Statistica Sinica **20** (2010), 1097-1123

STATISTICAL MODELLING OF NONLINEAR LONG-TERM CUMULATIVE EFFECTS

Efang Kong, Howell Tong and Yingcun Xia

University of Kent at Canterbury, London School of Economics and National University of Singapore

Abstract: In epidemiology, bio-environmental research, and many other scientific areas, the possible long-term cumulative effect of certain factors has been well acknowledged, air pollution on public health, exposure to radiation as a possible cause of cancer, among others. However, there is no known statistical method to model these effects. To fill this gap, we propose a semi-parametric time series model, called the functional additive cumulative time series (FACTS) model, and investigate its statistical properties. We develop an estimation procedure that combines the advantages of kernel smoothing and polynomial spline smoothing. As two case studies, we analyze the effect of air pollutants on respiratory diseases in Hong Kong, and human immunity against influenza in France. Based on the results, some important issues in epidemiology are addressed.

Key words and phrases: Cumulative effect, generalized additive model, local linear smoother, nonlinear time series, polynomial splines, single-index model.

1. Introduction

In epidemiology, cumulative effect refers to the fact that long-term exposures to harmful environments impair public health. The cumulative effect has been noted to be the main cause of many diseases for which short-term/individual effects may be insignificant. For example, continual exposure to air pollution affects the lungs of growing children and may aggravate or complicate medical conditions in the elderly (Galizia and Kinney (1999)). The extent to which an individual is harmed by air pollution usually depends on the total exposure to the damaging chemicals. Another example is the cumulative effect of ultraviolet radiation as a major cause of skin cancer (Young (1990)). Cumulative effects are also observed in many other areas, examples include loss of wetland habitats, climate change, and increased risk of flooding. In fact, assessing cumulative effects is an essential mission of the Environmental Protection Agency (*Report - Considering Cumulative Effects Under NEPA* `http://www.epa.gov`), the World Health Organization (WHO), and other similar organizations. Investigations on specific cumulative effects can be found in the existing literature. See, for example, Smith and Spaling (1995) and Ceriello et al. (2002), Dubé et al. (2006), and the references therein.

Long-term cumulative effects, although recognized as important, have not been properly modelled or quantified by existing methodologies, and are thus often ignored before they become serious and, by then, it is too late to act. In fact, most existing time series models focus on the effects of a few individual historical data points that might fail to capture the cumulative effect. As a motivating example, we consider the effect of air pollution on the number of daily hospital admissions in Hong Kong. Figure 1 presents several aspects of the data collected in Hong Kong from January 1, 1994 to December 31, 1998. To make our point, consider for the moment the effect of the daily average level of nitrogen dioxide ($\mathrm{NO}_{2,t}$, in *ppb*) on the number of daily hospital admissions of patients suffering from respiratory problems. The daily average level of NO_2 on any single day does not have much explanatory power over the daily number of admissions as suggested in Figure 1(a) and Figure 1(c). On the other hand, a much larger portion of its variation can be explained by the overall pollution level of NO_2 in the past 220 days, $\sum_{\tau=0}^{220} NO_{2,t-\tau}$, as shown in Figure 1(b) and Figure 1(d).

Here we adopt a semi-parametric approach to analyze cumulative effect. As suggested by Figure 1(b), the cumulative effect of NO_2 tends to be *nonlinear* and to increase more rapidly as the cumulation level increases. A nonparametric function is introduced as the link function; a nonparametric smoothing method is used for the estimation. In the example, the upper limit in the summation, $\sum_{\tau=1}^{D} NO_{2,t-\tau}$, is obtained by maximizing, with respect to D, the correlation coefficient between the summation (as a function of D) and the number of hospital admissions on day t. It is more appropriate to consider a set of data-driven weights in the summation, i.e. $\sum_{\tau=1}^{D} w_\tau NO_{2,t-\tau}$, where D should be data adaptive and the weight w_τ need to be estimated subject to $\sum_{\tau=1}^{D} w_\tau = 1$. In the Hong Kong data example, it is very likely that, besides NO_2, other pollutants and weather conditions contribute to the variation in the number of hospital admissions. To incorporate these factors, one option is to adopt an additive structure, e.g., Hastie and Tibshirani (1990) and Dominici et al. (2002). As suggested by Figure 1(b), higher levels of cumulative pollution tend to result in more hospital admissions. This motivates us to impose a monotone constraint on the link function. Similar assumptions could also be imposed on the weight function w_τ, if suggested by empirical evidence.

Among the various smoothing methods of estimating a semi-parametric model, polynomial spline and kernel smoothing appear to be dominant, with respective advantages. Thus polynomial spline smoothing is more convenient for

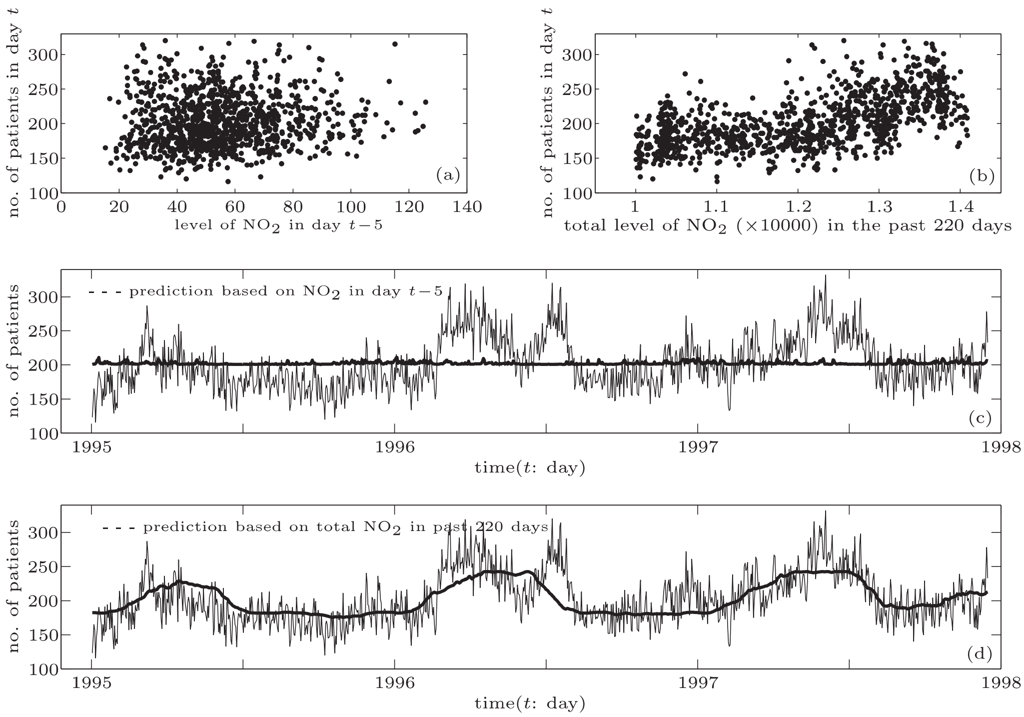

Figure 1. The level of NO_2 and number of patients suffering from respiratory diseases in Hong Kong. Panel (a) is the plot of the number of patients on day t against the level of NO_2 on day $t-5$, the latter day having the largest correlation coefficient with the number of patients among all past days; (b) is that against the total/cumulated levels of NO_2 over the past 220 days. (c) is the fitted number of patients based on the level of NO_2 on day $t-5$ using a kernel smoothing; (d) is that based on the total/cumulated level of NO_2. over the past 220 days

incorporating global constraints on the functions, while kernel smoothing is more convenient for local Taylor expansion and approximation. In this paper we use polynomial spline for function estimation and kernel smoothing for local approximation in estimating the cumulative weights. By doing so, computations are simplified to a standard quadratic programming problem for which very efficient and fast algorithms exist.

The rest of the paper is organized as follows. In Section 2, we propose the functional additive cumulative time series (FACTS) model. Through a penalized spline smoothing approach, we derive the asymptotic properties of the penalized least squares estimator of the new model. To implement the estimation, a semiparametric procedure that combines polynomial spline and kernel smoothing is developed in Section 4. In Section 5, we are back to the study of the cumulative effect of air pollution on respiratory diseases in Hong Kong. Some of the questions posed by WHO are answered based on the study. Section 6 is another case study on infectious diseases and human immunity, a proper modeling of the two being essential for policy decision making relating to vaccination and disease control, see e.g., Ferguson, Galvani and Bush (2003). Based on the weekly influenza cases in France (www.sentiweb.org), a decreasing pattern of immunity is revealed.

2. A Semi-Parametric Model for the Cumulative Effect

Suppose $Y_{t_i} \stackrel{def}{=} Y(t_i)$, $i = 1, \ldots, n$, is a (discrete) response time series and $\{\mathbf{Z}(t), \mathbf{X}(t)\}, t \geq 0$, are multivariate (continuous) time series with $\mathbf{Z}(t) = (1, Z_1(t), \ldots, Z_p(t))^\top$ and $\mathbf{X}(t) = (X_1(t), \ldots, X_q(t))^\top$. The semiparametric additive model of interest has

$$E\{Y(t_i)|Z(s), X(s), s \leq t_i\} = \beta^\top \mathbf{Z}(t_i) + g_1(X_1(t_i - \tau_1)) + \cdots + g_q(X_q(t_i - \tau_q)), \quad (2.1)$$

where β is an unknown $p-$dimensional parameter vector, $g_k, k = 1, \ldots, q$, are unknown link functions and $\tau_k, k = 1, \ldots, q$, are lags. See, e.g., Hastie and Tibshirani (1990), Liu and Stengos (1999), and Dominici et al. (2002).

As noticed from the Hong Kong data, the effect of pollution on any single day is not significant. However, persistent pollution over a relatively long period can explain much of the variation in the number of daily hospital admissions. In other words, cumulative effects result from individually minor but collectively significant covariates over a period of time. However, in (2.1) it is assumed that the expected value of $Y(t)$ depends on only a finite number of the historical values of $\mathbf{X}(t)$ without reference to the cumulative and continuous effects discussed above. Specifically, for the Hong Kong data, empirical study suggests that the semiparametric additive model (2.1) typically does not lead to a good fit; see Figure 5(a) in Section 6. A straightforward extension of model (2.1) to enlarge the number of additive components is infeasible, as the resulted model tends to be plagued by unstable estimation and difficult interpretation in practice.

We propose to model the cumulative effect of a single covariate, $X_1(\tau)$ say, by $\int_0^\Delta X_1(t - \tau)\theta(\tau)d\tau$ for some $\Delta > 0$, where $\theta(\tau) \geq 0$ is a weight function defined over $[0, \Delta]$. When incorporated with the additive structure, one has

$$Y(t_i) = \mathbf{Z}^\top(t)\beta^0 + \sum_{k=1}^{q} g_k \left(\int_0^\Delta X_k(t_i - \tau)\theta_k(\tau)d\tau \right) + \varepsilon(t_i), \ i = 1, \ldots, n, \quad (2.2)$$

where $\varepsilon(t_i)$ is a martingale difference with $E(\varepsilon(t_i)^2|\mathbf{Z}(t), \mathbf{X}(s) : s \leq t_i) = \sigma^2$, and $g_k(\cdot)$ and $\theta_k(\cdot) > 0, \ k = 1, \ldots, q$, are unknown smooth functions with

$$E\{g_k \left(\int_0^\Delta X_k(t - \tau)\theta_k(\tau)d\tau \right)\} = 0, \qquad \int_0^\Delta \theta_k(\tau)d\tau = 1, \ k = 1, \ldots, q, \quad (2.3)$$

for identification purposes. Alternative identification conditions can be imposed depending on the purpose of the modeling. We call (2.2) the functional additive cumulative time series (FACTS) model. In it, $g_k(\cdot)$ is referred to as an effect or link function, $\theta_k(\cdot)$ is a weight function.

