Discussion Paper No. 501

A NEWLY DEVELOPED JAPANESE PNEUMONIA AND INFLUENZA MORTALITY MODEL AND STATISTICAL ANALYSIS OF EXCESS MORTALITY BY STOCHASTIC FRONTIER ESTIMATION

Nahoko Sindo
Masako Ii
Yasushi Ohkusa
and
Kiyosu Taniguchi

February 2000

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

A Newly Developed Japanese Pneumonia and Influenza Mortality Model and Statistical Analysis of Excess Mortality by Stochastic Frontier Estimation

Nahoko Sindo, Jikei University School of Medicine and National Institute for Infectious Diseases

Masako Ii, Yokohama National University

Yasushi Ohkusa, Osaka University

and

Kiyosu Taniguchi, National Institute for Infectious Diseases

Abstract

Models which estimate influenza mortality probably overestimate this mortality because their baseline estimates include only nonepidemic weeks. Because deaths due to noninfluenza illnesses rises during weeks in which influenza does, the models incorrectly conclude that all deaths during the influenza season are due to influenza. We present a stochastic frontier estimation model which better reflects influenza mortality.

Key Words : pneumonia, influenza, excess mortality, models, statistical.

In Japan, the influenza epidemic of the 1997-98 season was the most severe of the decade. There were more then 500,000 influenza-like illness cases and 50,000 pneumonia and influenza deaths. Like other industrialized countries, there is a growing elderly population in Japan and they are the ones most affected by influenza epidemics. Over 90% of influenza deaths occur among those aged 65 or older(1,2). Accurately determining the magnitude of mortality caused by influenza highlights the impact of influenza on the aged population, and may increase the commitment of resources to lessen this morbidity.

To assess the severity of influenza epidemics, "excess mortality", the number of deaths actually recorded in excess of the number expected on the basis of past seasonal experience, has been used as a major index. This is because the number of deaths caused by acute respiratory diseases sharply rises during influenza epidemic, although most of them are without pathological diagnosis and few are virologicaly confirmed. In 1963,Dr.Robert E.Serfling graphically presented expected numbers of seasonal pneumonia-influenza mortality with a regression method. He estimated influenza and pneumonia deaths using Fourier Series with Linear Trend(3). Subsequently, Choi and Thacker applied a seasonal ARIMA (autoregressive integrated moving average) model to estimate numbers of deaths attributable to influenza and pneumonia in the absence of an epidemic(4). They stated that the Serfling method tends to underestimate the expected numbers of deaths since data from epidemic weeks are omitted in deriving the regression equation of expected death. Their model included data for epidemic weeks, but used data from past nonepidemic weeks to replace them. They arbitrarily defined epidemic weeks as ones in which there were rises in mortality beyond what was expected, along with widespread reporting of influenza-like illness and laboratory evidence of infection. The Serfling method is currently used with some modifications by the Centers for Disease Control and Prevention and also has been applied to several influenza studies by Who and other countries(5).

We believe it is conceptually incorrect to include only nonepidemic weeks in these models. Because deaths due to noninfluenza illnesses rise during weeks in which influenza does, the models will overestimate mortality because they incorrectly assume that all deaths during the influenza season are due to influenza. We present a new pneumonia and influenza mortality model, based on stochastic frontier estimation. This model draws the base line curve of expected deaths without excluding any data during the year. In addition, by doing this, there is no need to make an arbitrary decision about epidemic periods. This model provides numerical information about how rarely, from a statistical point of view, epidemics of influenza mortality occur.

The estimation equation for pneumonia and influence death at year t, P_t, is

$$\log P_t = \alpha + \gamma_1 D_{95} + \gamma_2 T_t + \gamma_3 D_{95} T_t + \gamma_4 T_t^2 + \gamma_5 D_{95} T_t^2 + \sum \beta_i M_{it} + \varepsilon_t \tag{1}$$

where D_{95} is a dummy variable which takes one after January 1995 and zero otherwise, because ICD-10 (International Statistical Classification of Diseases and Related Health Problem s, Tenth Revision) was used since January 1995 on Japanese medical certificates of death. The number of pneumonia deaths in (10,995) was 77.8% of those in 1994 (14,129). This decrease is explained as an artifactual change due to mortality and morbidity coding rules stated in ICD-10 to chose underlying cause of death[6]. For example, a patient with a cerebral hemorrhage followed by pneumonia would be coded as a pneumonia death in 1994 but a cerebral hemorrhage death in 1995. Pneumonia and influenza mortality includes ICD-9 code 480-487 or ICD-10 code J10-J18 in this study. A variable T_t is linear time trend. The third degree polynomials are not significant (0.74 and 0.95), therefore time trend variables for the third and higher degree polynomial are not included in the model.

