

JSPS Grants-in-Aid for Creative Scientific Research
Understanding Inflation Dynamics of the Japanese Economy
Working Paper Series No.31

Goodness-of-Fit Test for Price Duration Distributions

Takayasu Matsuoka

March 10, 2009

Research Center for Price Dynamics
Institute of Economic Research, Hitotsubashi University
Naka 2-1, Kunitachi-city, Tokyo 186-8603, JAPAN
Tel/Fax: +81-42-580-9138
E-mail: sousei-sec@ier.hit-u.ac.jp
<http://www.ier.hit-u.ac.jp/~ifd/>

Goodness-of-Fit Test for Price Duration Distributions

Takayasu Matsuoka*

Graduate School of Economics, Osaka University

March 6, 2009

Abstract

Is the actual price-setting behavior of an individual commodity item consistent with the assumptions of a sticky-price model? Part of the question may formally be addressed by performing a goodness-of-fit test for price duration distributions. For each of the 429 items in the Japanese retail price data for 2000–2005, we fitted the standard parametric models with or without unobserved heterogeneity to the data and tested the goodness of fit. We found that 8.6 percent of the tested items cannot reject the hypothesis that the underlying distribution is exponential, which corresponds to the time-dependent pricing model of Calvo (1983).

JEL classification codes: D40, E31, C41

Key words: Sticky prices, Hazard function, Goodness-of-fit test

*I would like to thank Kanemi Ban, Tsutomu Watanabe, Mototsugu Shintani, and the participants in the study meeting at the Statistics Bureau, Ministry of Internal Affairs and Communications for helpful comments. Any remaining errors are my own responsibility. This study is financially supported by the research fellowships of the Japan Society for the Promotion of Science for Young Scientists. Address for correspondence: 1-7 Machikaneyama, Toyonaka, Osaka 560-0043, Japan. Tel.: +81 90 5009 4141. E-mail: fgp019mt@mail2.econ.osaka-u.ac.jp.

1 Introduction

This paper examines the distributional assumption of the duration of price spells. It forms part of an attempt to construct a formal theory dealing with sticky prices, because existing sticky-price models in macroeconomics explicitly formulate the mechanism of a firm's price change by assuming that the length of price spells follows a certain distribution. One example is the Calvo (1983) model, which assumes that the probability of a firm's price change is determined exogenously and does not change over time. This assumption implies that price spell durations have an exponential distribution with a constant hazard rate. The other example is the Dotsey, King, and Wolman (1999) model, which assumes a fixed cost of adjusting price. This model predicts a monotonically increasing hazard function when the general level of prices continues upward.

In this paper, we examine the distributional assumption using a Pearson-type goodness-of-fit test for censored data, which is based on a maximum likelihood estimation of parametric hazard function (Akritas (1988) and Hjort (1990)). Intuitively, the test statistic compares, on a span-by-span basis, the observed number and expected number of price changes implied by the hypothesized hazard function. Compared to the Wald test frequently used for a specification check for parametric model, our test statistic is quite general, in the sense that it does not require the explicit statement of alternative models that nest the null model. Thus, the hypothesized hazard function is not limited to exponential, but also includes Weibull and other kinds of parametric hazard function. By performing the goodness-of-fit test, we will investigate whether actual price-setting behavior is consistent with the implications of sticky-price models.

Empirical studies on price stickiness have recently focused on the voluminous price dataset underlying the computations of the Consumer Price Index (CPI) and produced a rapidly growing literature. (See, for example, Bils and Klenow (2004); Nakamura and Steinsson (2008) for the United States; Dhyne et al. (2006) for the European Union; and Higo and Saita (2007) for Japan.¹) Since then, researchers have discovered stylized facts at the individual item levels as well as at the aggregated level.² The interest of our analysis lies in

¹This subject has been studied extensively in many countries other than those listed above. See also Álvarez (2007) for a useful survey.

²A remarkable finding at the aggregated level is that the hazard function exhibits a downward slope. Many authors attribute the discrepancy between this empirical finding and the implication of sticky-price models to heterogeneity in price-setting behavior. Using the finite mixture model, Álvarez et al. (2005) and Ikeda and Nishioka (2007) show that the aggregation of several price-setters facing different hazard functions results in a

the process of price changes at the item level. The reason is that the component items of the CPI are highly heterogeneous and, thereby, the estimates of the hazard rate from the pooled sample of several items are influenced by the heterogeneity effect. In this paper, we conduct the goodness-of-fit test for each of the 429 items available in the *Monthly Report on the Retail Price Survey* for 2000–2005.

For the inference of the unknown distribution at the item level, previous studies frequently examine how the nonparametric hazard rate changes with time.³ For example, by doing this Higo and Saita (2007) classify the item-level hazard functions into three categories: flexible type, Taylor type, and decreasing hazard type. However, it can be problematic to draw conclusions about the distributional assumptions on the basis of the graphical representation. One reason is that in some items, we are quite uncertain about which of these shapes characterizes the process of price change: decreasing, constant, increasing, or a mixture of them. More seriously, the precision of the estimate of the hazard rate gets lower as the elapsed time becomes longer and, therefore, the judgment can change according to how we evaluate the hazard rates at the longer duration, which is less precisely estimated. The other reason is that unobserved heterogeneity in price-setting behavior may exist at the item level. In that case, we cannot judge from the graphical representation whether the decreasing hazard function exhibits true duration dependence or merely the heterogeneity effect.

In contrast to the classification of Higo and Saita (2007), we draw a clear distinction as to whether an item rejects the hazard function with a particular shape. For the inference of the unknown distribution, we hypothesize two parametric models: the exponential model and the Weibull model. Consequently, our analysis formally addresses such questions as sample of how many items are from exponential distribution with a constant hazard rate or question whether an item that retains the hypothesis of Weibull distribution with monotonically increasing hazard rates exists or not. Furthermore, in order to account for the unobserved heterogeneity, we model the gamma-distributed heterogeneity that appears multiplicatively in the hazard func-

decreasing hazard function. Nakamura and Steinsson (2008) and Matsuoka (2007) reach the same conclusion using a random effect model in which the item-level heterogeneity is incorporated.

³Exceptionally, Fougère, Le Bihan, and Sevestre (2007) test whether the hazard rates are constant over the elapsed time since the last price change as a specification test for exponential model. They use a piecewise-constant hazard model and test the equality of hazard rates over the span of time. According to their testing procedure, we can confirm our goodness-of-fit test results for exponentiality, such as items that cannot the exponential hypothesis have in common a low frequency of price changes.

tion. Fitting the exponential-gamma mixture model and the Weibull-gamma mixture model to the data, we investigate whether the unobserved heterogeneity, to which we presumably attribute the city characteristics or the pricing strategy of an outlet, can account for the process of price changes at the item level.

Our test results can be summarized in the following four findings. First, 8.6 percent of the tested items (in percentage share of CPI weight) cannot reject the hypothesis that the underlying distribution is exponential, which corresponds to the time-dependent model of Calvo (1983). Second, those items that retain the hypothesis of exponentiality have in common a low frequency of price changes. The frequency of price changes for these items is limited to 11 percent per month. Some items in the subgroups such as Eating out in the Food sector, Repairs and maintenance in the Housing sector, and Recreational goods and services in the Reading and recreation sector lead to the nonrejection of the hypothesis of exponentiality. The results are also verified by the Shapiro and Wilk (1972) test for exponentiality. Third, the well-fitting Weibull hazard models for our data have a duration-dependent parameter that is nearly equal to or less than one, which provides little support for monotonically increasing hazard function at the item level. Fourth, the performance of the model with unobserved heterogeneity is worse than that of the model without unobserved heterogeneity. In view of the results, it is not easy to attribute the reason for the decreasing hazard estimates simply to the unobservable characteristics of the survey cities.

The remainder of the paper is organized as follows. Section 2 formulates the process of price change using the notion of counting processes, on which our test statistic is based. In Section 3, we show the principle of our goodness-of-fit test and the related graphical method using nonparametric estimates of the cumulative hazard rate. In Section 4, we summarize the properties of our four hypothesized models. In Section 5, we show the nonparametric test for an exponential hypothesis by Shapiro and Wilk (1972). In Section 6, after we describe our price data, we discuss the results of the goodness-of-fit test. Section 7 concludes our analysis.

2 Counting processes and Martingales

Our test statistic is based on the theory of counting process.⁴ A counting process is a stochastic process whose value counts the number of events. It is an increasing (right continuous) process with jumps of size 1. In order to

⁴See Kalbfleisch and Prentice (2002) and Andersen et al. (1993) for details.

construct the goodness-of-fit test statistic, a good starting point is to describe two counting processes: the failure process and the at-risk process.

In our analysis, “failure” means an observed price change of an item. The price change of the i th item ($i = 1, \dots, I$) in the j th city ($j = 1, \dots, J(i)$) is defined as $P_{t^*}^{ij} \neq P_{t^*-1}^{ij}$, where t^* is a calendar time. In the following analysis, we do not focus on the calendar time when price changes occur, but on the elapsed time over which price remain unchanged, that is, the time between two consecutive price changes. Formally, the elapsed time of the k th price spell of the i th item in the j th city, t_{ijk} ($k = 1, \dots, K(i, j)$), is defined as follows: let t_{ijk}^* be the calendar time when the k th event of price change occurs in the j th city. The event indicates the onset of the risk of a price change for the k th price spell. Suppose our observation of the k th spell is completed, i.e., we observe the price spell until the $k + 1$ th (next) price change occurs. Then we have the elapsed time of the k th price spell, $t_{ijk} \equiv t_{ijk+1}^* - t_{ijk}^*$.

Next suppose that we can no longer observe the k th price spell by reason other than its price change and this incomplete state of the observation arises at (calendar) time c_{ijk}^* . In this case, we say that the k th price spell is right censored at time c_{ijk}^* . Right censoring can occur, for example, when a price spell ends with an item substitution, or when the price change has not occurred by the end of the observation period. The censoring time c_{ijk} is also defined as the elapsed time since the last price change and therefore $c_{ijk} \equiv c_{ijk}^* - t_{ijk}^*$.

Now we are in a position to understand the failure and the at-risk processes. Suppose that the k th price spell of the i th item in the j th city remains unchanged for t_{ijk} . The failure process, in other words, the price change process, is defined as

$$N_{ijk}(t) \equiv 1(t_{ijk} \leq t). \quad (1)$$

The process of price change is a sequence of zero and one. For example, a price spell with a five-month duration can be represented by the sequence of $\{0 \ 0 \ 0 \ 0 \ 1\}$. The at-risk process of price changes can be written as

$$Y_{ijk}(t) \equiv 1(t_{ijk} \geq t, c_{ijk} \geq t), \quad (2)$$

where c_{ijk} denotes the censoring time. The at-risk process takes one as long as we observe that the price does not change just before time t . Hence, if $Y_{ijk}(t) = 1$, the price spell is exposed to the risk of price change. These two processes are the basic building blocks on which our test statistic is constructed.

Superposing these individual processes, we obtain the number of observed price spells changed in $[0, t]$, $N_i(t) = \sum_j \sum_k N_{ijk}(t)$ and the number of price spells at risk just before time t , $Y_i(t) = \sum_j \sum_k Y_{ijk}(t)$. Denoting a time just before t by t^- , we define $dN_i(t) \equiv N_i(t^- + dt) - N_i(t^-)$ as the increment of $N_i(t)$ over the small interval $[t, t + dt)$. Now we are able to construct a model that describes the transition of price spells. The essential part of the model is the hazard function $\lambda_i(t)$ which, multiplied by $Y_i(t)$, implies the expected number of price changes in the i th item at time t . The model can be written as follows:

$$E[dN_i(t)|\mathcal{F}_{t-}] = Y_i(t)\lambda_i(t)dt, \quad (3)$$

where \mathcal{F}_{t-} is an information set during the period $[0, t)$, which is referred to as the filtration of counting process.

For the model defined in Equation (3), consider the following process

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda_i(s)ds, \quad \text{for all } t \geq 0. \quad (4)$$

The process $M_i(t)$ is called a counting process martingale⁵ and satisfies the condition $E[dM_i(t)|\mathcal{F}_{t-}] = 0$ for all t . The second term on the right-hand side is called the compensator of the counting process, which equals the cumulative sum of the expected number of price changes in $[0, t]$. Denoting the compensator by $E_i(t)$, Equation (4) can be written compactly as $M_i(t) = N_i(t) - E_i(t)$. The property of a martingale process plays a crucial role in our test statistic.

3 Tests for distributional assumptions

In this section, we will show the two kinds of model checking. One is a graphical check that compares the theoretical cumulative value of the hazard function with the nonparametric estimates of cumulative hazard rate, which is called the Nelson–Aalen estimator.⁶ The other is the goodness of fit test statistic on which our analysis is based. They are conceptually equivalent but the latter is more rigorous. In the following description, we shall omit the subscript i to simplify the notation while all quantities are calculated on an item basis.

