Statistical Modelling of Nonlinear Long-Term Cumulative Effects

Efang Kong
EURANDOM, the Netherlands
Howell Tong
London School of Economics, UK
Yingcun Xia
National University of Singapore, 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, such as 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 in this paper a semi-parametric time series model, called the functional additive cumulative time series (FACTS) model, and investigate statistical properties. We develop an estimation procedure that combines the advantages of kernel smoothing and polynomial splines smoothing. As two case studies, we analyze the effect of air pollutants on respiratory diseases in Hong Kong and the human immunity against influenza in France. Based on the results, some of the important issues in epidemiology are addressed.

Keywords: 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 by cumulation. The cumulative effect has been noted to be the main cause of many diseases although 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). A sunburn develops when the amount of ultraviolet radiation exposure is greater than what can be protected against by the skin's melanin. Skin responds to cumulative sun exposure by thickening and hardening, resulting in leathery skin and wrinkles later in life and the risk of skin cancer. Cumulative effects are also observed in many other areas besides epidemiology. 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 of USA (*Report - Considering Cumulative Effects Under NEPA*; http://www.epa.gov), World Health Organization (WHO) and other similar organizations. Investigations on specific cumulative effects can be found in the existing literature. See, for example, Smith & 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 too late to act. In fact, most existing time series models focus on the effects of a few individual history data points, which 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. Fig. 1 presents several aspects of the data collected in Hong Kong from January 1, 1994 to December 31, 1998. The data includes a set of variables. To make our point, we for the moment merely consider the effect of the daily average level of nitrogen dioxide ($NO_{2,t}$, 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 Fig.1(a) and Fig.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 Fig.1(b) and Fig.1(d).

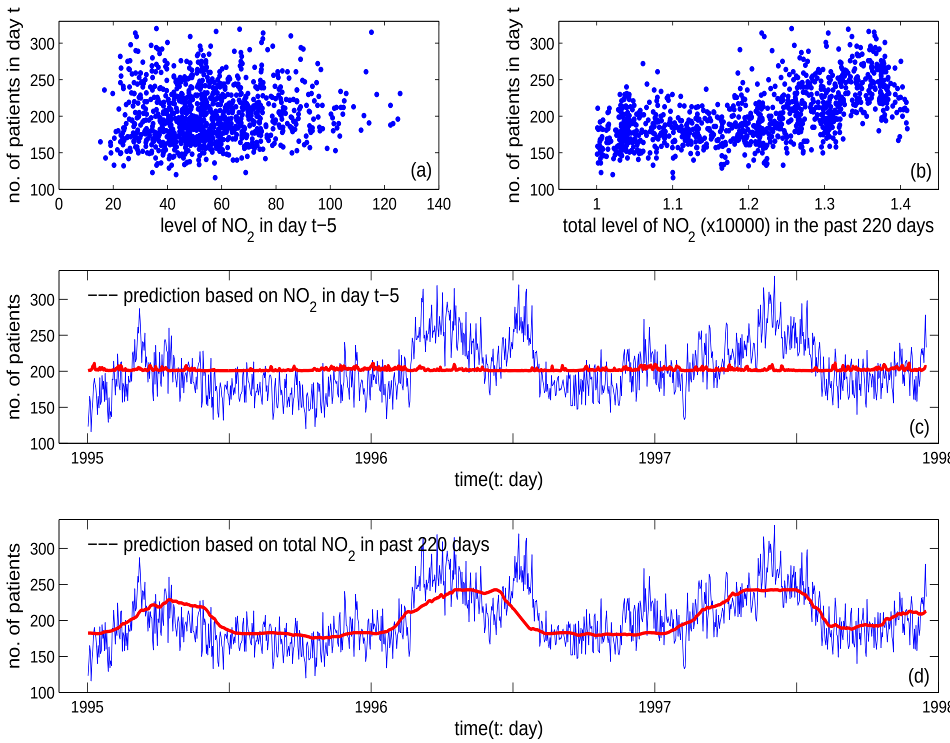

Figure 1: *The level of* NO_2 *and number of patients suffering from respiratory diseases in Hong Kong. Panel (a) is the plot of the number of patients on day* t *against the level of* NO_2 *on day* $t-5$*, the latter day having the largest correlation coefficient with the number of patients among all 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*

