Calvo, Taylor and the estimated hazard function for price changes

D. Fougère[*], H. Le Bihan[†], P. Sevestre[‡§]

November 2004 Preliminary version

Abstract

This paper analyzes price setting behavior estimating duration models on a large dataset of French consumer price quotes. The unconditional estimated hazard function for price change is found to be decreasing, in contradiction with standard models of nominal rigidity like those of Calvo and Taylor.

We find that accounting for sectoral heterogeneity and time-varying covariates (inflation) provide baseline hazard functions that are more in line with theoretical models.

JEL Codes: E31, C41.
Keywords: Sticky Price, Hazard function, Duration models.

[*]CREST-ENSAE and CNRS. denis.fougere@ensae.fr.
[†]Banque de France. herve.lebihan@banque-france.fr
[‡]Université Paris XII-Val de Marne and Banque de France. sevestre@univ-paris12.fr
[§]This paper is part of a Eurosystem research project about the persistence of inflation in the Euro area (Inflation Persistence Network). We are grateful to the INSEE for communicating the data used in this paper. We thank T.Kamionka and J.Mairesse, as well as participants at the June 2004 Panel Data Conference held in College Station, Texas and seminars at the ECB (IPN) and the University Paris XII for helpful remarks. This paper does not necessarily represent the views of the Banque de France.

1 Introduction

Nominal rigidities are a major determinant of the response of the economy to exogenous shocks and in particular to monetary policy actions. This points to the need of a better understanding of the pricing behavior of firms and/or outlets. Uncontroversially, a microeconometric approach should be well suited to shed light on such an issue.

Early empirical microeconomic analyses of nominal rigidities had to rely on rather small and restricted datasets (e.g. see Cecchetti, 1986, Kashyap, 1995). Interest in the use of panel data about microeconomic prices at the firm or outlet level has sharply increased in the very recent period, in part due to availability of larger datasets. Bils and Klenow (2002) provide results on price stickiness for a major part of the US CPI categories. As regards the euro area, some stylized facts characterizing prices rigidity in several countries have been produced by Aucremanne and Dhyne (2004), Baudry et al. (2004), Dias et al. (2004), Veronese et al. (2004) [1]. The average duration of price spells in the euro area appears to be around about 3 quarters . However, there exist a great deal of heterogeneity, in particular across different types of goods. Prices in the services sector last longer (about one year) while those in the energy and food sectors change more frequently. Moreover, prices appear to change more frequently in large outlets such as supermarkets than in more "traditional" corner shops.

The present paper pursues a further step in microeconometric analysis of prices rigidity by looking at the empirical relevance of the existing theoretical models of pricing behaviors such as those proposed by Taylor (1980), Calvo (1983), Sheshinski and Weiss (1983), or Dotsey, King and Wolman (1999). In order to estimate models that "formalize" the predictions of those theoretical models in terms of statistically observable variables, the paper proceeds by focusing on the hazard function for price changes. The hazard function approach has been widely used in unemployment analysis, but so far, to our knowledge, rarely employed to characterize price rigidity.

The aim of this paper is thus to estimate duration models for explaining the duration of consumer prices at the microeconomic level and to test whether those estimates are consistent with the predictions from the theoretical models of pricing behavior such as those proposed by Calvo, Taylor and others.[2] We make use of a large dataset consisting of price quotes in French outlets over the period July 1994 to February 2003. We show that, not accounting for heterogeneity, the unconditional empirical hazard functions are decreasing, thus being in contradiction with the theoretical predictions. However, taking

[1] see Dhyne et al. (2004) for a summary

[2] Related papers that follow the same approach are Alvarez et al. (2004), Aucremanne and Dhyne (2004,b), Dias et al. (2004)

account of heterogeneity, results are more in line with an "amended" form of Calvo's model.

2 Theoretical patterns of the hazard function for price changes

This section provides a short review of the implications of the main models of price setting behavior used in monetary economics (see e.g. Taylor, 1998 , and Burriel et al., 2004 for more detailed discussions) in terms of the hazard function for price changes. In our context, the hazard function is the probability of adjusting prices conditional on the elapsed duration since last price change.

Models of price rigidity can broadly be classified into two categories: time-dependent and state dependent models. Time dependent models assume that prices changes take place at fixed or random intervals. One prominent model is Taylor (1980)'s staggered contracts model that has been used to describe both the price and wage adjustment patterns. The model assumes that prices and wages are negotiated for fixed periods, say one year. As a consequence, the hazard function should be zero for the first quarters and exhibit one spike with value one at the duration of the contract. One may expect several spikes in the hazard function, if contracts of different lengths coexist in the economy. A widespread alternative in the monetary policy literature is Calvo's model. In this model, each firm has a constant probability of changing its price. As a consequence, the hazard function is flat. A variant that encompasses both Taylor's and Calvo's schemes is the truncated Calvo model, in which there is a maximum duration for prices. The hazard is flat up to this maximum value, then is equal to one for this maximal duration (see Wolman, 1999).

State-dependent models predict on the contrary that the probability of a price change varies according to the state of the economy. State-dependency with infrequent price change typically emerges from menu cost models. Such models imply that a firm will change its price if the (absolute) deviation of its current price from the optimal price is larger than the menu cost, i.e. the fixed cost of changing price. Sheshinski and Weiss (1983) have analyzed this model. The probability of a price change is predicted to decrease with the size of the menu cost, while the size of the price change will increase with that of the menu cost. Generally, the probability of a price change is predicted to increase with trend inflation. More recently Dotsey, King, and Wolman (1999) have proposed a model that generalizes Calvo's model by incorporating state-dependent pricing into a truncated Calvo model. In their model, firms face a random menu cost. Only the firms with relatively low menu costs will choose to adjust. The hazard rate will be an increasing function of time since the previous change, since firms with prices set a long time ago are more likely to observe relative price in excess of the menu cost. An higher steady state inflation will lead to more rapid erosion of relative prices and to hence to a steeper hazard function.

It is to be noted that one common feature of all those price models is that the hazard function is a non-decreasing function of elapsed duration (except perhaps for spikes in the hazard function). Also note that one widely used model of price setting is not considered here, namely Rotemberg's (1982) quadratic adjustment cost model. Indeed, this model predicts frequent price adjustments of a small magnitude. Such a pattern is at variance with the evidence of lumpy adjustment typically found with micro data in most sectors .

3 The econometric modelling of price spells

As mentioned just above, a way to discriminate among the different theoretical models of pricing behavior and to assess their respective empirical relevance is to consider the shape of the hazard function of price changes, i.e. the probability of occurrence of a price change after a given duration has elapsed. The approach we adopt here is that of duration models. Indeed, from elementary price quotes of the French CPI we are able to compute durations, i.e. the length of time during which a price was not changed. We start this section by summarizing the duration analysis useful for the present paper.[3]

An alternative but related approach would be to model the probability of price change, using variants of the logit or tobit model, in the vein of Cechetti (1986). See Aucremanne and Dhyne (2004) for the implementation of such an approach. One advantage of duration analysis is that, despite the discrete nature of observations, we are able to use continuous time duration models and allow for the fact that price changes do occur at any point in time during a month.

3.1 The econometric analysis of durations: a reminder

Let us consider a price spell i, beginning at time $t_{i0} = 0$ and lasting until time $t_{i,end}$ where the date $t_{i,end}$ is not known exactly, although it is observed to be located somewhere between dates $t-1$ and t. For the convenience of presentation, we shall consider that dates t and $t-1$ correspond to the end of months t and $t-1$ respectively: then, all what we know is that the price spell has lasted more than $t-1$ months and at most t months. Then, we aim at characterizing the probability for a price change to occur after some time has elapsed since the previous price change.

[3] See Kiefer (1988) and Lancaster (1997) for comprehensive expositions of duration analysis.

To do so, let us first consider that there is no censoring[4] and define the probability for a spell to last at least $t-1$ months (the so-called survivor function) as:

$$\begin{aligned}\Pr(T &\geq t-1) = S(t-1) \\ &= 1 - F(t-1)\end{aligned}$$

where $F(.)$ and $S(.)$ are the cumulative density function and the survivor function of T respectively. Then, the probability that a spell will end during month t_i is given by

$$\begin{aligned}\Pr(t-1 &< T \leq t) = F(t) - F(t-1) \\ &= S(t-1) - S(t).\end{aligned}$$

What we are interested in is the hazard function, i.e. the conditional probability for a spell to end during month t, given it was still on-going at the end of $t-1$:

$$\begin{aligned}h(t) &= \Pr(t-1 < T \leq t \mid T > t-1) \\ &= \frac{\Pr(t-1 < T \leq t)}{\Pr(T > t-1)} \\ &= \frac{S(t-1) - S(t)}{S(t-1)} \\ &= 1 - \frac{S(t)}{S(t-1)}.\end{aligned}$$

Because of the relationship between the survivor function and the cumulated hazard ($\ln(S(t) = -H(t)$, e.g. see Kalbfleisch and Prentice (2002)), this can be written as

$$\begin{aligned}h(t) &= 1 - \exp(-H(t)) \ / \ \exp(-H(t-1)) \\ &= 1 - \exp[H(t) - H(t-1)] \\ &= 1 - \exp(-\int_{t-1}^{t} \lambda_\tau \ d\tau) \\ &= 1 - \exp(-\gamma(t))\end{aligned}$$

where λ_τ is the instantaneous hazard rate, which is equal, in a continuous time setting, to the risk of a spell to finish at time τ. $h(t)$ is the probability for a price spell to last t months, i.e. to have finished at any point in time τ between the beginning and the end of that month. It must be recognized here that the parameters characterizing the baseline hazard λ are not identified here, unless we make some assumption about the behavior of this baseline hazard over the

[4] The question of censoring will be delt with in the next section.

period $]t-1,t]$. The simplest assumption is to consider that it is constant over such an interval: $\lambda_\tau = \lambda_t, \forall\ t-1 < \tau \leq t$, although it can vary over different intervals, indexed by t[5]. Then,

$$\begin{aligned} h(t) &= 1-\exp(-\int_{t-1}^{t} \lambda_\tau \ d\tau) \\ &= 1-\exp(-\lambda_t) \end{aligned}$$

Accounting for time discretization, the probability for a spell to last t months is then given by

$$\begin{aligned} f(t) &= \Pr(t-1 < T \leq t\ ) \\ &= h(t) \times S(t-1) \\ &= [1-\exp(-\lambda_t)] \times \exp[-H(t-1)] \\ &= [1-\exp(-\lambda_t)] \times \exp[-\sum_{s=1}^{t-1} \lambda_s] \end{aligned}$$

since

$$\begin{aligned} \ln(S(t-1)) &= -H(t-1) \\ &= -\sum_{s=1}^{t-1} \lambda_s. \end{aligned}$$

The general form of the log-likelihood, assuming that there is no censoring, is written as

$$\begin{aligned} \ln L &= \sum_{i=1}^{N} \ln(f_i(t_i)) \\ &= \sum_{i=1}^{N} [\ln(h_i(t_i) + \ln(S_i(t_i-1))] \\ &= \sum_{i=1}^{N} [\ln(h_i(t_i) - H_i(t_i-1))] \end{aligned}$$

where $\lambda_i(t_i)$ is the hazard rate of spell i, $S_i(t_i)$ the survivor function and $H_i(t_i - 1)$ the cumulated hazard function. Given the assumption that the hazard function is piecewise constant over monthly intervals ($\lambda_\tau = \lambda_t, \forall t-1 < \tau \leq t, t = 1, ..., T_{\max}$), this log-likelihood can be written as:

$$\ln L = \sum_{i=1}^{N} \ [\ln(1-\exp(-\lambda_t)) \ - \sum_{s=1}^{t_i-1} \lambda_s \ ]$$

[5] This amounts to specify a piecewise constant hazard model.