Based on the discussion in Section 1, if warranted by empirical evidence, a monotonicity constraint could be imposed on either $\theta_k(\cdot)$ or $g_k(\cdot)$, $k = 1, \ldots, q$, or both. To obtain estimates at (2.2) that are consistent with such a constraint is an important goal of this paper.

Proposition 2.1. *Suppose* $\mathbf{X}$ *is a multivariate stationary process with continuous joint probability density functions. Let* $\tilde{\mathbf{Z}}(t) = E\{\mathbf{Z}(t)|\mathbf{X}(\tau), \tau \leq t\}$. *If* $E[\{\mathbf{Z}(t) - \tilde{\mathbf{Z}}(t)\}\{\mathbf{Z}(t) - \tilde{\mathbf{Z}}(t)\}^\top]$ *is invertible, then model* (2.2) *is identifiable: if there exists another set of parameters* $\tilde{\beta}^0$ *and functions* $\tilde{g}_k(\cdot), \tilde{\theta}_k(\cdot)$ *such that* (2.2) *and* (2.3) *hold, then* $\tilde{\beta}^0 \equiv \beta^0, \quad \tilde{g}_k(\cdot) \equiv g_k(\cdot), \quad \tilde{\theta}_k(\cdot) \equiv \theta_k(\cdot)$, $k = 1, \ldots, q$.

The FACTS model is closely linked with functional regression models. See for example Ramsay and Silverman (1997) and James and Silverman (2005). If $X_k(t)$ has a step sample path, then the integration component of model (2.2) reduces to a summation, leading to the discretized version

$$Y(t) = \mathbf{Z}^\top(t)\beta^0 + \sum_{k=1}^{q} g_k\Big(\sum_{\ell=1}^{D} X_k(t-\ell)\theta_k(\ell)\Big) + \varepsilon_t, \tag{2.4}$$

where $\theta_k(\ell), \ell = 1, \ldots, D,\ k = 1, \ldots, p$ are unknown parameters. This is a partially linear additive single-index model. A special case is the partially linear single-index model; see e.g., Carroll et al. (1997) and Yu and Ruppert (2002).

3. Estimation of FACTS Model

As in Yu and Ruppert (2002), we adopt a spline smoothing approach to estimate both the unknown link function $g_k(\cdot)$ and the weight function θ_k. Suppose for each $k = 1, \ldots, q$, there exist vectors η_k^0 and γ_k^0 such that, approximately, $g_k(\nu) = \mathbf{A}(\nu)^\top \eta_k^0$, and $\theta_k(\tau) = \mathbf{B}(\tau)^\top \gamma_k^0$, where $\mathbf{A}(\nu)$ and $\mathbf{B}(\tau)$ are finite $r-$dimensional bases functions, e.g., cubic splines. Let $b = (b_1, \ldots, b_r)^\top = \int_0^\Delta \mathbf{B}(\tau)d\tau$, a column vector of length r with first component b_1 nonzero. Then the second equation in (2.3) can be approximately rewritten as

$$b^\top \gamma_k^0 = 1, \quad 1 \leq k \leq q. \tag{3.1}$$

Write $\gamma_k^0 = (\gamma_{k,1}^0, \ldots, \gamma_{k,r}^0)^\top$ and define

$$X_{t_i}^k = \int_0^\Delta X_k(t_i - \tau)\mathbf{B}(\tau)d\tau, \quad \mathbf{v}_i = (\mathbf{Z}(t_i)^\top, X_{t_i}^1, \ldots, X_{t_i}^q),$$
$$\xi = (\beta^\top, \eta_1^\top, \ldots, \eta_q^\top, \gamma_1^\top, \ldots, \gamma_q^\top)^\top,$$

where $\eta_i, \gamma_i,\ i = 1, \ldots, q$ are $r \times 1$ vectors. Define the mean function

$$m(\mathbf{v}_i;\xi) = \mathbf{Z}^\top(t_i)\beta + \sum_{k=1}^{q} \eta_k^\top \mathbf{A}(\gamma_k^\top X_{t_i}^k). \tag{3.2}$$

Existing methods, such as the penalized spline method in Yu and Ruppert (2002), can be used to estimate $\xi^0 = (\beta^{0^\top}, \eta_1^{0^\top}, \ldots, \eta_q^{0^\top}, \gamma_1^{0^\top}, \ldots, \gamma_q^{0^\top})^\top$ if the value of $X_{t_i}^k$ is available. However, this is usually not the case in practice, as quite often $\{\mathbf{X}(t)\}$ can only be observed at discrete time points, although not necessarily with the same frequency as Y_{t_i}. Suppose $\{\mathbf{X}(\tilde{t}_j), j = 1, \ldots,\}$ is the observed discrete time series of $\{\mathbf{X}(t)\}$. If as specified in (A2) in the Appendix, $\max_{j\geq 1} |\tilde{t}_j - \tilde{t}_{j+1}|$ is sufficiently small relative to n, the total number of observations on $Y(t)$, then based on the continuous property of $\mathbf{B}(\cdot)$ and the sample path of $X_k(\cdot)$, we can approximate $X_{t_i}^k$ by

$$X_{t_i}^{n,k} = \sum_{j:t_i-\Delta<\tilde{t}_j\leq t_i} X_k(\tilde{t}_j)(\tilde{t}_j - \tilde{t}_{j-1})\mathbf{B}(t_i - \tilde{t}_j),$$

which, when substituted for $X_{t_i}^k$ in (3.2), leads to the approximation

$$m_n(\mathbf{v}_i;\xi) = \mathbf{Z}^\top(t_i)\beta + \sum_{k=1}^{q} \eta_k^\top \mathbf{A}(\gamma_k^\top X_{t_i}^{n,k}). \tag{3.3}$$

Parameter ξ^0 can thus be estimated by the penalized least squares estimator (PLSE) that minimizes

$$Q_{n,\lambda}(\xi) \overset{def}{=} n^{-1} \sum_{i=D+1}^{n} \Big\{Y_{t_i} - m_n(\mathbf{v}_i;\xi)\Big\}^2 + \lambda_n \delta^\top \Sigma\, \delta, \tag{3.4}$$

where $D = \min\{i | t_i - \tilde{t}_1 \geq \Delta\}$, $\delta = (\eta_1^\top, \ldots, \eta_q^\top, \gamma_1^\top, \ldots, \gamma_q^\top)^\top$, λ_n is a penalty parameter, and Σ is an appropriate positive semidefinite symmetric matrix; see Yu and Ruppert (2002).

3.1. Re-parameterization and asymptotic properties

The constraint (3.1) on γ_k makes reparameterization necessary in proving the consistency and asymptotic normality of the PLSE. Let $\tilde{b} = (b_2, \ldots, b_r)^\top$, $\phi_k = (\phi_{k,1}, \ldots, \phi_{k,r-1})^\top$, and

$$\gamma_k(\phi_k) = (b_1^{-1}(1 - \tilde{b}^\top \phi_k), \phi_{k,1}, \ldots, \phi_{k,r-1})^\top. \tag{3.5}$$

Let $\phi_k^0 = (\gamma_{k,2}^0, \ldots, \gamma_{k,r}^0)^\top$ denote a subvector of γ_k^0. By (3.1), we have $\gamma_k(\phi_k^0) = \gamma_k^0$. It is easy to see that $\gamma_k(\phi_k)$ is infinitely differentiable in a neighborhood of

ϕ_k^0, with the gradient matrix given by

$$\gamma_k^{(1)}(\phi_k) = \left(-b_1^{-1}\tilde{b} \vdots \mathbf{I}_{r-1}\right)^\top, \quad k = 1, \ldots, q,$$

where $\mathbf{I}_r$ is the $r \times r$ identity matrix. Following the above notations, we take

$$\xi_\gamma \equiv \xi = (\beta^\top, \eta_1^\top, \ldots, \eta_q^\top, \gamma_1^\top, \ldots, \gamma_q^\top)^\top, \quad \xi_\phi = (\beta^\top, \eta_1^\top, \ldots, \eta_q^\top, \phi_1^\top, \ldots, \phi_q^\top)^\top,$$

where ξ_ϕ is of lower dimension than ξ_γ. Consequently, the regression mean function (3.2) and its approximation (3.3) can be reparameterized as

$$m(\mathbf{v}_i; \xi_\phi) = \mathbf{Z}^\top(t_i)\beta + \sum_{k=1}^{q} \eta_k^\top \mathbf{A}(\gamma_k(\phi_k)^\top X_{t_i}^k),$$
$$m_n(\mathbf{v}_i; \xi_\phi) = \mathbf{Z}^\top(t_i)\beta + \sum_{k=1}^{q} \eta_k^\top \mathbf{A}(\gamma_k(\phi_k)^\top X_{t_i}^{n,k}).$$

To state the asymptotic results, let

$$m_\eta^{(1)}(\mathbf{v}_i; \xi_\phi) = \begin{pmatrix} \mathbf{A}\left(\gamma_1^\top(\phi_1) X_{t_i}^1\right) \\ \vdots \\ \mathbf{A}\left(\gamma_q^\top(\phi_q) X_{t_i}^q\right) \end{pmatrix},$$
$$m_\phi^{(1)}(\mathbf{v}_i; \xi_\phi) = \begin{pmatrix} {\eta_1}^\top \mathbf{A}^{(1)}\left(\gamma_1^\top(\phi_1) X_{t_i}^1\right) \gamma_1^{(1)}(\phi_1)^\top X_{t_i}^1 \\ \vdots \\ {\eta_k}^\top \mathbf{A}^{(1)}\left(\gamma_q^\top(\phi_q) X_{t_i}^q\right) \gamma_q^{(1)}(\phi_q)^\top X_{t_i}^q \end{pmatrix}.$$

Then the gradient of $m(\mathbf{v}_i; \xi_\phi)$ with respect to parameter ξ_ϕ is

$$m^{(1)}(\mathbf{v}_i; \xi_\phi) = \begin{pmatrix} \mathbf{Z}(t_i) \\ m_\eta^{(1)}(\mathbf{v}_i; \xi_\phi) \\ m_\phi^{(1)}(\mathbf{v}_i; \xi_\phi) \end{pmatrix},$$

and the Jacobian matrix of ξ_γ with respect to ξ_ϕ is given by

$$\mathbf{J}(\phi) = \xi_\gamma^{(1)}(\xi_\phi) = \begin{bmatrix} \mathbf{I}_{qr+p} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \gamma_1^{(1)}(\phi_1) & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \gamma_q^{(1)}(\phi_q) \end{bmatrix}, \tag{3.6}$$

where $\mathbf{O}$ is the $r \times (r-1)$ matrix with entries zero. Let

$$\Omega(\xi) = \lim_n \frac{1}{n} \sum_{i=1}^{n} m^{(1)}(\mathbf{v}_i; \xi_\phi) m^{(1)}(\mathbf{v}_i; \xi_\phi)^\top. \tag{3.7}$$

Theorem 3.1. *Under* (A1)−(A7) *in the Appendix, if the smoothing parameter* $\lambda_n = o(n^{-1/2})$, *the PLSE* $\hat{\xi}_\gamma = (\hat{\beta}^\top, \hat{\eta}_1^\top, \ldots, \hat{\eta}_q^\top, \hat{\gamma}_1^\top, \ldots, \hat{\gamma}_q^\top)^\top$ *with constraints* $b^\top \hat{\gamma}_k = 1,\ k = 1, \ldots, q$, *is strongly consistent and*

$$\sqrt{n}(\hat{\xi}_\gamma - \xi^0) \xrightarrow{D} N\Big\{\mathbf{0}, \sigma^2 \mathbf{J}(\phi^0)\Omega^{-1}(\xi_\phi^0)\mathbf{J}^\top(\phi^0)\Big\}.$$

Remark. Note that the above asymptotic distribution is obtained without a monotonicity constraint imposed on either the estimated link function or the weight function $\hat{g}_k(\nu) = \mathbf{A}(\nu)^\top \hat{\eta}_k^0,\quad \hat{\theta}_k(\tau) = \mathbf{B}(\tau)^\top \hat{\gamma}_k^0,\ k = 1, \ldots, q$. However, as the PLSE estimator is strongly consistent, the aforementioned estimated function without constraint will automatically satisfy the nonnegative requirement, with probability 1, if n is large enough. Therefore, the same asymptotic property holds for estimators both with and without constraint. More details on constrained minimization can be found in e.g., Liew (1976).