In this paper, we shall adopt a semi-parametric approach to analyze the cumulative effect, with the following technical considerations. (1) Unknown link functions. As suggested by Fig.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 will be used for the estimation. (2) The pattern of cumulation weight. In the above 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 w_τ need to be estimated subject to $\sum_{\tau=1}^{D} w_\tau = 1$. (3) Multivariate analysis. In the Hong Kong data example, it is very likely that besides NO_2, other pollutants and weather conditions also contribute to the variation in the number of hospital admissions. To incorporate all these factors, one option is to adopt an additive structure. See, e.g., Hastie & Tibshirani (1996) and Dominici et al (2002). (4) Monotonicity. As suggested by Fig.1(b), higher level of cumulative pollution should result in more hospital admissions. This motivates us to impose 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, the polynomial spline smoothing and the kernel smoothing appear to be dominant with respective advantages. For example, the polynomial spline smoothing is more convenient for incorporating global constraints on the functions, while the kernel smoothing is more convenient for local Taylor expansion and approximation. In this paper we shall use polynomial splines for function estimation and kernel smoothing for local approximation in estimating the cumulative weights. By doing so, the computations are simplified as a standard quadratic programming problem, for which there exist very efficient and fast algorithms.

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 semi-parametric procedure which combines polynomial splines smoothing 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 the 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 is essential for policy decision making relating to vaccination and disease control. See e.g. Ferguson et al (2003). Based on the weekly notified influenza cases in

France (www.sentiweb.org), a decreasing pattern of immunity is revealed.

2 A semi-parametric model for the cumulative effect

Suppose $Y_{t_i} \overset{def}{=} Y(t_i),\ i = 1, \cdots, n$, is the (discrete) response time series and $\{\mathbf{Z}(t), \mathbf{X}(t)\}, t \geq 0$, are multivariate (continuous) time series with $\mathbf{Z}(t) = (1, Z_1(t), \cdots, Z_p(t))^\top$ and $\mathbf{X}(t) = (X_1(t), \cdots, X_q(t))^\top$. An important model concerning the relationship between Y_{t_i} and $\{\mathbf{Z}(t), \mathbf{X}(t)\}$ is the semiparametric additive model

$$E\{Y(t_i)|Z(s), X(s),\ s \leq t_i\} = \beta^\top \mathbf{Z}(t_i) + g_1(X_1(t_i - \tau_1)) + \cdots + g_q(X_q(t_i - \tau_q)), \tag{1}$$

where β is an unknown $p-$dimensional parameter vector, $g_k, k = 1, \cdots, q$, are unknown link functions and $\tau_k, k = 1, ..., q$, are lags. See, e.g., Hastie & Tibishirani (1993), Liu and Stengos (1999) and Dominici et al (2002).

As we have noticed from the Hong Kong data, effect of pollution on any single day is not significant. However, persisting pollution over a relatively long period can explain much of the variation of 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 (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 (1) typically does not lead to a good fit; see Fig.5(a) in Section 6. A straightforward extension of model (1) by simply enlarging the number of additive components is infeasible, as the resulted model tends to be plagued by unstable estimation and difficult interpretation in practice.

In this paper, we propose to model the cumulative effect of a single covariate, $X_1(\tau)$ say , by its weighted integral over a finite interval, namely

$$\int_0^\Delta X_1(t - \tau)\theta(\tau)d\tau,$$

for some $\Delta > 0$, where $\theta(\tau) \geq 0$ is the weight function defined over $[0, \Delta]$. This, when incorporated with the additive structure, leads to the following model

$$Y(t_i) = \mathbf{Z}^\top(t)\beta^0 + \sum_{k=1}^{q} g_k\left(\int_0^\Delta X_k(t_i-\tau)\theta_k(\tau)d\tau\right) + \varepsilon(t_i),\ i=1,\cdots,n, \tag{2}$$

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

$$E\{g_k\left(\int_0^\Delta X_k(t-\tau)\theta_k(\tau)d\tau\right)\} = 0, \quad \int_0^\Delta \theta_k(\tau)d\tau = 1,\ k=1,\cdots,q, \tag{3}$$

for identification purposes. Alternative identification conditions can be imposed depending on the purpose of the modeling. We call model (2) the functional additive cumulative time series (FACTS) model. $g_k(.)$ is referred to as the effect function or the link function and $\theta_k(.)$ the weight function.

Based on the discussion in Section 1, if warranted by empirical evidence, monotonicity constraint could be imposed on either $\theta_k(.)$ or $g_k(.),\ k=1,\cdots,q$ or both. To obtain estimates of model (2) which are consistent with such a constraint is also an important feature of this paper.