However, it may be unduly restrictive to assume the hazard to be constant over all individuals, irrespective of their observed or unobserved characteristics. One has indeed to worry about the possible existence of heterogeneity across spells.

3.2 Accounting for heterogeneity

There are many reasons that can explain why the duration of price spells can vary across products, outlets and over time. Indeed, outlets have their own pricing policy, depending on the type of product they sell, on the characteristics of their customers and on the competition with other retailers. Differences in the evolution of costs in different sectors and in the production and merchandizing technologies may also contribute to explain differences in the pricing behaviors across different types of goods.

3.2.1 Observed heterogeneity with time-constant covariates.

Examples of such possible covariates are the type of good or service (e.g. food, gasoline, clothes, services, etc.) or the outlet category (hypermarket, general store, traditional "corner shop", etc.) which can be seen as intrinsic characteristics of the spell. In our case, prices of some categories of products are assumed to last longer than others, whatever the type of outlet (for example, prices of energy do not last very long while, on the contrary, prices of services last often around one year). Similarly, some outlets are assumed to set prices for longer periods than others, for any product they sell. Indeed, it seems that hypermarkets and supermarkets change their prices more frequently than smaller shops. Then, it seems reasonable, at least in a first step of the analysis, to consider that those characteristics affect the hazard at different durations in a proportional way. The instantaneous hazard rate is assumed to be :

$$\lambda_i(t_i) = \lambda_{t_i} \exp(z_i \beta).$$

Then,

$$\begin{aligned} h_i(t_i) &= 1 - \exp(-\int_{t_i-1}^{t_i} \lambda_i(\tau)\ d\tau) \\ &= 1 - \exp(-\int_{t_i-1}^{t_i} [\lambda_\tau \exp(z_i\beta)]\ d\tau) \\ &= 1 - \exp\left[-\exp(z_i\beta) \times \int_{t_i-1}^{t_i} \lambda_\tau\ d\tau\right] \\ &= 1 - \exp(-\lambda_{t_i}) \times \exp(-\exp(z_i\beta)) \end{aligned}$$

and

$$\begin{aligned} f_i(t_i) &= h_i(t_i) \times S_i(t_i - 1) \\ &= [1 - \exp(-\lambda_{t_i}) \times \exp(-\exp(z_i\beta))] \times [-\exp(z_i\beta) \sum_{s=1}^{t_i-1} \lambda_s]. \end{aligned}$$

3.2.2 Observed heterogeneity with time-varying covariates

In the case where the covariates are time-varying, one has to take into account their path over the course of the spell, since this is the full "history" of those variables that may explain the end of the spell. In our case, this is particularly clear when one thinks about inflation. It is the cumulated inflation at each point in time during the spell that creates an increasing gap between the unchanged price of a product in a given outlet with either the general price level and/or the general price of the same product in other outlets as a whole or in direct competitors of the outlet. Taking inflation at the end of the spell only would create a bias since longer spells would be systematically associated with high inflation.

There are a few other time-varying covariates that we should include in our models; namely, we include dummies for changes in the VAT rate as well as for the Euro cash change-over. Then, assuming the hazard to be

$$\lambda_i(t_i) = \lambda_{t_i} \exp(z_i(t_i)\alpha).$$

one gets:

$$\begin{aligned} h_i(t_i) &= 1 - \exp\left[-\int_{t_i-1}^{t_i} \lambda_i(\tau)\ d\tau\right] \\ &= 1 - \exp\left[-\int_{t_i-1}^{t_i} \lambda_\tau \exp(z_i(\tau)\alpha)\ d\tau\right]. \end{aligned}$$

Then, given that $\lambda(\tau)$ as well as the variables $z_i(\tau)$ do not vary over the time interval $]t-1, t] : z_i(\tau) = z_{it},\ \lambda_i(\tau) = \lambda_t,\ \forall\ t - 1 < \tau \leq t$, one can write

$$h_i(t_i) = 1 - \exp(-\lambda_{t_i}) \times \exp(-\exp(z_{it}\alpha))$$

and

$$\begin{aligned} f_i(t_i) &= h_i(t_i) \times S_i(t_i - 1) \\ &= [1 - \exp(-\lambda_{t_i}) \times \exp(-\exp(z_{it}\alpha))] \times [-\sum_{s=1}^{t_i-1} \lambda_s \exp(z_{is}\alpha)]. \end{aligned}$$

3.2.3 Unobserved heterogeneity

We have assumed until now that the heterogeneity across price spells can be fully captured by covariates, either constant or time-varying. It may also be the case that there are other factors affecting the duration of spells in a systematic way but which we do not observe. It may then be worth it to account for such unobserved heterogeneity. The most usual way to do it is to assume that this heterogeneity factor has a particular distribution, of which we estimate the characteristic moments. Consider that the hazard is given by

$$\lambda_i(t_i) = \theta_i \; \lambda_{t_i} \exp(z_i\beta + z_i(t_i)\alpha)$$

where θ_i is a random variable independent of z_i and $z_i(t_i)$. Then, the survivor function accounting for unobserved heterogeneity is:

$$S(t_i \mid \theta_i) = S(t_i)^{\theta_i}.$$

To estimate the model, we have to "integrate out" this heterogeneity term, given its distribution $\mu(\theta)$:

$$S(t_i) = \int_0^{\infty} S(t_i)^{\theta_i} \mu(\theta) \; d\theta.$$

Assuming that θ is Gamma distributed with mean 1 and variance σ^2 we get:

$$\begin{aligned} S(t_i) &= [1 - \sigma^2 \ln(S(t_i)]^{-1/\sigma^2} \\ &= [1 + \sigma^2 H(t_i)]^{-1/\sigma^2}. \end{aligned}$$

Then, the log-likelihood can be written as (e.g. see Meyer (1990)):

$$\begin{aligned} \ln L &= \sum_{i=1}^{N} \ln(f_i(t_i)) \\ &= \sum_{i=1}^{N} [\ln(S_i(t_i - 1)) - \ln(S_i(t_i))] \\ &= \sum_{i=1}^{N} \ln([1 + \sigma^2 H(t_i - 1)]^{-1/\sigma^2}) - \sum_{i=1}^{N} \ln([1 + \sigma^2 H(t_i)]^{-1/\sigma^2}) \\ &= \sum_{i=1}^{N} \ln([1 + \sigma^2 \sum_{s=1}^{t_i - 1} \lambda_s \exp(z_{is}\alpha + z_i\beta)]^{-1/\sigma^2})] \\ &\quad - \sum_{i=1}^{N} \ln([1 + \sigma^2 \sum_{s=1}^{t_i} \lambda_s \exp(z_{is}\alpha + z_i\beta)]^{-1/\sigma^2})] \end{aligned}$$

3.3 Accounting for censoring

Censoring is a major issue when analyzing durations in general as in our particular context. Indeed, there are many reasons that explain why our observations of prices are censored:

First, the observation period is restricted by the database availability. This makes it very likely to observe product prices in a given outlet after the current price of the product was actually set and/or before that price ceases to exist. Indeed, the probability that the first spell in a price trajectory is left-censored is high, as is the one of the last spell being right-censored.

Second, the sampling of products and outlets by the statistical agency is also likely to generate some censoring. Indeed, the statistical institute may decide to discard a specific product from the "representative" CPI basket due to changes in technology or consumer behavior that shrink the expenditure share on these products (e.g. black and white TV sets) although they may still be sold in outlets. Then, the last price spell of such a product will be right-censored. Conversely, products may be included in the CPI basket and their price observed after they were actually available for consumers. This will generate left-censoring of the price spell.

Third, outlets and firms may decide to stop selling a product which price was followed by the statistical agency. Then, the procedure which is often adopted by statistical agencies in charge of computing the CPI consists in replacing the "old product" by another one, either a close substitute in the same outlet or the same product but in another outlet. It is then very likely that the price of the "replacing product" was set before the first price observation for this product. We then have left-censoring of the price spell for this new product.

Individual price data also feature attrition, i.e. individuals dropping out of the sample. Two sources of attrition are the following:

First, products have life-cycles: "old products" disappear from the market and "new ones" appear. The time-series of prices observations for a specific product is very likely to be interrupted at some point during the observation period. In this respect, it is important to precisely identify when the replacement of a product by a close substitute occurs, as one can wonder about the correct interpretation of the price change generally associated with this replacement. Note that in the dataset used in this paper, product replacement is identified by a specific flag (at the end of some price trajectories). Hence product replacements can be distinguished from pure price changes.

Second, outlets or firms may close, which obviously interrupts the time-series of prices observations for all products sold by the outlet or the firm.

In the present context, attrition is considered as a "normal" spell ending and thus, no censoring is assumed to characterize such spells.