⁵The first and second term on the right-hand side of Equation (4) are both the sum of individual processes, which is called a superposed process. The important property of the superposed process is that the sum of individual martingale process is also a martingale: Suppose $E[dM_{ijk}(t)|\mathcal{F}_{t-}] = 0$ for all t , $j = 1, \dots, J(i)$, $k = 1, \dots, K(i, j)$, then $M_i(t) = \sum_j \sum_k M_{ijk}(t)$ is a martingale with respect to \mathcal{F}_t (See Kalbfleisch and Prentice (2002)).

⁶Andersen et al. (1993) illustrate the details of the graphical check.

3.1 Graphical check using the Nelson-Aalen estimator

The correct specification of a certain parametric hazard function can be checked by graphically comparing the Nelson–Aalen estimator⁷

$$\hat{A}(t) = \int_0^t \{J(s)/Y(s)\}dN(s) \quad (5)$$

with the parametric cumulative hazard function

$$A(t; \theta) = \int_0^t J(s)\lambda(s; \hat{\theta})ds, \quad (6)$$

where $J(s) \equiv 1(Y(s) > 0)$ so that we can define the processes only in the regions where a price spell is observed. In Equation (6), $\lambda(s; \hat{\theta})$ is a parametric hazard function with the maximum likelihood estimator $\hat{\theta}$. We will specify the hypothesized hazard functions in Section 4. If the difference $\hat{A}(t) - A(t; \hat{\theta})$ is considerably large over $t \in [0, T]$, where T denotes the largest of the observation time, the parametric model is judged as a poor one. The goodness-of-fit test described below gives a criterion for this kind of graphical check.

3.2 Goodness-of-fit test

Now we describe the goodness of fit test for the censored data. Our null hypothesis is the following composite hypothesis

$$H_0 : \lambda(t) \in \mathcal{L} = \{\lambda(t; \theta); \theta \in \Theta \subset \mathcal{R}^p\}, \quad (7)$$

that is, the true hazard function $\lambda(t)$ belongs to the parametric family of hypothesized hazard functions \mathcal{L} . When the null hypothesis holds, $\lambda(t)$ can be specified by $\lambda(t; \theta_0)$, where θ_0 is the p -dimensional true parameter vector. Since θ_0 is unknown, we employ the maximum likelihood estimator $\hat{\theta}$ corresponding to the hypothesized hazard function. This test statistic, which is based on the maximum likelihood estimator, is originally derived by Akritas

⁷Equation (5) is expressed by using the notion of the Lebesgue–Stieltjes integral. When the integrator $N(s)$ is a right continuous step function, it will have many jumps at each of the points a_1, a_2, \dots , where $\Delta N(a_n) = N(a_n) - N(a_n^-) > 0$. In this case, the Stieltjes integral becomes

$$\int_0^t \{J(s)/Y(s)\}dN(s) = \sum_{n:0 < a_n \leq t} J(s) \frac{\Delta N(a_n)}{Y(a_n)}.$$

See Fleming and Harrington (1991) for detail.

(1988) but its general framework is proposed by Hjort (1990). We follow Hjort (1990) and present the principle of the test.

Hjort (1990) derives the limit distribution of the following process

$$H_n(t) = \sqrt{n} \int_0^t K_n(s) J(s) \{(1/Y(s)) dN(s) - \lambda(s; \hat{\theta}) ds\}, \quad (8)$$

where $K_n(t)$ is an almost surely bounded weighting process.⁸ When $K_n(t) = 1$, the statistic reduces to $H_n(t) = \sqrt{n} \{\hat{A}(t) - A(t; \hat{\theta})\}$, which compares the Nelson–Aalen estimates and the parametric cumulative hazard function. Setting $K_n(t) = Y(t)/n$, the statistic now becomes $H_n(t) = \{N(t) - E(t; \hat{\theta})\}/\sqrt{n}$ and resolves into the comparison between the observed and expected number of price changes. Our analysis employs the latter case.

As we mentioned above, the test statistic compares these quantities on a span-by-span basis. Let $0 = a_0 < \dots < a_m = T$ be a division of analysis time into m cells $I_l = (a_{l-1}, a_l]$, $l = 1, \dots, m$. We then define

$$Q_{n,l} = \frac{1}{\sqrt{n}} \left[N(a_{l-1}, a_l] - \int_{I_l} Y(s) \lambda(s; \hat{\theta}) ds \right] = \frac{1}{\sqrt{n}} (N_l - E_l), \quad (9)$$

and $Q_n = (Q_{n,1}, \dots, Q_{n,m})'$. The numerator of $Q_{n,l}$ is the difference between the observed and expected number of events in the l th interval.⁹ Hjort (1990) shows that

$$Q_n \xrightarrow{\mathcal{D}} N(0, R), \quad (10)$$

as $n \rightarrow \infty$. On this basis we obtain the following test statistic

$$X_n^2 = Q_n' \hat{R}^- Q_n, \quad (11)$$

where \hat{R}^- is the generalized inverse of any consistent estimator of the covariance matrix R . The $m \times m$ covariance matrix R can be written in the following form

$$R = D - S' \Sigma^{-1} S. \quad (12)$$

Note that Equation (12) implies the generalized inverse of R

$$R^- = D^{-1} + D^{-1} S' G^- S D^{-1}, \quad (13)$$

⁸The derivation of the limit distribution requires a fundamental convergence theorem, the so-called martingale central limit theorem. For the details of the theorem, see Kalbfleisch and Prentice (2002). The formal proof of the weak convergence of the process in Equation (8) is given by Hjort (1990) and is also available in Andersen et al. (1993).

⁹Hjort (1990) proposes that we should choose the m cells so that each cell contains at least five observations.

where G^- is the $p \times p$ generalized inverse of $G = \Sigma - SD^{-1}S'$. The component parts of R are as follows: D is a diagonal matrix with elements $d_l = \int_{I_l} y(s)\lambda(s; \theta_0)ds$, where $y(s)$ is a probability limit of $Y(t)/n$; S is the $p \times m$ matrix (b_1, \dots, b_m) , in which $b_l = \int_{I_l} y(s)\psi(s; \theta_0)\lambda(s; \theta_0)ds$, where $\psi(s; \theta_0)$ is the p -dimensional score vector of $\log\lambda(s; \theta_0)$; and $p \times p$ covariance matrix $\Sigma = \int_0^T y(s)\psi(s; \theta_0)\psi(s; \theta_0)'\lambda(s; \theta_0)ds$. These quantities are naturally estimated by

$$\hat{d}_l = E_l/n = n^{-1} \int_{I_l} Y(s)\lambda(s; \hat{\theta})ds, \quad (14)$$

$$\hat{b}_l = n^{-1} \int_{I_l} Y(s)\psi(s; \hat{\theta})\lambda(s; \hat{\theta})ds, \quad (15)$$

$$\hat{\Sigma} = n^{-1} \int_0^T Y(s)\psi(s; \hat{\theta})\psi(s; \hat{\theta})'\lambda(s; \hat{\theta})ds. \quad (16)$$

Using Equations (9) and (13)-(16), we can simplify the test statistic

$$\begin{aligned} X_n^2 &= Q_n' \hat{D}^{-1} Q_n + Q_n' \hat{D}^{-1} \hat{S}' \hat{G}^- \hat{S} \hat{D}^{-1} Q_n \\ &= \sum_{l=1}^m \frac{(N_l - E_l)^2}{E_l} + V_n' \hat{G}^- V_n, \end{aligned} \quad (17)$$

where $V_n = \sqrt{n} \sum_{l=1}^m \{(N_l - E_l)/E_l\} \hat{b}_l$ and $\hat{G}^- = (\hat{\Sigma} - \sum_{l=1}^m \hat{b}_l \hat{b}_l' / \hat{d}_l)^-$.

Under the null hypothesis (7), the test statistic converges in distribution to chi-squared distribution with degrees of freedom $df = \text{Rank}(R)$, which is equal to the number of cells, m . For a large value of the statistic, say $X_n^2 \geq \chi_{\alpha, m}^2$, where $\chi_{\alpha, m}^2$ is the upper α critical point of a chi-squared distribution with m degrees of freedom, we reject the hypothesis.¹⁰

[Table 1 about here.]

Table 1 illustrates the procedure of the goodness-of-fit test at the item level. Suppose we wish to test the hypothesis that the price duration of Coffee (eating out) is a sample from an exponential distribution. The transition of price spells is summarized in the statistics shown in the first four columns. From this data, we obtain the maximum likelihood estimator that corresponds to the exponential model $\hat{\theta}$ (in this case, $\hat{\theta} = 0.054$, thus the mean duration is 18.5 months). We then divide the observation period into

¹⁰When the dimension of the parameter vector $p \geq 2$, the computation of the generalized inverse of \hat{R} is quite intractable. Hjort (1990) points out a slightly conservative test procedure that rejects H_0 if $X_{0,n}^2 = \sum_{l=1}^m (N_l - E_l)^2 / E_l \equiv \sum_{l=1}^m Z_l \geq \chi_{\alpha, m}^2$. Except for a one-dimensional case, we follow his remark and conduct the conservative version of the test.

m cells. Following the remark shown in Footnote 9, when we observe less than five price changes at a point in the interval, we add the number to the number of price changes at the next point until the total number of price changes in the cell reaches at least five. The boxes in the table indicate that process. For example, we observe 7(=2+5) complete price spells that end with a price change in the first box, which form the observed price changes in the 10th cell, N_{10} . Consequently, the number of cells m amounts to 21.

From the calculations shown in the last column in the table, we obtain $X_{0,n}^2 = \sum_{l=1}^m Z_l = 22.18$. The second term on the right-hand side of Equation (17) is calculated to be 1.28. Therefore, we have the test statistic for exponentiality

$$X_n^2 = 22.18 + 1.28 = 23.46, \quad (18)$$

which has approximate chi-squared distribution with $m = 21$ degrees of freedom. The associated p-value is 0.320, which leads us to retain the null hypothesis of exponentiality.

4 Parametric models

In order to make the inferences about unknown price duration distributions, we test whether the underlying hazard function belongs to a certain parametric family as shown in the hypothesis (7). For the hypothesized hazard function, we consider the following four parametric models: exponential; Weibull; exponential-gamma mixture; and Weibull-gamma mixture. In this section, we will discuss the features of these models.¹¹

4.1 Models without unobserved heterogeneity

Model 1. Exponential model

$$\lambda(t; \theta) = \mu. \quad (19)$$

The simplest parametric model is the exponential model with a constant hazard rate. In this case, the parameter is one-dimension, i.e., $\theta = \mu$. This model has strong economic implication because it corresponds to the price-setting behavior of Calvo's (1983) model, where price changes occur according to the Poisson process with an incidence rate of μ , which is referred to as the frequency of price changes. It is not difficult to confirm that the maximum likelihood estimator of μ is identical to the frequency of price changes.¹²

¹¹For the details of parametric hazard functions, see Cameron and Trivedi (2005) and Klein and Moeschberger (2003).

¹²See, for example, Fleming and Harrington (1991).

Model 2. Weibull model

$$\lambda(t; \theta) = \mu \alpha t^{\alpha-1}. \quad (20)$$

A natural generalization of the exponential model is the Weibull model, which allows the hazard function to change monotonically over time. The parameter is $\theta = (\mu, \alpha)'$, which consists of the scale parameter μ and the shape parameter α . The shape parameter α evaluates the duration dependence since the shape of the hazard function is monotonically increasing (decreasing) if α is more (less) than 1. For $\alpha = 1$, the Weibull distribution is an exponential distribution and exhibits no duration dependence. If the price spells of some items are from the Weibull distribution with $\alpha > 1$, we may support the state dependent model of Dotsey, King, and Wolman (1999), in which the hazard function of price changes monotonically increase under the premise that prices continue to rise.¹³

4.2 Models with unobserved heterogeneity

The models we mentioned so far presuppose that the samples within an item are homogeneous in terms of the intensity of price changes. However, there is some reason that unobserved heterogeneity exists even at the item level. As Higo and Saita (2007) state, unobserved heterogeneity may arise because of the differences in a city attribute or the pricing strategy of individual outlet. According to this remark, we also fit the model that allows heterogeneity across cities. Generally, the individual hazard function can be written as

$$\lambda_{jk}(t|v_j) = v_j \lambda_0(t), \quad (21)$$

where v_j is the unobserved heterogeneity that is specific to the j th city and $\lambda_0(t)$ is the baseline hazard function. As we can see from Equation (21), the heterogeneity effect v_j appears multiplicatively and thus shifts the individual hazard function proportional to its baseline hazard. We assume that v_j is independently and identically distributed as Gamma(1, δ). For the baseline hazard $\lambda_0(t)$, we consider exponential and Weibull hazard functions. We will show the aggregated hazard function for each specification below.

