E C O N O M I C S   B U L L E T I N

Semi-nonparametric estimation of regression-based survival models

Hiroaki Masuhara
Section of Health Economics, Department of Gerontological Policy, National Center for Geriatrics and Gerontology (NCGG)

Abstract

To estimate survival data with unobserved heterogeneity, this paper proposes the generalized lognormal survival analysis using Hermite polynomials and the Box-Cox transformation. The General Social Survey (GSS) in 2002 demonstrates good performance of the proposed model.

This study is supported by the Research Grant for Longevity Sciences (18C-7) from the Ministry of Health, Labor and Welfare, in Japan.
Citation: Masuhara, Hiroaki, (2007) "Semi-nonparametric estimation of regression-based survival models." *Economics Bulletin*, Vol. 3, No. 61 pp. 1-12
Submitted: October 12, 2007. **Accepted:** November 14, 2007.
URL: http://economicsbulletin.vanderbilt.edu/2007/volume3/EB-07C20101A.pdf

1. Introduction

Survival analysis is widespread in applied econometrics such as labor economics, industrial organizations, health economics, and population economics. Many survival models, including the most popular Cox's proportional hazard model, do not explicitly assume unobserved heterogeneity. It is empirically difficult to separate the effect of duration dependence from those of unobserved heterogeneity. However, in social science, the existence of omitted variables is inevitable, and it is always inadequate to control population heterogeneity. Therefore, the model without unobserved heterogeneity overestimates (or underestimates) the degree of negative (positive) duration dependence in the hazard, even if this model has consistency of coefficients.

One way to introduce unobserved heterogeneity into survival models is to assume that the heterogeneity is multiplicative to the hazard function and follows a gamma distribution. This method is simple and has a closed form solution. However, this is a parametric method, and the distribution assumption of heterogeneity is rather important. Furthermore, we do not determine the econometric interpretation of this method if the probability density function is not an exponential or Weibull distribution.

In this paper, we propose new semi-nonparametric survival models that generalize unobserved heterogeneity, as well as a dependent variable of the lognormal survival model. First, we generalize the log-transformed dependent variable using the Box-Cox transformation, which contains various function forms. Second, we generalize the normally distributed unobserved heterogeneity using Hermite polynomials, which include a normal distribution as a special case. In the empirical application, this paper compares the performance of the proposed models, using the General Social Survey (GSS) in 2002 following Winkelmann and Boes (2006).

This paper is organized as follows. Section 2 proposes the semi-nonparametric regression-based survival model. Section 3 depicts the application of the fertility data, and Section 4 presents our concluding remarks.

2. The model

We consider the standard survival analysis. Suppose that a random variable, $T_i, i = 1, \ldots, N$, has a continuous probability distribution $f(t_i)$, where t_i is a realization of T_i. The cumulative distribution function of this variable takes the following form: $F(t_i) = \int_0^{t_i} f(s_i)\, ds_i$. We define a censoring indicator $c_i = 1$ if the observation is censored, and $c_i = 0$ if the observation is uncensored. Further, the log-likelihood function is obtained as follows: $\ln L = \sum_{i=1}^{N} (1 - c_i) \ln f(t_i) + c_i \ln [1 - F(t_i)]$.

In the lognormal survival model, a logarithmic survival variable consists of a linear index of independent variables and an additive normally distributed error term: $\ln t_i = x_i'\beta + \varepsilon_i$, where x_i is a $K \times 1$ covariate vector of covariates and $\beta \sim K \times 1$ is a parameter to be estimated. Although the lognormal model clearly assumes heterogeneity, it has some disadvantages. First, the transformation of the dependent variable t_i is quite arbitrary. The log transformation may not always be the best choice. Second, this model is parametric; we assume the normally distributed unobserved heterogeneity and estimate the parameters using the maximum likelihood (ML) method. However, the incorrect specification of the error term causes the inconsistency of the ML. Therefore, we require a more flexible method to estimate survival data.

First, we generalize the log-transformed dependent variable using the Box-Cox transformation, such as $\left(t_i^\lambda - 1\right)/\lambda = x_i'\beta + \varepsilon_i$, where λ is a parameter of the Box-Cox transformation. In this model, when $\lambda \to 0$, the left-hand side (LHS) is $\ln(t_i)$, and when $\lambda = 1$, the LHS is $t_i - 1$. The Box-Cox transformation contains the log-linear and linear models as a special case.