Thus, in our context, we have right censored spells, left censored spells as well as both right and left-censored spells. In the case or right censoring, things are rather simple. Indeed, what we know about a right-censored spell at time t_i is that its (unobserved) duration will be at least of t_i months. In other words, the spell is still on-going at the end of the $(t_i - 1)^{th}$ period. Thus, we have, for right censored spells:

$$g(t_i) = S(t_i - 1)$$

Then, defining a dummy variable δ_i taking value 1 when the spell is uncensored, one can write the log-likelihood as (e.g. see Meyer (1990)):

$$\begin{aligned}
\ln L &= \sum_{i=1}^{N} \ln(f_i(t_i)) + \sum_{i=1}^{N} \delta_i \ln(g_i(t_i)) \\
&= \sum_{i=1}^{N} \ln([1 + \sigma^2 \sum_{s=1}^{t_i - 1} \lambda_s \exp(z_{is}\alpha + z_i\beta)]^{-1/\sigma^2})] \\
&\quad -\delta_i \sum_{i=1}^{N} \ln([1 + \sigma^2 \sum_{s=1}^{t_i} \lambda_s \exp(z_{is}\alpha + z_i\beta)]^{-1/\sigma^2})]
\end{aligned}$$

Many of our spells are also left-censored or both right and left-censored (i.e. the price spell of the product actually begins before the statistical agency starts to observe the product). As shown in the appendix, the statistical treatment of such spells induces more difficulties than accounting for the sole right-censoring. However, the sample we have is made of thousands of spells for similar products in quite similar outlets - i.e. we have many spells that concern a given particular product sold in a given type of outlet (e.g. in small corner grocery shops, or in supermarkets). Discarding the left censored spells will result in a substantial loss of information nor produce a selection bias as the left-censoring is independent of the duration of price spells. Contrarily to what may happen in the context of unemployment duration studies for example, left-censoring in our case does not concern a particular sub-population with specific characteristics[6]. Then, discarding those left-censored spells does not lead to a selection bias as there are many other uncensored spells relating to the same products/outlets in the sample. After having checked for this with a small simulation study, we have then decided to discard left-censored (and right and left-censored) spells from our econometric sample.

[6] The only exception is the category of clothes where there are very specific pricing and renewing of items policies that are not independent from each other: changes in prices are not necessarily frequent but changes in the items often occur every 6 months (the "winter/summer collection pattern" is important for this particular group of products.

Finally, given that some of our spells have quite long durations, we have decided to impose a constant hazard beyond the 25th month. This choice corresponds to a length of more than 2 years and only about 5% of our spells are concerned. Although this is not negligible, it appeared very difficult to get satisfactory estimates of specific duration parameters beyond this point.[7] Then, the log-likelihood function can be split in four sub-groups of observations:

1) G1: Uncensored spells with length t_i less than $T\max = 25$.

2) G2: Uncensored spells with length t_i equal to or longer than $T\max = 25$.

3) G3: Right-censored spells with length t_i less than $T\max = 25$.

4) G4: Right-censored spells with length t_i equal to or longer than $T\max = 25$.

Then, the log-likelihood we have maximized can be written as

$$\ln L = \sum_{i \in G1} f_{i1}(t_i) + \sum_{i \in G2} f_{i2}(t_i) + \sum_{i \in G3} g_{i3}(t_i) + \sum_{i \in G4} g_{i4}(t_i)$$

with

$$\begin{aligned} f_{i1}(t_i) \quad &= \quad \sum_{i=1}^{N} \ln([1+\sigma^2 \sum_{s=1}^{t_i - 1} \lambda_s \exp(z_{is}\alpha + z_i\beta)]^{-1/\sigma^2})] \\ &\quad - \sum_{i=1}^{N} \ln([1+\sigma^2 \sum_{s=1}^{t_i} \lambda_s \exp(z_{is}\alpha + z_i\beta)]^{-1/\sigma^2})] \end{aligned}$$

$$\begin{aligned} f_{i2}(t_i) \quad &= \quad \sum_{i=1}^{N} \ln([1+\{\sigma^2 \sum_{s=1}^{T\max - 1} \lambda_s \exp(z_{is}\alpha + z_i\beta)\} \\ &\qquad + \{\sigma^2(t_i - T_{\max} + 1).\lambda_{T\max} \exp(z_{iT\max}\alpha + z_i\beta)]^{-1/\sigma^2})] \\ &\quad -\delta_i \sum_{i=1}^{N} \ln([1+\{\sigma^2 \sum_{s=1}^{T\max} \lambda_s \exp(z_{is}\alpha + z_i\beta)\} \\ &\qquad + \{\sigma^2(t_i - T_{\max}).\lambda_{T\max} \exp(z_{iT\max}\alpha + z_i\beta)]^{-1/\sigma^2} \end{aligned}$$

$$g_{i3}(t_i) = \sum_{i=1}^{N} \ln([1+\sigma^2 \sum_{s=1}^{t_i - 1} \lambda_s \exp(z_{is}\alpha + z_i\beta)]^{-1/\sigma^2})]$$

[7] In fact, this difficulty occurred even sooner in some cases where, for some sub-groups in our sample, spells longer than one year may be very rare. This is the case for energy goods for example as well as for spells correponding to sales or promotions.

$$
\begin{aligned}
g_{i4}(t_i) &= \sum_{i=1}^{N} \ln([1+\{\sigma^2 \sum_{s=1}^{T\max-1} \lambda_s \exp(z_{is}\alpha + z_i\beta)\} \\
&\qquad + \{\sigma^2(t_i - T_{\max}+1).\lambda_{T\max} \exp(z_{iT\max}\alpha + z_i\beta)]^{-1/\sigma^2})].
\end{aligned}
$$

The advantage of this specification is that it can fit almost any empirical hazard. Moreover, it is quite easy to implement a test of the theoretical models by testing:

- Calvo's model: $\lambda_1 = \lambda_2 = \lambda_3 = ... = \lambda_T = \lambda_{T+1}$

- Truncated Calvo's model: $\lambda_1 = \lambda_2 = \lambda_3 = ... = \lambda_T$ and $\lambda_{T+1} = 1$ if the maximum duration of prices is set at $T+1$

- Taylor's model: $\lambda_1 = \lambda_2 = \lambda_3 = ... = \lambda_T = \lambda_{T+1} = 0$ except λ_n if contracts last for n periods.

Obviously, it is very unlikely that any of these models will fit the data as there are strong heterogeneities across agents in their pricing behavior. A well-known consequence of such an heterogeneity is that, by not accounting for such an heterogeneity, one can estimate a decreasing hazard function while in fact, there are 2 groups of individuals with both a constant hazard but which is set at different levels.

4 The duration of price spells in France: basic facts

The data used for our econometric analysis, i.e. the price quotes from the CPI, is quite original, both as regards its contents and its size. We briefly document the data below, referring to Baudry et al. (2004) for additional details and evidence.

4.1 Dataset: 13 millions price quotes from the French CPI.

Data are monthly price quotes from the CPI collected by the INSEE (Institut National de la Statistique et des Etudes Economiques). The methodology of data collection is described in INSEE (1998). The sample contains CPI records from 1994:7 to 2003:2. These data cover around 65% of CPI (Fresh foods, Rents, Purchase of cars, and Administred prices are not covered). The number of price quotes is for the full sample around 13 million observations, and around 9.7 for the pre 2001 period. In Appendix 1 we document how some data issues were dealt with (euro cash-changeover, missing prices, etc.).

The breakdown of records by sectors is presented in Table 1.

Table 1: Number of price records by category of goods/services

Type of good	Number of obs.	% in database	weight in database	weight in CPI
Unprocessed food	1,655,484	12.51	9.87	8.69
Processed food	2,443,456	18.47	15.51	11.93
Non-energy manuf.goods	6,496,559	49.10	36.75	32.12
Energy	345,512	2.62	7.87	8.39
Services	2,246,977	16.98	30.00	38.87
Unidentified (out of CPI)	42,351	0.32	-	-
Total	13,230,339	100.00	100.00	100.00

Note: The weights in the database are obtained by summing up the weight of the elementary groups covered in the database. The last column reports the weight of each subindex in the overall CPI, including items (e.g. rents) not available in the data set.

With each individual record the information recorded is : the price level, an individual product code (outlet and product category), the year and month of record, a "type of record" code (indicating whether the price record is a regular one, whether it correspond to sales, whether the item was temporarily unavailable so that the price was replaced, etc.).

The structure of the database as regards the type of outlet is given in Table 2, while that according to the nature of the price quote is given in Table 3.

Table 2: Price records by type of outlet

Type of outlet	Number of obs.	% in database
Hypermarket	2,629,270	19.87
supermarket	1,802,583	13.62
hard discount store	155,483	1.18
convenience store	317,233	2.41
general store	483,165	3.65
department store	483,981	3.66
large-area specialist	1,679,584	12.69
traditional corner-shop	3,611,926	27.30
market	62,257	0.47
services	1,896,128	14.33
others	108,729	0.82
total	13,230,339	100.00

Table 3: Breakdown of price records by type of observation

Type of observation	Number of obs.	% in database
Normal observation	9,463,849	71.53
Sales	100,584	0.76
Temporary rebate	254,542	1.92
Temporary replacement	129,687	0.98
Price not observed (outlet closed, etc.)	1,405,579	10.62
New product / outlet	121,549	0.92
Replacing product	509,223	3.85
Replaced product	499,008	3.77
Quarterly observation	746,318	5.64
total	13,230,339	100.00

"Missing" prices appear not to be uncommon. Those unobserved prices are worth being precisely considered as their treatment can significantly alter the duration of price spells. In that respect, we have adopted an approach which differs from that of the statistical institute. When a product is temporarily absent from an outlet (or when an outlet is temporarily closed), INSEE evaluates the missing price by applying to the base price the average increase of prices of the same kind of products in the same area. Then, although this approach is perfectly suited for computing an aggregate price index, such a evaluation cannot be regarded as a proxy for the unobserved price.

Our approach then differs from that of INSEE and can be described as follows: temporarily unobserved prices (dated t) are replaced by the previous price when that previous price (dated t-1) is equal to the subsequent price (dated t+1). More formally when $P_{j,t,k}$ is a "pseudo-observation" (in fact has been replaced) and that $P_{j,t-1,k}$ =$P_{j,t+1,k}$, we set $P_{j,t,k} = P_{j,t-1,k}$.

4.2 Some stylized facts

A set of descriptive statistics about price spells is reported in Table 4 .

Table 4: Observed durations of price spells

Population	All price spells	Price spells averaged by elementary group of goods	Price change frequency measure of duration
Mean	5.28	7.14	8.38
Median	3	5.90	6.19
Min	1	1.22	
25^{th}percentile	1	4.54	4.30
75^{th}percentile	7	9.10	11.60
Max	104	33.93	
Standard deviation	6.73		
Number of obs.	2377682	1328	104

The number of observed trajectories (one trajectory is a sequence of prices of one product in one given outlet) is $K = 754,220$. The average length of an observed trajectory is $\overline{L} = 16.65$ months. The average number of spells per trajectory is 3.15 so the overall average unweighted duration over all price spells is evaluated to be $\overline{T} = 5.28$ months. The distribution around this mean is very asymmetric : the median duration of spells is 3 months. There is a very high mode at duration 1 month. Also there is a very long right-side tail with 25% of spells lasting more than 7 months.

The average duration clearly over-weights the products with short durations (since for a given trajectory length, a larger number of spells is observed). Consolidating first at the elementary group level, and using the CPI weights to aggregate, the weighted average duration of price spells is 7.14 months. Restricting to the "pre-euro" period (pre-2001) does not modify results much, yielding a weighted average duration of 7.07 months.