Model 3. Exponential-gamma mixture model

¹³A full examination of the Dotsey, King, and Wolman (1999) model requires careful consideration of the firm's relative price, which is beyond the scope of our analysis.

$$\lambda(t; \theta) = \mu[1 + \mu\delta t]^{-1}. \quad (22)$$

Here, we specify the baseline hazard function as constant over time, that is, $\lambda_0(t) = \mu$. The parameters μ and δ denote the baseline hazard rate and the (normalized) variance of gamma distribution, respectively. The value of δ indicates the degree of heterogeneity among individuals. In model 3, even if the individual hazard functions are constant overtime, in other words, they exhibit no duration dependence, the aggregated hazard function can be decreasing according to the value of δ . This is an effect of aggregating across heterogeneous individuals.

Model 4. Weibull-gamma mixture model

$$\lambda(t; \theta) = \mu\alpha t^{\alpha-1}[1 + (\mu\delta t^\alpha)]^{-1}. \quad (23)$$

The baseline hazard function of the last model is Weibull, which allows the individual hazard functions to monotonically change over time. The parameters are the scale parameter μ , the duration dependence parameter α , and the heterogeneity variance parameter δ . This model can account for both the heterogeneity effect and duration dependence. A well-fitting Weibull-gamma mixture model can explain the interaction of both effects: The aggregated hazard function is decreasing due to the heterogeneity effect while the individual hazard function exhibits positive duration dependence.

5 Shapiro–Wilk test for exponentiality

We have seen general framework for testing distributional assumptions for the censored data so far. As for a test of the composite hypothesis of exponential distribution, there is another class of nonparametric tests. Among these tests, Shapiro and Wilk (1972) propose a basic and fairly simple test for exponentiality.¹⁴ The principle of the test, which is similar to their famous test for normality (Shapiro and Wilk (1965)), is to evaluate the adequacy of the linear regression of the ordered observations on the expected values of the order statistics.

Let $x_1 \leq x_2 \leq \dots \leq x_N$ denote the N order statistics based on the durations of complete price spells that end with a price change. Then the test statistic

$$W = \frac{(\bar{x} - x_1)^2}{(N-1)\hat{\sigma}^2}, \quad (24)$$

¹⁴An example of the application to economic issues is the investigation of duration dependence in the American business cycle by Diebold and Rudebusch (1990).

where $\bar{x} = \sum_{i=1}^N x_i/N$ and $\hat{\sigma}^2 = \sum_{i=1}^N (x_i - \bar{x})^2/N$. The important property of the test statistic is that the distribution is invariant to the parameter values of exponential distribution and depends only on the sample size of complete spells, n_c . Shapiro and Wilk (1972) report the simulated percentage points of the test statistic for sample size $n_c = 3, \dots, 100$. In order to test items, whose number of complete price spells exceeds 100, we extend the statistical table of the percentage points for $n_c > 100$. We obtain the null distribution by generating 5000 samples of the test statistic for each sample size of an item.¹⁵ The test statistic is to be used as a two-tailed test and, therefore, we reject the hypothesis at a significance level α if $W \leq h$ or $k \leq W$, where h and k are the $(\alpha/2)100$ th and $(1 - \alpha/2)100$ th percentiles of the null distribution, respectively.

6 Empirical Analysis

6.1 Data

We employ a sample of the retail prices underlying the computation of the Japanese CPI. These prices are collected on a monthly basis by the Statistics Bureau, Ministry of Internal Affairs and Communications, and appear in the *Monthly Report on the Retail Price Survey*. Prices are reported for each city with a prefectural government and for cities with a population of 150,000 or more. The analysis covers the period from January 2000 to December 2005.

Preferably, we include as many items as possible, but three groups of items are left out of consideration so as not to create a severe bias in our analysis. First, we exclude items that were newly listed as a component item of the CPI and that were not used for calculation of the CPI during the observation period.¹⁶ Second, we exclude seasonal items such as fruits, vegetables, or clothing in which the maximum length of price spells is fairly short due to the right-censoring. Third, we remove items that the statistical agency does not survey every month. These items include, for example, PTA membership fees that are surveyed in April and September every year.

In addition, we exclude two kinds of price data from the original data set. First, we exclude left-censored price spells, whose duration is unknown to us due to the censoring of the starting time. Second, we exclude the price

¹⁵Shapiro and Wilk (1972) generate 5000 samples of the test statistic for $n_c \leq 50$ and $[250,000/n_c]$ samples for $50 < n_c \leq 100$.

¹⁶If a new product replaces an old one, we do not exclude the price data of both products, instead, we treat the price spell of the old product as a right-censored spell, censored at the time when the item substitution occurs.

spells whose number of collected prices in the survey exceeds four. Reported prices in the *Monthly Report on the Retail Price Survey* are an arithmetic mean of prices collected at individual outlets. The number of prices collected in each city ranges from one to forty-two, depending on the characteristics of the items and the size of the city. This exclusion limits the maximum number of summand prices to four and, thus, certainly makes our price data represent the process of price changes at the individual outlets within a survey district.¹⁷

As a consequence of these exclusions, we consider 429 items, covering 60.1 percent of the Japanese CPI in 2000. Table 2 shows the summary statistics.

[Table 2 about here.]

6.2 Results and Discussions

The overall results of the goodness-of-fit test are summarized in Table 3. We report the results for each group of expenses in the Japanese CPI to find out how the goodness-of-fit of hypothesized models varies over groups.¹⁸

The second column in the table shows the frequency of price changes by group. The frequency of price changes in a group is calculated as follows: Let \mathcal{G}_q be the q th item group, the frequency in the group is $\sum_{i \in \mathcal{G}_q} w_i \hat{\mu}_i$, where $\hat{\mu}_i$ is the parameter estimates of the exponential model, which is identical to the frequency of price changes as we have seen in Section 4, and w_i is the i th item's CPI weight in 2000, which satisfies $\sum_{i=1}^{429} w_i = 1$. The figures clearly show that the degree of price stickiness differs across item groups, supporting the finding of many authors including Bils and Klenow (2004).

[Table 3 about here.]

In the exponential model (Model 1), 352 items reject the hypothesis of exponentiality at the one percent significance level. In the course of calculation, we cannot calculate the test statistic X_n^2 for 19 items because the number of complete spells is too small and thereby we cannot form any single cell. For the rest of the 25 items, which amount to 8.6 percent in weighted

¹⁷On these exclusions of items and price spells, we follow Higo and Saita (2007) in most respects. A slight difference lies in the last exclusion, in which they use the price data of 55 middle-sized cities. In fact, the number of prices collected in those cities is also less than or equal to four. We include, however, in addition to price data of those middle-sized cities, price data that are surveyed in large-sized cities, but whose number of collected prices is less than or equal to four.

¹⁸Table 5 in the Appendix shows the test results by item.

share of nonrejected items, we can conclude that exponentiality holds.¹⁹ Table 3 shows that most items in the subgroups of Eating out in the Food sector and Repairs and maintenance in the Housing sector and some items in the Reading and recreation sector lead to the nonrejection of an exponential hypothesis.

It is notable that such items have in common a low frequency of price changes. This fact is clearly illustrated in Figure 1, in which we superpose the distribution of the parameter estimates of the exponential model for the items that lead to nonrejection (denoted by the black bar) on the distribution for all 429 items (gray bar). We can see from Figure 1 that most parameter values of the well-fitting exponential models distribute in the region up to 0.125, whereas in the higher region they do not, except for one item, Camera, with the parameter value $\hat{\mu}=0.288$.

[Figure 1 about here.]

The results from the exponentiality test by Shapiro and Wilk (1972) also verify the evidence. Figure 2 depicts the same parameter distribution as the previous figure, but the black bars in Figure 2 denote the number of items that retain the exponentiality in the Shapiro and Wilk (1972) test. The figure illustrates that the items with the higher value of $\hat{\mu}$ uniformly reject the hypothesis at the one percent significance level. The maximum of $\hat{\mu}$ for the items that retain the hypothesis is 0.11, which implies the frequency of price changes for these items is limited to 11 percent per month.

[Figure 2 about here.]

As for the Hjort (1990) test, we can verify the results using the conceptually equivalent method that compares the Nelson–Aalen estimate $\hat{A}(t)$ with the parametric cumulative hazard function $A(t; \hat{\theta})$. We choose four items with a different value of parameter estimate $\hat{\mu}$ in Figure 3: (a) Coffee (eating out) with $\hat{\mu} = 0.054$, (b) Babies’ clothes with $\hat{\mu} = 0.202$, (c) Butter with $\hat{\mu} = 0.399$, and (d) Detergent, laundry with $\hat{\mu} = 0.604$. As we have seen in the last part of Section 3.2, Coffee (eating out) retains the null of exponentiality. Panel (a) shows a typical appearance of both plots when the hypothesis of exponentiality holds. Specifically, the Nelson–Aalen estimate in Panel (a) fluctuates around the predicted value of the cumulative hazard

¹⁹Specifically, we cannot reject the null of exponentiality in 58 (=429–(352+19)) items but the figure includes 33 items with high price flexibility. In these flexible items, the test statistic takes a necessarily small value, because the observed and expected number of price changes is nearly equal in each cell over the interval. This can be understood as a trivial case in the sense that we cannot reject any assumed parametric models. Thus, we subtract 33 items from 58 items and report 25 items in Table 3.

function. We illustrate cases of the rejection in Panels (b)-(d). The predicted values in these panels fail to capture the shapes of the Nelson-Aalen estimate especially in the longer duration.

[Figure 3 about here.]

In Table 3, there is a difference in the number of retained items between the Shapiro and Wilk (1972) test and the Hjort (1990) test for exponentiality. One reasonable explanation for the difference is that the power of the Shapiro–Wilk test against local alternatives is weaker than that of the Hjort test. To confirm this explanation, we conduct power analysis through Monte Carlo simulations. In the power analysis, we compare the performance of both tests for testing the null hypothesis of exponentiality H_0 , when the true distribution is Weibull with shape parameter α . As we have shown in Section 4, when $\alpha < 1$ ($\alpha > 1$), the parametric hazard function slopes downward (upward) and $\alpha = 1$ corresponds to H_0 . Table 4 presents the simulated powers with 1 percent significance level for the randomly censored samples with sizes $n = 100$ and 200 , and censoring percentages of 25% and 50%.²⁰ The simulated powers are computed by generating 1000 replications of samples. In addition, we examine two cases in which the Hjort (1990) test statistic includes 5 and 10 cells.

Table 4 shows that assuming the different number of cells in the Hjort (1990) test does not substantially alter the simulated powers for each samples. Comparing the powers of the Hjort (1990) test and the Shapiro and Wilk (1972) test, we find that neither test dominates the other throughout the chosen range of shape parameter, from 0.5 to 1.5: in the range $\alpha < 1$, the Hjort test dominates the Shapiro–Wilk test, whereas in the range $\alpha > 1$, it does not. However, the range most relevant to our discussion is $\alpha < 1$. This is because too often the Shapiro–Wilk test tends to accept the null of exponentiality in those items, whose Weibull shape parameter estimate is less than one, and in which the Hjort test cannot reject the Weibull model.²¹ For the 200 samples with 50%-censored simulated data from Weibull distribution with $\alpha = 0.7$, for example, the simulated power of the Shapiro–Wilk test is 0.697, which is less than that of the Hjort test by 18 percentage points. Considering the test results from the Weibull model, as discussed in detail below, the fact that the Hjort test has relatively high power against Weibull alternatives with $\alpha < 1$ may account for the difference in the test results.

²⁰The number of samples and the censoring percentage are chosen so that the simulated samples adequately represent the item in which the results of the two tests are contradictory: The average number of price spells and the censoring percentage of those items are 183 and 51%, respectively.

²¹See Figure 4 and Table 5 in the Appendix.

[Table 4 about here.]

The Weibull model (Model 2) fits well for 54 items, the weighted share of which amounts to 14.6 percent. Now the interest lies in the shape parameter α of the model. If the parameter α exceeds one and the goodness of fit is plausible, we then conclude that the hazard rate of that item tends to increase over time. Figure 4 shows, however, there exist no such items in our dataset. The items that retain the hypothesis that the underlying distribution is Weibull exhibit a decreasing or constant pattern of hazard rates. We can find three items suitable for the Weibull fit (denoted in black) in the region where α exceeds one, but the value of the parameter is nearly equal to one. Thus, the hazard function of these items does not sharply increase over time. In fact, the three items above also retain the hypothesis of exponentiality. When α is close to one, an item that cannot reject a Weibull hypothesis tends to retain an exponential hypothesis as well, which accounts for the results in Table 3 that both hypotheses hold in some items in the group such as Eating out or Repairs and maintenance. The result is highly predictable because Weibull models encompass exponential models.

[Figure 4 about here.]

Figure 5 illustrates the graphical check for Weibull models. In the same manner as Figure 2, we show the case of nonrejection in Panel (a), Women’s haircut charges with $\hat{\mu} = 0.058$ and $\hat{\alpha} = 0.758$, and the cases of rejection in Panels, (b) Gas with $\hat{\mu} = 0.130$ and $\hat{\alpha} = 0.998$, (c) Microwave ovens with $\hat{\mu} = 0.366$ and $\hat{\alpha} = 1.21$, and (d) Gasoline (regular) with $\hat{\mu} = 0.371$ and $\hat{\alpha} = 1.51$. Note that the estimated Weibull hazard function of four items has different implications as to duration dependence: (a) negative duration dependence; (b) no duration dependence; and (c)–(d) positive duration dependence. The Weibull model for Women’s haircut charges in Panel (a) has a concave cumulative hazard function, which means the hazard rates decrease over time. In Panel (a) we observe that the step function of the Nelson–Aalen estimate hovers around the cumulative hazard function, while in Panels (b)–(d) we do not observe the same.