Second, we generalize the error term. If the relation between the error term and the survival random variable is quasi-linear, like the lognormal model, we obtain the following conditions using the Jacobian of the transformation $dt_i/d\varepsilon_i$: $f(\varepsilon_i) = f(t_i)\,dt_i/d\varepsilon_i$ and $F(\varepsilon_i) = F(t_i)$. Therefore, we concentrate the distribution of ε_i. Following Gallant and Nychka (1987), who proposed the semi-nonparametric series based on a normal distribution, we approximate the unknown error term using Hermite polynomials.[1] The approximated density is obtained as follows:

$$f(\varepsilon_i) = \frac{1}{P}\left(\sum_{k=0}^{K}\alpha_k\varepsilon_i^k\right)^2 \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) \equiv \frac{f^*(\varepsilon_i)}{P}, \tag{1}$$

where σ is a standard deviation parameter, α_k is a parameter of a Hermite series to be estimated, and $P = \int_{-\infty}^{\infty} f^*(\varepsilon_i)\,d\varepsilon_i$ ensures integration to 1 by scaling density. Moreover, we impose $\alpha_0 = 1$ to ensure the identification of the parameters. This Hermite series density contains not only a normal distribution as a special case but also fat-tailed and twin peak distributions.

We discuss two generalizations for survival data and term the above model as the Box-Cox transformed semi-nonparametric (BC-SN) model. This model generalizes not only the error terms using Hermite polynomials but also the dependent variable using the transformation of t_i, and includes

[1] See Gabler *et al.* (1993), van der Klaauw and Koning (2003), and Stewart (2004) for further details.

the lognormal model as a special case. In estimating survival data, the calculation of $F(\varepsilon_i) = \int_{-\infty}^{\varepsilon_i} f(\nu_i)\, d\nu_i$ is required. Fortunately, this integral has a closed-form solution and is easy to evaluate.[2]

The log-likelihood function is usually maximized by gradient-based methods such as Newton, BFGS, or BHHH algorithms. However, these algorithms are sensitive to initial values and contain the problem of local maxima. It is difficult to estimate parameters α_k of the semi-nonparametric model correctly, and it is practically impossible to attempt several initial conditions. To avoid the problem of local maxima, we use the stochastic evolution algorithm (StocE), which attains the global maxima with a high probability (Saab and Rao, 1990; Sait and Youssef, 2000).

Let the parameter vector of this density be $\theta = [\beta', \sigma, \lambda, \alpha_1, \alpha_2, \ldots]'$. Further, the StocE algorithm obtains the parameters as follows:

1) Set the initial parameters p_0, θ_0, R, p_{up}, r and $t \leftarrow 0$. Calculate the log-likelihood $\ln L_0$ under θ_0.

2) Set $\theta_{\mathrm{best}} = \theta_0$ and $\ln L_{\mathrm{best}} = \ln L_0$.

3) Generate $\hat{\theta}$ in the neighborhood of θ; e.g., calculate $\hat{\theta} = \theta_t + a \cdot u_1$ and $\ln \hat{L}\left(\hat{\theta}\right)$, where a is a positive constant number and u_1 is a uniform random vector on $[-1, 1]$.

4) If $\ln \hat{L} - \ln L_t > -p_t \cdot u_2$, where u_2 is a uniform random number, then $\theta_{t+1} = \hat{\theta}$; otherwise, $\theta_{t+1} = \theta_t$. Calculate $\ln L(\theta_{t+1})$.

5) If $\ln L(\theta_{t+1}) = \ln L(\theta_t)$, then set $p_{t+1} = p_t + p_{\mathrm{up}}$; otherwise, $p_{t+1} = p_0$.

6) If $\ln L(\theta_{t+1}) > \ln L_{\mathrm{best}}$, then $\theta_{\mathrm{best}} = \theta_{t+1}$, $\ln L_{\mathrm{best}} = \ln L_{t+1}$, and $r = r - R$; otherwise, $r = r + 1$. Set $t \leftarrow t + 1$.

7) Repeat steps 3 to 6 until $r > R$.

