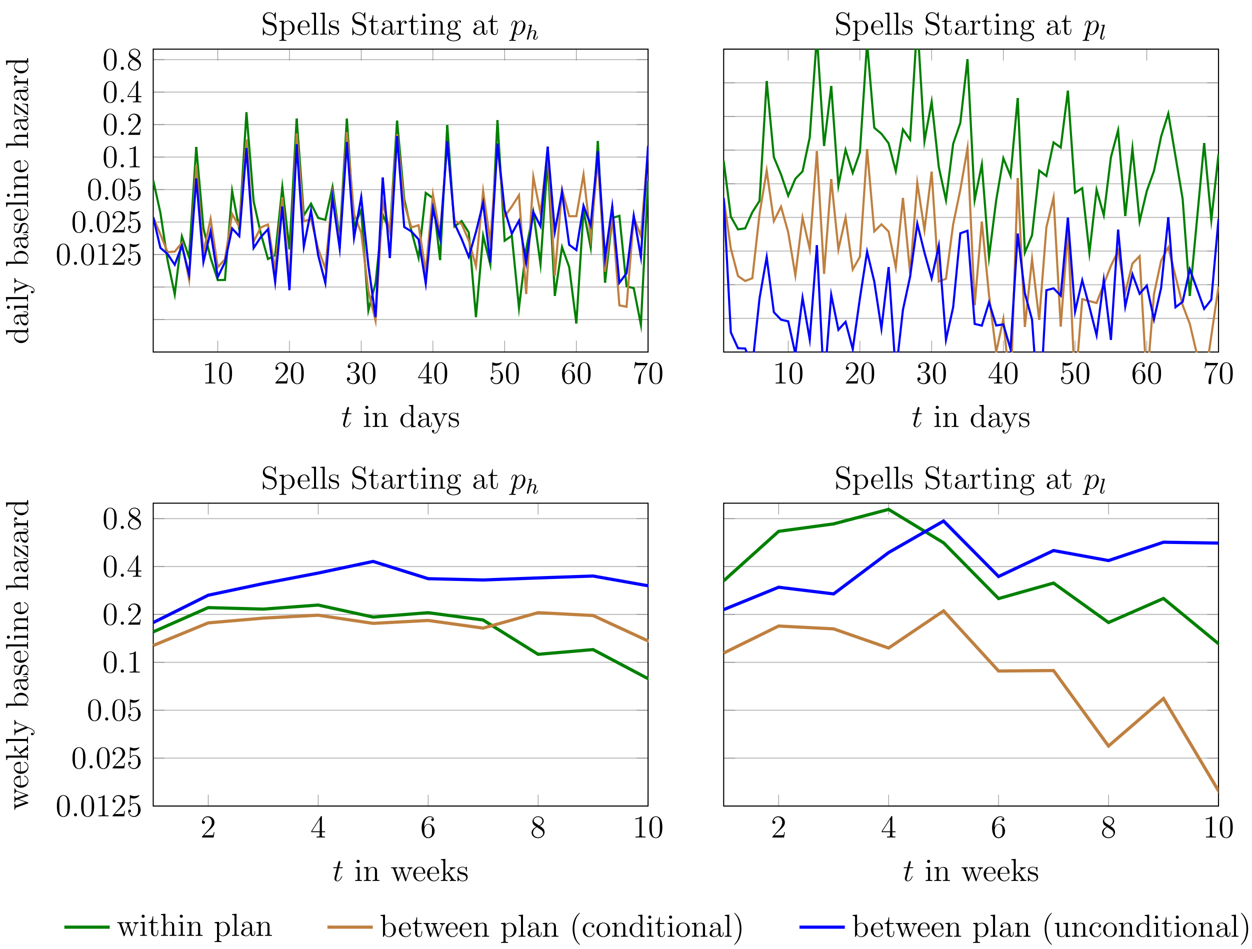

Figure 18: Baseline hazard of changing a price within a plan and between plans for Online Micro Price Data, starting at high and low price of a plan, using daily and weekly data, log scale. The top row shows daily data, the bottom row weekly data. The green line shows the within-plan hazard (part of competing risks), the brown lines shows the hazard to a new plan (competing risks); both are conditional. The blue line shows the unconditional hazard of choosing a new plan.

L Trigonometric Simplifications

Proof of Lemma 2. We start with the product-to-sum identity for sine functions, $\sin(A)\sin(B) = \frac{1}{2}\big(\cos(A-B) - \cos(A+B)\big)$. Applying this to our expression, we get:

$$\sin\left(\frac{n}{N+1}j\pi\right)\sin\left(\frac{n'}{N+1}j\pi\right) = \frac{1}{2}\left(\cos\left(\frac{n-n'}{N+1}j\pi\right) - \cos\left(\frac{n+n'}{N+1}j\pi\right)\right)$$

Next we sum this expression over j from 1 to N:

$$\sum_{j=1}^{N}\sin\left(\frac{n}{N+1}j\pi\right)\sin\left(\frac{n'}{N+1}j\pi\right) = \frac{1}{2}\left(\sum_{j=1}^{N}\cos\left(\frac{n-n'}{N+1}j\pi\right) - \sum_{j=1}^{N}\cos\left(\frac{n+n'}{N+1}j\pi\right)\right) \tag{O.32}$$

We next prove a preliminary result by applying Lagrange's trigonometric identity:

$$\sum_{j=1}^{N}\cos(j\theta) = -\frac{1}{2} + \frac{\sin((N+\frac{1}{2})\theta)}{2\sin(\frac{\theta}{2})},$$

which holds whenever θ that is not a multiple of 2π (so $\sin(\theta) \neq 0$).