[Figure 5 about here.]

Finally, we discuss the results of the model with unobserved heterogeneity. As we have shown in Section 4, if samples are from a heterogeneous population, the estimated hazard rates decrease over time due to the heterogeneity effect. As Higo and Saita (2007) state, even if we estimate the hazard function by item, the estimates can be affected by the local characteristics of a city or an outlet. That is the reason why we estimate the

parametric hazard model with unobserved heterogeneity (Models 3 and 4). However, the performance of these models is worse than that of the models without heterogeneity: The weighted share of Models 3 and 4 is even smaller than those of Models 1 and 2. Most items that retain the hypothesis of a heterogeneous model also retain the hypothesis of a homogeneous model, and on comparing the numbers of nonrejected items by groups, the introduction of the gamma-distributed heterogeneity does not additionally increase the number of items of nonrejection. In view of these results, it is not easy to attribute the reason for the decreasing hazard estimates simply to the unobservable characteristics of the survey cities.

7 Concluding remarks

In this paper, we examine the distributional assumptions of price duration distributions. We use the goodness-of-fit test by Hjort (1990), in which we compare, on a span-by-span basis, the number of observed price changes and expected price changes implied by the hypothesized hazard function. For each of the 429 items available in the *Monthly Report on the Retail Price Survey* for 2000–2005, we fit the standard parametric hazard models with or without unobserved heterogeneity and test the goodness-of-fit.

We establish the following four facts. First, 8.6 percent of the tested items (in weighted share) cannot reject the hypothesis that the underlying distribution is exponential, which corresponds to the time-dependent model of Calvo (1983). Second, most of the items with a high frequency of price changes reject the hypothesis of exponentiality at the one percent significance level. This finding is also confirmed by the Shapiro and Wilk (1972) test for exponentiality. Third, a well-fitting Weibull hazard model has a decreasing or almost constant hazard function. In our dataset, there is no item that retains a Weibull hypothesis and that exhibits a sharply increasing hazard function. Fourth, the introduction of unobserved heterogeneity that is specific to a survey city does not improve the performance of the model without unobserved heterogeneity at the item level.

References

- Akritas, M. G. (1988) “Pearson-Type Goodness-of-Fit Tests: The Univariate Case”, *Journal of the American Statistical Association*, Vol. 83, No. 401, pp. 222–230.

- Álvarez, L. J., P. Burriel, and I. Hernando (2005) “Do Decreasing Hazard Functions for Price Changes Make Any Sense?” *ECB Working Paper Series*, No. 461.
- Álvarez, L. J. (2007) “What Do Micro Price Data Tell Us on the Validity of the New Keynesian Philips Curve?”, mimeo.
- Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993) *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.
- Bils, M. and P. J. Klenow (2004) “Some Evidence on the Importance of Sticky Prices”, *Journal of Political Economy*, Vol. 112, No. 5, pp. 947–985.
- Calvo, G. (1983) “Staggered Prices in a Utility-Maximizing Framework”, *Journal of Monetary Economics*, 12 (3), pp. 383–398.
- Cameron, A. C. and P. K. Trivedi (2005) *Microeconometrics: Methods and Applications*, Cambridge: Cambridge University Press.
- Dhyne, E., L. J. Álvarez, H. L. Bihan, G. Veronese, D. Dias, J. Hoffmann, N. Jonker, P. Lünemann, F. Rumler, and J. Vilmunen (2006) “Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data”, *Journal of Economic Perspectives*, Vol. 20, No. 2, pp. 171–192.
- Diebold, F. X. and G. D. Rudebusch (1990) “A Nonparametric Investigation of Duration Dependence in the American Business Cycle”, *Journal of Political Economy*, Vol. 98, No. 3, pp. 596–616
- Dotsey, M., R. G. King, and A. L. Wolman (1999) “State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output”, *The Quarterly Journal of Economics*, 114, pp. 655–690.
- Fleming, T. R. and Harrington, D. P. (1991) *Counting Processes and Survival Analysis*, New York: Wiley.
- Fougère, D., H. Le Bihan, and P. Sevestre (2007) “Heterogeneity in Consumer Price Stickiness: A Microeconomic Investigation”, *Journal of Business & Economic Statistics*, Vol. 25, No. 3, pp. 247–264.
- Higo, M. and Y. Saita (2007) “Price-setting in Japan: Evidence from Individual Retail Price Data”, *Bank of Japan Working Paper Series*, No.07-E-20.

- Hjort, N. L. (1990) “Goodness of Fit Tests in Models for Life History Data Based on Cumulative Hazard Rate”, *The Annals of Statistics*, Vol. 18, No. 3, pp. 1221–1258.
- Ikeda, D. and S. Nishioka (2007) “Price Setting Behavior and Hazard Functions: Evidence from Japanese CPI Micro Data”, *Bank of Japan Working Paper Series*, No.07-E-19.
- Kalbfleisch, J. D. and R. L. Prentice (2002) *The statistical Analysis of Failure Time Data*, 2nd ed., New York: Wiley.
- Klein, J. P. and M. Moeschberger (2003) *Survival Analysis*, 2nd ed., New York: Springer.
- Matsuoka, T. (2007) “Unobserved Heterogeneity in Price-Setting Behavior: a Duration Analysis Approach”, *Understanding Inflation Dynamics of the Japanese Economy Working Paper Series*, No.16.
- Nakamura, E. and J. Steinsson (2008) “Five facts about prices: a reevaluation of menu cost models”, *Quarterly Journal of Economics*, Vol. 123, No. 4, pp. 1415–1464.
- Shapiro, S. S. and M. B. Wilk (1965) “An Analysis of Variance Test for Normality”, *Biometrika*, 52, pp. 591–611.
- Shapiro, S. S. and M. B. Wilk (1972) “Variance Test for the Exponential Distribution”, *Technometrics*, Vol. 14, No. 2, pp. 355–370.

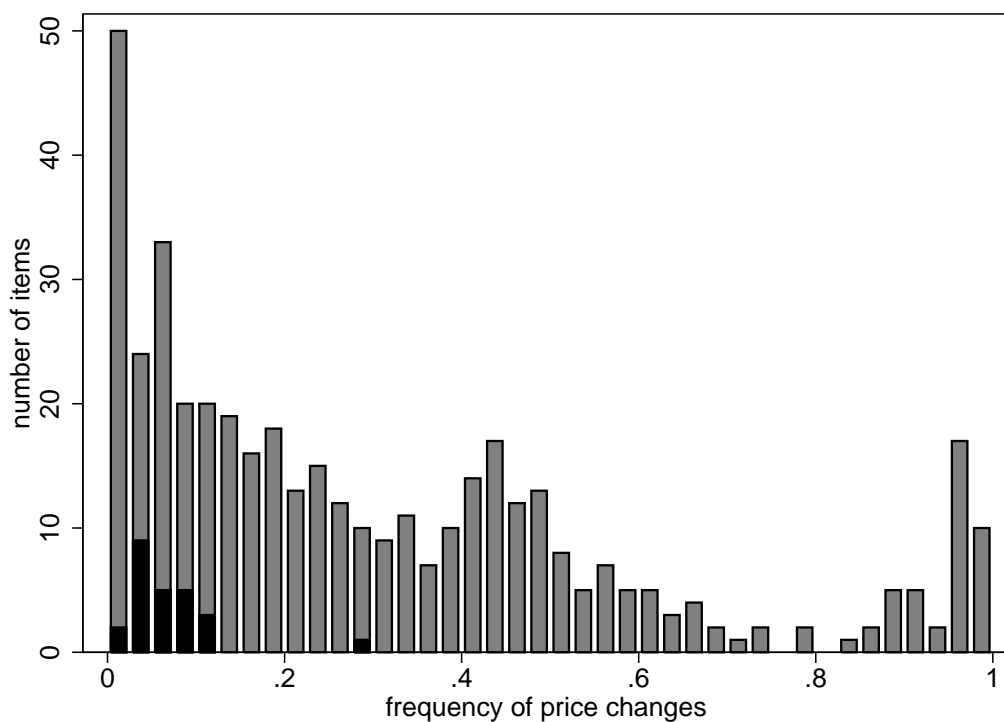


Figure 1: Distribution of $\hat{\mu}$: The black bar denotes the number of items that retain the hypothesis of exponentiality. The test statistic is the Hjort (1990) test of goodness-of-fit. The total number of items is 429. The retail price data are from Jan 2000–Dec 2005.

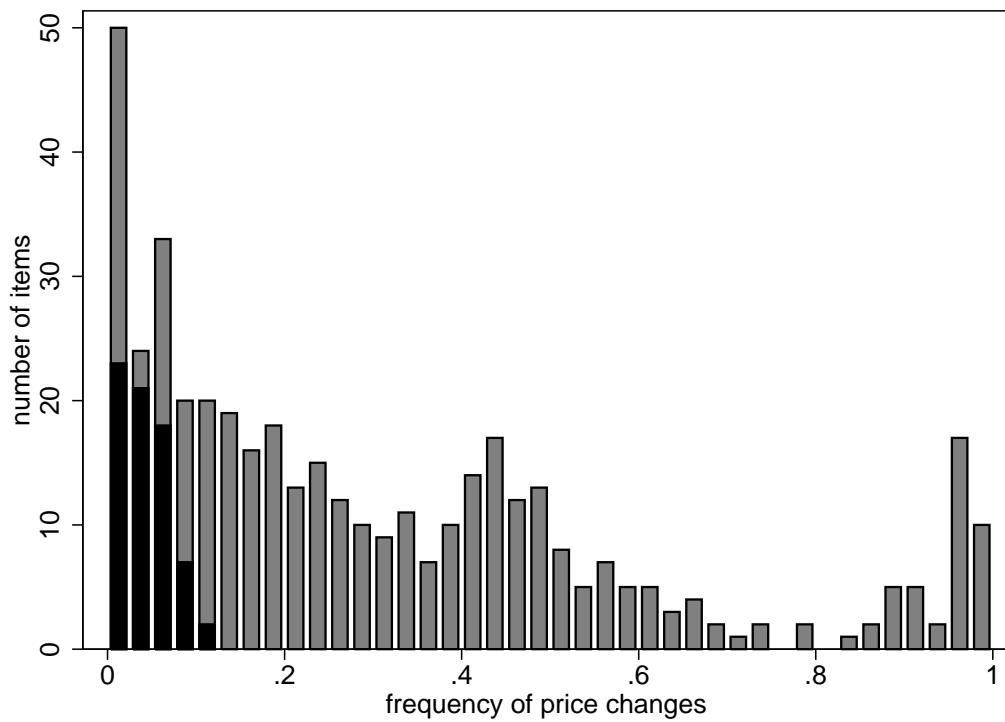


Figure 2: Distribution of $\hat{\mu}$: The black bar denotes the number of items that retain the hypothesis of exponentiality. The test statistic is the Shapiro–Wilk (1972) test for Exponentiality. The total number of items is 429. The retail price data are from Jan 2000–Dec 2005.

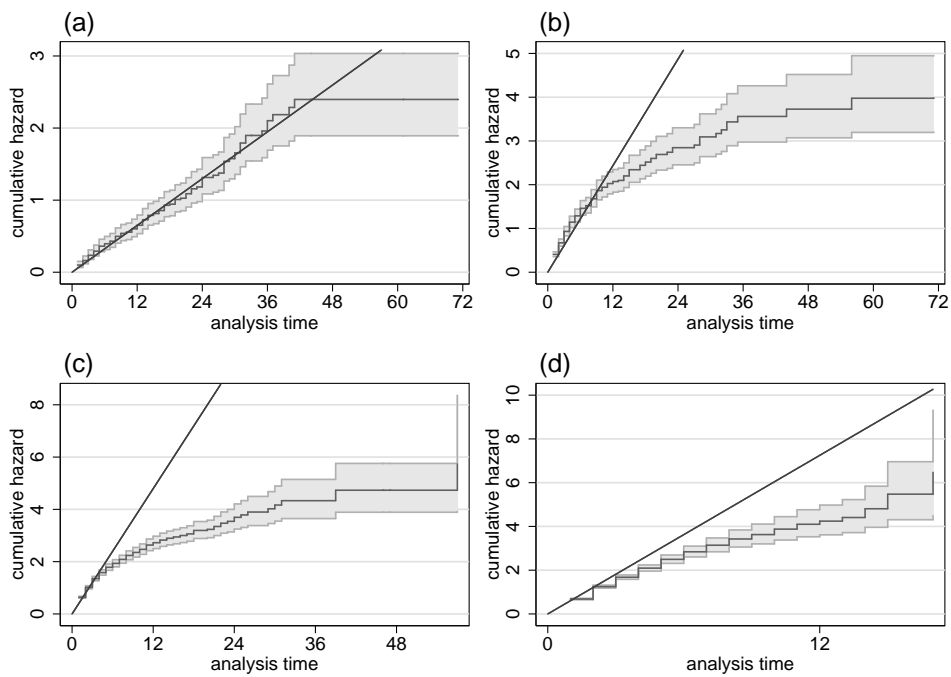


Figure 3: Graphical check for exponentiality: Nelson–Aalen estimate with 95 percent confidence band (step function) and the exponential cumulative hazard function (straight line). (a) Coffee (eating out) with $\hat{\mu} = 0.054$, (b) Babies' clothes with $\hat{\mu} = 0.202$, (c) Butter with $\hat{\mu} = 0.399$, and (d) Detergent, laundry with $\hat{\mu} = 0.604$.