Proposition 2.1 *Suppose $\mathbf{X}(t)$ is a multi-variate stationary process with a continuous joint probability density function. Every set of q continuous functions of t can be a sample path of $\mathbf{X}(t)$. Let $\tilde{\mathbf{Z}}(t) = E\{\mathbf{Z}(t)|\mathbf{X}(\tau), \tau \le t\}$. If $E[\{\mathbf{Z}(t)-\tilde{\mathbf{Z}}(t)\}\{\mathbf{Z}(t)-\tilde{\mathbf{Z}}(t)\}^\top]$ is invertible, then model (2) is identifiable. In other words, if there exists another set of parameters $\tilde{\beta}^0$ and functions $\tilde{g}_k(.), \tilde{\theta}_k(.)$ such that (2) and (3) hold, then*

$$\tilde{\beta}^0 \equiv \beta^0, \quad \tilde{g}_k(.) \equiv g_k(.), \quad \tilde{\theta}_k(.) \equiv \theta_k(.),\ k=1,..,q.$$

The FACTS model is closely linked with functional regression models. See for example Ramsay & Silverman (pp. 88, 1997) and James & Silverman (pp.567, 2005). If $X_k(t)$ has a step sample path, then the integration component of model (2) reduces to a summation, leading to a discretized version of the FACTS model, i.e.

$$Y(t) = \mathbf{Z}^\top(t)\beta^0 + \sum_{k=1}^{q} g_k\Big(\sum_{\ell=1}^{D} X_k(t-\ell)\theta_k(\ell)\Big) + \varepsilon_t, \tag{4}$$

where $\theta_k(\ell), \ell = 1, ..., D, k = 1, ..., p$ 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 (1996) and Yu & Ruppert (2002).

3 Estimation of FACTS model

Similar to Yu & Ruppert (2002), we adopt a spline smoothing approach to estimate both the unknown link function $g_k(.)$ and the weight function θ_k. Suppose for each $k = 1, \cdots, q$, there exist two vectors η_k^0 and γ_k^0, such that we have approximately

$$g_k(\nu) = \mathbf{A}(\nu)^\top \eta_k^0, \quad \theta_k(\tau) = \mathbf{B}(\tau)^\top \gamma_k^0,$$

where $\mathbf{A}(\nu)$ and $\mathbf{B}(\tau)$ are two finite $r-$dimensional bases function, e.g. cubic splines. Define $b = (b_1, ..., b_r)^\top = \int_0^\Delta \mathbf{B}(\tau) d\tau$, an $r-$dimensional column vector with its first component b_1 nonzero, which can always be realized by rearranging the order of the r basis functions. Then the second equation in (3) can be approximately rewritten as

$$b^\top \gamma_k^0 = 1, \quad 1 \le k \le q. \tag{5}$$

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

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

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

$$m(\mathbf{v}_i; \xi) = \mathbf{Z}^\top(t_i)\beta + \sum_{k=1}^q \eta_k^\top \mathbf{A}(\gamma_k^\top X_{t_i}^k). \tag{6}$$

Existing methods, such as the penalized spline method in Yu and Ruppert (2002), can be used to estimate $\xi^0 = (\beta^{0^\top}, \eta_1^{0^\top}, \cdots, \eta_q^{0^\top}, \gamma_1^{0^\top}, \cdots, \gamma_q^{0^\top})^\top$, 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 $\{\mathbf{X}(\tilde{t}_j), j = 1, \cdots, \}$ is the observed discrete time series of $\{\mathbf{X}(t)\}$. If as specified in (A2) in the Appendix, $\max_{j \ge 1} |\tilde{t}_j - \tilde{t}_{j+1}|$ is sufficiently small relative to n, the total number of observations on $Y(t)$, then based on the continuous property of $\mathbf{B}(.)$ and the sample path of $X_k(.)$, 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 (6), leads to the approximated regression mean function

$$m_n(\mathbf{v}_i; \xi) = \mathbf{Z}^\top(t_i)\beta + \sum_{k=1}^{q} \eta_k^\top \mathbf{A}(\gamma_k^\top X_{t_i}^{n,k}). \tag{7}$$

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

$$Q_{n,\lambda}(\xi) \stackrel{def}{=} n^{-1} \sum_{i=D+1}^{n} \Big\{ Y_{t_i} - m_n(\mathbf{v}_i; \xi) \Big\}^2 + \lambda_n \delta^\top \Sigma\, \delta, \tag{8}$$

where $D = \min\{i | t_i - \tilde{t}_1 \geq \Delta\}$, $\delta = (\eta_1^\top, \cdots, \eta_q^\top, \gamma_1^\top, \cdots, \gamma_q^\top)^\top$, 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 (5) on γ_k makes reparameterization necessary in proving the consistency and asymptotic normality of the PLSE. Let $\tilde{b} = (b_2, \cdots, b_r)^\top$ and $\phi_k = (\phi_{k,1}, \cdots, \phi_{k,r-1})^\top$. Define

$$\gamma_k(\phi_k) = (b_1^{-1}(1 - \tilde{b}^\top \phi_k), \phi_{k,1}, \cdots, \phi_{k,r-1})^\top. \tag{9}$$

Let $\phi_k^0 = (\gamma_{k,2}^0, ..., \gamma_{k,r}^0)^\top$, which is a subvector of γ_k^0. By (5), 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