A disturbance term is defined as

$$\varepsilon_t = \upsilon_t + |\omega_t| \tag{2}$$

where υ_t and ω_t are mutually independent random variables, and $\upsilon_t \sim N(0, \mu^2)$ and $\omega_t \sim N(0, \xi^2)$. A random variable υ_t is purely stochastic disturbance and ω_t represents the magnitude of the epidemic which is a non-negative deviation term from expected number of deaths. The equation form (1) is selected by statistical inference.

The probability density function of this model is

$$f(\varepsilon_t) = \frac{2}{(\mu^2+\xi^2)^{0.5}}\phi(\frac{\varepsilon_t}{(\mu^2+\xi^2)^{0.5}})\Phi(\varepsilon_t\frac{\xi/\mu}{(\mu^2+\xi^2)^{0.5}}) \tag{3}$$

where ε_t are disturbance terms in equation (1), and defined by the equation (2). ϕ is probability density function for standard normal distribution and Φ is its cumulative distribution function. The likelihood function is defined as the production of probability density function over time : $L(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_t) = \prod f(\varepsilon_t)$. The estimators for α, γ's, β's, ξ/μ and $(\mu^2+\xi^2)^{0.5}$ are chosen so as to maximize the likelihood function. The obtained estimators are maximum likelihood estimators. This model is known in economics as stochastic frontier estimation [7,8]. In economics, ω_t is often defined as an inefficiency, and stochastic frontier estimation is used to measure this inefficiency. For example, in estimating production function the deviation from the most efficient production is defined as inefficient, and in estimating cost function the deviation from the least cost is defined as inefficient. In this paper, we defined ω_t as the magnitude of epidemic. In this way we can estimate the excess mortality without making an arbitrary assumption of what epidemic or non-epidemic weeks are. For example, the referred models arbitrarily define non-epidemic weeks and estimate the excess mortality during the epidemic weeks using these " non-epidemic weeks " as baseline.

For the years 1998 - February 1999, the largest monthly excess mortality from pneumonia and influenza occurred in January, 16.8 % more than that in December, while in September it was the smallest (26.1 % less) (Table 1 (a)). Excess mortality from pneumonia and influenza grew at the rate of 1.1 % per year until 1995 (Linear Time Trend in Table 1 (a)). The number of total death in January was also the largest, 7.2 % more than that in December (Table 1 (b)).

The epidemic threshold is usually defined as an upper limit of 95 % confidence interval for stochastic disturbance terms, υ_t. However, υ_t are originated from measurement errors and not from the magnitude of epidemics. The magnitude of an epidemic should be measured by ω_t . Therefore, in this paper, threshold is defined as a summation of estimated number of death and, $\overline{\omega}$ which is defined in the following equation :

$$\int_0^{\overline{\omega}} \int_{-\infty}^{\infty} \hat{f}(\upsilon + |\omega|) d\upsilon d\omega = 0.95 \tag{4}$$

where, $\hat{f}$ is evaluated using μ and ξ estimated in equation (3).

Thus, in our model the actual number of deaths from pneumonia and influenza was above the threshold in 1990, 1993, 1995, 1997, 1998 and 1999(Figure 1 (a)). Excess mortality was 1, 951 in January 1990, 2,191 in February 1993, 2, 641 February 1995, 5, 032 in January 1997, 2, 820 in February 1998, and 6, 798 in January 1999. For these months the probability values of excess mortality occurrence are less than 0.00001, which is calculated by the following equation :

$$\int_{-\infty}^{\infty} \int_{\hat{\varepsilon}-\upsilon}^{\infty} \hat{f}(\upsilon + |\omega|) d|\omega| d\upsilon \tag{5}$$

In the figure 1 (b), the threshold indicates that actual number of total deaths is above the threshold in 1990, 1993,1995,1997,1998 and 1999. Excess mortality was 7,465 on January 1990, 7,192 in January 1993, 18,184 in January 1995, 12,405 in January 1997, 6,569 in February 1998, and 22,503 in January 1999. For these months the probability values of excess mortality occurrence are less than 0.00001 when calculated by equation (5).