The StocE algorithm resembles the simulated annealing (SA) method, which is a stochastic technique using the Metropolis algorithm. If the annealing process is slow, the SA algorithm is similar to a random search and its convergence speed is slow. The StocE algorithm eliminates the inefficient path with a probability $p_t u_2$. Therefore, in general, the StocE algorithm is faster than the SA. However, the StocE algorithm does not ensure the convergence of the global maximum, because the sequence of the StocE is not a perfect Markov chain. Nonetheless, in practice, the StocE is effective.

[2]Due to space constraints, we do not provide the calculation results of $F(\varepsilon_i)$. For further details, please contact the author.

3. An application to fertility

We present the results of the application of the proposed model. In this example, we analyze the factors that affect the time until a woman bears her first child, using the data obtained from the GSS in 2002, an annual or biannual cross-section survey that began in 1972.[3] Following Winkelmann and Boes (2006), we regard the variable age at the time of the first child's birth as duration and employ women under the age of 40. The total number of observations is $1,371$. There are two types of women: Type A includes women who have had their first child (the number of uncensored sample is 1,154); Type B includes women who are childless under the age of 40 (the number of right-censored samples is 217). Moreover, this survey considers the number of years of formal schooling, the number of siblings, four dummy variables, namely, those in the low-income group at age 16 (less than average income), those who were urban residents at age 16, those who are white, and those who are immigrants.

We investigate the result of the BC-SN model but find that the parameter λ is nearly zero (and not significant). Further, we estimate the semi-nonparametric (LN-SN) model with log transformation ($\lambda = 0$). For comparison purposes, we include the standard lognormal survival model and the lognormal model with gamma distributed unobserved heterogeneity (LN-G). Table 1 shows the estimated results of the four models and also presents the values of the log-likelihood, Akaike's information criteria (AIC), and Bayesian information criteria (BIC). The maximum value of the log-likelihood and the minimum value of the AIC is the LN-SN model (the reason for the positive log-likelihood is $\sigma < 1$). The minimum value of the BIC is the LN-G model. Further, we use Vuong's (1989) test to select the unique model between the LN-SN and LN-G models. Let $f^{(1)}$ be the likelihood of model 1 and $f^{(2)}$ be that of model 2. Under the null hypothesis that both the models are equivalent ($E\left[\ln f^{(1)} - \ln f^{(2)}\right] = 0$), the test statistic

$$\frac{\sum_{i=1}^{N}\left[\ln f^{(1)} - \ln f^{(2)}\right]}{\sqrt{\sum_{i=1}^{N}\left[(\ln f^{(1)} - \ln f^{(2)})^2 - \left(\sum_{i=1}^{N}\ln f^{(1)} - \ln f^{(2)}\right)^2/N\right]}} \tag{2}$$

follows a standard normal distribution. If the statistic exceeds the critical value c, model $f^{(1)}$ is better than model $f^{(2)}$. If the statistic is smaller than $-c$, model $f^{(2)}$ is better than model $f^{(1)}$. The test statistic for the LN-SN

[3] The data "kids.asc" is downloadable from
http://www.uzh.ch/sts/research/publications/microdata/index.html.

model against the LN-G model is 1.252, and its value at the significant level is 0.105.[4] Hence, there is (weak) evidence that the LN-SN model is the best of the four.

In Table 1, we find three features of the estimated parameters. First, the estimated parameters of the four models closely resemble each other. Second, the significant level of the estimated parameters also resemble each other. Third, the values of the estimated parameters of the BC-SN and LN-SN models are midway of those of the lognormal and LN-G models. The reason for these features is that the BC-SN, LN-SN, and LN-G models are based on and extended to the standard lognormal model.

However, there are large differences among the four models. For example, in the lognormal model, the value of *low income at age 16* is -0.021 and is not statistically significant; in all the remaining models, its value ranges from -0.031 to -0.038 and is statistically significant at the 5% level. Moreover, the value of *white* is 0.091 in the LN-G model and 0.088 in the LN-SN model. The difference between these two models is small but negligible in this data.

Figure 1 shows the estimated density of ε_i. The solid line is the LN-SN model,[5] the dotted line is the lognormal model, and the dashed line is the LN-G model. The density of the LN-SN model is skewed to the left and has a fatter tail. However, the density of the LN-G model has the fattest tail among the three models.