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

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

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

where ξ_ϕ is $q-$dimension lower than ξ_γ. Consequently, the regression mean function (6) and its approximation (7) can be reparameterized as

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

To state the asymptotic results, let

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

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

$$\mathbf{J}(\phi) = \xi_\gamma^{(1)}(\xi_\phi) = \begin{bmatrix} \mathbf{I}_{qr+p} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \gamma_1^{(1)}(\phi_1) & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \gamma_q^{(1)}(\phi_q) \end{bmatrix}, \tag{10}$$

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

$$\Omega(\xi) = \lim_n \frac{1}{n} \sum_{i=1}^{n} m^{(1)}(\mathbf{v}_i;\xi_\phi) m^{(1)}(\mathbf{v}_i;\xi_\phi)^\top. \tag{11}$$

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

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

Remark Note that the above asymptotic distribution is obtained without monotonicity constraint imposed on either the estimated link function or the weight function

$$\hat{g}_k(\nu) = \mathbf{A}(\nu)^\top \hat{\eta}_k^0, \quad \hat{\theta}_k(\tau) = \mathbf{B}(\tau)^\top \hat{\gamma}_k^0,\ k = 1, \cdots, q.$$

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 parameter ξ^0 through minimizing (8) with respect to ξ with λ_n fixed as 0. There are two reasons for such a choice of λ_n. Firstly, as indicated in Theorem 3.1, for the estimate to be asymptotically normal, we need $\lambda_n = o(n^{-1/2})$. Secondly in a Monte Carlo study by Yu and Ruppert (2002), they found that the confidence bands using $\lambda = 0$ were often closer to the Monte Carlo bands than that using the true value of λ. Moreover, based on both simulation study and real data analysis in this paper, such prefixed value of λ_n has not caused any serious over-fitting problem.

The implementation of minimizing (8) is not easy, 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 is not the converging point of 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 $k = 1, 2, ..., q$. Thus, to estimate the FACTS model we need to estimate β and $\eta_k, k = 1, ..., q$. The minimization in (8) is a simple linear regression estimation in solving β and $\eta_k, k = 1, ..., q$. Denote them by $\beta^{(0)}$ and $\eta_k^{(0)}, k = 1, ..., q$, respectively. Thus, g_k is estimated by $g_k^{(0)}(.) = \mathbf{A}(.)^\top \eta_k^{(0)}$. After this initial step, we can follow the idea of the back-fitting algorithm to estimate model (2). See Hastie & 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

$$y_{t_i}^{n,k} = g_k(\gamma_k^\top X_{t_i}^{n,k}) + \epsilon_{t_i}^{n,k}, \tag{12}$$

where $g_k(.) = \eta_k^\top \mathbf{A}(.)$. 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 (12). See for example Härdle & Stoker (1989), Yu & Ruppert (2002), Yin & Cook (2004) 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 & Tong (1993), 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

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

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 $x = \tilde{X}_j, j = 1, \cdots, n$. Thus, our estimation procedure is to minimize

$$\sum_{j=1}^n \hat{\sigma}^2(\tilde{X}_j | \tilde{\gamma}, \tilde{\eta}) = \sum_{j=1}^n \sum_{i=1}^n \Big[\tilde{y}_i - \tilde{\eta}^\top \{\mathbf{A}(\tilde{\gamma}^\top \tilde{X}_j) + \mathbf{A}'(\tilde{\gamma}^\top \tilde{X}_j)\tilde{\gamma}^\top \tilde{X}_{ij}\}\Big]^2 w_{ij} \tag{13}$$

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 minimization (13) easily as follows. Note that with fixed w_{ij}, the minimization of (13) can be decomposed into two separate quadratic programming problems with unknown parameters $\tilde{\eta}$ and $\tilde{\gamma}$ respectively. We can solve (13) 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 (13) is equivalent to

$$\min_{\tilde{\eta}} \ \tilde{\eta}^\top D_n \tilde{\eta} - 2C_n^\top \tilde{\eta}, \tag{14}$$

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 (13) with respect to $\tilde{\gamma}$ is equivalent to

$$\min_{\tilde{\gamma}} \ \tilde{\gamma}^\top D_n' \tilde{\gamma} - 2{C_n'}^\top \tilde{\gamma}, \tag{15}$$