One can also get an indirect measure of the duration of price spells by using the relationship between durations and the frequency of price changes. The weighted average frequency of price changes over the baseline period is 0.189. The estimator based on inverting the weighted average frequency is $\overline{T}^{\overline{F},H} = (1/\overline{F}^{W}) = 5.29$, or under the "continuous -time assumption" $\overline{T}^{\overline{F},H} = -1/\ln(1-\overline{F}^{W}) = 4.77$. The weighted average of implied durations (continuous time) is $T^{\overline{F},W} = 8.38$. As expected $T^{\overline{F},W}$ is both larger than the inverse of the aggregate frequency ($\overline{T}^{\overline{F},H}$) and larger than the direct average of durations. The latter inequality results from censoring (note that the discrete time assumption used in the direct measurement of durations tends to attenuate the discrepancy). The weighted median of inverse frequencies is 6.20. The difference between the median and the weighted average is due to some frequencies of price changes F_j being close to zero for a few elementary groups. Overall, our best estimate for the average duration of prices is thus around 8 months. However such a statistic is a very crude summary of the data since there is considerable heterogeneity across sectors.

Table5: Duration of price spells by group of products						
	All price spells	Food	Durable goods	Clothes	Other manuf.	Services
Mean	7.14	5.23	5.44	5.56	5.28	10.64
Median	5.90	4.78	5.16	5.07	5.29	11.07
Min	1.22	2.57	2.60	1.94	1.22	2.19
25^{th}percentile	4.54	4.06	4.37	4.64	1.49	7.39
75^{th}percentile	9.10	5.52	6.14	5.99	7.04	12.50
Max	33.93	13.84	9.73	11.95	19.71	33.93
Standard deviation	4.04	2.06	1.57	1.38	3.28	4.10
# of elem. groups	1328	409	58	225	371	265

Note: price spells are averaged by elementary groups

Price rigidity, as measured by the direct measure average duration of price spells, strongly vary across sectors (Table 5). The main relevant contrast seems to be between services and other types of goods. The weighted average duration of a price is twice larger in the service sector (10.64 month) than in the manufacturing sectors (durable goods, clothing, other manufactured goods) and in the food sector (around 5 months).

4.3 The unconditional hazard rate functions for price changes

We now turn to testing the predictions of pricing models. Given that our database contains about 2.3 millions price spells, it is not possible to expect to estimate duration models on such a huge database. We have then proceeded to a random sampling to extract a 2% sample from this database. This leaves us with a sample of more than 40,000 spells, covering various categories of goods/services and types of outlets.

We have first estimated a piece-wise constant hazard model on the whole sample as well as on the different sub-groups corresponding to the different types of goods in order to get a rough idea of the shape of the hazard function. Because the maximum possible length is very high, we have constrained the hazard to be constant after 25 months[8] and the hazard rate is thus plotted until $t = 25$. The hazard is presented in Figure 1.

As Figure 1 clearly shows, the hazard rate of price changes seems to be decreasing, which is in strong opposition with the theoretical predictions of the

[8] Let us remind that, overall, only about 5% of spells last for more than 24 months.

models presented in section 2. To interpret this result our favored option is to rely on heterogeneity in pricing behavior. Indeed, it is well-known that aggregating distinct populations of which each can be characterized by a constant but different hazard leads to a "pooled hazard" which appears to be decreasing (see e.g. Kiefer, 1988). Then, it is worth to have a further look at possible heterogeneities in our population. Indeed, as Figures 2 to 6 indicate, the hazard function clearly differs from one category of good to the other.

From Figures 2 to 6, several conclusions emerge:

1) On one hand, we still get decreasing hazard functions for food and manufactured goods. The shape of the hazard functions for durables show a clear quarterly pattern which may be due to the periodicity of many price quotes since, for those goods, many price quotes are made on a quarterly basis. The pattern of the hazard for clothes is very peculiar. One can observe a huge spike at 6 months, which undoubtedly corresponds to the seasonal pattern of clothing items.

2) On the other hand, the hazard function for services shows an almost constant hazard except for a very significant spike at 12 months. One can infer from these estimates that the pricing behavior in the services sector is a mix between a "Taylor type" behavior with a 12 months duration (people revise their prices every year) and a "Calvo type" behavior where the probability of changing prices is about 15%, irrespective of the time elapsed since the previous price change. It remains to find out what are the distinctive characteristics of those two sub-populations.

The main challenge is then to see whether a better account for observed and unobserved heterogeneity would allow us to identify more clearly hazard patterns that might be in better accordance with the theoretical models. We have decided to focus on manufactured goods. Indeed, the clothing items and the durable goods appear to be affected by very specific phenomena that we leave for future research.

5 Accounting for heterogeneity and covariates : econometric results

A first characteristic that clearly makes a strong difference across spells is the existence of sales or rebates. Indeed, those spells that correspond to such events are very likely to be short lived, typically less than 3 months. Moreover, the impact of covariates such as the cumulated inflation over the spell do not play the same role: "sales price spells" do not end because the cumulated inflation has reached a threshold during the spell but because sales are temporary by nature. Then, in order to limit the consequences of such an heterogeneity, we have decided, in a first step, to discard price spells corresponding to sales and rebates from our sample.

5.1 Modelling time-varying heterogeneity

Table 6 below provides estimates of the impact of various factors that might affect the likelihood of a price variation, conditional on the duration elapsed since the previous price change. Except for the "September" dummy, all coefficients appear to be significant.

The probability of changing prices in January is higher than what it is for the rest of the year. This is a calendar effect, part of which may be related to the end of price spells associated with sales. The Euro cash change-over has had a significant effect, though rather limited. However, as we will see in the next section, there was some heterogeneity in the effect of the change-over on the probability of changing prices across different types of outlets.

The two variables that appear to have the stronger effect are the dummies accounting for the two VAT changes that occurred in 1995 and 2000. It is interesting to note that there seems to be a near symmetry of the effect of an increase (1995) or a decrease (2000) in the VAT rate; at least as regards the probability of changing prices.

Finally, as could be expected, there are strong differences across different types of outlets: Hyper- and supermarkets appear to change prices more often than the more traditional outlets. Manufactured goods sold in "services outlets" (such as "shampoo" from an hairdresser for example) appear to have more stable prices than when they are sold in other types of outlets. This issue will be more thoroughly analyzed in the next section.

Table 6: Estimates of a piece-wise hazard model with covariates

	All outlets
Euro	0.226**
TVA_1995	1.401***
TVA_2000	1.351***
January	0.378***
September	0.109
cumulated inflation	2.329**
Hypermarkets	ref
Supermarkets	-0.206***
Conven. store+hard discount	-0.499***
General store	-0.221***
Department store	-0.412***
Large area specialists	-0.351***
Traditional corner shop	-0.588***
Services	-0.794***
Number of obs.	4331
Wald: $\lambda_1 = \lambda_2 = ... = \lambda_{12}$	88.3 (p=0.00)
Wald: $\lambda_2 = \lambda_3 = ... = \lambda_{11}$	14.43 (p=0.11)

A consequence of the inclusion of covariates in the model is that the hazard function is slightly "flattened" as compared to the unconditional one. In particular, it is likely that inflation plays a role in such changes as inflation over the course of the spell is a "competitor" to duration dependence for explaining the occurrence of a price change after some time has elapsed.

Then, one cannot reject that the hazard function is constant over the first year, once the particular peaks observed for the 1 month and 12 months durations have been excluded. This can be considered a a rough indication that mixing three populations made of, respectively, "flexible price outlets", "Taylor 12 months outlets" and "Calvo outlets" might be an acceptable way to characterize the heterogeneity in the pricing behavior of outlets .

5.2 Accounting for outlet behaviors heterogeneity

5.2.1 Estimating specific models for different types of outlets

Using an outlet dummy in a proportionnal hazard set-up may not be the most relevant option to account for heterogeneity, for instance if the whole baseline hazard differs across outlets. The estimation results we get for the remaining sample of manufactured goods price spells and by splitting this sample according to the type of outlet are given in the Table 7.

Table 7: Estimates of a piece-wise hazard model with covariates

	Hyper and supermarkets	Other types of outlets
Euro	0.068	0.362**
TVA_1995	1.138***	1.729***
TVA_2000	1.246***	1.409***
January	0.210***	0.541***
September	0.083	0.132*
cumulated inflation	10.412***	4.818**
Hypermarkets	ref	-
Supermarkets	-0.212***.	-
Conven. store+hard discount	-	ref.
General store	-	0.280**
Department store	-	0.081
Large area specialists	-	0.146*
Traditional corner shop	-	-0.100
Services	-	-0.309***
Number of obs.	2230	2101
Wald: $\lambda_1 = \lambda_2 = ... = \lambda_{12}$	75.63 (p=0.00)	34.79 (p=0.00)
Wald: $\lambda_2 = \lambda_3 = ... = \lambda_{11}$	30.86 (p=0.00)	4.59 (p=0.87)

Notes:

1) cumulated inflation at the COICOP5 level

2) significance: *** at 0.01, ** at 0.05, * at 0.10, otherwise not significant

Several results are worth noting. The most noticeable one is that heterogeneity across different types of outlets is clearly confirmed. The estimates at the bottom of the first column show indeed that prices change more often in hypermarkets, then come the supermarkets (small and large). At the opposite, prices remain unchanged for longer periods in services outlets and in traditional corner shops. Indeed, the baseline hazard is significantly higher in the former (about 0.2 on average during the first 12 months) than in the latter category of outlets (about 0.12 over the same period of time).

Another difference relates to the impact of the Euro cash change-over and to that of changes in the VAT rate. It appears that hyper- and supermarkets do not react to those changes as fast as smaller outlets do. This is particularly striking as regards the Euro cash change-over. The dummy is not significant when one restricts the sample to hyper- and supermarkets. This is an indication that those outlets have indeed stuck to the commitment they took in the year before the change-over to freeze their prices for several months. We still have to check whether they did change their prices well in advance and/or caught up their "backlog" afterwards. In much more moderate way, the impact of VAT in large outlet type appears to be lower than the same impact for more traditional stores: since hazard is raised by a factor 1.409 for small outlet in the occurence of a VAT change, while it is raised by a factor 1.246. Note that, since the baseline hazard is higher in supermarket, it is closer to the upper bound of 1, which puts more restriction on the parameter on VAT change.