Fix an integer k with $-2N \leq k < 0$ or $0 < k \leq 2N$, so $k/(2(N+1))$ is not an integer. Then

$$\sum_{j=1}^{N}\cos\left(\frac{kj\pi}{N+1}\right) = -\frac{1}{2} + \frac{\sin\left(\frac{(2N+1)k\pi}{2(N+1)}\right)}{2\sin\left(\frac{k\pi}{2(N+1)}\right)} = -\frac{1}{2} + \frac{\sin\left(k\pi - \frac{k\pi}{2(N+1)}\right)}{2\sin\left(\frac{k\pi}{2(N+1)}\right)} = -\frac{1+(-1)^k}{2}, \tag{O.33}$$

where the first equation is Lagrange's trigonometric identity; the second equation breaks the term inside the numerator into two pieces and applies the definition of α to one of them; and the third uses the fact that $\sin(k\pi - \alpha\pi) = \sin(\alpha\pi)$ if k is an odd integer and $\sin(k\pi - \alpha\pi) = -\sin(\alpha\pi)$ if k is an even integer.

We use equation (O.33) to evaluate the two summations in equation (O.32), considering two cases separately.

Inequality: $n \neq n'$. First let $k = n - n'$ and then let $k = n + n'$. Since $n, n' \in \{1, \ldots, N\}$ and $n \neq n'$, this satisfies the restrictions that $-2N \leq k < 0$ or $0 < k \leq 2N$. Equation (O.33) implies

$$\sum_{j=1}^{N}\cos\left(\frac{n-n'}{N+1}j\pi\right) = -\frac{1+(-1)^{n-n'}}{2} \quad \text{and} \quad \sum_{j=1}^{N}\cos\left(\frac{n+n'}{N+1}j\pi\right) = -\frac{1+(-1)^{n+n'}}{2}.$$

Thus when $n \neq n'$,

$$\frac{1}{2}\left(\sum_{j=1}^{N}\cos\left(\frac{n-n'}{N+1}j\pi\right)-\sum_{j=1}^{N}\cos\left(\frac{n+n'}{N+1}j\pi\right)\right)$$
$$=\frac{1}{2}\left(\left(-\frac{1+(-1)^{n-n'}}{2}\right)-\left(-\frac{1+(-1)^{n+n'}}{2}\right)\right)=\frac{(-1)^{n-n'}}{4}\left(-1+(-1)^{2n'}\right)=0,$$

where the last equation uses $(-1)^{2n'}=1$ since n' is an integer.

Equality: $n=n'$. When $k=n-n'=0$, we cannot apply equation (O.33), which holds only for non-integer k. Instead, we evaluate it directly:

$$\sum_{j=1}^{N}\cos\left(\frac{n-n}{N+1}j\pi\right)=N,$$

where we use $\cos(0)=1$. For $k=n+n'=2n\in\{2,\dots,2N\}$, we apply equation (O.33) directly:

$$\sum_{j=1}^{N}\cos\left(\frac{2n}{N+1}j\pi\right)=-\frac{1+(-1)^{2n}}{2}=-1,$$

where the last equation uses $(-1)^{2n}=1$. Thus when $n=n'$,

$$\frac{1}{2}\left(\sum_{j=1}^{N}\cos\left(\frac{n-n'}{N+1}j\pi\right)-\sum_{j=1}^{N}\cos\left(\frac{n+n'}{N+1}j\pi\right)\right)=\frac{1}{2}(N-(-1))=\frac{N+1}{2}.$$

This completes the proof. ■

Proof of Lemma 3. We use the other Lagrange trigonometric identity:

$$\sum_{n=1}^{N}\sin(n\theta)=\frac{\cos(\frac{\theta}{2})-\cos(\frac{(2N+1)\theta}{2})}{2\sin(\frac{\theta}{2})}.$$

Setting $\theta=k\pi/(N+1)$, we apply this to our expression:

$$\sum_{n=1}^{N}\sin\left(\frac{n}{N+1}k\pi\right)=\frac{\cos\left(\frac{k\pi}{2(N+1)}\right)-\cos\left(\frac{(2N+1)k\pi}{2(N+1)}\right)}{2\sin\left(\frac{k\pi}{2(N+1)}\right)}=\frac{\cos\left(\frac{k\pi}{2(N+1)}\right)-\cos\left(k\pi-\frac{k\pi}{2(N+1)}\right)}{2\sin\left(\frac{k\pi}{2(N+1)}\right)},$$

where the second equation algebraically manipulates the second term in the numerator. Now if k is even, $\cos(k\pi - \alpha) = \cos(-\alpha) = \cos(\alpha)$ for all α. Thus when k is even we have

$$\sum_{n=1}^{N} \sin\left(\frac{n}{N+1}k\pi\right) = \frac{\cos\left(\frac{k\pi}{2(N+1)}\right) - \cos\left(\frac{k\pi}{2(N+1)}\right)}{2\sin\left(\frac{k\pi}{2(N+1)}\right)} = 0.$$

In contrast, k is odd, $\cos(k\pi - \alpha) = -\cos(-\alpha) = -\cos(\alpha)$ for all α. Thus when k is odd we have

$$\sum_{n=1}^{N} \sin\left(\frac{n}{N+1}k\pi\right) = \frac{\cos\left(\frac{k\pi}{2(N+1)}\right) + \cos\left(\frac{k\pi}{2(N+1)}\right)}{2\sin\left(\frac{k\pi}{2(N+1)}\right)} = \cot\left(\frac{k\pi}{2(N+1)}\right).$$

This completes the proof. ■