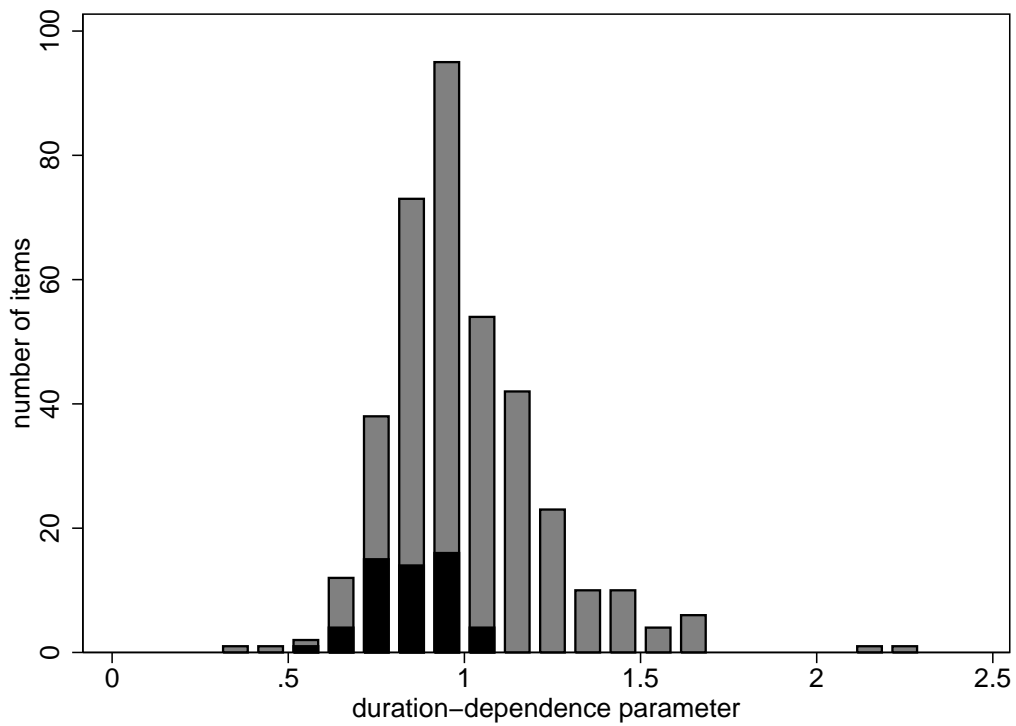


Figure 4: Distribution of $\hat{\alpha}$: The black bar denotes the number of items that retain the hypothesis that the underlying distribution is Weibull distribution. Same data as Figure 1.

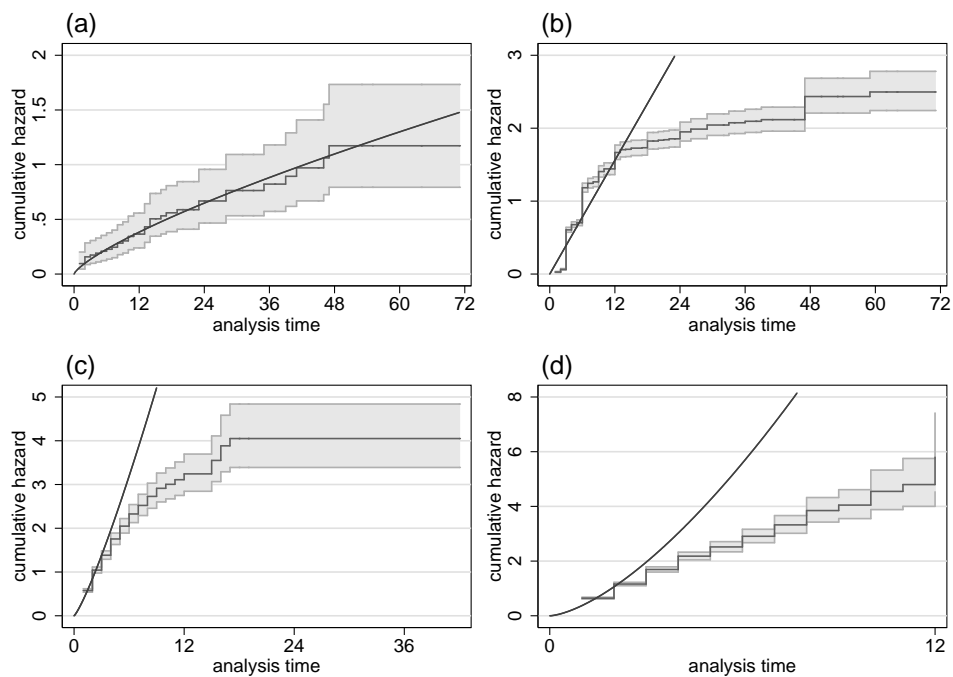


Figure 5: Graphical check for Weibull hazard model: Nelson–Aalen estimate with 95 percent confidence band (step function) and the Weibull cumulative hazard function (smooth line). (a) Women’s haircut charges with $\hat{\mu} = 0.058$ and $\hat{\alpha} = 0.758$, (b) Gas with $\hat{\mu} = 0.130$ and $\hat{\alpha} = 0.998$, (c) Microwave ovens with $\hat{\mu} = 0.366$ and $\hat{\alpha} = 1.21$, and (d) Gasoline (regular) with $\hat{\mu} = 0.371$ and $\hat{\alpha} = 1.51$.

Duration t	Number of price spells			Cell l	Goodness-of-fit test		
	$Y(t)$	$\Delta N(t)$	$\Delta C(t)$		N_l	E_l	Z_l
1	231	23	0	1	23	12.5	8.8
2	208	13	3	2	13	11.3	0.3
3	192	14	3	3	14	10.4	1.3
4	175	10	6	4	10	9.5	0.0
5	159	11	1	5	11	8.6	0.7
6	147	5	1	6	5	8.0	1.1
7	141	5	2	7	5	7.6	0.9
8	134	9	3	8	9	7.2	0.4
9	122	5	1	9	5	6.6	0.4
10	116	2	5	10	7	12.2	2.2
11	109	5	4				
12	100	5	4	11	5	5.4	0.0
13	91	7	0	12	7	4.9	0.9
14	84	5	1	13	5	4.5	0.0
15	78	2	4	14	5	8.1	1.2
16	72	3	0				
17	69	5	1	15	5	3.7	0.4
18	63	1	0	16	5	6.8	0.5
19	62	4	2				
20	56	1	7	17	7	8.0	0.1
21	48	3	1				
22	44	3	2				
23	39	1	0	18	6	4.2	0.8
24	38	5	0				
26	33	1	0	19	7	5.2	0.6
27	32	1	0				
28	31	5	0				
29	26	1	0	20	6	3.9	1.1
30	25	2	1				
31	22	3	0				
32	19	2	0	21	8	6.3	0.4
33	17	0	1				
35	16	1	1				
36	14	2	0				
37	12	1	1				
40	10	1	0				
41	9	1	0				
44	8	0	1				
61	7	0	2				
71	5	0	5				

Table 1: Goodness-of-fit test for exponentiality: Coffee (eating out). $Y(t)$, the number of price spells at risk just before t ; $\Delta N(t)$ the number of price spells end with a price change at t ; $\Delta C(t)$, the number of price spells censored at t ; N_l , the observed number of price changes in the l th cell; E_l , the expected number of price changes implied by the exponential distribution with hazard rate $\hat{\theta} = 0.054$. Z_l is defined as $(N_l - E_l)^2/E_l$.

Number of items	429
Number of price records	1,460,823
Number of price spells	477,210
Number of complete spells	450,445
Number of right-censored spells	26,765
Share of right-censored spells	5.6%

Table 2: Retail price data used for the calculation of the Japanese CPI from Jan 2000–Dec 2005.

Groups	Frequency of price changes	Hjort (1990) test for goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality	Number of items tested
		Exponential	Weibull	Exponential- Gamma	Weibull- Gamma		
Food							
Cereals	0.440	0	0	0	0	0	13
Fish & shellfish	0.741	0	0	0	0	0	29
Meat	0.477	0	0	0	0	0	10
Dairy products & eggs	0.477	0	0	0	0	1	8
Vegetables & seaweeds	0.727	0	0	0	0	0	36
Fruits	0.946	0	0	0	0	0	5
Oils, fats & seasonings	0.432	0	0	0	0	1	16
Cakes & candies	0.325	0	0	0	0	0	17
Cooked food	0.333	0	0	0	0	0	14
Beverages	0.338	0	0	0	0	1	14
Alcoholic beverages	0.250	0	0	0	0	0	13
Eating out	0.073	9	12	7	5	12	17
Housing							
Repairs & maintenance	0.036	6	9	4	4	9	12
Fuel, Light & Water Charges							
Electricity	0.070	0	0	0	0	0	1
Gas	0.112	0	0	0	0	0	4
Furniture & Household Utensils							
Household durables	0.464	0	1	0	0	0	15
Interior furnishings	0.227	0	0	0	0	0	5
Bedding	0.261	0	0	0	0	0	4
Domestic utensils	0.154	1	2	0	0	3	15
Domestic non-durables	0.481	0	0	0	0	0	8
Domestic services	0.006	0	0	0	0	1	3
Cloths & Footwear							
Japanese clothing	0.088	0	0	0	0	0	2
Clothing	0.256	0	0	0	0	0	4
Shirts & sweaters	0.215	0	0	0	0	0	1
Underwear	0.178	0	0	0	0	0	13
Cloth & thread	0.115	0	0	0	0	0	3

Groups (<i>continued</i>)	Frequency of price changes	Hjort (1990) test for goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality	Number of items tested
		Exponential	Weibull	Exponential-Gamma	Weibull-Gamma		
Other clothing	0.161	0	1	0	0	0	6
Services related to clothing	0.049	0	1	0	0	1	2
Medical Care							
Medicines & health fortification	0.098	0	3	0	0	4	11
Medical supplies & appliances	0.252	1	1	2	1	1	9
Medical services	0.008	0	1	0	0	1	5
Transportation & Communication							
Public transportation	0.001	0	0	0	0	1	9
Private transportation	0.323	2	4	1	2	7	12
Communication	0.023	0	1	0	1	2	4
Education							
Tutorial fees	0.062	0	0	0	0	1	1
Reading & Recreation							
Recreational durables	0.516	1	0	1	0	0	5
Recreational goods	0.292	3	7	1	1	6	31
Recreational services	0.075	2	5	1	1	13	19
Miscellaneous							
Personal care services	0.026	0	3	0	1	6	7
Toilet articles	0.286	0	1	0	0	1	12
Personal effects	0.165	0	2	0	0	1	9
Other	0.005	0	0	0	0	1	5
Total	0.247	25	54	17	16	74	429
	-	(8.6)	(14.6)	(6.1)	(5.0)	(21.3)	-

Table 3: The number of items that retain the hypothesis that the underlying distribution belongs to a specified family of distributions. Entries for groups are the subgroups of the ten major groups in the Japanese CPI. The frequency of price changes is calculated as $\sum_{i \in \mathcal{G}_q} w_i \hat{\mu}_i$, where \mathcal{G}_q is the q th item group and w_i is i th item's CPI weight in 2000 with $\sum_{i=1}^{429} w_i = 1$ ($\hat{\mu}_i$ is defined in the text). The weighted share (in percentage) is in the parentheses in the last row.