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 (15) is $\tilde{\gamma}^{(\ell+1)} = \{D_n'\}^{-1} C_n'$. Set $\ell = \ell + 1$. Repeat (14) and (15) until convergence. Denote the final values by $\tilde{\tilde{\gamma}}$ and $\tilde{\tilde{\eta}}$ respectively. Finally, we update $g_k^{(L)}(.)$ by $g_k^{(L+1)}(.) \stackrel{def}{=} \mathbf{A}(.)^\top \tilde{\tilde{\eta}}$ 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 link function or the weight function, monotone estimate of concerned function 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 (13) is realized through solving the two quadratic programming problems below alternately. With initial value $\tilde{\gamma}^{(\ell)} = \gamma_k^{(L)}$ and $\ell = 0$, we solve

$$\min_{\tilde{\eta}} \ \tilde{\eta}^\top D_n \tilde{\eta} - 2C_n^\top \tilde{\eta}, \quad \text{subject to}: \ \tilde{\eta}_2, ..., \tilde{\eta}_r \geq 0, \tag{16}$$

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

$$\min_{\tilde{\gamma}} \ \tilde{\gamma}^\top D_n' \tilde{\gamma} - 2{C_n'}^\top \tilde{\gamma}, \quad \text{subject to}: \quad \tilde{\gamma} \geq 0, \tag{17}$$

and denote the solution to (17) by $\tilde{\gamma}^{(\ell+1)}$. Set $\ell = \ell + 1$. Repeat (16) and (17) until convergence. Although we do not have a closed form for the solutions of (16) and (17), they are typically quadratic programming problems. There are many efficient algorithms. See for example Nocedal & 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, namely the number of the knots of the spline base function and the bandwidth h, need to be selected. As we will 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 et al (1995). For a pre-specified number of knots, the knots are placed at equally spaced sample quantiles of the predictors $(\tilde{\gamma}^{(\ell)})^\top \tilde{X}_i, i = 1, ..., n$. We can estimate the link function $g_k(.)$ by $\mathbf{A}(.)^\top \hat{\eta}$, where $\hat{\eta} = (\mathcal{A}_n^\top \mathcal{A}_n)^{-1} \mathcal{A}_n^\top Y$ with $\mathcal{A}_n = (\mathbf{A}(\tilde{X}_1^\top \tilde{\gamma}^{(\ell)}), ..., \mathbf{A}(\tilde{X}_n^\top \tilde{\gamma}^{(\ell)}))^\top$. The fitted value of the response at the n points are $\hat{Y} = \mathcal{A}_n(\mathcal{A}_n^\top \mathcal{A}_n)^{-1}\mathcal{A}_n^\top Y$. Following Craven & 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 & 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 et al (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 computation.

4 Statistical simulations of finite samples

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

Consider the following model

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

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 $P(Z_{kt} = 1) = P(Z_{kt} = 0) = 0.5,\ k = 1, 2, 3$, $X_1(t) = \sin(t) + 1, X_2(t) = \sin(3t) + 1$, 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 $\tau \geq 0$. Monotone decreasing property is imposed upon estimators of the weight functions $\theta_1(\tau)$ and $\theta_2(\tau)$, while Monotone increasing property is imposed upon the link function $g_1(v)$.

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

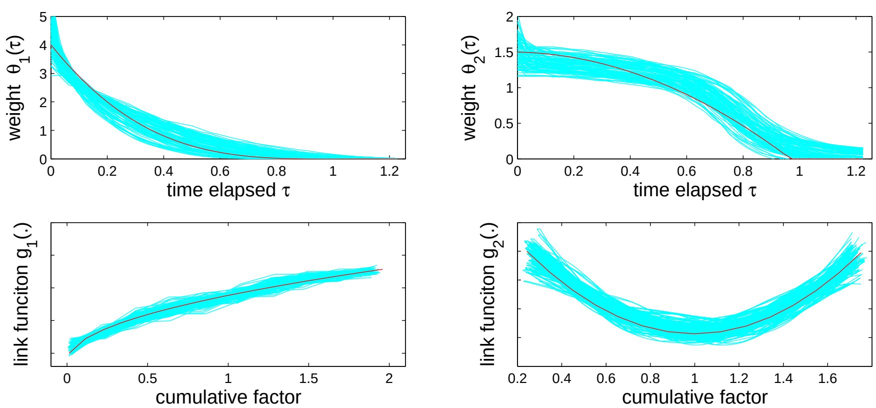

Figure 2: *Results for Example 4. The black curve in each panel is the true link functions g_k or the true weight functions θ_k; the cyan/grey curves are the corresponding estimated functions.*

Example 4.1 In this example, $\mathbf{X}(t)$ has step sample paths with equal steps, and all the steps are observed once. The integration in the FACTS model can be written as summations. Thus, the model can be simplified as

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

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 $\theta_k(\tau)$ are given by the black curves as shown in the first 3 panels of Fig.3. Covariates $Z_{1,t_i}, Z_{2,t_i}, i = 1, \cdots, n \overset{IID}{\sim} P(Z_{1,t_i} = 1) = P(Z_{1,t_i} = 0) = 0.5,\ X_{1,t_i} = 0.8X_{1,t_{i-1}} + e_{1,t_i},\quad X_{2,t_i} = 0.6X_{2,t_{i-1}} + 0.3X_{2,t_{i-2}} + e_{2,t_i}$, $X_{3,t_i} = -0.5X_{3,t_{i-1}} + e_{3,t_i}$ where $\{\varepsilon_{t_i}\}$, $\{e_{1,t_i}\}$, $\{e_{2,t_i}\}$ and $\{e_{3,t_i}\}$ are IID N(0,1).