We believe that , compared to the two referred models, stochastic frontier estimation has at least four advantages in estimating excess mortality. One is that the model does not overestimate mortality caused by influenza. Indeed, by using in its base line all data from epidemic weeks, it is likely, if anything, to underestimate mortality. Ideally, we would like to know the proportion of all deaths reported as ICD-10 J10-18 actually caused by influenza. Unfortunately, neither we norm to our knowledge, other countries have this information. Until such information is available, models will need to be used. A second advantage is that stochastic frontier estimation does not depend on arbitrarily defined epidemic weeks. Third, this estimation model gives numerical information that shows how rare epidemics that cause greatly increased mortality are. Fourth is utilizing long term past data for making base line curve while the current CDC model is looking back data for only five-year period.

This models should be applicable for weekly analysis and useful as a prediction tool. We will be to get national weekly mortality data in the near future, and hope to promptly predict outbreaks of influenza by making use of this mortality model.

Acknowledgments

The authors thank Dr. Mark white and Dr. George Schmid for their comments.

References

[1] Statistics and Information Department, Minister's Secretariat, Ministry of Health and welfare. Causes of death-Table 3. Death by Causes (the condensed list of causes of death for Japan), sex and month of occurrence : Japan, 1998-1997. (Vital Statistics Mortality Data, Total death, influenza death, pneumonia death), *VITAL STATISTICS OF JAPAN*, vol.3, 1997-1999.

[2] Statistics and Information Department, Minister's Secretariat, Ministry of Health and welfare. *Monthly Vital Statistics Report* (preliminary), Total death, influenza deaths, pneumonia death,1998 Jan.-1999 Feb.

[3] Serfling RE.," Methods for current statistical analysis of excess pneumonia-influenza deaths," *Public Health Rep* 1963; 78 : 494-506.

[4] Choi K and Thacker SB.,"An evaluation of influenza mortality surveillance, 1962-1979", *I. Am J Epidemiology* 1981; 113 :215-226.

[5] Assad F, Cockburn WC, Sundaresan TK.,"Use of excess mortality from respiratory diseases in the study of influenza," *Bull WHO* 1973; 49 : 219-233.

[6] Statistics and Information Department, Minister's Secretariat, Ministry of Health and welfare. Reference table. *VITAL STATISTICS OF JAPAN*, vol.1, 1995; p456.

[7] Aigner AD, Lovell K and Schmidt P.,"Formulation and estimation of stochastic frontier production function models," *Journal of Econometrics* 1977; pp.21-37.

[8] Jondrow J, Lovell K, Materov S and Schmidt P.,"On estimation of technical inefficiency in the stochastic frontier production function model," *Journal of Econometrics* 1982 ; pp.233-239.

Table 1: Stochastic Frontier Estimation for Mortality, January 1988 - February 1999

(a) Pneumonia and Influenza

Variable	Estimated Coefficient	t-statistics
Constant	8.20787	107.946
Linear Time Trend	0.010526	4.25308
Dummy for after 1995	-0.154352	-2.39021
Linear Time Trend $\times$ Dummy for after 1995	-0.078245	-5.0944
Linear Time Trend squared	-0.000043	-1.89429
Linear Time Trend squared $\times$ Dummy for after 1995	0.000329	4.98485
Monthly Dummies		
January	0.167703	3.27628
February	0.096200	1.90556
March	0.092937	1.77739
April	-0.026695	-0.419345
May	-0.071048	-1.09857
June	-0.199961	-3.29451
July	-0.156862	-2.32106
August	-0.170525	-2.70645
September	-0.260593	-3.98653
October	-0.183964	-2.69883
November	-0.109000	-1.48611
$(\mu^2 + \xi^2)^{0.5}$	0.148406	14.6737
ξ/μ	4.16483	2.61146
Number of observations	134	
Log likelihood	-166.329	

(b) Total Death

Variable	Estimated Coefficient	t-statistics
Constant	11.1335	691.067
Linear Time Trend	0.001473	14.0684
Monthly Dummies		
January	0.072121	4.25377
February	-0.038273	-2.2639
March	0.003039	0.174893
April	-0.088863	-4.47695
May	-0.109695	-5.23517
June	-0.193285	-9.25166
July	-0.162045	-8.86165
August	-0.166225	-9.18318
September	-0.206316	-9.58921
October	-0.111206	-4.96559
November	-0.084749	-4.0317
$(\mu^2+\xi^2)^{0.5}$	0.056938	13.3600
ξ/μ	3.31355	2.93399
Number of observations	134	
Log likelihood	-290.511	