3.2. A semi-parametric implementation

In this section, we focus on estimating ξ^0 through minimizing (3.4) with respect to ξ, with λ_n fixed as 0. There are two reasons for such a choice of λ_n. First, as indicated in Theorem 3.1, for the estimate to be asymptotically normal we need $\lambda_n = o(n^{-1/2})$. Second, the Monte Carlo study of Yu and Ruppert (2002) found that confidence bands using $\lambda = 0$ resemble the Monte Carlo confidence bands more than those using the true value of λ. Moreover, based on both simulation and data analysis in this paper, such a prefixed value of λ_n has not caused any serious over-fitting problem.

Minimizing (3.4) is in no way trivial, especially if any monotonicity constraint is imposed. As noted in Yu and Ruppert (2002), the performance of their estimation algorithm quite often depends on the starting value and, in some cases, the least squares estimator does not result from the iterations unless the distribution of the predictor is close to normal. Since there are many efficient algorithms for quadratic programming problems, we propose to transform the minimization problem into two separate quadratic programming problems, and to obtain the estimator by iterating between the two programming problems.

Let L denote the number of iterations. As an initial step with $L = 0$, we select $\gamma_k^{(0)}$ such that $\theta_k(t_i - t_{i-\tau}) \propto 1 - \tau/D$ for $k = 1, 2, \ldots, q$. Thus, to estimate the FACTS model we need to estimate β and $\eta_k, k = 1, \ldots, q$. The minimization in (3.4) is a simple linear regression estimation for β and $\eta_k, k = 1, \ldots, q$. Denote them by $\beta^{(0)}$ and $\eta_k^{(0)}, k = 1, \ldots, q$, respectively. Thus, g_k is estimated by $g_k^{(0)}(\cdot) = \mathbf{A}(\cdot)^\top \eta_k^{(0)}$. After this initial step, we can follow the idea of the back-fitting algorithm to estimate model (2.2). See Hastie and Tibshirani (1990) for more details. Here we only discuss how to update the nonlinear components in the model.

Let $y_{t_i}^{n,k} = Y_{t_i} - \mathbf{Z}(t_i)^\top \beta^{(L)} - \sum_{\iota \neq k} g_\iota^{(L)}((X_{t_i}^{n,\iota})^\top \gamma^{(L)})$. To update the estimators of g_k and γ_k, we can consider a nominal single-index model

$$y_{t_i}^{n,k} = g_k(\gamma_k^\top X_{t_i}^{n,k}) + \epsilon_{t_i}^{n,k}, \tag{3.8}$$

where $g_k(\cdot) = \eta_k^\top \mathbf{A}(\cdot)$. For ease of exposition, denote $y_{t_i}^{n,k}, X_{t_i}^{n,k}, \gamma_k, \eta_k$, and $\epsilon_{t_i}^{n,k}$ by $\tilde{y}_i, \tilde{X}_i, \tilde{\gamma}, \tilde{\eta}$, and $\tilde{\epsilon}_i$ respectively. Without constraints, many easily implemented estimation methods are available for model (3.8). See for example Härdle and Stoker (1989), Yu and Ruppert (2002), Yin and Cook (2005), and Xia (2006), among others. Consider a local expansion of $g_k(\tilde{\gamma}^\top \tilde{X}_i)$ at x. If $(\tilde{X}_i - x)^\top \tilde{\gamma} = o(1)$, we have the Taylor expansion

$$\begin{aligned} g_k(\tilde{\gamma}^\top \tilde{X}_i) &= g_k(\tilde{\gamma}^\top x) + g_k'(\tilde{\gamma}^\top x)(\tilde{X}_i - x)^\top \tilde{\gamma} + O[\{(\tilde{X}_i - x)^\top \tilde{\gamma}\}^2] \\ &= \tilde{\eta}^\top \{\mathbf{A}(\tilde{\gamma}^\top x) + \mathbf{A}'(\tilde{\gamma}^\top x)(\tilde{X}_i - x)^\top \tilde{\gamma}\} + O[\{(\tilde{X}_i - x)^\top \tilde{\gamma}\}^2]. \end{aligned}$$

Following Fan, Yao and Tong (1996), for given $\tilde{\gamma}$ and $\tilde{\eta}$, the local discrepancy or conditional variance $\sigma^2(x) = E[\tilde{\epsilon}_i^2 | X_i = x]$ can be estimated by the local linear smoother as

$$\hat{\sigma}^2(x|\tilde{\gamma}, \tilde{\eta}) = \sum_{i=1}^n \Big[\tilde{y}_i - \tilde{\eta}^\top \{\mathbf{A}(\tilde{\gamma}^\top x) + \mathbf{A}'(\tilde{\gamma}^\top x)(\tilde{X}_i - x)^\top \tilde{\gamma}\}\Big]^2 K((\tilde{X}_i - x)^\top \tilde{\gamma}) \Big/ \sum_{i=1}^n K((\tilde{X}_i - x)^\top \tilde{\gamma}),$$

where $K(v)$ is a symmetric probability density function, h is a bandwidth, and $K_h(v) = h^{-1} K(v/h)$. Obviously, the best approximation of $\tilde{\gamma}$ and $\tilde{\eta}$ should minimize the overall discrepancy for all $x = \tilde{X}_j, j = 1, \ldots, n$. Thus, our estimation procedure is to minimize

$$\sum_{j=1}^n \hat{\sigma}^2(\tilde{X}_j|\tilde{\gamma}, \tilde{\eta}) = \sum_{j=1}^n \sum_{i=1}^n \Big[\tilde{y}_i - \tilde{\eta}^\top \{\mathbf{A}(\tilde{\gamma}^\top \tilde{X}_j) + \mathbf{A}'(\tilde{\gamma}^\top \tilde{X}_j)\tilde{\gamma}^\top \tilde{X}_{ij}\}\Big]^2 w_{ij} \tag{3.9}$$

with respect to $\tilde{\gamma}$ and $\tilde{\eta}$, where $w_{ij} = K(\tilde{\gamma}^\top \tilde{X}_{ij}) / \sum_{i=1}^n K(\tilde{\gamma}^\top \tilde{X}_{ij})$, and $\tilde{X}_{ij} = \tilde{X}_i - \tilde{X}_j$. A similar idea was used in Xia et al. (2002).

Without constraints, we can implement the minimization of (3.9) easily, as follows. Note that with fixed w_{ij}, the minimization can be decomposed into two separate quadratic programming problems with unknown parameters $\tilde{\eta}$ and $\tilde{\gamma}$, respectively. We can solve (3.9) by iteration as follows. Set the number of iteration

$\ell = 0$. With initial value $\tilde{\gamma}^{(0)} = \gamma_k^{(L)}$ and $w_{ij}^{(\ell)} = K(\tilde{X}_{ij}^\top \tilde{\gamma}^{(\ell)}) / \sum_{i=1}^n K(\tilde{X}_{ij}^\top \tilde{\gamma}^{(\ell)})$, the minimization in (3.9) is equivalent to

$$\min_{\tilde{\eta}} \ \tilde{\eta}^\top D_n \tilde{\eta} - 2C_n^\top \tilde{\eta}, \tag{3.10}$$

where

$$D_n = \sum_{j=1}^n \sum_{i=1}^n w_{ij}^{(\ell)} \mathbf{A}_{ij}^{(\ell)} (\mathbf{A}_{ij}^{(\ell)})^\top, \quad C_n = \sum_{j=1}^n \sum_{i=1}^n w_{ij}^{(\ell)} \mathbf{A}_{ij}^{(\ell)} \tilde{y}_i,$$

with $\mathbf{A}_{ij}^{(\ell)} = \mathbf{A}(\tilde{X}_j^\top \tilde{\gamma}^{(\ell)}) + \mathbf{A}'(\tilde{X}_j^\top \tilde{\gamma}^{(\ell)}) X_{ij}^\top \tilde{\gamma}^{(\ell)}$. The solution is $\tilde{\eta}^{(\ell)} = D_n^{-1} C_n$. With the updated $\tilde{\eta}^{(\ell)}$, minimizing (3.9) with respect to $\tilde{\gamma}$ is equivalent to

$$\min_{\tilde{\gamma}} \ \tilde{\gamma}^\top D_n' \tilde{\gamma} - 2{C_n'}^\top \tilde{\gamma}, \tag{3.11}$$

where

$$D_n' = \sum_{j=1}^n \sum_{i=1}^n w_{ij}^{(\ell)} \mathbf{C}_{ij}^{(\ell)} (\mathbf{C}_{ij}^{(\ell)})^\top, \quad C_n' = \sum_{j=1}^n \sum_{i=1}^n w_{ij}^{(\ell)} \mathbf{C}_{ij}^{(\ell)} [y_i - (\tilde{\eta}^{(\ell)})^\top \mathbf{A}(\tilde{X}_j^\top \tilde{\gamma}^{(\ell)})],$$

with $\mathbf{C}_{ij}^{(\ell)} = (\tilde{\eta}^{(\ell)})^\top \mathbf{A}'(\tilde{X}_j^\top \tilde{\gamma}^{(\ell)}) \tilde{X}_{ij}$. The solution to (3.11) is $\tilde{\gamma}^{(\ell+1)} = \{D_n'\}^{-1} C_n'$. Set $\ell = \ell + 1$. Repeat (3.10) and (3.11) until convergence. Denote the final values by $\tilde{\tilde{\gamma}}$ and $\tilde{\tilde{\eta}}$ respectively. Finally, we update $g_k^{(L)}(\cdot)$ by $g_k^{(L+1)}(\cdot) \stackrel{def}{=} \mathbf{A}(\cdot)^\top \tilde{\tilde{\eta}}$ and $\gamma_k^{(L)}$ by $\gamma_k^{(L+1)} \stackrel{def}{=} \tilde{\tilde{\gamma}}$.

In situations where it is deemed reasonable to assume monotonicity for either the link function or the weight function or both, monotone estimates can be obtained by applying estimation procedures discussed above with bases function $\mathbf{A}(\nu)$ and $\mathbf{B}(\tau)$ chosen from the monotone spline bases (Ramsay (1988)), for a monotone function can always be constructed as nonnegative linear combination of monotone spline bases. In this case, the minimization of (3.9) is realized again through alternatively solving two quadratic programming problems. With initial value $\tilde{\gamma}^{(\ell)} = \gamma_k^{(L)}$ and $\ell = 0$, we solve

$$\min_{\tilde{\eta}} \ \tilde{\eta}^\top D_n \tilde{\eta} - 2C_n^\top \tilde{\eta}, \quad \text{subject to}: \ \tilde{\eta}_2, \ldots, \tilde{\eta}_r \geq 0, \tag{3.12}$$

where $\tilde{\eta} = (\tilde{\eta}_1, \ldots, \tilde{\eta}_r)^\top$ and denote the solution by $\tilde{\eta}^{(\ell)}$. Solve

$$\min_{\tilde{\gamma}} \ \tilde{\gamma}^\top D_n' \tilde{\gamma} - 2{C_n'}^\top \tilde{\gamma}, \quad \text{subject to}: \quad \tilde{\gamma} \geq 0, \tag{3.13}$$

and denote the solution to (3.13) by $\tilde{\gamma}^{(\ell+1)}$. Set $\ell = \ell + 1$. Repeat (3.12) and (3.13) until convergence. Although we do not have a closed form for the solutions of (3.12) and (3.13), they are typically quadratic programming problems. There are many efficient algorithms. See for example Nocedal and Wright (1999).