There are 2 unknown parameters and 6 unknown functions. With sample size $n = 500$ and $D = 100$, it is obviously difficult to obtain efficient estimates unless useful prior knowledge is available. However, if we were informed that $g_3(.)$ is monotone increasing, and all the weight functions are monotone decreasing, we could estimate the functions with unexpected degree of accuracy; as showed by Fig. 3. As for parameters, the mean and standard deviation of estimated β are respectively $(1.0020, -1.0016)^\top$ and $(0.0324, 0.0300)^\top$.

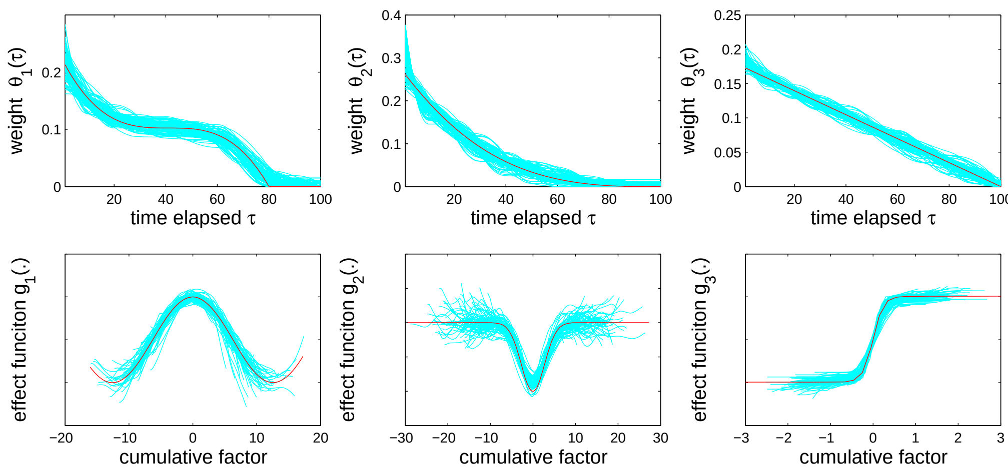

Figure 3: *Calculation results for Example 4.1. The black curve in each panel is the true link function g_k or the true weight parameters $\theta_k, k = 1, 2, 3$; the cyan/grey curves are the corresponding estimated functions.*

5 Effects of air pollution on the respiratory diseases

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

$$\begin{aligned}\log(y_{t_i}) \quad = \quad & \sum_{d=1}^{7} \beta_d D_{t_i,d} + g_1(\sum_{\tau=0}^{D} \theta_{1,\tau} N_{t_{i-\tau}}) + g_2(\sum_{\tau=0}^{D} \theta_{2,\tau} S_{t_{i-\tau}}) + g_3(\sum_{\tau=0}^{D} \theta_{3,\tau} P_{t_{i-\tau}}) \\ & + g_4(\sum_{\tau=0}^{D} \theta_{4,\tau} O_{t_{i-\tau}}) + g_5(\sum_{\tau=0}^{D} \theta_{5,\tau} T_{t_{i-\tau}}) + g_6(\sum_{\tau=0}^{D} \theta_{6,\tau} H_{t_{i-\tau}}) + \varepsilon_t, \qquad (18)\end{aligned}$$

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; $D_{d,t_i}, d = 1, ..., 7$, 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 $i \geq \tau > 0$.