On the contrary, large outlets seem to adjust more promptly to the cumulated inflation than smaller ones. Given the way the cumulated inflation is considered here (it is measured as the difference to the average changes in the prices of the same type of products), this may be seen as a faster adjustment to changes in the competitiveness of the outlet as compared to the others. Those large outlets may have a better capacity to observe those changes and to react to them than it is the case for smaller-"traditional" shops. This may be related to the last interesting conclusion that can be drawn from those estimates: the seasonality in pricing behaviors (which is quite strongly marked in previous results; e.g. see Baudry et al. (2004)) seems to be stronger in small outlets than in large ones. Indeed, the seasonal dummies we have included in the model take higher values and are more significant for the former than they are for the latter.

On the whole, although this has to be confirmed by further analysis our results indicate a stronger propensity of hyper- and supermarkets to use "state-dependent" pricing rules[9] while the smaller and more traditional outlets would more "time-dependent" price-makers.

Finally, going back to our main issue, that is the shape of the hazard function, our conclusions are mixed. On the one hand, the shapes of the hazard functions for these two groups of outlets clearly confirm their heterogeneity. The hazard function for hyper- and supermarkets is decreasing, which is in contradiction with pricing theories. Whether this is still due to some ignored heterogeneity is still an open question to check. On the other hand the hazard function for smaller outlets is more in line with theory. Here again, we cannot reject its constancy along the first 12 months except for the spikes at 1 and 12 months. As in the case of services that we previously considered, it seems that there would be a mix of three populations: two with fixed length contracts (either short ones or long ones - i.e. for 12 months) and another one better described by Calvo's model. Estimating mixture models, in the line of Alvarez et al. (2004), is a promising avenue to formally assess this hypothesis. Our research agenda is to estimate mixture model incorporating time-varying covariates.

6 Conclusion

To be completed.

[9] An exception would be the Euro cash change-over.

7 References

Álvarez, L., Burriel, P. and I. Hernando, (2004): "Do Decreasing Hazard Functions Make Sense?", Banco de España, mimeo.

Aucremanne L. and Dhyne E. (2004a) "How Frequently do price changes? Evidence based on the micro data underlying the Belgian CPI", ECB Working Paper 331.

Aucremanne, L. and E. Dhyne (2004b), "The frequency of price changes: a panel data approach to the determinants of price changes", National Bank of Belgium, mimeo

Baudry L., Le Bihan H., Sevestre P., Tarrieu S. (2004) "Price Rigidity in France. Some evidence from consumer price micro data" ECB Working Paper 384.

Bils M. and Klenow P. (2002) "Some evidence on the Importance of Sticky prices", NBER Working paper 9069.

Calvo, G.A. (1983) "Stagerred prices in a utility maximising framework", Journal of Monetary Economics, 12.

Cecchetti S. (1986) "The frequency of price adjustement", Journal of Econometrics, 31, 255-274.

Dhyne E., Álvarez L., Le Bihan H., Veronese G., Dias D., Hoffman J., Jonker N., Lunnemann P., Rumler F., Vilmunen J. (2004) "Price setting in the euro area: Some stylised facts from Individual Consumer Price Data", mimeo.

Dias M., Dias D., Neves P. (2004) "Stylised feature of price setting behaviour inPortugal: 1992-2001" ECB Working Paper 332.

Dias, M., C. Robalo Marques and J.M.C. Santos Silva (2004): "Testing price setting theories:evidence from Portuguese micro data", Bank of Portugal, mimeo.

Dotsey M, King R. and Wolman A (1999)"State-dependent pricing and the general equilibrium dynamics of money and output" Quarterly Journal of Economics, 114, 655-690.

Kashyap A.K.(1995) "Sticky Prices: New Evidence From Retail Catalogs" Quarterly Journal of Economics, 110, 245-274.

Kalbfleisch and Prentice (2002) Statistical Analysis of failure Time Data

Klenow P. and Kryvtsov O. (2003) "State-Dependent or Time Dependent Pricing: Does It Matter for Recent US Inflation?", Manuscript, Federal Reserve Bank of Mineapolis.

Kiefer N. (1988) "Economic Duration Data and Hazard Functions", Journal of Economic Literature, vol. 26, pp 646-679.

Lancaster T. (1997) The econometric analysis of duration data, Cambridge University Press

Meyer B. (1990) "Unemployment Insurance and Unemployment Spells" Econometrica, 54, 757-782.

INSEE (1998) Pour comprendre l'Indice des prix, INSEE Méthodes,$81-82$.

Rotemberg J. (1982) "Monopolistic Price Adjustment and Aggregate Output" Review of Economic Studies 49, 517-531.

Sheshinski E. and Weiss Y. "Optimum Pricing under Stochastic Inflation" Review of Economic Studies, 50, 513-529.

Taylor J. (1980) "Aggregate Dynamics and Staggered Contracts" Journal of Political Economy, 88, 1-22.

Taylor J. (1998) "Staggered Price and Wage Setting in Macroeconomics" Handbook of Macroeconomics, vol I, J. Taylor and M.Woodford, ed. chapter 15.

Weiss, Y. (1993) "Inflation and price adjustment: a survey of findings from Micro-data", in E. Sheshinski and Y.Weiss, editors, Optimal Pricing, and the cost of price adjustment, MIT Press.

Wolman A (1999) "Sticky Prices, Marginal Cost, and the Behavior of Inflation", Federal Reserve Bank of Richmond, Economic Quarterly, Vol 85/4, Fall.

Wolman A (2000) "The frequency and cost of individual price adjustment", Federal Reserve Bank of Richmond, Economic Quarterly, Vol 86/4, Fall.

8 Appendix 1 - Specific data issues

Dealing with the euro cash change-over

When we consider full sample 1994:7 to 2003:2, price are collected in euros from 2002:1 onwards. We deal with this numéraire issue by dividing all prices prior to 2002:1 by 6.55957, the official Franc/euro exchange rate, without rounding [notice that, had we rounded prices, then some subsequent but different prices expressed in French Francs would have been rounded to the same price in euros, thus spuriously merging different price spells]. At the time of euro cash changeover (january 2002) however, all price were set in euro rounding up to the second decimal. In order to construct price spells we adopted the following rule: if the price (in euro) in december 2001 rounded up to the secon decimal is equal to the price in january 2002, then the two prices are considered as part of the same spell.

Dealing with corrections for temporarily unobserved prices.

When a product is temporarily absent from an outlet (or when an outlet is temporarily closed) the INSEE evaluates price, in order to compute the price index, by applying the avera increase in the same kind of products in the same area. The price evaluated by the INSEE cannot be regarde as a proxy for the unobserved price (it aims at producing an unbiased aggregate index).

Our approach is the following : temporarily unobserved prices (dated t) are replaced by previous price when previous price (dated t-1) is equal to subsequent price (dated t+1).

More formally when $P_{j,t,k}$ is a "pseudo-observation" (in fact has been replaced) and that $P_{j,t-1,k} = P_{j,t+1,k}$, we set $P_{j,t,k} = P_{j,t-1,k}$.

Note that our treatment is different from that adopted by INSEE, which replaces an unobserved price by the average in the same area. The INSEE, in fact, has to evaluate prices in "real time".(at date t) and by definition cannot use the same rule as us.

Dealing with sales, rebates, changes in taxes.

In the data we can identify whether the observed price corrresponds to sales, or temporary rebates. The proportion of price quote that are sales is 0.76% and temporary reduction amount to 1.92%

Also note that two major specific event occuring during the observed sample: in August 1995 the normal VAT rate was raised from 18.6% to 20.6% and in April 2000 this rate was lowered to 19.6%.

In the baseline analysis, we considered these price change as regular.

To summarize, our baseline data is 1994:7 to 2000:12, where "pseudo observations" (corresponding to temporarily unobserved prices) are corrected accordiing to the rule described above.

9 Appendix 2: Analysing the duration of left-censored (and possibly also right-censored) durations

The analysis of left-censored spells such as the first spell in figure 1 is more complex. Left-censoring may only affect the first spell of a price trajectory, so that we denote it as t_{1j}. As in the case of right censoring, we know that the true unobserved duration t_{1j}^{*} is greater or equal than the observed one t_{1j}. However, taking this into account cannot be done as simply as in the case of right-censoring as in that case, we do not know the starting date of the spell. This has the important implication that, except in the particular case where the hazard rate function is constant, the probability of lasting one more period differs from one spell to the other, depending on the unknown starting date of the spell. We then have to integrate over all possible starting dates of the spell, from $-\infty$ to the beginning of the observation period.

From now on, let us omit the indexes 1 and j for the sake of simplicity. Let us denote by S the random variable defined as the duration of the unobserved part of the spell. As just said above, one must integrate the joint distribution $l(s,t)$ with respect to s:

$$h(t) = \int_0^\infty l(s,t)\ ds.$$

We can write the joint probability $l(s,t)$ that the unobserved part of the spell lasts for s periods and that the observed one lasts for t periods as $l(s,t) = m(t \mid s) \times f_s(s)$ where $f_s(s)$ is the unconditional density of the unobserved part of the spell.

Now, $m(t \mid s)$ is the probability of the observed part of the spell lasting t given the unobserved part lasted s. It is then equal to the joint probability of the unobserved part of spell lasting s and that of the observed part lasting t, divided by the probability that the unobserved part lasted s, which is exactly the survivor function evaluated at s. Indeed, Since the former joint probability is nothing but the probability of the "full" spell lasting $s+t$, we can write:

$$m(t \mid s) = f(t+s)/(1 - F_s(s)).$$

In other words, we adjust the probability of a spell lasting $t+s$ by taking account that, for this to be possible, one must restrict to spells that survived s periods.