Regarding the convergence of the algorithm proposed here, Xia (2006) proved that it converges at a geometric rate under mild conditions in the case of no constraint. Furthermore, he showed that the asymptotic efficiency is the same as in the case of parametric estimation methods if the covariates are normally distributed. The same efficiency is applicable to estimation under constraints, which follows from the arguments in Liew (1976).

3.3. Selection of pilot parameters

Suitable values of two pilot parameters, the number of the knots of the spline base function and the bandwidth h, need to be selected. As we now explain, we only need to choose the number of knots, since the bandwidth h can be decided based on a well-known result of the optimal bandwidth and the plug-in idea in Ruppert, Sheather and Wand (1995). For a pre-specified number of knots, the knots are placed at equally-spaced sample quantiles of the predictors $(\tilde{\gamma}^{(\ell)})^\top \tilde{X}_i, i = 1, \ldots, n$. We can estimate the link function $g_k(\cdot)$ by $\mathbf{A}(\cdot)^\top \hat{\eta}$, where $\hat{\eta} = (\mathcal{A}_n^\top \mathcal{A}_n)^{-1} \mathcal{A}_n^\top Y$ with $\mathcal{A}_n = (\mathbf{A}(\tilde{X}_1^\top \tilde{\gamma}^{(\ell)}), \ldots, \mathbf{A}(\tilde{X}_n^\top \tilde{\gamma}^{(\ell)}))^\top$. The fitted value of the response at the n points is $\hat{Y} = \mathcal{A}_n(\mathcal{A}_n^\top \mathcal{A}_n)^{-1} \mathcal{A}_n^\top Y$. Following Craven and Wahba (1979), we define the generalized cross-validation as

$$GCV(N) = \frac{||\hat{Y} - Y||^2/n}{(1 - tr(S_n)/n)^2},$$

where $S_n = \mathcal{A}_n(\mathcal{A}_n^\top \mathcal{A}_n)^{-1} \mathcal{A}_n^\top$. The selected number of knots minimizes $GCV(N)$. As noticed in Yu and Ruppert (2002), the possible range for N can be 5-10 in minimizing $GCV(N)$. When N is selected, the bandwidth can be calculated by the plug-in method proposed by Ruppert, Sheather and Wand (1995),

$$h = \Big[\frac{4 \int K^2(v) dv \hat{\sigma}^2}{\int K(v) v^2 dv \bar{m}_n^2 n}\Big]^{1/5},$$

where $\hat{\sigma}^2 = ||\hat{Y} - Y||^2/n$ and

$$\bar{m}_n^2 = n^{-1} \sum_{i=1}^{n} \{g_k''(\tilde{X}_i^\top \tilde{\gamma}^{(\ell)})\}^2 = n^{-1} \sum_{i=1}^{n} \{\mathbf{A}''(\tilde{X}_i^\top \tilde{\gamma}^{(\ell)})^\top \eta^{(\ell)}\}^2.$$

Another bandwidth selection approach is the simple rule-of-thumb of Silverman (1986). By the rule, the bandwidth is $h = c_n n^{-1/5}$, where $c_n = 1.06\{\sum_{i=1}^n (\tilde{X}_i^\top \tilde{\gamma}^{(\ell)} - \bar{c})^2/n\}^{1/2}$ and $\bar{c} = n^{-1} \sum_{i=1}^n \tilde{X}_i^\top \tilde{\gamma}^{(\ell)}$. This rule has proved efficient in our computations.

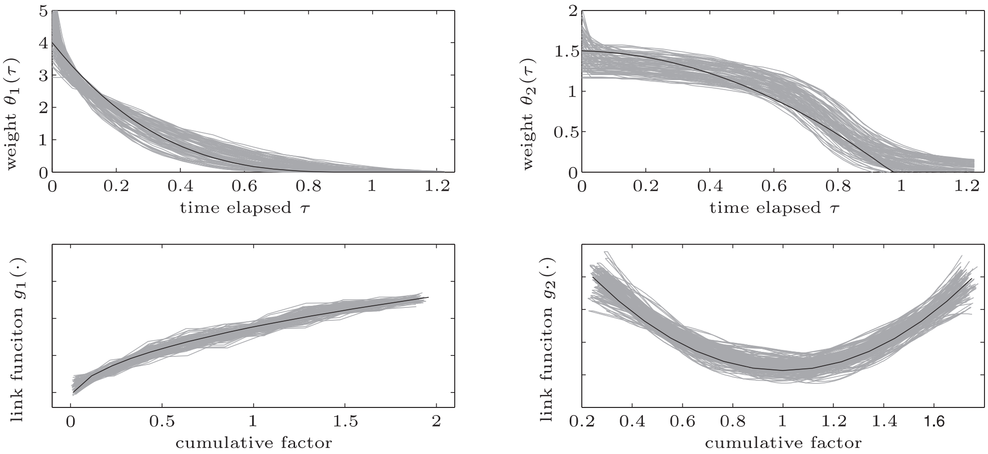

Figure 2. Results for Example 4. The black curve in each panel is the true link functions g_k or the true weight functions θ_k; the grey curves are the corresponding estimated functions.

4. Statistical Simulations of Finite Samples

To assess the performance of our algorithm with finite sample sizes, we considered two simulated examples. In the first, the $\mathbf{X}(t)$ are deterministic smooth functions, while the sample paths of $\mathbf{X}$ in the second example are step functions.

Consider the following model

$$Y_t = \beta^\top \mathbf{Z}_t + g_1\Big(\int_0^1 X_1(t-\tau)\theta_1(\tau)d\tau\Big) + g_2\Big(\int_0^1 X_2(t-\tau)\theta_2(\tau)d\tau\Big) + 0.5\varepsilon_t,$$

where $\varepsilon_t \sim N(0,1)$ and $\beta = (0.5, -1, 0.5)^\top$. Covariates $\mathbf{Z}_t = (Z_{1t},\ Z_{2t},\ Z_{3t})$ are independent random vectors with independent elements and $P(Z_{kt} = 1) = P(Z_{kt} = 0) = 0.5,\ k = 1, 2, 3,\ X_1(t) = \sin(t) + 1, X_2(t) = \sin(3t) + 1$, and

$$g_1(v) = v^{1/2} - 0.62,\ g_2(v) = (v-1)^2 - 0.25,\ \theta_1(\tau) = 4(1-\tau)_+^3,\ \theta_2(\tau) = 1.5(1-\tau^2)_+,$$

where $\tau \geq 0$. A monotone decreasing requirement is imposed on estimators of the weight functions $\theta_1(\tau)$ and $\theta_2(\tau)$, and on the link function $g_1(v)$.

Five hundred equally spaced observations were drawn from $[0, 16\pi]$. With 100 replications, the mean and standard deviation of the estimated β were respectively, (0.5034, -1.0042, 0.4955) and (0.0403, 0.0362, 0.0425). The estimator was quite accurate and stable. The estimated weight functions $\theta_k(\cdot)$ and link functions $g_k(\cdot)$ are shown in Figure 2. All functions were estimated with reasonable accuracy.

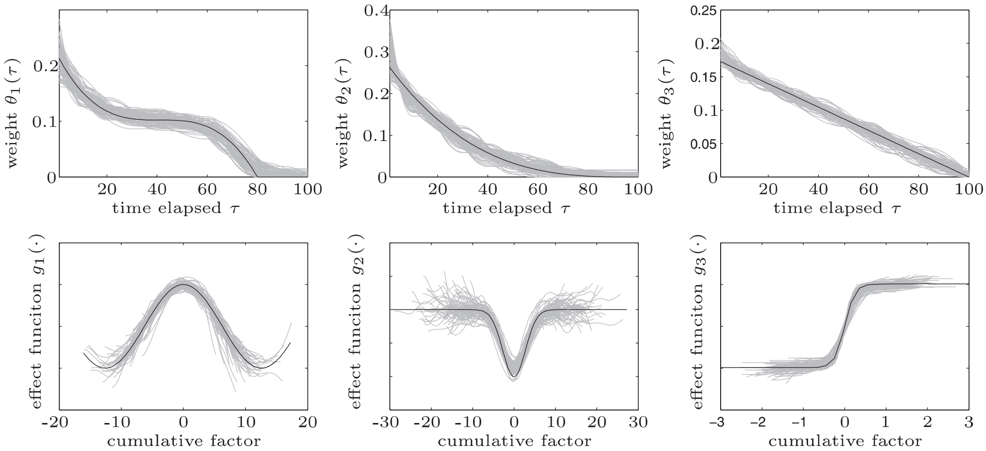

Figure 3. Calculation results for Example 4.1. The black curve in each panel is the true link function g_k or the true weight parameters $\theta_k, k = 1, 2, 3$; the grey curves are the corresponding estimated functions.

Example 4.1. In this example, **X** has step sample paths with equal steps, and all steps are observed once. The integration in the FACTS model can be written as summations. Thus, the model was

$$Y_{t_i} = \beta_1 Z_{1,t_i} + \beta_2 Z_{2,t_i} + g_1(\sum_{\tau=0}^{D} \theta_1(t_i - t_{i-\tau}) X_{1,t_{i-\tau}}) + g_2(\sum_{\tau=0}^{D} \theta_2(t_i - t_{i-\tau}) X_{2,t_{i-\tau}})$$
$$+ g_3(\sum_{\tau=0}^{D} \theta_3(t_i - t_{i-\tau}) X_{3,t_{i-\tau}}) + 0.2\varepsilon_t,$$

where $\beta_1 = 1$, $\beta_2 = -1$, $g_1(v) = \cos(3v) - 0.54$, $g_2(v) = 1 - \exp(-v^2) - 0.47$, $g_3(v) = 2\exp(40v)/\{1 + \exp(40v)\} - 0.5$, and the weight functions $\theta_k(\tau)$ were given by the black curves as shown in the first three panels of Figure 3. Covariates $Z_{1,t_i}, Z_{2,t_i}, i = 1, \ldots, n$, were IID with $P(Z_{1,t_i} = 1) = P(Z_{1,t_i} = 0) = 0.5$, while $X_{1,t_i} = 0.8X_{1,t_{i-1}} + e_{1,t_i}$, $X_{2,t_i} = 0.6X_{2,t_{i-1}} + 0.3X_{2,t_{i-2}} + e_{2,t_i}$, $X_{3,t_i} = -0.5X_{3,t_{i-1}} + e_{3,t_i}$ where $\{\varepsilon_{t_i}\}$, $\{e_{1,t_i}\}$, $\{e_{2,t_i}\}$, and $\{e_{3,t_i}\}$ were IID N(0,1).

There are two unknown parameters and six unknown functions. With $n = 500$ and $D = 100$, it is obviously difficult to obtain efficient estimates unless useful prior knowledge is available. For example, if we were aware that $g_3(\cdot)$ was monotone increasing and all weight functions were monotone decreasing, we could estimate the functions with unexpected degree of accuracy; as shown by Figure 3. As for parameters, the mean and standard deviation of estimated β were respectively, $(1.0020, -1.0016)^\top$ and $(0.0324, 0.0300)^\top$.