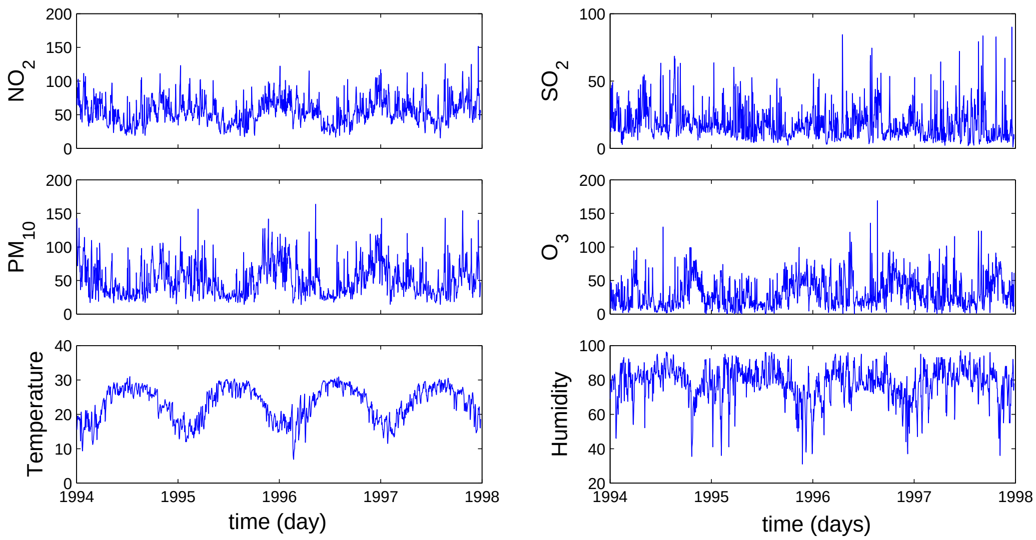

Figure 4: *The observed time series of average levels of* SO_2 *(in ppb, parts per billion),* NO_2 *(in ppb),* PM_{10} *(in* $\mu g/m^3$*), ozone (in ppb), average of temperature (in* ^oC*) and average of humidity (in %), in Hong Kong.*

With $D = 300$, FACTS model (18) gives fitted values as shown in Fig.5(b). As a comparison, the fitted values of the additive model with lags that give the best fit are shown in Fig.5(a). These figures suggest that FACTS model (18) 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 with $D > 300$. We found that the estimated weights are quite stable in that they tend to be practically zero after some specific lag. The estimation results are shown in Fig.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 are not serious sufferers. At the

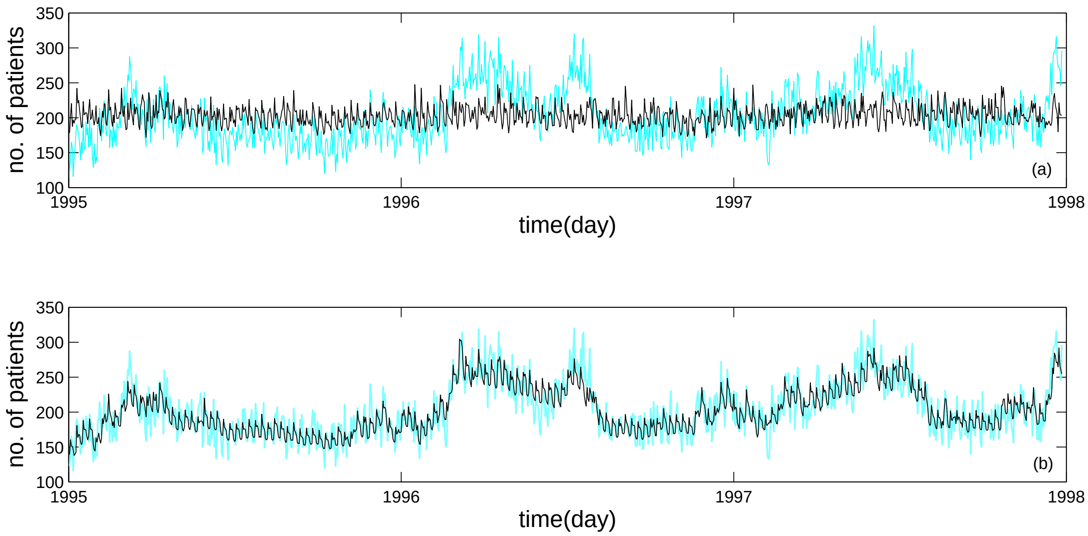

Figure 5: *Fitted results of the additive model and FACTS model. In panel (a), the blue line is the observed numbers of hospital admissions and the black line is the fitted numbers by the additive model. In panel (b), the blue line is the observed number of hospital admissions and the black line the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

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 $(\beta_1, \cdots, \beta_7)$ in model (18) are 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, indicating 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 due to human behavior rather than the pollutants as noticed in Forster & 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 1990's to reduce the level of SO_2;