Now, $f_s(s)$ the probability for a spell to last s periods can be expressed as the product of having started s periods before, $k(-s)$, by the probability of being still alive after s periods, $1 - F_s(s)$:

$$f_s(s) = k(-s) \times (1 - F_s(s)).$$

Now, one can assume that the so-called "entry rate" $k(-s)$,which measures the probability that a price change did occur s periods before, is constant and does not depend on $s : k(-s) = k, \forall s$. Then,

$$l(s,t) = \frac{f(t+s)}{1 - F_s(s)}.k.(1 - F_s(s)) \ .$$

which leads to,

$$\begin{aligned} h(t) &= k \int_0^\infty f(s+t)\ ds \\ &= \frac{1 - F(t)}{E(T)}. \end{aligned}$$

where the entry rate k is approximated by the inverse of the duration of spells (T) *in the population*. Indeed, defining $v = t + s$, the integral in the expression above can be written as

$$\int_0^\infty f(s+t)\ ds = \int_t^\infty f(v)\ dv = 1 - F(t).$$

9.0.2 Analysing the duration of right and left-censored durations

Having both right and left-censoring appears to be not that uncommon in our sample, as shown in table CCC above. It is then worth trying to generalize the arguments above for finding the distribution of spells that are both right- and left-censored. Based on the argument that the probability of observing a right-censored spell lasting t periods is equal to the probability that a spell has survived t periods at least $(g(t) = 1 - F(t))$, one can write that, knowing that a right-censored spell lasted s periods before the beginning of the observation period, the probability of being still going on after t periods in the observation period is given by:

$$m(t \mid s) = (1 - F(t+s))/(1 - F_s(s)$$

where $1 - F(t+s)$ is the probability that a spell will last at least $t + s$ periods due to right-censoring, and where dividing by $1 - F_s(s)$ is necessary to account for left-censoring as we condition on s and thus, restrict to those spells still alive after s periods. Now, one has to integrate over all possible values of s. Then, as seen in the previous section:

$$\begin{aligned} \widetilde{h}(t) &= \int_0^\infty l(s,t)\ ds \\ &= \int_0^\infty m(t \mid s)\ f_s(s)\ ds \end{aligned}$$

with $f_s(s) = (1 - F_s(s))/E(T)$. Then,

$$\widetilde{h}(t) = \frac{1}{E(T)} \int_0^\infty [1 - F(t+s)]\ ds.$$

Then, setting $v = t + s$, one can rewrite this expression as

$$\widetilde{h}(t) = \frac{1}{E(T)} \int_t^\infty [1 - F(v)]\ dv.$$

Figure 1: Hazard rate (all products/services)

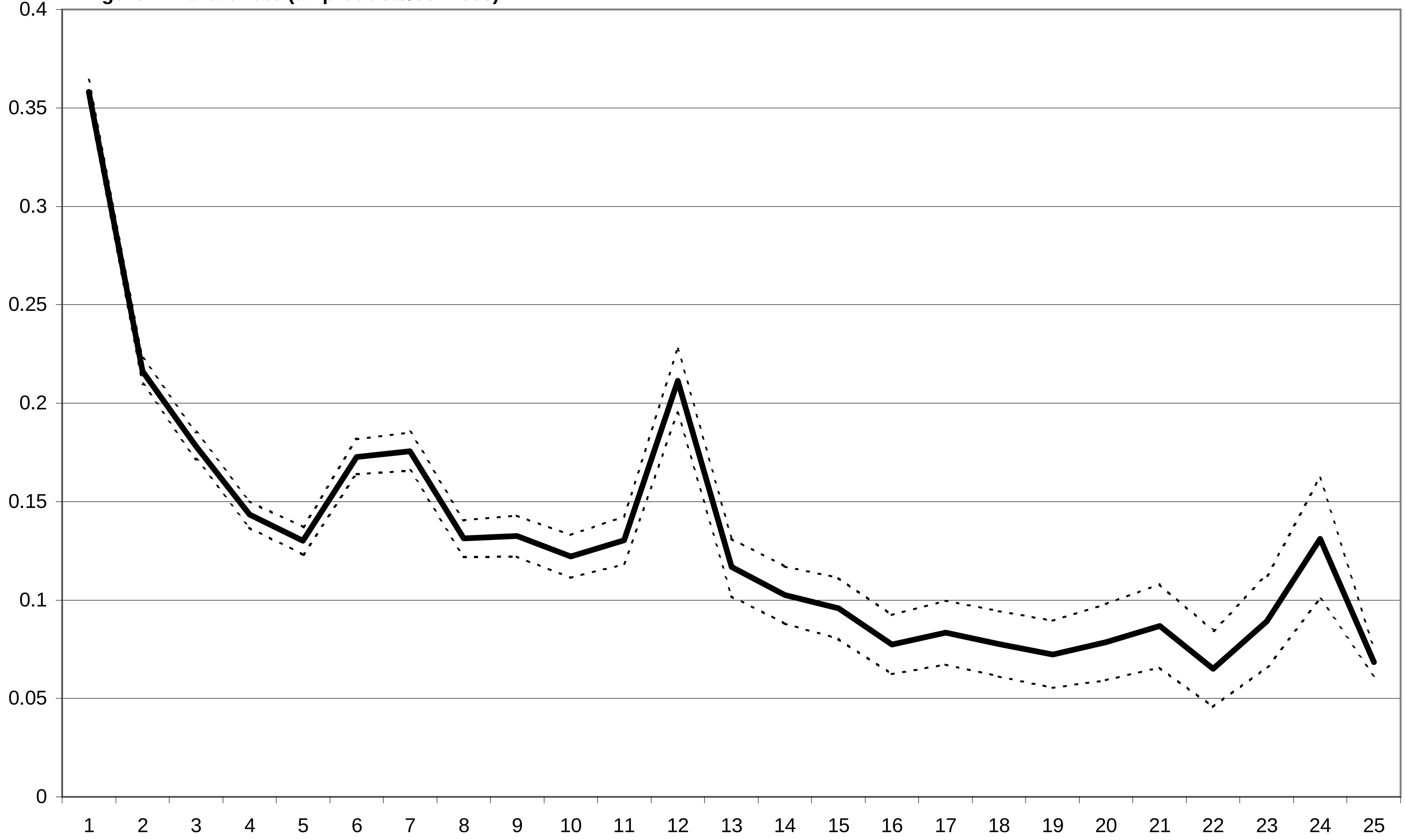

Figure 2: Hazard rate (food products)

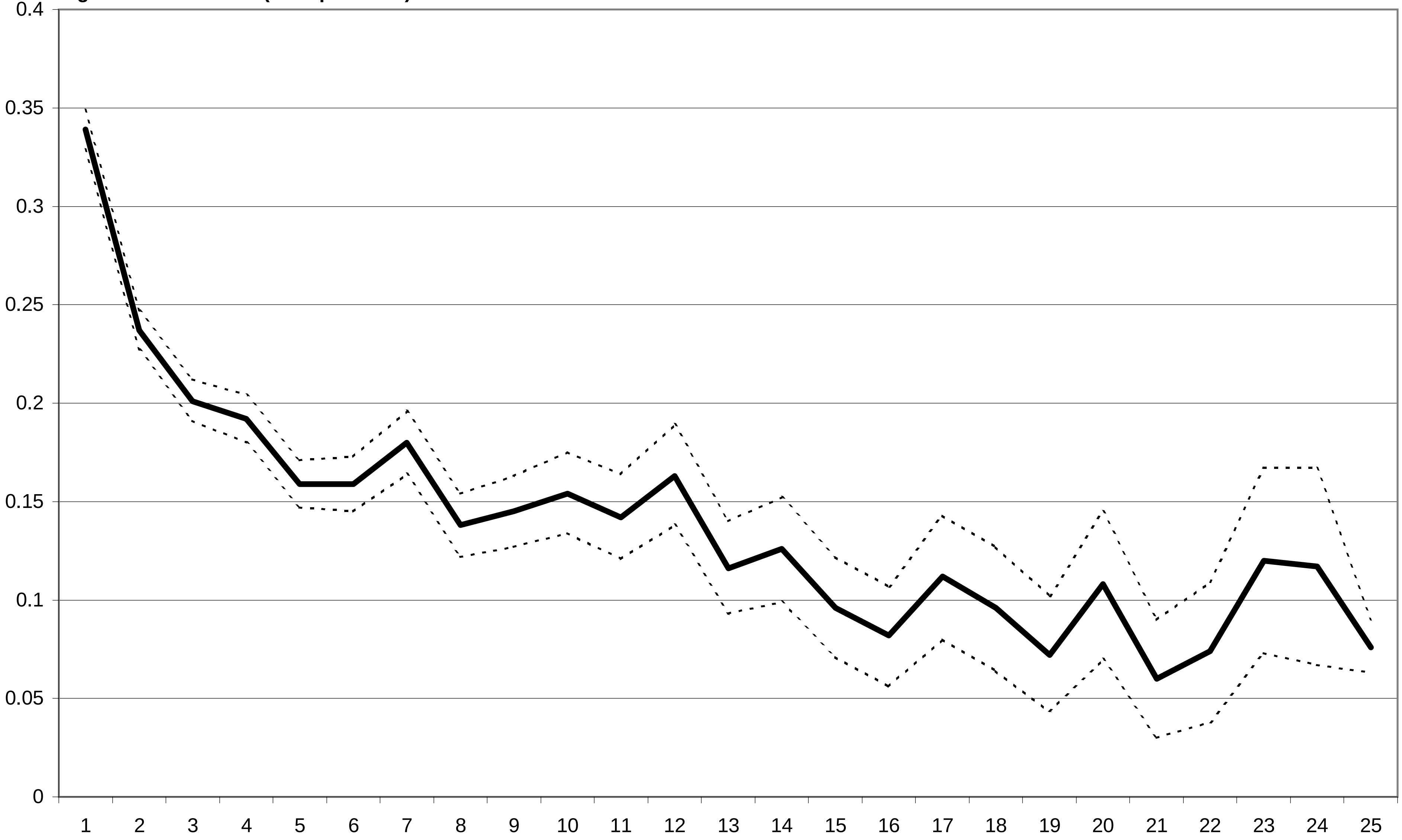

Figure 3: Hazard rate (clothes)