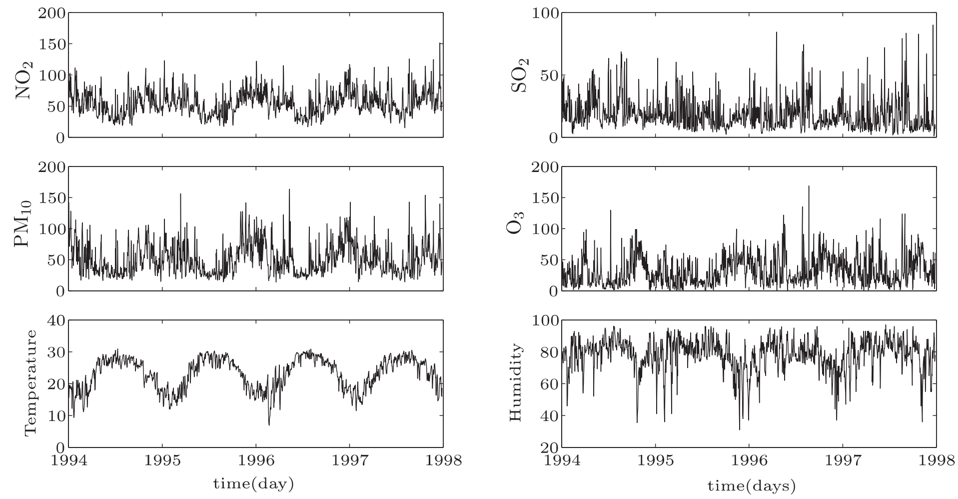

Figure 4. The observed time series of average levels of SO_2 (in ppb, parts per billion), NO_2 (in ppb), PM_{10} (in $\mu g/m^3$), ozone (in ppb), average of temperature (in ^oC) and average of humidity (in %), in Hong Kong.

5. Effects of Air Pollution on the Respiratory Diseases

As applications of the proposed methodology, we first consider the motivating example of the cumulative effect of air pollution on the respiratory diseases in Hong Kong. The data has been analyzed in the literature, e.g., Cai, Fan and Li (2000a), Cai, Fan and Yao (2000b) and Fan and Zhang (1999). However, they did not consider the cumulative effect. All associated pollutants and weather conditions are shown in Figure 4, while the number of hospital admissions of patients suffering from respiratory diseases is shown in Figure 1(c). We take y_{t_i} = log(number of hospital admissions of patients suffering from respiratory diseases on day i) to render the distribution closer to symmetry. For simplicity, we assume that the population remained largely unchanged. Since the data were observed at equal time intervals, we consider the discrete FACTS model

$$\log(y_{t_i}) = \sum_{d=1}^{7} \beta_d D_{t_i,d} + g_1(\sum_{\tau=0}^{D} \theta_{1,\tau} N_{t_{i-\tau}}) + g_2(\sum_{\tau=0}^{D} \theta_{2,\tau} S_{t_{i-\tau}}) + g_3(\sum_{\tau=0}^{D} \theta_{3,\tau} P_{t_{i-\tau}})$$
$$+ g_4(\sum_{\tau=0}^{D} \theta_{4,\tau} O_{t_{i-\tau}}) + g_5(\sum_{\tau=0}^{D} \theta_{5,\tau} T_{t_{i-\tau}}) + g_6(\sum_{\tau=0}^{D} \theta_{6,\tau} H_{t_{i-\tau}}) + \varepsilon_t, \quad (5.1)$$

where $N_{t_i}, S_{t_i}, P_{t_i}, O_{t_i}, T_{t_i}$, and H_{t_i} are, respectively, the average levels of NO_2, SO_2, PM_{10}, O_3, temperature and humidity on day i; $D_{d,t_i}, d = 1, \ldots, 7$, are dummy variables representing the day of the week. Here, $\theta_{k,\tau} = \theta_k(\Delta_\tau)$ with $\Delta_\tau = t_i - t_{i-\tau}$ for all $i \geq \tau > 0$.

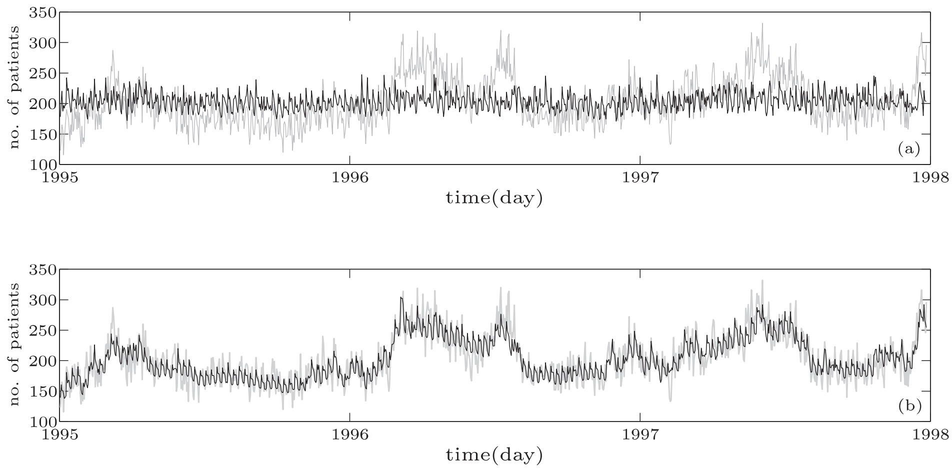

Figure 5. Fitted results of the additive model and FACTS model. In panel (a), the grey line is the observed numbers of hospital admissions and the black line is the fitted numbers by the additive model. In panel (b), the grey line is the observed number of hospital admissions and the black line the fitted numbers by the FACTS model.

With $D = 300$, (5.1) had fitted values as shown in Figure 5(b). As a comparison, the fitted values of the additive model with lags that give the best fit are shown in Figure 5(a). These figures suggest that (5.1) can capture the main signatures of the effects of pollution on the respiratory diseases in Hong Kong, while the additive model is wide off the mark. We have tried different $D > 300$ and found estimated weights quite stable in that they tended to be practically zero after some specific lag. The estimation results are shown in Figure 6. Instead of single-past-day effects as noticed in Dominici et al. (2002), all the pollutants and adverse weather conditions exhibit cumulative effects for the Hong Kong data, in that a weighted average of pollutants and weather conditions over the past 50-300 days has a strong effect on hospital admission. As we can see, FACTS can throw light on how the pollutants affect the diseases. In Hong Kong, most of the admitted patients of respiratory diseases were not serious sufferers. At the notified levels of the relevant pollutants in Hong-Kong, the necessity for their admission is mainly the result of *cumulative exposure* rather than single-day exposure to the pollutants or adverse weather conditions. It is therefore not surprising that the single-day-effect given by the additive model explains only 16.8% of the variation of the hospital admission in contrast to the 75.4% explained by the FACTS model.

The parameters $(\beta_1, \ldots, \beta_7)$ in (5.1) were estimated as (5.38, 5.33, 5.33, 5.32, 5.25, 5.24, 5.24). Their corresponding standard errors are all around 0.0075, in-

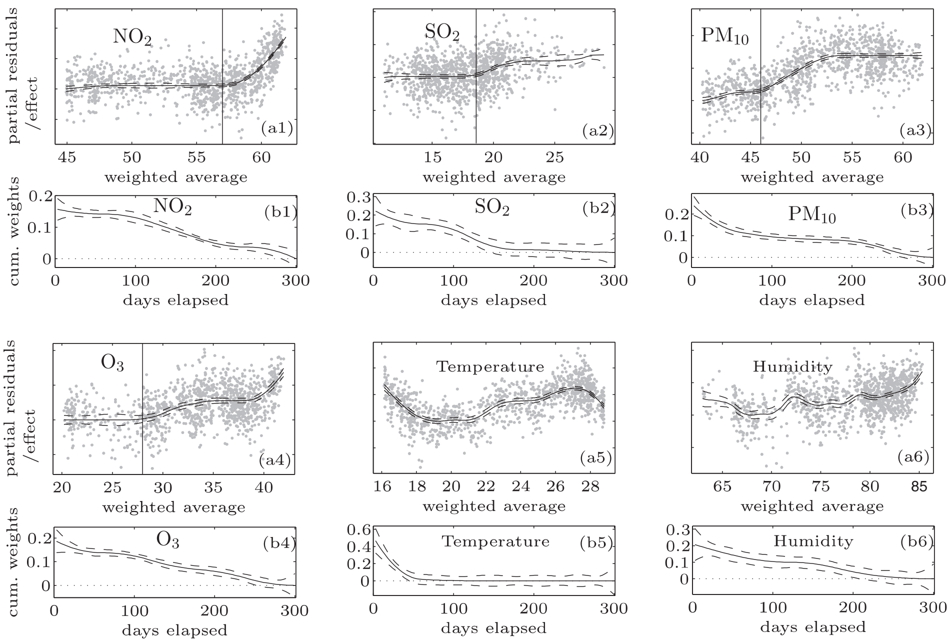

Figure 6. Calculation results for the pollution and respiratory diseases. (a1)−(a6) are the estimated cumulative effect functions (the central line) and corresponding 95% pointwise confidence intervals (upper and lower lines) for the pollutants and weather conditions; (b1)−(b6) are correspondingly the estimated accumulation weight functions and 95% confidence intervals. The solid vertical lines are roughly the thresholds above which the effect starts to appear.

dicating that the day-of-the-week effect is significant: high admission at the beginning of the week and low admission at the weekend. Of course, the "weekend-effect" is mainly due to human behavior rather than the pollutants, as noticed in Forster and Solomon (2003); the weather and pollution levels in Hong Kong show no such effect. Pollutants NO_2, O_3, PM_{10}, and weather conditions demonstrate strong adverse effects on health, while the effect of SO_2 is relatively small due to the measures taken by the Hong Kong Government in the 1990's to reduce the level of SO_2; see Hedley et al. (2002).

Based on the estimated effect functions $g_k(\cdot)$, there are thresholds at which the effects start to increase; see Figure 6(a1)−(a4). The thresholds are listed in Table 1. For NO_2 and PM_{10}, our thresholds roughly coincide with the National Ambient Air Quality Standards (NAAQS) in USA. However, our analysis suggests that the cumulative effects of SO_2 and O_3 start to increase at much lower levels than stipulated by NAAQS. Epidemiological studies suggested that O_3 affects the forced vital capacity at a much lower level; see Abelson

Table 1. Threholds and National Ambient Air Quality Standards (NAAQS)

Standards	NO_2	SO_2	O_3	PM_{10}
FACTS model	56ppb	18ppb	28ppb	$46\mu g/m^3$
NAASQ*	53ppb	30ppb	80ppb	$50\mu g/m^3$
	(annual)	(annual)	(24 hours)	(annual)

* these standards can be found at *http://www.epa.gov/air/criteria.html*

(1997). Delfino et al. (1997) also suggested a threshold at 29ppb of O_3 for old people. Sunyer et al. (2003) demonstrated statistically that SO_2 starts to increase asthma hospital admissions at a very low level. In other words, our analysis lends support to those epidemiology studies that suggest that the effect of SO_2 and O_3 on the respiratory diseases is far more pronounced than suggested by the current NAAQS. Therefore, we urge that the NAAQS standards be revised to lower levels in the interest of public health.

For the weather conditions, both an unusually cold season and an unusually hot season can aggravate the diseases; see Figure 6(a5). This statistical observation is consistent with the medical observation that unduly cold weather or unduly hot weather is not favorable to disease sufferers: the transmission rate for viruses and diseases is higher in cold season thus exacerbating other diseases; hot weather increases the risk of dehydration and other adverse effects ; see Rastogi, Tanveer and Gupta (1998) and McGeehin and Mirabelli (2001). The wetter weather causes more hospital admissions of respiratory diseases; see Figure 6(a6). This statistical evidence is also consistent with the biological understanding that wetter weather makes easier fungal colonization, thus worsening the air quality and causing health problems; see Ezeonu et al. (1994).