$n = 100$						
α	25% censored			50% censored		
	Hjort test		Shapiro– Wilk test	Hjort test		Shapiro– Wilk test
	$m = 5$	$m = 10$		$m = 5$	$m = 10$	
0.5	1.000	1.000	0.988	1.000	0.998	0.919
0.6	0.982	0.982	0.876	0.908	0.942	0.692
0.7	0.744	0.801	0.586	0.588	0.631	0.355
0.8	0.270	0.344	0.222	0.170	0.251	0.125
0.9	0.056	0.095	0.049	0.033	0.089	0.023
1.0 (H_0)	0.009	0.030	0.009	0.019	0.061	0.016
1.1	0.026	0.043	0.030	0.031	0.065	0.021
1.2	0.078	0.108	0.128	0.059	0.098	0.081
1.3	0.182	0.195	0.355	0.123	0.176	0.177
1.4	0.388	0.358	0.622	0.244	0.284	0.353
1.5	0.623	0.585	0.790	0.367	0.415	0.464

$n = 200$						
α	25% censored			50% censored		
	Hjort test		Shapiro– Wilk test	Hjort test		Shapiro– Wilk test
	$m = 5$	$m = 10$		$m = 5$	$m = 10$	
0.5	1.000	1.000	1.000	1.000	1.000	1.000
0.6	1.000	1.000	0.995	0.999	0.999	0.960
0.7	0.974	0.979	0.898	0.872	0.879	0.697
0.8	0.559	0.529	0.467	0.349	0.365	0.273
0.9	0.062	0.074	0.081	0.053	0.062	0.060
1.0 (H_0)	0.007	0.012	0.007	0.008	0.021	0.008
1.1	0.037	0.032	0.072	0.027	0.039	0.036
1.2	0.179	0.153	0.361	0.086	0.099	0.207
1.3	0.507	0.455	0.716	0.300	0.293	0.486
1.4	0.869	0.781	0.935	0.594	0.539	0.747
1.5	0.977	0.948	0.995	0.836	0.769	0.917

Table 4: Simulated powers of the Hjort (1990) test and the Shapiro and Wilk (1972) test. The true distribution is Weibull with shape parameter α , and the null hypothesis of exponentiality, H_0 , corresponds to $\alpha = 1.0$. The test size is 0.01. The number of replications is 1000. m denotes the number of cells in the Hjort (1990) test statistic.

Appendix

No.	Item	Frequency of price changes	Hjort (1990) test of goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality
			Exponential	Weibull	Exponential-Gamma	Weibull-Gamma	
[Food]							
1	Rice A (domestic), “Koshihikari”	0.435	0.000	0.000	0.000	0.000	reject
2	Rice B (domestic), non-blended rice excluding “Koshihikari”	0.486	0.000	0.000	0.000	0.000	reject
3	Blended rice	0.382	0.000	0.000	0.000	0.000	reject
4	Glutinous rice	0.340	0.000	0.000	0.000	0.000	reject
5	White bread	0.469	0.000	0.000	0.000	0.000	reject
6	Bean-jam buns	0.276	0.000	0.000	0.000	0.000	reject
7	Boiled noodles	0.373	0.000	0.000	0.000	0.000	reject
8	Dried noodles	0.427	0.000	0.000	0.000	0.000	reject
9	Spaghetti	0.519	0.000	0.000	0.000	0.000	reject
10	Instant noodles	0.523	0.000	0.000	0.000	0.000	reject
11	Uncooked Chinese noodles	0.577	0.000	0.000	0.000	0.000	reject
12	Wheat flour	0.482	0.000	0.000	0.000	0.000	reject
13	“Mochi”, rice-cakes	0.411	0.000	0.000	0.000	0.000	reject
14	Tuna fish	0.864	0.000	0.000	0.000	0.000	reject
15	Horse mackerel	0.975	0.999†	-	-	-	reject
16	Sardines	0.973	0.975†	-	-	-	reject
17	Flounder	0.963	0.145†	-	-	-	reject
18	Salmon	0.829	0.000	0.000	0.000	0.000	reject
19	Mackerel	0.955	0.931†	-	-	-	reject
20	Saury	0.916	0.088†	-	-	-	reject
21	Sea bream	0.899	0.000	0.000	0.000	0.000	reject
22	Yellowtail	0.916	0.000	-	-	-	reject
23	Cuttlefish	0.969	0.984†	-	-	-	reject
24	Octopus	0.796	0.000	0.000	0.000	0.000	reject
25	Prawns	0.870	0.000	0.000	0.000	0.000	reject
26	Short-necked clams	0.778	0.000	0.000	0.000	0.000	reject
27	Scallops	0.882	0.000	0.000	0.000	0.000	reject
28	Salted salmon	0.700	0.000	0.000	0.000	0.000	reject
29	Salted cod roe	0.539	0.000	0.000	0.000	0.000	reject
30	“Shirasu-boshi”, dried young sardines	0.636	0.000	0.000	0.000	0.000	reject
31	Dried horse mackerel	0.742	0.000	0.000	0.000	0.000	reject
32	Dried sardines	0.659	0.000	0.000	0.000	0.000	reject
33	“Niboshi”, dried small sardines	0.397	0.000	0.000	0.000	0.000	reject
34	Capelin	0.714	0.000	0.000	0.000	0.000	reject
35	“Agekamaboko”, fried fish-paste patties	0.521	0.000	0.000	0.000	0.000	reject
36	“Chikuwa”, baked fish-paste bars	0.437	0.000	0.000	0.000	0.000	reject
37	“Kamaboko”, steamed fish-paste cakes	0.400	0.000	0.000	0.000	0.000	reject
38	Dried bonito fillets	0.279	0.000	0.000	0.000	0.000	reject
39	Pickled fish	0.663	0.000	0.000	0.000	0.000	reject
40	Fish prepared in soy sauce	0.301	0.000	0.000	0.000	0.000	reject
41	Canned fish	0.248	0.000	0.000	0.000	0.000	reject
42	“Shiokara”, salted fish guts	0.261	0.000	0.000	0.000	0.000	reject
43	Beef (loin)	0.516	0.000	0.000	0.000	0.000	reject
44	Beef (shoulder)	0.559	0.000	0.000	0.000	0.000	reject
45	Beef (imported)	0.729	0.000	0.000	0.000	0.000	reject
46	Pork (loin)	0.425	0.000	0.000	0.000	0.000	reject
47	Pork (shoulder)	0.467	0.000	0.000	0.000	0.000	reject
48	Chicken	0.437	0.000	0.000	0.000	0.000	reject
49	Liver	0.268	0.000	0.000	0.000	0.000	reject

No.	Item	Frequency of price changes	Hjort (1990) test of goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality
			Exponential	Weibull	Exponential- Gamma	Weibull- Gamma	
50	Ham	0.368	0.000	0.000	0.000	0.000	reject
51	Sausages	0.390	0.000	0.000	0.000	0.000	reject
52	Bacon	0.316	0.000	0.000	0.000	0.000	reject
53	Fresh milk (delivered)	0.051	0.000	0.000	0.000	0.000	*
54	Fresh milk (sold in stores)	0.364	0.000	0.000	0.000	0.000	reject
55	Powdered milk	0.213	0.000	0.000	0.000	0.000	reject
56	Yogurt	0.656	0.000	0.000	0.000	0.000	reject
57	Butter	0.399	0.000	0.000	0.000	0.000	reject
58	Cheese	0.428	0.000	0.000	0.000	0.000	reject
59	Cheese (imported)	0.130	0.000	0.000	0.000	0.000	reject
60	Hen eggs	0.880	0.179†	-	-	-	reject
61	Cabbage	0.989	0.997†	-	-	-	reject
62	Spinach	0.998	1.000†	-	-	-	reject
63	Chinese cabbage	0.986	0.984†	-	-	-	reject
64	Welsh onions	0.992	0.990†	-	-	-	reject
65	Lettuce	0.995	0.977†	-	-	-	reject
66	Broccoli	0.997	1.000†	-	-	-	reject
67	Bean sprouts	0.338	0.000	0.000	0.000	0.000	reject
68	Asparagus	0.976	0.971†	-	-	-	reject
69	Sweet potatoes	0.969	0.481†	-	-	-	reject
70	White potatoes	0.957	0.451†	-	-	-	reject
71	Taros	0.968	0.656†	-	-	-	reject
72	Radishes	0.985	0.989†	-	-	-	reject
73	Carrots	0.974	0.316†	-	-	-	reject
74	Burdocks	0.964	0.011†	-	-	-	reject
75	Onions	0.958	0.560†	-	-	-	reject
76	Lotus roots	0.972	0.498†	-	-	-	reject
77	“Naga-imo” yams	0.921	0.009	-	-	-	reject
78	Tomatoes	0.992	0.991†	-	-	-	reject
79	Green peppers	0.961	0.961†	-	-	-	reject
80	“Shiitake”, Japanese mushrooms, fresh	0.961	0.332†	-	-	-	reject
81	“Enokidake”, mushrooms	0.916	0.341†	-	-	-	reject
82	“Shimeji”, mushrooms	0.940	0.993†	-	-	-	reject
83	“Azuki”, red beans	0.313	0.000	0.000	0.000	0.000	reject
84	“Shiitake”, Japanese mushrooms, dried	0.266	0.000	0.000	0.000	0.000	reject
85	Laver	0.417	0.000	0.000	0.000	0.000	reject
86	“Wakame”, seaweed	0.349	0.000	0.000	0.000	0.000	reject
87	Dried tangle	0.436	0.000	0.000	0.000	0.000	reject
88	Bean curd	0.297	0.000	0.000	0.000	0.000	reject
89	Fried bean curd	0.405	0.000	0.000	0.000	0.000	reject
90	“Natto”, fermented soybeans	0.499	0.000	0.000	0.000	0.000	reject
91	“Konnyaku”, devil’s-tongue jelly	0.374	0.000	0.000	0.000	0.000	reject
92	“Umeboshi”, pickled plums	0.418	0.000	0.000	0.000	0.000	reject
93	Pickled radishes	0.476	0.000	0.000	0.000	0.000	reject
94	Pickled Chinese cabbage	0.603	0.000	0.000	0.000	0.000	reject
95	Sliced vegetables pickled in soy sauce	0.415	0.000	0.000	0.000	0.000	reject
96	Tangle prepared in soy sauce	0.491	0.000	0.000	0.000	0.000	reject
97	Grapefruits	0.953	0.592†	-	-	-	reject
98	Oranges	0.952	0.955†	-	-	-	reject
99	Lemons	0.960	0.785†	-	-	-	reject
100	Bananas	0.938	0.107†	-	-	-	reject
101	Kiwi fruits	0.951	0.431†	-	-	-	reject
102	Edible oil	0.486	0.000	0.000	0.000	0.000	reject
103	Margarine	0.424	0.000	0.000	0.000	0.000	reject
104	Salt	0.020	0.000	0.001	0.000	0.000	*

No.	Item	Frequency of price changes	Hjort (1990) test of goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality
			Exponential	Weibull	Exponential- Gamma	Weibull- Gamma	
105	Soy sauce	0.381	0.000	0.000	0.000	0.000	reject
106	Soybean paste	0.436	0.000	0.000	0.000	0.000	reject
107	Sugar	0.434	0.000	0.000	0.000	0.000	reject
108	Vinegar	0.240	0.000	0.000	0.000	0.000	reject
109	Worcester sauce	0.359	0.000	0.000	0.000	0.000	reject
110	Ketchup	0.498	0.000	0.000	0.000	0.000	reject
111	Mayonnaise	0.400	0.000	0.000	0.000	0.000	reject
112	Jam	0.497	0.000	0.000	0.000	0.000	reject
113	Instant curry mix	0.537	0.000	0.000	0.000	0.000	reject
114	Instant dried soup	0.429	0.000	0.000	0.000	0.000	reject
115	Flavor seasonings	0.492	0.000	0.000	0.000	0.000	reject
116	“Furikake”, granular flavor sea- sonings	0.493	0.000	0.000	0.000	0.000	reject
117	Liquid seasonings	0.453	0.000	0.000	0.000	0.000	reject
118	“Yokan”, sweet bean jelly	0.061	0.000	0.007	0.000	0.000	reject
119	“Manju”, bean-jam cakes	0.254	0.000	0.000	0.000	0.000	reject
120	“Daifukumochi”, rice cakes stuffed with sweetened bean jam	0.214	0.000	0.000	0.000	0.000	reject
121	“Kasutera”, sponge cakes	0.118	0.000	0.000	0.000	0.000	reject
122	Cakes	0.208	0.000	0.000	0.000	0.000	reject
123	Jelly	0.434	0.000	0.000	0.000	0.000	reject
124	Pudding	0.477	0.000	0.000	0.000	0.000	reject
125	Cream puffs	0.218	0.000	0.000	0.000	0.000	reject
126	“Sembei”, Japanese rice crackers	0.524	0.000	0.000	0.000	0.000	reject
127	“Sembei”, Japanese wheat crack- ers	0.377	0.000	0.000	0.000	0.000	reject
128	Biscuits	0.660	0.000	0.000	0.000	0.000	reject
129	Potato chips	0.596	0.000	0.000	0.000	0.000	reject
130	Candies	0.550	0.000	0.000	0.000	0.000	reject
131	Chocolate	0.349	0.000	0.000	0.000	0.000	reject
132	Ice cream	0.227	0.000	0.000	0.000	0.000	reject
133	Peanuts	0.370	0.000	0.000	0.000	0.000	reject
134	Chewing gum	0.072	0.000	0.000	0.000	0.000	reject
135	Box lunch	0.096	0.000	0.000	0.000	0.000	reject
136	Rice balls	0.167	0.000	0.000	0.000	0.000	reject
137	Bread like sandwiches put cooked food between bread	0.394	0.000	0.000	0.000	0.000	reject
138	Frozen pilaf	0.493	0.000	0.000	0.000	0.000	reject
139	“Kabayaki”, broiled eels	0.632	0.000	0.000	0.000	0.000	reject
140	Salad	0.433	0.000	0.000	0.000	0.000	reject
141	Croquettes	0.339	0.000	0.000	0.000	0.000	reject
142	Cutlets	0.228	0.000	0.000	0.000	0.000	reject
143	Fried chicken	0.471	0.000	0.000	0.000	0.000	reject
144	Ch(i)aotzu	0.467	0.000	0.000	0.000	0.000	reject
145	Frozen croquettes	0.439	0.000	0.000	0.000	0.000	reject
146	Cooked curry	0.436	0.000	0.000	0.000	0.000	reject
147	“Mazegohan no moto”, prepared materials to boiled rice with as- sorted ingredients	0.440	0.000	0.000	0.000	0.000	reject
148	Boiled beans	0.458	0.000	0.000	0.000	0.000	reject
149	Green tea (“Bancha”)	0.261	0.000	0.000	0.000	0.000	reject
150	Green tea (“Sencha”)	0.316	0.000	0.000	0.000	0.000	reject
151	Black tea	0.339	0.000	0.000	0.000	0.000	reject
152	Instant coffee	0.409	0.000	0.000	0.000	0.000	reject
153	Coffee beans	0.465	0.000	0.000	0.000	0.000	reject
154	Coffee beverages	0.317	0.000	0.000	0.000	0.000	reject
155	Fruit juice	0.509	0.000	0.000	0.000	0.000	reject
156	Beverages which contains juice	0.271	0.000	0.000	0.000	0.000	reject