4. Conclusion

This paper proposes new semi-nonparametric survival models that generalize both an explanatory variable and unobserved heterogeneity. The former is Box-Cox transformation and the latter is a Hermite series. In an example using the GSS data, the Box-Cox transformation does not work well. However, the LN-SN model overcomes the other models, except for the BIC, and shows a good performance. Therefore, there is (weak) evidence that the LN-SN model is the best of the four. Moreover, the coefficients of the models, except for the lognormal model, resemble each other but the difference is not negligible.

In both econometric and statistic interpretations, the results may show a small difference between the LN-G and semi-nonparametric models. This view is incorrect. The LN-G model does not have a clear econometric interpretation because this model assumes multiplicative heterogeneity not to

[4]The tests for the LN-SN model, the BC-SN model, and the LN-G model against the lognormal model are 6.327, 6.320, and 4.999, respectively. This means that the lognormal assumption is strongly rejected.

[5]The densities of the LN-SN and BC-SN models are almost identical. Thus, we omit the latter in Figure 1.

the probability density function but to the *hazard* function. That is, the LN-G model explicitly assumes multi-heterogeneity. Therefore, in applied econometrics, the semi-nonparametric model proposed in this paper is an important method to estimate a demand or supply function because it assumes only single and additively separable heterogeneity.

References

Gallant, A.R., and D.W. Nychka (1987) "Semi-nonparametric Maximum Likelihood Estimation" *Econometrica* **55**, 363-390.

Gabler, S., F. Laisney, and M. Lechner (1993) "Semiparametric Estimation of Binary-choice Models with an Application to Labor-force Participation" *Journal of Business and Economic Statistics* **11**, 61-80.

Saab, Y.G., and V.B. Rao (1990) "Stochastic Evolution: A Fast Effective Heuristic for Some Genericlayout Problems" Design Automation Conference 1990 Proceedings 27th ACM/IEEE.

Sait, S.M., and H. Youssef (2000) *Iterative Computer Algorithms with Applications in Engineering: Solving Combinatorial Optimization Problems*, IEEE Computer Society: California.

Stewart, M.B. (2004) "Semi-nonparametric Estimation of Extended Ordered Probit Models" *Stata Journal* **4**, 27-39.

Van der Klaauw, B., and R.H. Koning (2003) "Testing the Normality Assumption in the Sample Selection Model with an Application to Travel Demand" *Journal of Business and Economic Statistics* **21**, 31-42.

Vuong, Q.H. (1989) "Likelihood Ratio Tests for Model Selection and Nonnested Hypotheses" *Econometrica* **57**, 307-333.

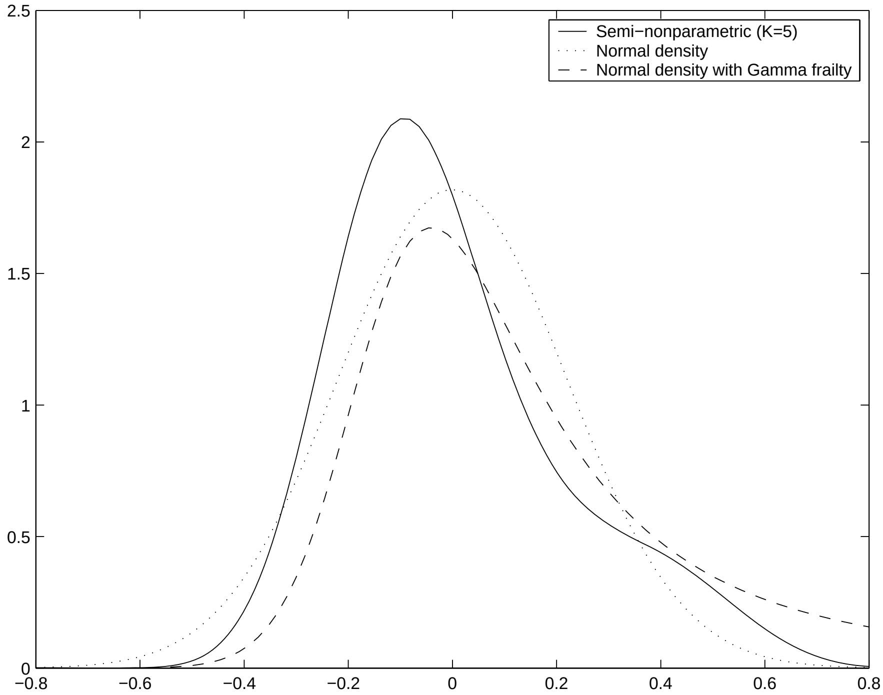
Winkelmann, R. and S. Boes (2006) *Analysis of Microdata*, Springer: Berlin.