6. Decay of Immunity Against Influenza

Influenza is an infectious disease arising as a series of seasonal epidemics. A weekly notified time series of influenza-like cases in France is shown in Figure 7. In human influenza (type A), immunity to re-infection is finite, particularly because the virus undergoes a combination of year-to-year antigenic drift and occasional dramatic shift in haemagglutinin and neuraminidase surface protein; see Nicholson, Webster and Hay (1998). Pease (1987) conjectured that immunity to influenza decays linearly with time elapsed; Couch and Kasel (1983) argued that immunity lasts for more than four years. However, there have been few quantitative investigations of how immunity decays with time since recovery. It is not difficult to see that the decaying of immunity is a “cumulative” procedure. In the following, we focus on FACTS modeling of the decaying pattern of human immunity against this particular disease.

Let $p(\tau)$ denote the probability that a host is susceptible at time τ after his/her last recovery from the disease. For simplicity, we assume the value of

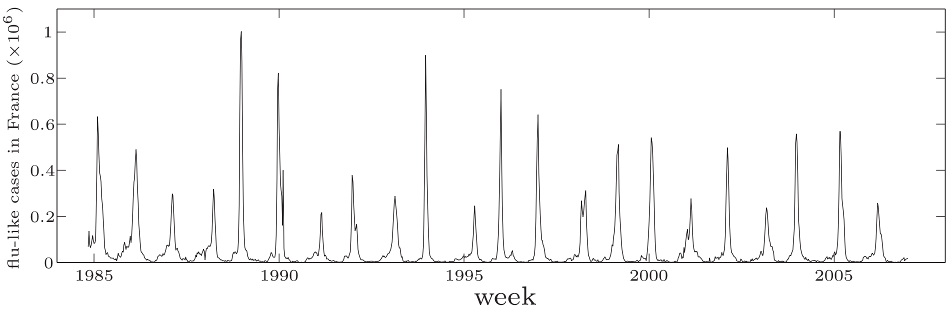

Figure 7. Weekly notified cases of flu-like diseases in France

$p(\tau)$ depends only on the time, τ, elapsed after the last recovery from infection. Note that $p(\tau)$ is an increasing function. Suppose that p_0 is the limit of $p(\tau)$ as $\tau \to \infty$. We call $\kappa(\tau) = 1 - p(\tau)/p_0$ the *relative immunity*. It is easy to see that $0 \le \kappa(\tau) \le 1$ and $\kappa(\tau)$ is a decreasing function of τ. Let $y_{t-\tau}$ denote the number of hosts infected at time point $t - \tau$; each of them has the probability $p(\tau)$ at time t to be re-infected, i.e. their immunity at t is $\kappa(\tau)$. We assume that when recovering from the infection at time t, a person's immunity is built up from this infection alone. The expected number of susceptibles (within these $y_{t-\tau}$ hosts) is $y_{t-\tau}p(\tau) = p_0 y_{t-\tau}(1 - \kappa(\tau))$. If a person has never been infected before, we may simply take him/her as having been infected in the remotest past. Among the population $N = \int_0^\infty y_{t-\tau} d\tau$, which is again assumed to be constant, the total number of susceptibles at t, is then

$$S_t = \int_0^\infty y_{t-\tau} p(\tau) d\tau = p_0 \{N - \int_0^\infty y_{t-\tau} \kappa(\tau) d\tau\}.$$

The general susceptible-infected-recovered-susceptible (SIRS) mechanism suggests the model

$$\frac{dy_t}{dt} = \beta_t y_t^\alpha S_t^\gamma; \tag{6.1}$$

see Liu, Hethcote and Levin (1987), Anderson and May (1991), and Finkenstädt and Grenfell (2000). In the model, β_t describes the seasonal effect. We can make the model more flexible by replacing S_t^α with $\nu(S_t)$ or $\mu(\int_0^\Delta y_{t-\tau}\kappa(\tau)d\tau\})$, where $\nu(\cdot)$ and $\mu(\cdot)$ are unknown link functions. See Xia, Gog and Grenfell (2005). The function $\mu(\cdot)$ describes the functional relation between the expected number of immune hosts and expected cases in the next time unit.

In practice, we can only observe the dynamics at discrete time. Let y_{t_i} be the cases in a time period $t_i - t_{i-1}$; the time period $t_i - t_{i-1}$ is usually one week.

Following Finkenstädt and Grenfell (2000) and Xia, Gog and Grenfell (2005), (6.1) can be approximated by

$$y_{t_i} = \beta_{t_i} y_{t_{i-1}}^{\alpha} \mu(\sum_{\tau=0}^{D} y_{t_{i-\tau}} \kappa(t_i - t_{i-\tau})).$$

With approximately 52 weeks in a year, dummy variables $D_{k,t}$ are employed to describe weekly seasonal variations in infection rate: $D_{k,t_i} = 1$ if $k = t_i(\text{mod}52)$, 0 otherwise. Write $\beta_{t_i} = \exp\{\sum_{\tau=1}^{52} \varrho_k D_{k,t_\tau}\}$, where ϱ_k are seasonal force parameters. A convenient stochastic model is then a discrete time FACTS model

$$\log(y_{t_i}) = \sum_{i=1}^{52} \varrho_k D_{k,t_i} + \alpha \log(y_{t_{i-1}}) + \tilde{\mu}(\sum_{\tau=0}^{D} y_{t_{i-\tau}} \kappa(t_i - t_{i-\tau})) + \varepsilon_{t_i}, \quad (6.2)$$

where $\tilde{\mu}(\cdot) = \log\{\mu(\cdot)\}$.

Results based on the weekly notified influenza cases in France are shown in Figure 8. We have that $\hat{\alpha} = 0.93$ (SE = 0.013) and that the variance of ε_{t_i} is 0.21. Note that the variance of $z_{t_i} = \log(y_{t_i})$ is 2.46. The proportion of the variance of $\{z_{t_i}\}$ that can be explained by the model is $R^2 = 91.5\%$. Therefore FACTS model fit the dynamics quite well. We conclude that (I) the expected number of immune hosts has a significant negative effect on the number of cases in the next time unit; this is in line with the SIRS mechanism. However, the quantitative relation between y_{t_i}, $y_{t_{i-1}}$, and the expected number of susceptibles, as shown in Figs 8(a)−(b), is more complicated than that assumed in (6.1). (II) The estimated seasonal infection rates, as shown in Figure 8(c), are consistent with the general medical observation that in the winter, the forces of infection are stronger than those at other periods (Nicholson, Webster and Hay (1998)). (III) The epidemics in France have a decay function of immunity as shown in Figure 8(d). It is noteworthy that the decay pattern of immunity is different from the conjecture of Pease (1987). In the first few months, the recovered hosts have a high level of immunity; after that, the immunity decreases quickly and after about 8-12 months (say 50 weeks), the immunity level is relatively low. However, this low level of immunity will last for as long as another two years. To explain this particular patter of decay of the immunity, two factors emerge. (1) The fast decay of immunity at the beginning may partly reflect hospital notification biases–if individuals are rapidly re-infected, they may not have clinical notification. It might also reflect the fact that drift speed could in turn depend on the number of cases; see Boni et al. (2004). Our result is consistent with the claim that short-term immunity is ‘strain-transcending’; see Ferguson et al. (2003). (2) The subsequent slow decay pattern could reflect gradual effects of drift, though this is a complex picture because influenza-like illness subsumes influenza B, influenza

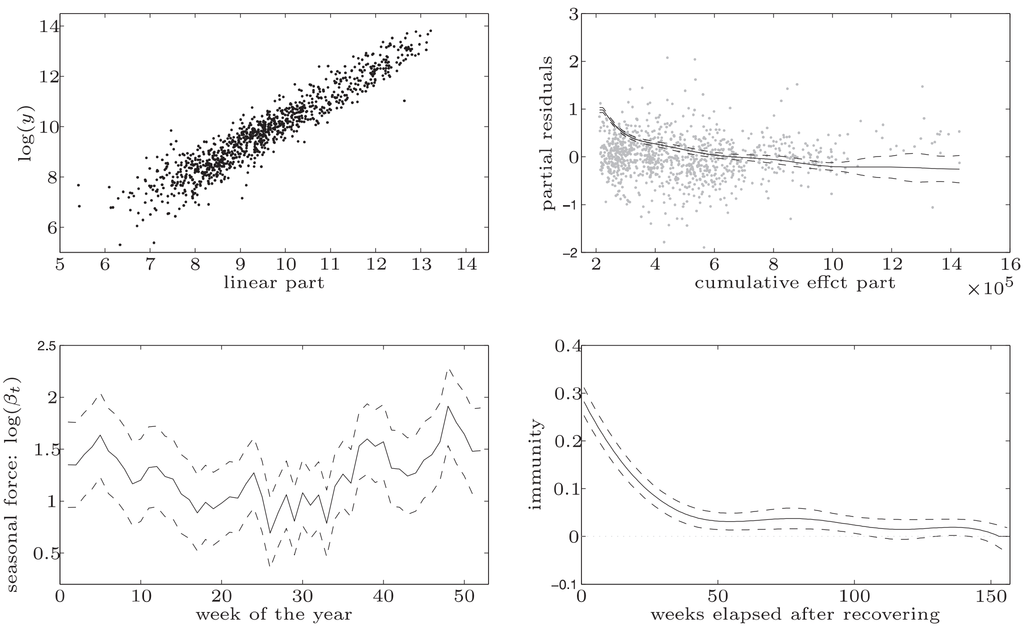

Figure 8. Results for the influenza data in France. Panel (a) is $\log(y_{t_i})$ plotted against the linear part in model (6.2). In panel (b), the dots are the corresponding partial residuals after removing the linear part in model (6.2). The x-axis is the expected immunized number $\sum_{\tau=0}^{D} y_{t_{i-\tau}}\kappa(t_i - t_{i-\tau})$. The solid line is the estimated link function, the dash lines are the 95% pointwise confidence intervals for $\tilde{\mu}(\cdot)$. Panel (c) is the estimated seasonal forces; the dash lines are their corresponding 95% confidence intervals. Panel (d) is the estimated decay function of immunity and its 95% confidence interval, represented by a solid line and dashed lines respectively.

A subtypes, and possibly other respiratory infections. The overall immunity can last as long as three years though it is relatively weak. In this sense, our results lend support to the arguments of Couch and Kasel (1983) and Murphy and Clements (1989).

Acknowledgement

The authors thank an associate editor and two referees for their very valuable comments. The research is supported by grant R-155-000-063-112, National University of Singapore.

Appendix

For ease of exposition, denote ξ_ϕ by ξ and the corresponding parameter space by Ξ. We need the following assumptions to prove the consistency of the least squares estimators.

(A1) The observations $\{\mathbf{Z}(t_i), \mathbf{X}(t_i), Y_{t_i}\}_{i=1}^n$ are strictly mixing. Specifically $X(t) = m(t) + B(t)$, where $m(\cdot)$ is continuous deterministic function of t and $B(\cdot)$ is the standard Brownian motion.

(A2) $\delta_{\mathbf{X}} := \max_{i\geq 1} |\tilde{t}_{i+1} - \tilde{t}_i| = O(n^{-1})$.

(A3) The parameter space $\boldsymbol{\Xi}$ is compact, and the mean function $m(\mathbf{v};\xi)$ is continuous on $\boldsymbol{\Xi}$ for any fixed $\mathbf{v}$.

(A4) The regression mean function $m(\mathbf{v};\xi)$ is twice continuously differentiable in a neighborhood of ξ^0.

(A5) $n^{-1}\sum_{i=1}^n \{m(\mathbf{v}_i;\xi) - m(\mathbf{v}_i;\xi^*)\}^2$ converges to a limit function uniformly in $\xi, \xi^* \in \boldsymbol{\Xi}$, and

$$Q(\xi) = \lim_n \frac{1}{n}\sum_{i=1}^n \{m(\mathbf{v}_i;\xi) - m(\mathbf{v}_i;\xi^0)\}^2 \qquad \text{(A.1)}$$

has a unique minimum at $\xi = \xi^0$.