No.	Item	Frequency of price changes	Hjort (1990) test of goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality
			Exponential	Weibull	Exponential- Gamma	Weibull- Gamma	
157	Vegetable juice	0.347	0.000	0.000	0.000	0.000	reject
158	Carbonated beverages	0.281	0.000	0.000	0.000	0.000	reject
159	Fermented lactic drinks, unsterilized (“Calpis”)	0.430	0.000	0.000	0.000	0.000	reject
160	Fermented lactic drinks, unsterilized (“Yakult”)	0.001	-	-	-	-	*
161	Sports soft drinks	0.439	0.000	0.000	0.000	0.000	reject
162	Mineral water	0.523	0.000	0.000	0.000	0.000	reject
163	“Sake A” (finest quality)	0.118	0.000	0.000	0.000	0.000	reject
164	“Sake B” (high quality)	0.167	0.000	0.000	0.000	0.000	reject
165	“Sake C” (medium quality)	0.287	0.000	0.000	0.000	0.000	reject
166	“Shochu”, distilled spirits	0.116	0.000	0.000	0.000	0.000	reject
167	Beer	0.286	0.000	0.000	0.000	0.000	reject
168	Beer (imported)	0.160	0.000	0.000	0.000	0.000	reject
169	Low-malt beer	0.331	0.000	0.000	0.000	0.000	reject
170	Whisky (imported)	0.300	0.000	0.000	0.000	0.000	reject
171	Whisky (43% vol. and over)	0.168	0.000	0.000	0.000	0.000	reject
172	Whisky (40% or more, but less than 43% vol.)	0.300	0.000	0.000	0.000	0.000	reject
173	Whisky (38% or more, but less than 40% vol.)	0.139	0.000	0.000	0.000	0.000	reject
174	Wine	0.246	0.000	0.000	0.000	0.000	reject
175	Wine (imported)	0.177	0.000	0.000	0.000	0.000	reject
176	Japanese noodles (eating out)	0.045	0.048*	0.307*	0.007	0.011*	*
177	Chinese noodles	0.047	0.000	0.064*	0.000	0.000	*
178	Spaghetti (eating out)	0.085	0.257*	0.320*	0.093*	0.061*	*
179	“Nigiri-zushi”, hand-rolled “Sushi”	0.072	0.000	0.008	0.000	0.000	reject
180	“Norimaki”, “Sushi” rolled in laver	0.075	0.004	0.009	0.000	0.000	*
181	Chicken & eggs on rice	0.047	0.001	0.185*	0.000	0.004	*
182	“Tendon”, prawns “Tempura” on rice	0.063	0.000	0.000	0.000	0.000	*
183	Curry & rice	0.064	0.107*	0.172*	0.038*	0.004	*
184	Bowl of rice topped with seasoned beef	0.136	0.000	0.000	0.000	0.000	reject
185	Ch(i)aotzu (eating out)	0.062	0.014*	0.059*	0.002	0.001	*
186	Hamburg steaks	0.102	0.059*	0.135*	0.010*	0.002	reject
187	Fried prawns	0.076	0.192*	0.282*	0.052*	0.028*	*
188	Lunch for children	0.082	0.029*	0.030*	0.027*	0.016*	*
189	Hamburgers	0.142	0.000	0.000	0.000	0.000	reject
190	Sandwiches	0.080	0.068*	0.219*	0.033*	0.048*	*
191	Pizza	0.061	0.000	0.185*	0.000	0.000	reject
192	Coffee (eating out)	0.054	0.320*	0.551*	0.025*	0.007	*
[Housing]							
193	Bathtubs	0.037	0.000	0.002	0.000	0.002	*
194	Toilet seat with a hot douche	0.293	0.000	0.000	0.000	0.000	reject
195	Hot-water supply equipment	0.047	0.171*	0.582*	0.186*	0.582*	*
196	Board	0.061	0.149*	0.265*	0.133*	0.116*	*
197	Paint	0.072	0.000	0.045*	0.000	0.001	reject
198	“Tatami” reupholstering	0.042	0.000	0.062*	0.000	0.000	*
199	Plastering	0.029	0.021*	0.056*	0.002	0.002	*
200	Gardening	0.029	0.010*	0.027*	0.012*	0.010*	*
201	Sheet glass replacement	0.031	0.000	0.259*	0.000	0.004	*
202	“Fusuma”, sliding doors reupholstering	0.017	0.856*	0.902*	0.890*	0.875*	*

No.	Item	Frequency of price changes	Hjort (1990) test of goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality
			Exponential	Weibull	Exponential- Gamma	Weibull- Gamma	
203	Carpentering	0.024	0.012*	0.026*	0.002	0.001	*
[Fuel, Light & Water Charge]							
204	Fire insurance premium	0.000	-	-	-	-	-
205	Electricity	0.070	0.000	0.000	0.000	0.000	reject
206	Gas	0.130	0.000	0.000	0.000	0.000	reject
207	Kerosene	0.425	0.000	0.000	0.000	0.000	reject
208	Water charges	0.005	0.000	0.000	0.000	0.000	reject
209	Sewerage charges	0.010	0.000	0.000	0.000	0.000	reject
[Furniture & Household Utensils]							
210	Microwave ovens	0.464	0.000	0.000	0.000	0.000	reject
211	Electric rice-cookers	0.553	0.000	0.000	0.000	0.000	reject
212	Electric pots	0.471	0.000	0.000	0.000	0.000	reject
213	Gas cooking tables	0.248	0.000	0.000	0.000	0.000	reject
214	Gas water heaters	0.152	0.000	0.000	0.000	0.000	reject
215	Refrigerators	0.683	0.000	0.000	0.000	0.000	reject
216	Vacuum cleaners	0.571	0.000	0.000	0.000	0.000	reject
217	Washing machines	0.635	0.000	0.000	0.000	0.000	reject
218	Sewing machines	0.115	0.001	0.014*	0.000	0.000	reject
219	Electric irons	0.472	0.000	0.000	0.000	0.000	reject
220	Room air conditioners	0.537	0.000	0.000	0.000	0.000	reject
221	Chests of drawers	0.208	0.000	0.000	0.000	0.000	reject
222	Wardrobes	0.188	0.000	0.000	0.000	0.000	reject
223	Sitting tables	0.193	0.000	0.000	0.000	0.000	reject
224	Kitchen cabinets	0.195	0.000	0.000	0.000	0.000	reject
225	Clocks	0.107	0.000	0.000	0.000	0.000	reject
226	Lighting apparatus	0.233	0.000	0.000	0.000	0.000	reject
227	Carpets	0.253	0.000	0.000	0.000	0.000	reject
228	“Goza”, rush floor coverings	0.196	0.000	0.000	0.000	0.000	reject
229	Curtains	0.226	0.000	0.000	0.000	0.000	reject
230	Beds	0.144	0.000	0.000	0.000	0.000	reject
231	Quilts	0.332	0.000	0.000	0.000	0.000	reject
232	Sheets	0.176	0.000	0.000	0.000	0.000	reject
233	Quilt covers	0.204	0.000	0.000	0.000	0.000	reject
234	Rice bowls	0.148	0.000	0.000	0.000	0.000	reject
235	Dishes	0.183	0.000	0.000	0.000	0.000	reject
236	Coffee cups & saucers	0.137	0.000	0.000	0.000	0.000	reject
237	Glasses	0.184	0.000	0.000	0.000	0.000	reject
238	Wine glasses	0.101	0.000	0.000	0.000	0.000	*
239	Sealed kitchenware	0.232	0.000	0.000	0.000	0.000	reject
240	Pans	0.160	0.000	0.000	0.000	0.000	reject
241	Pans (imported)	0.018	0.000	0.116*	0.000	0.000	*
242	Kettles	0.155	0.000	0.000	0.000	0.000	reject
243	Scrubbing brushes	0.303	0.000	0.000	0.000	0.000	reject
244	Shelves for microwave oven	0.134	0.000	0.000	0.000	0.000	reject
245	Fluorescent lamps	0.098	0.000	0.000	0.000	0.000	reject
246	Towels	0.140	0.000	0.000	0.000	0.000	reject
247	Vinyl hose	0.063	0.018*	0.274*	0.000	0.000	*
248	Clean water equipment	0.233	0.000	0.000	0.000	0.000	reject
249	Rolled toilet paper	0.459	0.000	0.000	0.000	0.000	reject
250	Liquid detergent, kitchen	0.571	0.000	0.000	0.000	0.000	reject
251	Detergent, laundry	0.604	0.000	0.000	0.000	0.000	reject
252	Food wrap	0.502	0.000	0.000	0.000	0.000	reject
253	Insecticide	0.209	0.000	0.000	0.000	0.000	reject
254	Moth repellent for clothes	0.393	0.000	0.000	0.000	0.000	reject

No.	Item	Frequency of price changes	Hjort (1990) test of goodness-of-fit				Shapiro-Wilk (1972) test for exponentiality
			Exponential	Weibull	Exponential- Gamma	Weibull- Gamma	
255	Fabric softener	0.611	0.000	0.000	0.000	0.000	reject
256	Fragrance	0.241	0.000	0.000	0.000	0.000	reject
257	Domestic help	0.010	0.000	0.000	0.000	0.000	reject
258	Charges for treatment of human waste	0.007	0.000	0.000	0.000	0.000	*
259	Charges for mop-rental	0.001	-	-	-	-	-
[Clothes & Footwear]							
260	Women's "Kimono"	0.092	0.000	0.002	0.000	0.000	reject
261	Women's "Obi"	0.078	0.000	0.002	0.000	0.000	reject
262	Men's slacks (jeans)	0.149	0.000	0.000	0.000	0.000	reject
263	Women's slacks (jeans)	0.162	0.000	0.000	0.000	0.000	reject
264	Boys' short pants	0.484	0.000	0.000	0.000	0.000	reject
265	Babies' clothes	0.203	0.000	0.000	0.000	0.000	reject
266	Men's business shirts (long sleeves)	0.215	0.000	0.000	0.000	0.000	reject
267	Men's undershirts (short sleeves)	0.125	0.000	0.000	0.000	0.000	reject
268	Men's briefs	0.120	0.000	0.000	0.000	0.000	reject
269	Brassieres	0.178	0.000	0.000	0.000	0.000	reject
270	Panties	0.076	0.000	0.000	0.000	0.000	reject
271	Slips	0.175	0.000	0.000	0.000	0.000	reject
272	Children's undershirts	0.217	0.000	0.000	0.000	0.000	reject
273	Men's shoes	0.177	0.000	0.000	0.000	0.000	reject
274	Women's shoes	0.225	0.000	0.000	0.000	0.000	reject
275	Children's shoes	0.186	0.000	0.000	0.000	0.000	reject
276	Canvas shoes (for adults)	0.144	0.000	0.000	0.000	0.000	reject
277	Canvas shoes (for children)	0.201	0.000	0.000	0.000	0.000	reject
278	Sandals	0.220	0.000	0.000	0.000	0.000	reject
279	"Zori", Japanese sandals	0.139	0.000	0.000	0.000	0.000	reject
280	Women's dress materials	0.159	0.000	0.000	0.000	0.000	reject
281	Men's suit materials	0.079	0.000	0.000	0.000	0.000	reject
282	Woolen yarn	0.101	0.000	0.000	0.000	0.000	reject
283	Hats & caps	0.049	0.000	0.050*	0.000	-	reject
284	Neckties	0.314	0.000	0.000	0.000	0.000	reject
285	Neckties (imported)	0.250	0.000	0.000	0.000	0.000	reject
286	Women's stockings	0.084	0.000	0.000	0.000	0.000	reject
287	Women's socks	0.227	0.000	0.000	0.000	0.000	reject
288	Belts	0.134	0.000	0.001	0.000	0.000	reject
289	Tailoring charges	0.041	0.000	0.005	0.000	0.000	reject
290	Laundry charges (men's business shirts)	0.050	0.000	0.174*	0.000	0.000	*
[Medical Care]							
291	Medicines for cold	0.126	0.000	0.000	0.000	0.000	reject
292	Antipyretic & analgesic medicines	0.087	0.000	0.001	0.000	0.000	reject
293	Gastrointestinal medicines	0.158	0.000	0.000	0.000	0.000	reject
294	Vitamins-A	0.102	0.000	0.000	0.000	0.000	reject
295	Vitamins-B	0.066	0.000	0.016*	0.000	0.000	reject
296	Health drinks	0.060	0.000	0.006	0.000	0.000	*
297	Dermal medicines	0.065	0.000	0.046*	0.000	0.000	*
298	Plasters	0.070	0.005	0.003	0.000	0.000	*
299	Eyewashes	0.184	0.000	0.000	0.000	0.000	reject
300	Breath fresheners	0.051	0.000	0.183*	0.000	0.000	*
301	Chinese medicines	0.109	0.000	0.000	0.000	0.000	reject
302	Disposable diapers	0.607	0.007	0.000	0.032*	0.000	reject