Table 1: Estimation results of age at first birth

	lognormal	LN-G	LN-SN	BC-SN
years of education	0.031	0.032	0.031	0.031
	(0.002)	(0.002)	(0.002)	(0.002)
number of siblings	−0.005	−0.003	−0.004	−0.004
	(0.002)	(0.002)	(0.002)	(0.002)
white	0.083	0.091	0.088	0.088
	(0.015)	(0.013)	(0.012)	(0.012)
immigrant	0.056	0.060	0.054	0.054
	(0.019)	(0.018)	(0.016)	(0.016)
low income at age 16	−0.021	−0.038	−0.031	−0.031
	(0.016)	(0.015)	(0.013)	(0.013)
lived in city at age 16	0.008	−0.004	−0.004	−0.004
	(0.013)	(0.011)	(0.011)	(0.011)
constant	2.696	2.615	2.693	2.693
	(0.038)	(0.036)	(0.032)	(0.032)
σ	0.219	0.150	0.214	0.214
	(0.005)	(0.007)	(0.004)	(0.004)
γ^{-1}		0.673		
		(0.085)		
λ				0.000
				(0.001)
α_1			−1.555	−1.555
			(0.065)	(0.065)
α_2			−0.171	−0.166
			(0.211)	(0.208)
α_3			15.489	15.493
			(0.379)	(0.379)
α_4			1.458	1.447
			(0.728)	(0.735)
α_5			−19.030	−19.039
			(1.139)	(1.139)
log-likelihood	−72.933	−8.699	0.534	0.440
AIC	161.865	35.398	24.932	27.119
BIC	203.652	82.407	92.835	100.245
number of regressors	8	9	13	14

Notes: Standard errors are in parentheses; LN-G, LN-SN, and BC-SN denote the lognormal model with gamma-distributed unobserved heterogeneity, the semi-nonparametric model with the log transformation, and the Box-Cox transformed semi-nonparametric, respectively; AIC $= -2\ln L + 2K$, BIC $= -2\ln L + K\ln N$, where L is the maximized likelihood, K is the number of parameters, and N is the number of observations ($N = 1,371$); γ is the parameter of the gamma frailty.

Figure 1: The estimated density function of ε_i



Appendix (Not for Publication)

A The Box-Cox transformed semi-nonparametric survival model

Following Gabler *et al.* (1993), Eq. (2) takes

$$f(\varepsilon_i) = \frac{1}{P}\left(\sum_{k=0}^{K}\alpha_k\varepsilon_i^k\right)^2 \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) \equiv \frac{f^*(\varepsilon_i)}{P}, \tag{A1}$$

where $\varepsilon_i = \left(t_i^\lambda - 1\right)/\lambda - x_i'\beta$ and

$$P = \int_{-\infty}^{\infty}\left(\sum_{k=0}^{K}\alpha_k\varepsilon_i^k\right)^2 \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right)d\varepsilon_i. \tag{A2}$$

We require algebraic computations of Eq. (A2). Van der Klaauw and Koning (2003) show the following recursion formulas:

$$I_k(a,b) = \int_a^b u^k \exp\left(-\left(\frac{u}{\delta}\right)^2\right)du. \tag{A3}$$

Equation (A3) obtains

$$I_j(-\infty,\infty) = \begin{cases} \delta\sqrt{\pi}, & j=0. \\ 0, & j=1,3,5,\dots \\ \frac{(j-1)\delta^2}{2}I_{j-2}(-\infty,\infty), & j=2,4,6,\dots \end{cases} \tag{A4}$$