(A6) The true parameter vector ξ^0 is an interior point of $\boldsymbol{\Xi}$.

(A7) $\Omega(\xi^0)$ exists and is nonsingular, with

$$\Omega(\xi) := \lim_n \frac{1}{n}\sum_{i=1}^n m^{(1)}(\mathbf{v}_i;\xi) m^{(1)}(\mathbf{v}_i;\xi)^\top \qquad \text{(A.2)}$$

and the $n^{-1}\sum_{i=1}^n \partial^2 m(\mathbf{v}_i;\xi)/(\partial\xi_{j_1}\partial\xi_{j_2})$, $j_1, j_2 = 1, \ldots, 2qr+p$, converge uniformly in $\boldsymbol{\Xi}$ in a neighbourhood of ξ^0.

Remark. To guarantee the consistency of the semi-parametric implementation, the observations in (A1) need to be strongly mixing. See for example Fan, Yao and Tong (1996) and Xia et al. (2002). (A2) is imposed to ensure the uniform convergence of $X_{t_i}^{n,k}$ to $X_{t_i}^k$ for all $i = 1, \ldots, n$ (Lemma A.1). Such a requirement is commonly made in dealing with a continuous-time model but with only discretized data available, especially in finance and biology, see e.g., Fan and Zhang (2003) and Fan and Jiang (2005). Assumptions (A3)−(A7) are similar to those in Yu and Ruppert (2002), and the uniform convergence assumption is needed to guarantee the continuity in ξ of the limit function.

Proof of Proposition 2.1. For ease of exposition, we only consider $q = 2$. Let $\tilde{\mathbf{Z}}(t) = E\{\mathbf{Z}^\top(t)|\mathbf{X}(s), s \leq t\}$. We have

$$E\{Y_t|\mathbf{X}(s), s \leq t\} = \tilde{\mathbf{Z}}^\top(t)\beta^0 + \sum_{k=1}^q g_k\left(\int_0^\Delta X_k(t-\tau)\theta_k(\tau)dt\right). \qquad \text{(A.3)}$$

Subtracting (2.2) by (A.3), $Y_t - E\{Y_t|\mathbf{X}(s), s \le t\} = \{\mathbf{Z}(t) - \tilde{\mathbf{Z}}(t)\}^\top \beta^0 + \varepsilon_t$. By the assumption of the invertibility of the matrix, it follows that

$$\beta^0 = [E\{\mathbf{Z}(t) - \tilde{\mathbf{Z}}(t)\}\{\mathbf{Z}(t) - \tilde{\mathbf{Z}}(t)\}^\top]^{-1} E[\{\mathbf{Z}(t) - \tilde{\mathbf{Z}}(t)\}\{Y_t - E(Y_t|\mathbf{X}(s), s \le t)\}].$$

Therefore, β^0 is uniquely determined. Let $\tilde{Y}_t = Y_t - \mathbf{Z}^\top(t)\beta^0$ and $m(x_1(s), x_2(s) : s \le t) = E\{\tilde{Y}_t | X_1(s) = x_1(s), X_2(s) = x_2(s) : s \le t\}$. It follows that

$$m(x_1(s), x_2(s) : s \le t) = \sum_{k=1}^{2} g_k\Big(\int_0^\Delta x_k(t-\tau)\theta_k(\tau)d\tau\Big).$$

For any sample paths $\tilde{x}_2(s)$ and $x_2(s)$, define

$$\Delta_2(\tilde{x}_2(t), x_2(t)) = \frac{m(x_1(s), \tilde{x}_2(s) : s \le t) - m(x_1(s), x_2(s) : s \le t)}{\int_0^\Delta \{\tilde{x}_2(t-\tau) - x_2(t-\tau)\}d\tau}.$$

Letting $\tilde{x}_2^a(s) = x_2(s) + h\{1 - (s-t)^2/h^2\}I(|s-t| < h)$, we have

$$\lim_{h\to 0} \Delta_2(\tilde{x}_2^a(t), x_2(t)) = \theta_2(0) g_2'\Big(\int_0^\Delta x_2(t-\tau)\theta_2(\tau)d\tau\Big). \tag{A.4}$$

Letting $\tilde{x}_2^b(s) = x_2(s) + h\{1 - (s-t+v)^2/h^2\}I(|s-t+v| < h)$, we have

$$\lim_{h\to 0} \Delta_2(\tilde{x}_2^b(t), x_2(t)) = \theta_2(v) g_2'\Big(\int_0^\Delta x_2(t-\tau)\theta_2(\tau)d\tau\Big).$$

It follows that $\theta_2(v)/\theta_2(0) = \lim_{h\to 0} \Delta_2(\tilde{x}_2^b(t), x_2(t)) / \lim_{h\to 0} \Delta_2(\tilde{x}_2^a(t), x_2(t))$ is determined by the conditional mean function $m(\cdot)$, which together with $\int \theta_2(v)dv = 1$ establishes the identifiability of the function $\theta_2(\cdot)$. Similarly, the other weight functions are identifiable.

By (A.4) and the identifiability of $\theta_2(\cdot)$, the derivative of $g_2(\cdot)$ is identifiable. By the first assumption in (2.3), $g_2(\cdot)$ is identifiable. Similar arguments can be applied to $g_1(\cdot)$.

Lemma A.1. *Under* (A1) *and* (A2), *we have*

$$X_{t_i}^{n,k} - X_{t_i}^k = O\Big(\big(\frac{n}{\log n}\big)^{-1/2}\Big) \quad a.s. \tag{A.5}$$

uniformly for all $k = 1, \dots, q$ *and* $i = 1, \dots, n$.

Proof. For $T \overset{def}{=} t_n$, Brownian motion $B(t), t \in (0,T)$ and any fixed $c > 1$, with probability one there exists $\delta > 0$, such that $|B(t) - B(t+h)| \le c|h/\log h|^{-1/2}$ for any $t \in [0,T)$ and $h < \delta$. Substitute n^{-1} for h and we have (A.5), as $m(t)$ is uniformly continuous on $[0,T]$.

Let $\Omega_n(\xi) = \lim_n n^{-1}\sum_{i=1}^n m_n^{(1)}(\mathbf{v}_i;\xi) m_n^{(1)}(\mathbf{v}_i;\xi)^\top$. The following Lemma shows that $\Omega_n(\xi)$ is a good approximation of $\Omega(\xi)$.

Lemma A.2. *Under* (A2)−(A4) *and* (A7), *we have* $\Omega_n(\xi_\phi) - \Omega(\xi_\phi) \to 0$ *and*

$$n^{-1}\sum_{i=1}^{n}\frac{\partial^2 m_n(\mathbf{v}_i;\xi_\phi)}{\partial\xi_{j_1}\partial\xi_{j_2}} - n^{-1}\sum_{i=1}^{n}\frac{\partial^2 m(\mathbf{v}_i;\xi_\phi)}{\partial\xi_{j_1}\partial\xi_{j_2}} \to 0$$

almost surely and uniformly in $\boldsymbol{\Xi}$ *in a neighborhood of* ξ^0, *for* $j_1,\ j_2 = 1,\ldots,$ $2qr+p$.

Proof. The result follows directly from Lemma A.1 and an application of the Cauchy-Schwartz Inequality.

Proof of Theorem 3.1. Following the proof of Yu and Ruppert (2002), write

$$\begin{aligned}
Q_{n,\lambda}(\xi) &= \frac{1}{n}\sum_{i=D+1}^{n}\Big\{Y_{t_i} - m_n(\mathbf{v}_i;\xi)\Big\}^2 + \lambda_n\delta^\top\Sigma\delta \\
&= \frac{1}{n}\sum_{i=D+1}^{n}\Big\{Y_{t_i} - m(\mathbf{v}_i;\xi^0) + m(\mathbf{v}_i;\xi^0) - m(\mathbf{v}_i;\xi) + m(\mathbf{v}_i;\xi) - m_n(\mathbf{v}_i;\xi)\Big\}^2 \\
&\quad + \lambda_n\delta^\top\Sigma\delta \\
&= \frac{1}{n}\sum_{i=D+1}^{n}\varepsilon_{t_i}^2 + \frac{2}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi^0) - m(\mathbf{v}_i;\xi)\}\varepsilon_{t_i} \\
&\quad + \frac{1}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi^0) - m(\mathbf{v}_i;\xi)\}^2 \\
&\quad + \frac{2}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi) - m_n(\mathbf{v}_i;\xi)\}\varepsilon_{t_i} + \frac{1}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi) - m_n(\mathbf{v}_i;\xi)\}^2 \\
&\quad + \frac{2}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi) - m_n(\mathbf{v}_i;\xi)\}\{m(\mathbf{v}_i;\xi^0) - m(\mathbf{v}_i;\xi)\} + \lambda_n\delta^\top\Sigma\delta \\
&= \frac{1}{n}\sum_{i=D+1}^{n}\varepsilon_{t_i}^2 + T_1 + T_2 + T_3 + T_4 + T_5 + T_6. \qquad \text{(A.6)}
\end{aligned}$$

All following convergence is uniform for all $\xi \in \boldsymbol{\Xi}$ almost surely, taken when $n \to \infty$ with $\delta_n \to 0$ unless otherwise stated. First note that by Lemma A.1, we have $T_k \to 0,\ k = 3,4,5$ under (A3)−(A5) and (A7). Under (A4) and (A5), the remaining terms can be handled in exactly the same manner as in Yu and Ruppert (2002). Therefore,

$$Q_{n,\lambda}(\xi) \to Q(\xi) + \sigma^2 \text{ uniformly for all } \xi \in \boldsymbol{\Xi}. \qquad \text{(A.7)}$$

The strong consistency of the $PLSE$ estimator $\hat{\xi}_{n,\lambda}$ thus follows from arguments parallel to those in Yu and Ruppert (2002).

As $\hat{\xi}_{n,\lambda}$ is a consistent estimate of ξ^0 and mimimizes

$$Q_{n,\lambda}(\xi) = \frac{1}{n}\sum_{i=D+1}^{n}\Big\{Y_{t_i} - m_n(\mathbf{v}_i;\xi)\Big\}^2 + \lambda_n\delta^\top\Sigma\delta,$$

Taylor expansion of $Q_{n,\lambda}(\xi)$ near ξ^0 yields

$$\mathbf{0} = \frac{\partial Q_{n,\lambda}}{\partial\xi}\Big|_{\hat{\xi}_{n,\lambda}} = \frac{\partial Q_{n,\lambda}}{\partial\xi}\Big|_{\xi^0} + \frac{\partial^2 Q_{n,\lambda}}{\partial\xi\partial\xi^\top}\Big|_{\tilde{\xi}}(\hat{\xi}_{n,\lambda} - \xi^0),$$

where $\tilde{\xi}$ is a vector between $\hat{\xi}_{n,\lambda}$ and ξ^0. Consequently, we have

$$\sqrt{n}(\hat{\xi}_{n,\lambda} - \xi^0) = -\Big\{\frac{\partial^2 Q_{n,\lambda}}{\partial\xi\partial\xi^\top}\Big|_{\tilde{\xi}}\Big\}^{-1}\sqrt{n}\frac{\partial Q_{n,\lambda}}{\partial\xi}\Big|_{\xi^0}.$$

It is sufficient to prove

$$\sqrt{n}\frac{\partial Q_{n,\lambda}}{\partial\xi}\Big|_{\xi^0} \xrightarrow{D} N(0, 4\sigma^2\Omega(\xi^0)), \tag{A.8}$$

$$\frac{\partial^2 Q_{n,\lambda}}{\partial\xi\partial\xi^\top}\Big|_{\tilde{\xi}} \xrightarrow{P} 2\Omega(\xi^0). \tag{A.9}$$