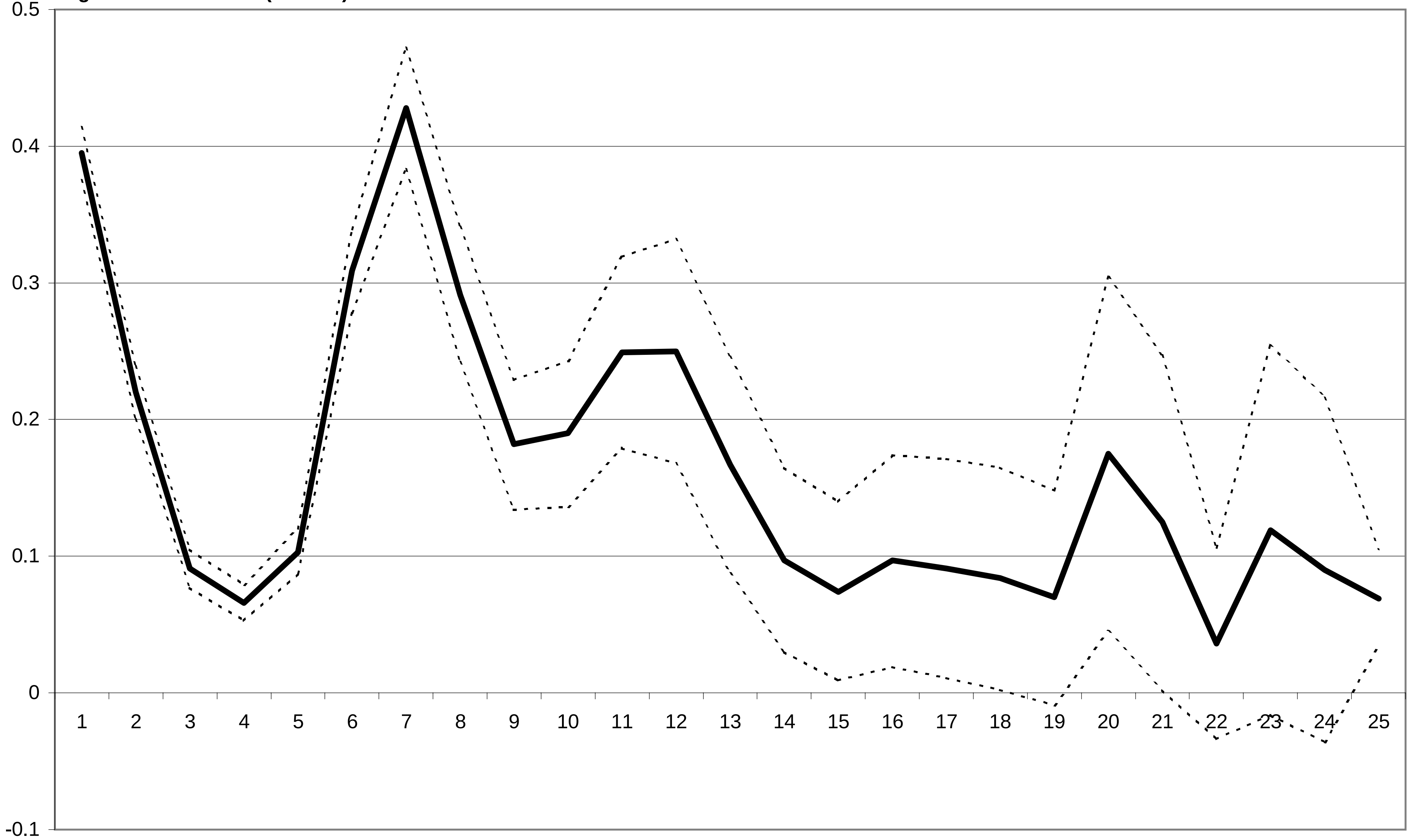

Figure 4: Hazard rate (manuf. goods)

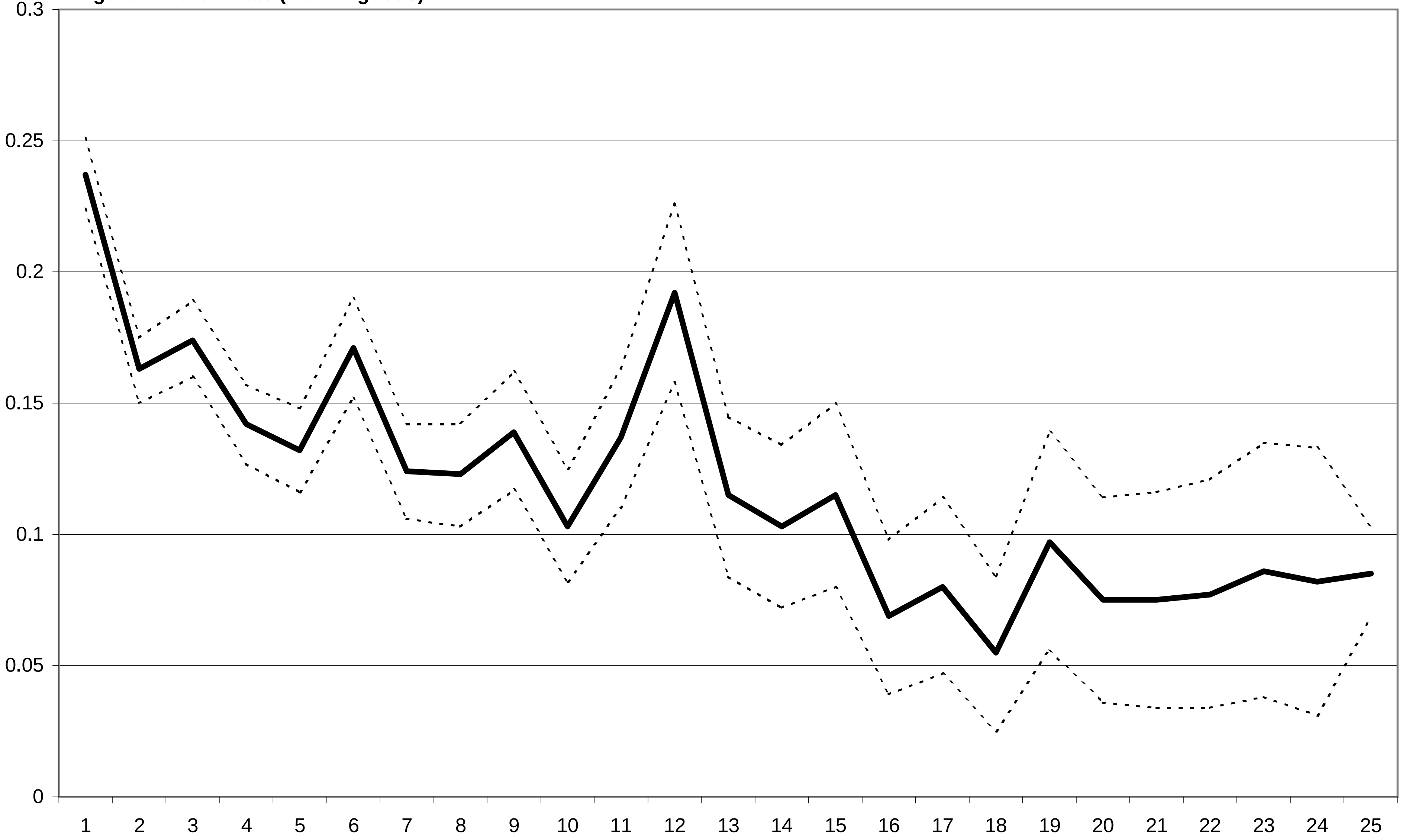

Figure 5: Hazard rate (durable goods)

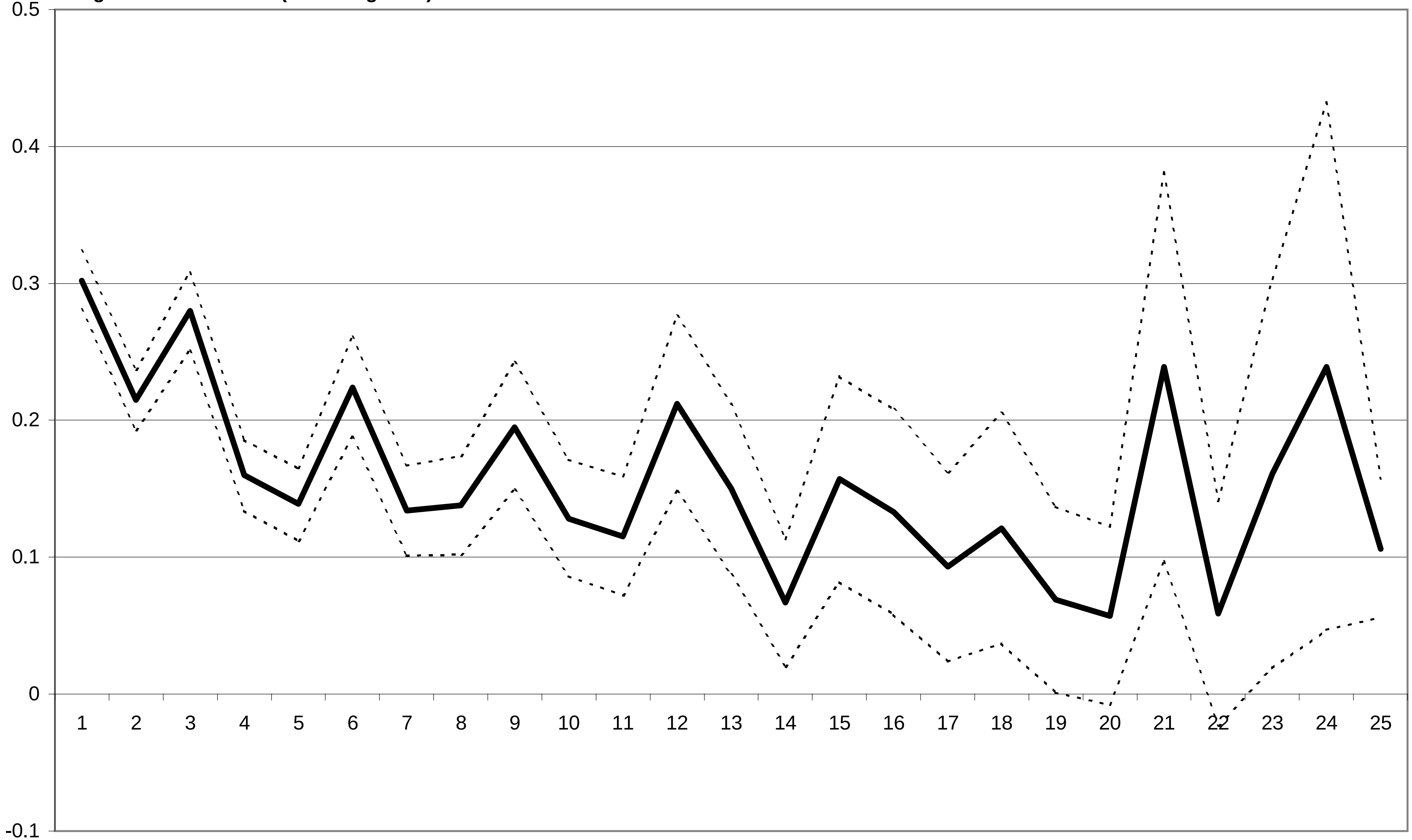

Figure 6: Hazard rate (services)