Substituting $\delta = \sqrt{2}\sigma$ into (A5) and calculating up to $j = 10$ yields

$$\begin{aligned} I_0 &= \sqrt{2}\sigma\sqrt{\pi}, & I_2 &= \frac{1}{2}\left(\sqrt{2}\sigma\right)^3\sqrt{\pi}, \\ I_4 &= \frac{3}{4}\left(\sqrt{2}\sigma\right)^5\sqrt{\pi}, & I_6 &= \frac{15}{8}\left(\sqrt{2}\sigma\right)^7\sqrt{\pi}, \\ I_8 &= \frac{105}{16}\left(\sqrt{2}\sigma\right)^9\sqrt{\pi}, & I_{10} &= \frac{945}{32}\left(\sqrt{2}\sigma\right)^{11}\sqrt{\pi}. \end{aligned}$$

When $K = 5$, we obtain the following relation after some algebraic manipulation.

$$\begin{aligned} \left(\sum_{k=0}^{5}\alpha_k\varepsilon_i^k\right)^2 &= \left(\alpha_0^2\right) + \varepsilon_i\left(2\alpha_0\alpha_1\right) + \varepsilon_i^2\left(2\alpha_0\alpha_2 + \alpha_1^2\right) + \varepsilon_i^3\left(2\alpha_0\alpha_3 + 2\alpha_1\alpha_2\right) \\ &\quad + \varepsilon_i^4\left(2\alpha_0\alpha_4 + 2\alpha_1\alpha_3 + \alpha_2^2\right) + \varepsilon_i^5\left(2\alpha_0\alpha_5 + 2\alpha_1\alpha_4 + 2\alpha_2\alpha_3\right) \\ &\quad + \varepsilon_i^6\left(2\alpha_1\alpha_5 + 2\alpha_2\alpha_4 + \alpha_3^2\right) + \varepsilon_i^7\left(2\alpha_2\alpha_5 + 2\alpha_3\alpha_4\right) \\ &\quad + \varepsilon_i^8\left(2\alpha_3\alpha_5 + \alpha_4^2\right) + \varepsilon_i^9\left(2\alpha_4\alpha_5\right) + \varepsilon_i^{10}\left(\alpha_5^2\right). \end{aligned} \tag{A5}$$

Substituting Eq. (A3), (A4), and (A5) into Eq. (A2) yields

$$\begin{aligned}
P =& \alpha_0^2 + \left(2\alpha_0\alpha_2 + \alpha_1^2\right)\frac{1}{2}\left(\sqrt{2}\sigma\right)^2 + \left(2\alpha_0\alpha_4 + 2\alpha_1\alpha_3 + \alpha_2^2\right)\frac{3}{4}\left(\sqrt{2}\sigma\right)^4 \\
&+ \left(2\alpha_1\alpha_5 + 2\alpha_2\alpha_4 + \alpha_3^2\right)\frac{15}{8}\left(\sqrt{2}\sigma\right)^6 \\
&+ \left(2\alpha_3\alpha_5 + \alpha_4^2\right)\frac{105}{16}\left(\sqrt{2}\sigma\right)^8 + \left(\alpha_5^2\right)\frac{945}{32}\left(\sqrt{2}\sigma\right)^{10}.
\end{aligned} \tag{A6}$$

Therefore, the probability density function of heterogeneity consists of Eq. (A1) and Eq. (A6). To ensure a zero mean, we impose the restriction $\alpha_0 = 1$ and $E(\varepsilon_i) = 0$. The expectation term $E(\varepsilon_i)$ takes the form as follows:

$$\begin{aligned}
E(\varepsilon_i) =& \frac{1}{P}\int_{-\infty}^{\infty} \varepsilon_i f^*(\varepsilon_i)\, d\varepsilon_i \\
=& \frac{1}{P}\left[2\alpha_1\left(2\alpha_0 I_2 + 2\alpha_2 I_4 + 2\alpha_4 I_6\right)\right. \\
&+ 2\alpha_3\left(2\alpha_0 I_4 + 2\alpha_2 I_6 + 2\alpha_4 I_8\right) \\
&\left.+2\alpha_5\left(2\alpha_0 I_6 + 2\alpha_2 I_8 + 2\alpha_4 I_{10}\right)\right].
\end{aligned} \tag{A7}$$