Figure 1(a): Number of Monthly Death from Pneumonia & Influenza

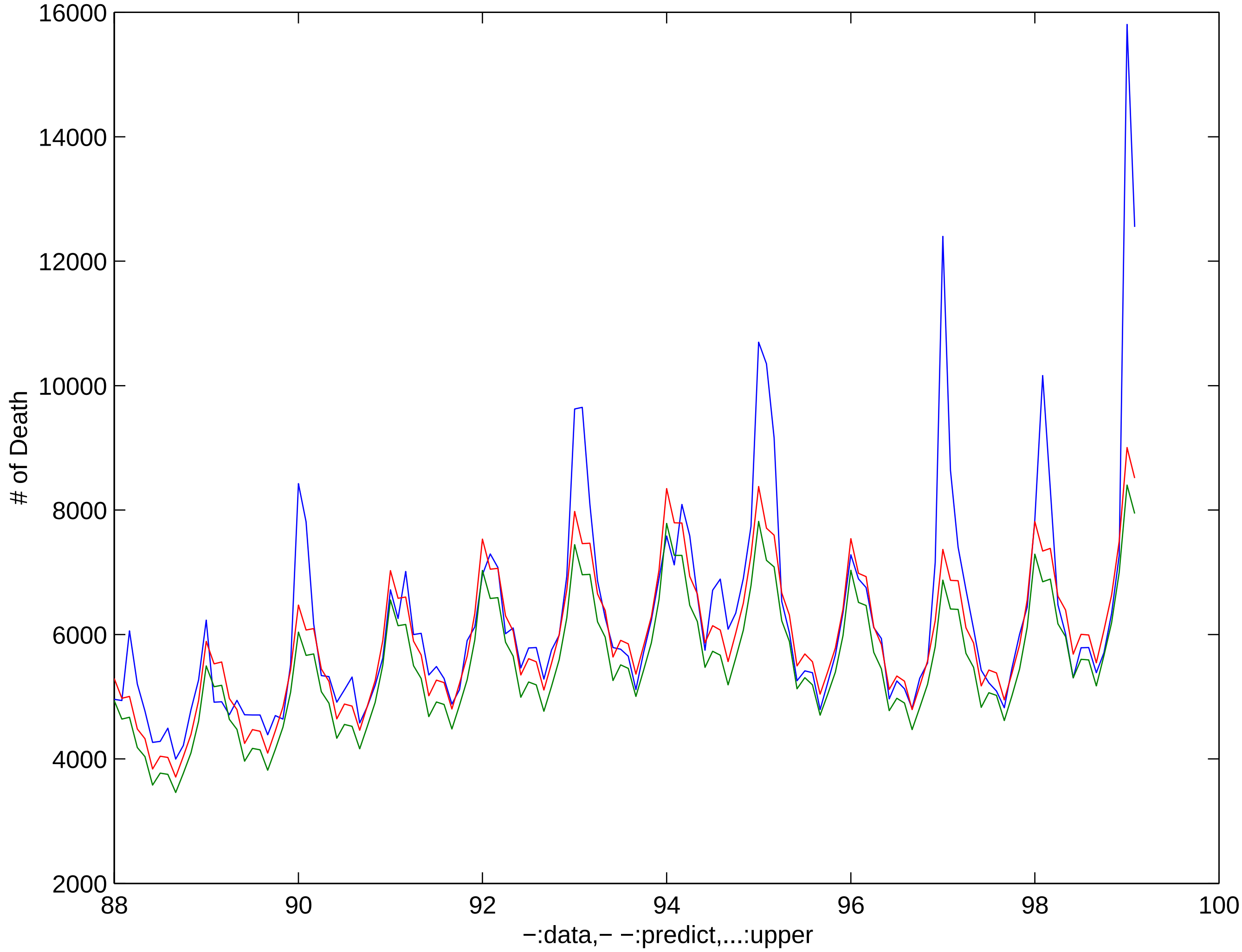

Figure 1(b): Number of Monthly Death, Total