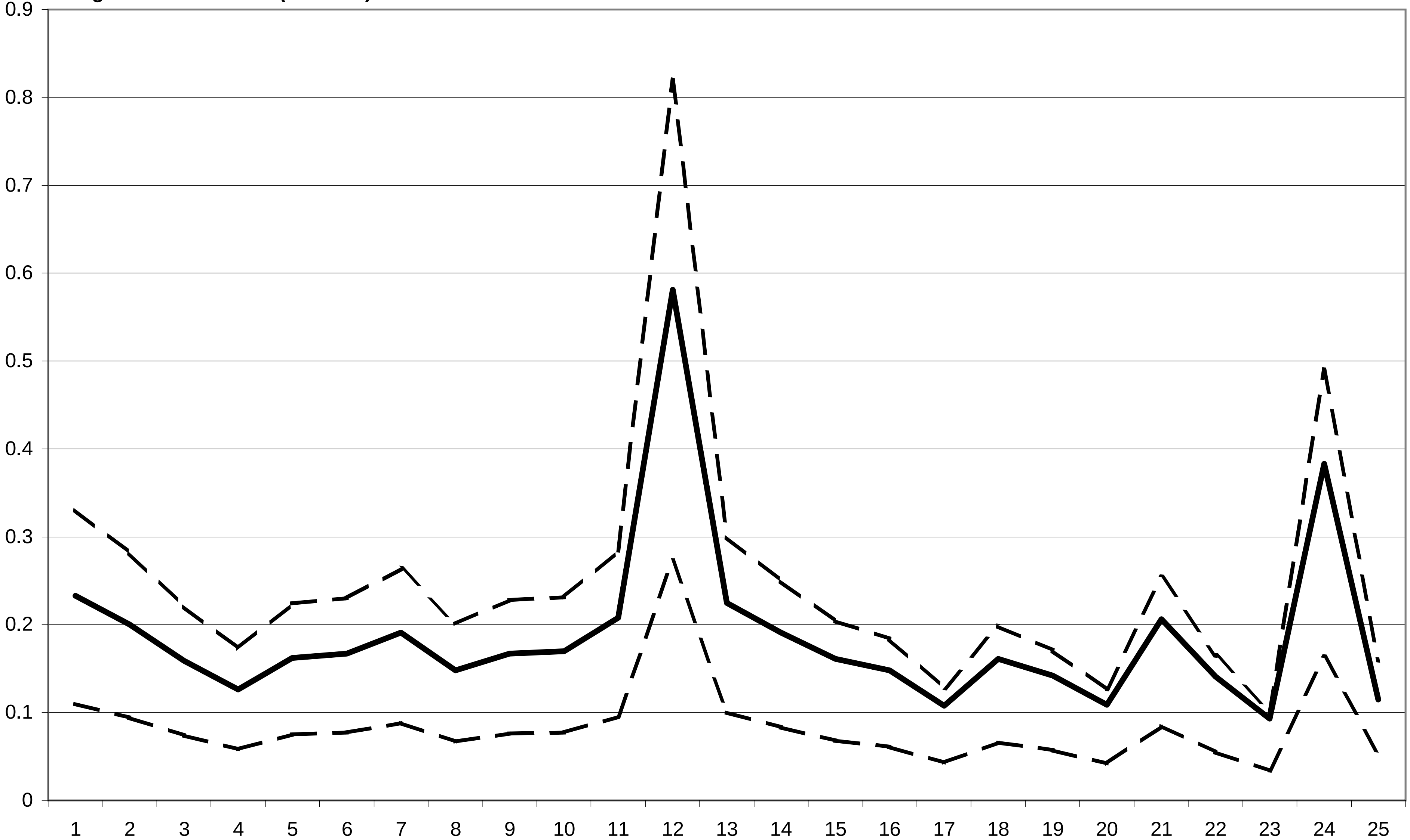

Figure 7: Baseline hazard for maufactured goods in all outlets

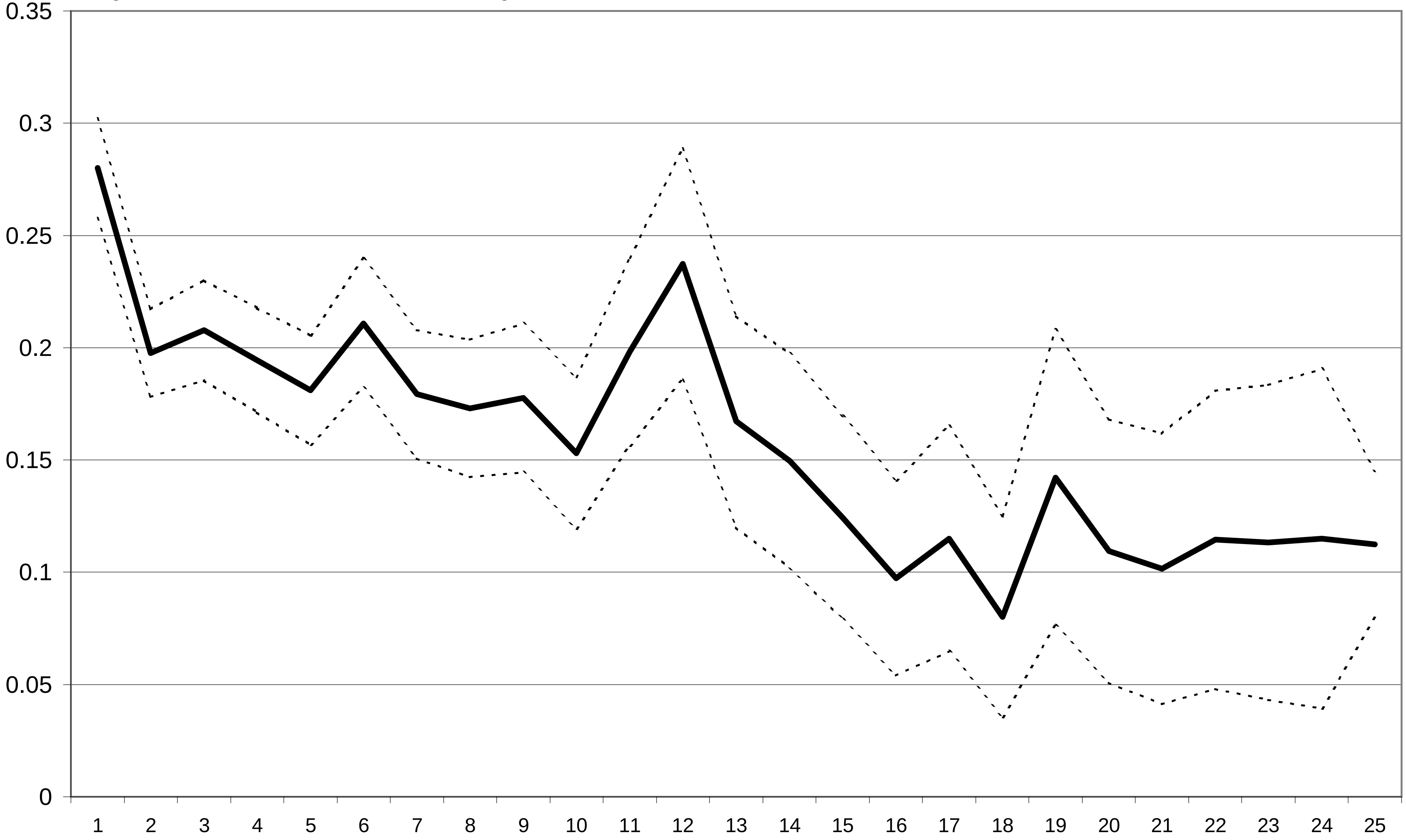

Figure 8: baseline hazard for manufactured goods in hyper and supermarkets

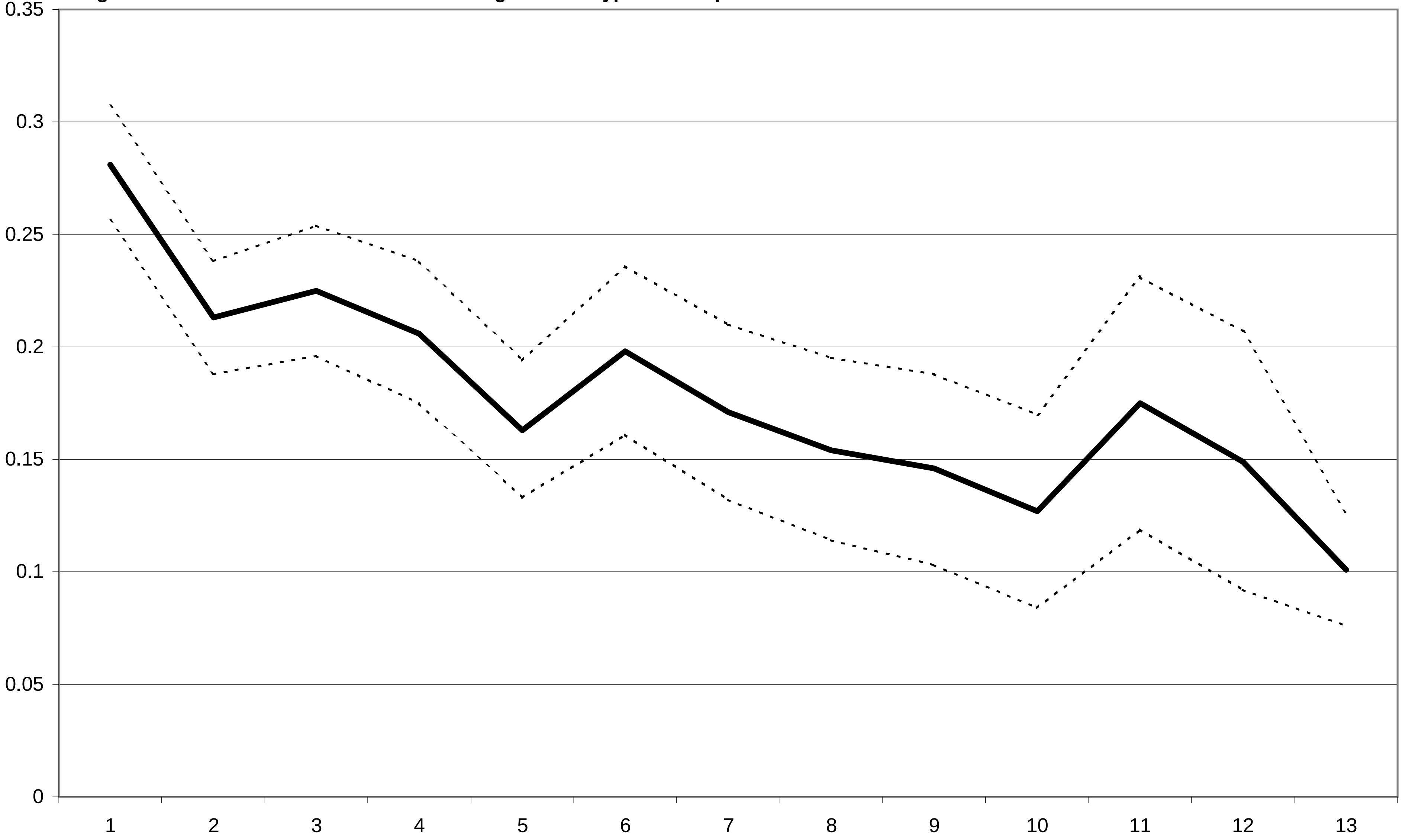

Figure 9: Baseline hazard for manuf. goods in ``smaller'' outlets