From the Eq. (A7), we obtain the following relation:

$$\alpha_5 = -\frac{\alpha_1\left(\alpha_0\delta^3 + \frac{3}{2}\alpha_2\delta^5 + \frac{15}{4}\alpha_4\delta^7\right) + \alpha_3\left(\frac{3}{2}\alpha_0\delta^5 + \frac{15}{4}\alpha_2\delta^7 + \frac{105}{8}\alpha_4\delta^9\right)}{\left(\frac{15}{4}\alpha_0\delta^7 + \frac{105}{8}\alpha_2\delta^9 + \frac{945}{16}\alpha_4\delta^{11}\right)}. \tag{A8}$$

Next, we present the calculation of $F(\varepsilon_i) = \int_{-\infty}^{\varepsilon_i} f(\nu_i)\, d\nu_i$. When $K = 5$, we obtain the following relation using Eq. (A3):

$$
\begin{aligned}
I_0(-\infty, \varepsilon_i) &= \sqrt{2\pi}\sigma\Phi\left(\frac{\varepsilon_i}{\sigma}\right), \\
I_1(-\infty, \varepsilon_i) &= -\sigma^2 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right), \\
I_2(-\infty, \varepsilon_i) &= -\sigma^2\varepsilon_i \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) + \sigma^2 I_0(-\infty, \varepsilon_i), \\
I_3(-\infty, \varepsilon_i) &= -\sigma^2\varepsilon_i^2 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) + 2\sigma^2 I_1(-\infty, \varepsilon_i), \\
I_4(-\infty, \varepsilon_i) &= -\sigma^2\varepsilon_i^3 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) + 3\sigma^2 I_2(-\infty, \varepsilon_i), \\
I_5(-\infty, \varepsilon_i) &= -\sigma^2\varepsilon_i^4 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) + 4\sigma^2 I_3(-\infty, \varepsilon_i), \\
I_6(-\infty, \varepsilon_i) &= -\sigma^2\varepsilon_i^5 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) + 5\sigma^2 I_4(-\infty, \varepsilon_i), \\
I_7(-\infty, \varepsilon_i) &= -\sigma^2\varepsilon_i^6 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) + 6\sigma^2 I_5(-\infty, \varepsilon_i), \\
I_8(-\infty, \varepsilon_i) &= -\sigma^2\varepsilon_i^7 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) + 7\sigma^2 I_6(-\infty, \varepsilon_i), \\
I_9(-\infty, \varepsilon_i) &= -\sigma^2\varepsilon_i^8 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) + 8\sigma^2 I_7(-\infty, \varepsilon_i), \\
I_{10}(-\infty, \varepsilon_i) &= -\sigma^2\varepsilon_i^9 \exp\left(-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right) + 9\sigma^2 I_8(-\infty, \varepsilon_i).
\end{aligned}
\tag{A9}
$$

Therefore,

$$
\begin{aligned}
F(\varepsilon_i) = &\int_{-\infty}^{\varepsilon_i} f(\nu_i)\, d\nu_i \\
= &\Big[\alpha_0^2 I_0(-\infty, \varepsilon_i) + 2\alpha_0\alpha_1 I_1(-\infty, \varepsilon_i) \\
&+ \left(2\alpha_0\alpha_2 + \alpha_1^2\right) I_2(-\infty, \varepsilon_i) + (2\alpha_0\alpha_3 + 2\alpha_1\alpha_2) I_3(-\infty, \varepsilon_i) \\
&+ \left(2\alpha_0\alpha_4 + 2\alpha_1\alpha_3 + \alpha_2^2\right) I_4(-\infty, \varepsilon_i) \\
&+ (2\alpha_0\alpha_5 + 2\alpha_1\alpha_4 + 2\alpha_2\alpha_3) I_5(-\infty, \varepsilon_i) \\
&+ \left(2\alpha_1\alpha_5 + 2\alpha_2\alpha_4 + \alpha_3^2\right) I_6(-\infty, \varepsilon_i) \\
&+ (2\alpha_2\alpha_5 + 2\alpha_3\alpha_4) I_7(-\infty, \varepsilon_i) + \left(2\alpha_3\alpha_5 + \alpha_4^2\right) I_8(-\infty, \varepsilon_i) \\
&+ 2\alpha_4\alpha_5 I_9(-\infty, \varepsilon_i) + \alpha_5^2 I_{10}(-\infty, \varepsilon_i)\Big] \times \frac{1}{P}\frac{1}{\sqrt{2\pi}\sigma}.
\end{aligned}
\tag{A10}
$$