No.	Item	Frequency of price changes	Hjort (1990) test of goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality
			Exponential	Weibull	Exponential-Gamma	Weibull-Gamma	
303	Sanitary napkins	0.550	0.000	0.000	0.000	0.000	reject
304	Bath preparation	0.414	0.000	0.000	0.000	0.000	reject
305	Contact lenses cleaning solution	0.112	0.000	0.002	0.000	0.000	reject
306	Spectacles	0.066	0.000	0.008	0.000	0.000	reject
307	Contact lenses	0.076	0.278*	0.339*	0.291*	0.123*	*
308	Bathroom scales	0.262	0.000	0.000	0.000	0.000	reject
309	Thermometers	0.058	0.000	0.000	0.000	0.000	reject
310	Sphygmomanometers	0.293	0.000	0.000	0.000	0.000	reject
311	Medical treatment	0.000	-	-	-	-	-
312	Delivery fees in national hospital	0.070	0.000	0.000	0.000	0.000	reject
313	Delivery fees in public hospital	0.051	0.000	0.000	0.000	0.000	reject
314	Charges for massage	0.016	0.000	0.039*	0.000	0.000	*
315	Fees for complete medical checkup	0.019	0.000	0.000	0.000	0.000	reject

[Transportation & Communication]

316	Railway fares (ordinary fares, excluding “Shinkansen”)	0.000	-	-	-	-	-
317	Railway fares (special fares, excluding “Shinkansen”)	0.000	-	-	-	-	-
318	Railway fares (students’ season tickets)	0.000	-	-	-	-	-
319	Railway fares (commuters’ season tickets)	0.000	-	-	-	-	-
320	Railway fares (ordinary passengers)	0.001	-	-	-	-	reject
321	Railway fares (students’ season tickets)	0.000	-	-	-	-	-
322	Railway fares (commuters’ season tickets)	0.001	-	-	-	-	-
323	Bus fares	0.003	-	-	-	-	*
324	Taxi fares	0.000	-	-	-	-	-
325	Bicycles	0.141	0.000	0.000	0.000	0.000	reject
326	Gasoline (regular)	0.583	0.000	0.000	0.000	0.000	reject
327	Gasoline (premium)	0.579	0.000	0.000	0.000	0.000	reject
328	Tires	0.077	0.000	0.000	0.000	0.000	*
329	Car wax	0.106	0.000	0.002	0.000	0.000	reject
330	Regular inspection	0.021	0.000	0.009	0.001	0.006	*
331	Muffler replacement	0.021	0.000	0.004	0.000	0.000	*
332	Puncture repairs	0.019	0.006	0.046*	0.008	0.046*	*
333	Motor oil replacement	0.040	0.013*	0.235*	0.000	0.000	*
334	Charges for garage rental	0.026	0.391*	0.699*	0.068*	0.043*	*
335	Charges for parking	0.014	0.000	0.107*	0.001	0.005	*
336	Charges for driving license	0.015	0.000	-	-	-	reject
337	Telephone charges	0.001	-	-	-	-	-
338	Mobile telephone charges	0.007	0.000	0.000	0.000	0.000	*
339	Forwarding charges	0.020	0.004	0.220*	0.006	0.220*	*
340	Communication equipments	0.604	0.000	0.000	0.000	0.000	reject

[Education]

341	Tutorial fees	0.062	0.000	0.000	0.000	0.000	*
-----	---------------	-------	-------	-------	-------	-------	---

[Reading & Recreation]

342	TV sets	0.568	0.000	0.000	0.000	0.000	reject
343	Stereo phonograph sets	0.548	0.000	0.000	0.000	0.000	reject

No.	Item	Frequency of price changes	Hjort (1990) test of goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality
			Exponential	Weibull	Exponential- Gamma	Weibull- Gamma	
344	Mobile audio equipment	0.596	0.000	0.000	0.000	0.000	reject
345	Video tape recorders	0.544	0.000	0.000	0.000	0.000	reject
346	Cameras	0.288	0.133*	0.000	0.250*	0.000	reject
347	Ball-point pens	0.036	0.057*	0.334*	0.052*	0.072*	*
348	Pencils	0.061	0.000	0.000	0.000	0.000	reject
349	Marking pens	0.023	0.000	0.083*	0.000	0.000	*
350	Notebooks	0.069	0.000	0.000	0.000	0.000	reject
351	Albums	0.047	0.042*	0.067*	0.002	0.001	*
352	Papers for office automation	0.092	0.000	0.000	0.000	0.000	reject
353	Cellophane adhesive tape	0.034	0.000	0.050*	0.000	0.000	*
354	Pencil cases	0.110	0.000	0.000	0.000	0.000	reject
355	Golf clubs	0.104	0.000	0.000	0.000	0.000	reject
356	Soccer balls	0.136	0.000	0.003	0.000	0.000	reject
357	Baseball gloves	0.156	0.000	0.000	0.000	0.000	reject
358	Tennis rackets	0.079	0.000	0.043*	0.000	0.000	reject
359	Tennis rackets (imported)	0.186	0.000	0.000	0.000	0.000	reject
360	Fishing rods	0.137	0.000	0.000	0.000	0.000	reject
361	Pants for exercise	0.197	0.000	0.000	0.000	0.000	reject
362	Swimming suits	0.182	0.000	0.000	0.000	0.000	reject
363	Video games, hardware	0.320	0.000	0.000	0.000	0.000	reject
364	Dolls	0.046	0.000	0.000	0.000	0.000	*
365	Toy cars	0.038	0.000	0.046*	0.000	0.000	*
366	Building blocks	0.122	0.000	0.000	0.000	0.000	reject
367	Cut flowers (Carnations)	0.885	0.009	0.000	0.009	0.000	reject
368	Cut flowers (Chrysanthemums)	0.903	0.605†	-	-	-	reject
369	Cut flowers (Roses)	0.899	0.006	0.000	0.006	0.000	reject
370	Films	0.102	0.014*	0.001	0.003	0.000	reject
371	Media for audio recording	0.162	0.000	0.000	0.000	0.000	reject
372	Video tapes	0.236	0.000	0.000	0.000	0.000	reject
373	Pet foods (dog foods)	0.409	0.000	0.000	0.000	0.000	reject
374	Pet foods (cat foods)	0.273	0.000	0.000	0.000	0.000	reject
375	Flowerpots	0.074	0.000	0.014*	0.000	0.000	reject
376	Gardening earth	0.175	0.000	0.000	0.000	0.000	reject
377	Dry batteries	0.118	0.000	0.000	0.000	0.000	reject
378	Lesson fees (English conversation school)	0.052	0.000	0.000	0.000	0.000	*
379	Lesson fees (calligraphy school)	0.018	0.000	0.045*	0.000	0.001	*
380	Lesson fees (music school)	0.023	0.000	0.001	0.000	0.000	*
381	Lesson fees (swimming school)	0.018	0.000	0.000	0.000	0.000	*
382	Lesson fees (dressmaking school)	0.021	0.000	0.000	0.000	0.000	reject
383	Lesson fees (cooking school)	0.030	0.012*	0.057*	0.001	0.001	*
384	Lesson fees, driving school	0.062	0.000	0.000	0.000	0.000	*
385	Admission, movies	0.018	0.000	0.000	0.000	0.000	reject
386	Charges for practicing golf	0.026	0.000	0.000	0.000	0.000	*
387	Charges for playing golf	0.255	0.000	0.000	0.000	0.000	reject
388	Tennis court charges	0.081	0.000	0.000	0.000	0.000	reject
389	Game charges, bowling	0.004	0.000	0.011*	0.000	0.000	*
390	Swimming pool charges	0.078	0.000	0.000	0.000	0.000	*
391	Admission fees to the art museum	0.250	0.000	0.000	0.000	0.000	reject
392	Game charges, mahjong	0.017	0.000	0.189*	0.000	0.000	*
393	“karaoke room” charges	0.106	0.316*	0.091*	0.241*	0.038*	*
394	Photo processing charges	0.051	0.000	0.000	0.000	0.000	*
395	Charges for video rental	0.099	0.000	0.001	0.000	0.000	reject
396	Veterinary surgeon fees	0.015	0.000	0.000	0.000	0.000	*
[Miscellaneous]							
397	Bathing charges (adults)	0.003	0.000	0.001	0.000	0.000	*

No.	Item	Frequency of price changes	Hjort (1990) test of goodness-of-fit				Shapiro–Wilk (1972) test for exponentiality
			Exponential	Weibull	Exponential- Gamma	Weibull- Gamma	
398	Bathing charges (children, 6–11 years)	0.001	-	-	-	-	*
399	Bathing charges (children, under 6 years)	0.000	-	-	-	-	-
400	Men’s haircut charges	0.020	0.001	0.072*	0.002	0.002	*
401	Permanent wave charges	0.039	0.000	0.000	0.000	0.000	*
402	Women’s haircut charges	0.026	0.007	0.318*	0.008	0.088*	*
403	Hair dyeing charges	0.030	0.001	0.106*	0.000	0.000	*
404	Electric shavers	0.412	0.000	0.000	0.000	0.000	reject
405	Electric shavers (imported)	0.358	0.000	0.000	0.000	0.000	reject
406	Toothbrushes	0.376	0.000	0.000	0.000	0.000	reject
407	Toilet soap	0.386	0.000	0.000	0.000	0.000	reject
408	Shampoo	0.412	0.000	0.000	0.000	0.000	reject
409	Hair rinse	0.342	0.000	0.000	0.000	0.000	reject
410	Toothpaste	0.453	0.000	0.000	0.000	0.000	reject
411	Hair liquid	0.170	0.000	0.000	0.000	0.000	reject
412	Hair tonic	0.154	0.000	0.000	0.000	0.000	reject
413	Face cream-B	0.070	0.000	0.016*	0.000	0.000	*
414	Toilet lotion	0.197	0.000	0.000	0.000	0.000	reject
415	Hair Dyeing	0.347	0.000	0.000	0.000	0.000	reject
416	Handbags	0.231	0.000	0.000	0.000	0.000	reject
417	Handbags (imported)	0.223	0.000	0.000	0.000	0.000	reject
418	Suitcases	0.183	0.000	0.000	0.000	0.000	reject
419	Rings	0.077	0.000	0.209*	0.000	0.000	reject
420	Wrist watches	0.070	0.000	0.000	0.000	0.000	reject
421	Wrist watches (imported)	0.113	0.000	0.000	0.000	0.000	reject
422	Repair charges of wrist watches	0.013	0.000	0.040*	0.000	0.000	*
423	Men’s umbrellas	0.181	0.000	0.000	0.000	0.000	reject
424	Handkerchiefs	0.041	0.000	0.001	0.000	0.000	reject
425	Nursery school fees	0.006	0.000	0.000	0.000	0.000	*
426	Charges for certificates of registered stamps	0.000	-	-	-	-	-
427	Charges for certificates of permanent registration	0.000	-	-	-	-	-
428	Charges for acquisition of passport	0.000	-	-	-	-	-
429	Day service fees of nursing care for the aged	0.019	0.000	0.000	0.000	0.000	reject

Table 5: Goodness-of-fit test results by item. Entries are the 429 items in the Japanese CPI for 2000–2005. P-values are reported for the Hjort (1990) test of goodness-of-fit. †, the Hjort (1990) test cannot reject the hypothesis due to the high price flexibility (see also Footnote 19 in the text); *, the item cannot reject the hypothesis that the underlying distribution belongs to a specified family of distributions. “reject” in the last column means the Shapiro and Wilk (1972) test rejects the null hypothesis of exponentiality at the one percent significance level.