To prove (A.8), notice that

$$\frac{\partial Q_{n,\lambda}}{\partial\xi} = -\frac{2}{n}\sum_{i=D+1}^{n}\Big\{Y_{t_i} - m_n(\mathbf{v}_i;\phi)\Big\}m_n^{(1)}(\mathbf{v}_i;\xi) + 2\lambda_n[0,\delta^\top\Sigma]^\top,$$

$$\frac{\partial Q_{n,\lambda}}{\partial\xi}\Big|_{\xi^0} = -\frac{2}{n}\sum_{i=D+1}^{n}\Big\{\varepsilon_{t_i} + \sum_{k=1}^{q}\eta_k^{0\top}[\mathbf{A}(\eta_k^{0\top}X_{t_i}^k) - \mathbf{A}(\eta_k^{0\top}X_{t_i}^{n,k})]\Big\}m_n^{(1)}(\mathbf{v}_i;\xi^0) + 2\lambda_n[0,\{\delta^0\}^\top\Sigma]^\top.$$

As $\lambda_n = o(n^{-1/2})$, the last term can be ignored. Under (A2) and (A3), it follows that $\mathbf{A}(\eta_k^{0\top}X_{t_i}^k) - \mathbf{A}(\eta_k^{0\top}X_{t_i}^{n,k}) = o(n^{-1/2})$, uniformly in i and k, which, together with Lemma A.2, leads to

$$\frac{2}{\sqrt{n}}\sum_{i=D+1}^{n}\sum_{k=1}^{q}\eta_k^{0\top}[\mathbf{A}(\eta_k^{0\top}X_{t_i}^k) - \mathbf{A}(\eta_k^{0\top}X_{t_i}^{n,k})]m_n^{(1)}(\mathbf{v}_i;\xi^0) \to 0. \tag{A.10}$$

By (A1) and the Central Limit Theorem for martingale differences, we have

$$\frac{1}{\sqrt{n}}\sum_{i=D+1}^{n}\varepsilon_{t_i}m_n^{(1)}(\mathbf{v}_i;\xi^0) \xrightarrow{D} N(0, \sigma^2\Omega(\xi^0)). \tag{A.11}$$

Combining (A.10) and (A.11) yields (A.8).

For (A.9), we have

$$\frac{\partial^2 Q_{n,\lambda}}{\partial\xi\partial\xi^\top}|_{\tilde{\xi}} = \frac{2}{n}\sum_{i=D+1}^{n} m_n^{(1)}(\mathbf{v}_i;\xi)m_n^{(1)}(\mathbf{v}_i;\xi)^\top|_{\tilde{\xi}}$$

$$-\frac{2}{n}\sum_{i=D+1}^{n}\left\{Y_{t_i} - m_n(\mathbf{v}_i;\xi)\right\}\frac{\partial^2 m_n(\mathbf{v}_i;\xi)}{\partial\xi_{j_1}\partial\xi_{j_2}}|_{\tilde{\xi}} + 2\lambda_n\Sigma$$

$$= 2\Omega_n(\tilde{\xi}) - \frac{2}{n}\sum_{i=D+1}^{n}\left\{Y_{t_i} - m(\mathbf{v}_i;\tilde{\xi})\right\}\frac{\partial^2 m_n(\mathbf{v}_i;\xi)}{\partial\xi_{j_1}\partial\xi_{j_2}}|_{\tilde{\xi}}$$

$$-\frac{2}{n}\sum_{i=D+1}^{n}\left\{m(\mathbf{v}_i;\tilde{\xi}) - m_n(\mathbf{v}_i;\tilde{\xi})\right\}\frac{\partial^2 m_n(\mathbf{v}_i;\xi)}{\partial\xi_{j_1}\partial\xi_{j_2}}|_{\tilde{\xi}} + 2\lambda_n\Sigma.$$

The first term on the right side goes to $2\Omega(\xi^0)$ by Lemma A.2, (A7), and the fact that $\tilde{\xi}\rightarrow\xi^0$ almost surely. The second term is $o_p(1)$ as argued in Yu and Ruppert (2002). By Lemma A.1 and (A3), the last two term are also $o_p(1)$. This completes the proof of (A.9).

References

Abelson, P. H. (1997). Proposed air pollutant standards. *Science* **277**, 15.

Anderson, R. M. and May, R. M. (1991). *Infectious Disease of Humans: Dynamics and Control.* Oxford University Press, Oxford.

Boni, M. F., Gog, J. R., Andreasen, V. and Christiansen, F. B. (2004). Influenza drift and epidemic size: the race the race between generating and escaping immunity. *Theor. Popul. Biol.* **65**, 179-191.

Cai, Z., Fan, J. and Li, R. Z. (2000a). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.* **95**, 888-902.

Cai, Z., Fan, J. and Yao, Q. (2000b). Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.* **95**, 941-956.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Ceriello, A., Taboga, C., Tonutti, L., Quagliaro, L., Piconi, L., Bais, B., Da, R. R. and Motz, E. (2002). Evidence for an independent and cumulative effect of postprandial hypertriglyceridemia and hyperglycemia on endothelial dysfunction and oxidative stress generation: effects of short- and long-term simvastatin treatment. *Circulation* **106**, 1211-1218.

Considering Cumulative Effects Under the National Environmental Policy Act (available at `http://ceq.eh.doe.gov/nepa/ccenepa/ccenepa.htm`).

Couch, R. B. and Kasel, J. A. (1983). Immunity to influenza in man. *Annu. Rev. Microbiol.* **37**, 529-549.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions, *Numerische Mathematik*, **31**, 377-403.

Delfino, R. J., Murphy-Moulton, A. M., Burnett, R. T., Brook, J. R. and Becklake, M. R. (1997). Effects of ozone and particulate air pollution on emergency room visits for respiratory illnesses in Montreal. *Am. J. Respir. Crit. Care Med.* **155**, 568-576.

Dominici, F., McDermott, A., Zeger, S. L. and Samet, J. M. (2002). On the use of generalized additive models in time series of air pollution and health. *Amer. J. Epidemiol.* **156**, 193-203.

Dubé, M., Johnson, B. Dunn, G., Culp, J., Cash, K., Munkittrick, K., Wong, I., Hedley, K., Booty, W., Lam, D., Resler, O. and Storey, A. (2006). Development of a new approach to cumulative effects assessment: a northern river ecosystem example *Environmental Monitoring and Assessment* **113**, 87-115.

Ezeonu, M. I., Noble, J. A., Simmons, R. B., Price, D. L., Crow, S. A. and Ahearn, D. G. (1994). Fungal production of volatiles during growth on fiberglass. *Appl. Environ. Microbiol.* **60**, 2149-2151.

Fan, J. and Jiang, J. (2005). Nonparametric inference for additive models. *J. Amer. Statist. Assoc.* **100**, 890-907.

Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83**, 189-196.

Fan, J. and Zhang, C. (2003). A re-examination of Stanton's diffusion estimations with applications to financial model validation . *J. Amer. Statist. Assoc.* **98**, 118-134.

Fan, J. and Zhang, W. Y. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27**, 1491-1518.

Ferguson, N. M., Galvani, A. P. and Bush, R. M. (2003). Ecological and immunological determinants of influenza evolution. *Nature* **422**, 428-433.

Finkenstädt, B. F. and Grenfell, B. T. (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Statist.* **49**, 187-205.

Forster, de F. P. M. and Solomon, S. (2003). Observations of a "weekend effect" in diurnal temperature range. *Proc. Nat. Acad. Sci.* **100**, 11225-11230.

Galizia, A. and Kinney, P. L. (1999). Long-term residence in areas of high ozone: associations with respiratory health in a nationwide sample of non-smoking young adults. *Environ. Health Perspect.* **107**, 675-679.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by method of average derivatives. *J. Amer. Statist. Assoc.* **84**, 986-995.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Hedley, A. J., Wong, C. M., Thach, T. Q., Ma, S., Lam, T. H. and Anderson, H. R. (2002). Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study. *The Lancet* **360**, 1646-1652.

James, G. M. and Silverman, B. W. (2005). Functional adaptive model estimation. *J. Amer. Statist. Assoc.*, **100**, 565-575.

Liew, C. K. (1976). Inequality constrained least-squares estimation. *J. Amer. Statist. Assoc.* **71**, 746-751.

Liu, R. M., Hethcote, H. W. and Levin, S. A. (1987). Dynamical behaviour of epidemiological models with nonlinear indendence rates. *J. Math. Biol.* **98**, 543-468.

Liu, Z. and Stengos, T. (1999). Non-linearities in Cross-Country Growth Regressions: A Semiparametric Approach. *J. Appl. Econometrics* **14**, 527-38.

McGeehin, M. A. and Mirabelli, M. (2001). The potential impacts of climate variability and change on temperature-realted morbidity and mortality in the United States. *Environ. Health Perspect.* **109**, 185-189.

Murphy, B. R. and Clements, M. L. (1989). The systemic and mucosal immune response of humans to influenza A virus. *Curr. Top. Microbiol. Immunol.* **146**, 107-116.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer, New York.

Nicholson, K. G., Webster, R. G. and Hay, A. J. (1998). *Textbook of Influenza*. Blackwell Science.

Pease, C. M. (1987). An evolutionary epidemiocological mechanism, with applications to type A influenza. *Theoretical Population Biology*, **31**, 422-452.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statist. Sci.* **3**, 425-461.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer, New York.

Rastogi, S. K., Tanveer, H. and Gupta, B. N. (1998). Pulmonary diseases caused by organic dusts in agricultural workers-a review. *Indian J. of Occup. and Envir. Medic.* **2**, 190-194.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression, *J. Amer. Statist. Assoc.* **90**, 1257-1270.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Smith, B. and Spaling, H. (1995). Methods for cumulative effects assessment. *Environmental Impact Assessment Review* **15**, 81-106.

Sunyer, J., Atkinson, R., Ballester, F., Le Tertre, A, Ayres, J. G., Forastiere, F., Forsberg, B., Vonk, J. M., Bisanti, L., Anderson, R. H., Schwartz, J., and Katsouyanni, K. (2003). Respiratory effects of sulphur dioxide: a hierarchical multicity analysis in the APHEA-2 study. *Occupational and Environmental Medicine* **60**, e2.

World Health Organization (2003). Reports on a WHO/HEI working group, Bonn, Germany. Available at `www.euro.who.int/document/e78992.pdf`.

Xia, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory* **22**, 1112-1137.

Xia, Y., Gog, J. and Grenfell, B. T. (2005). Semiparametric estimation of the duration of immunity from infectious disease time-series: influenza as a case study. *Appl. Statist.* **54**, 659-672.

Xia, Y., Tong, H., Li, W. K. and Zhu, L. (2002). An adaptive estimation of dimension reduction space (with discussion). *J. Roy. Statist. Soc. B* **64**, 363-410.

Young, A. R. (1990). Cumulative effects of ultraviolet radiation on the skin: cancer and photoaging. *Seminars in Dermatology* **9**, 25-31.

Yin, X. and Cook, R. D. (2005). Direction estimation in single-index regressions. *Biometrika* **92**, 371-384.

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97**, 1042-1054.

chool of Mathematics, Statistics and Actuarial Science, The University of Kent, Canterbury, Kent, CT2 7NZ, UK.

E-mail: E.Kong@kent.ac.uk

Department of Statistics, Columbia House, London School of Economics, Houghton Street, London WC2A 2AE, UK.

E-mail: h.tong@lse.ac.uk

Department of Statistics and Applied Probability, National University of Singapore, Singapore.

E-mail: staxyc@nus.edu.sg

(Received May 2008; accepted February 2009)