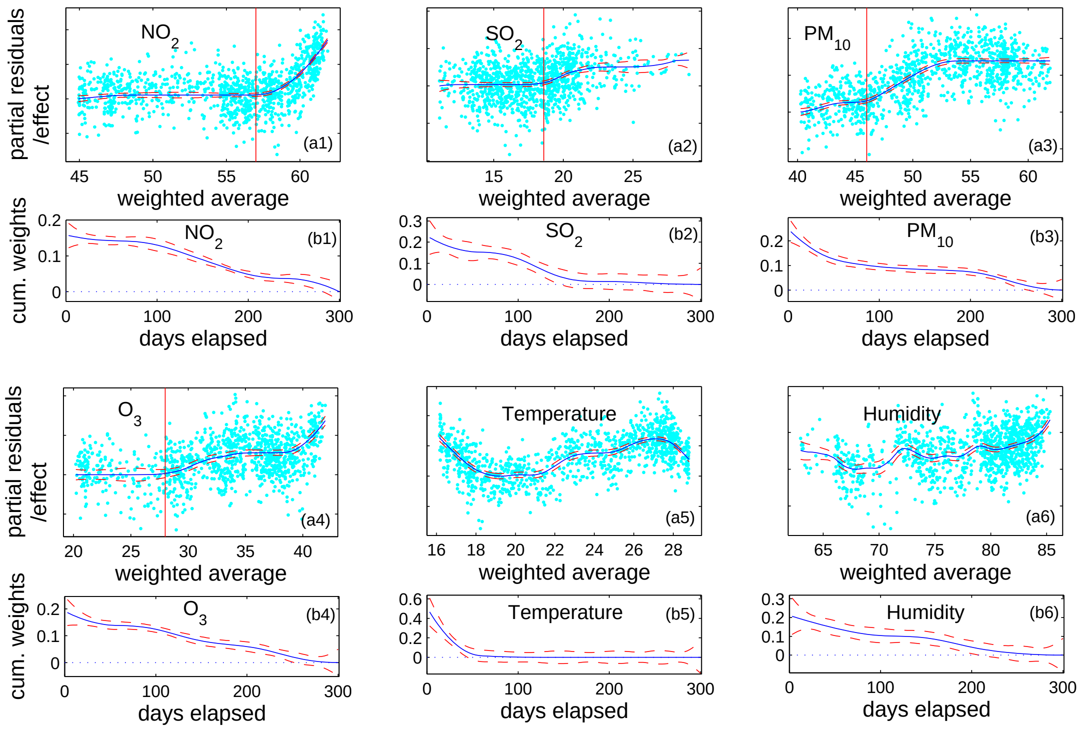

Figure 6: *Calculation results for the pollution and respiratory diseases. (a1)-(a6) are the estimated cumulative effect functions (the central line) and corresponding 95% pointwise confidence intervals (upper and lower lines) for the pollutants and weather conditions; (b1)-(b6) are correspondingly the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

see Hedley et al (2002).

Based on the estimated effect functions $g_k(.)$, there are thresholds at which the effects start to increase; see Fig.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 (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.

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 $http://www.epa.gov/air/criteria.html$

For the weather conditions, both an unusually cold season and an unusually hot season can aggravate the diseases; see Fig.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 et al (1998) and McGeehin & Mirabelli (2001). The wetter weather causes more hospital admissions of respiratory diseases; see Fig.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 a very important infectious disease arising as a series of seasonal epidemics. A weekly notified time series of influenza-like cases in France is shown in Fig.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 *et al.* (1998). Pease (1987) conjectured that immunity to influenza will decay linearly with time elapsed, while Couch & Kasel (1983) argued that immunity lasts for more than 4 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 will focus on FACTS modeling of the decaying pattern of human immunity against this particular disease.

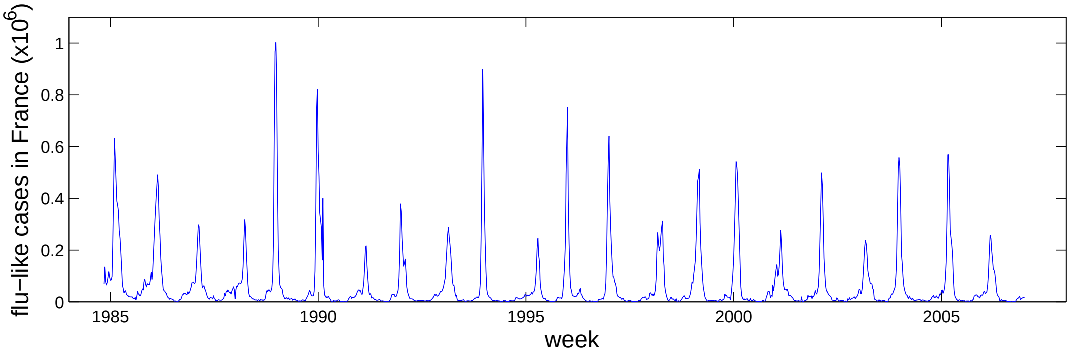

Figure 7: *Weekly notified cases of flu-like diseases in France*

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 make the following homogeneity assumption. The value of $p(\tau)$ 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 $\kappa(\tau) = 1 - p(\tau)/p_0$ *relative immunity*. It is easy to see that $0 \leq \kappa(\tau) \leq 1$ 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 aforementioned 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 following model

$$\frac{dy_t}{dt} = \beta_t y_t^\alpha S_t^\gamma; \tag{19}$$

see Liu et al (1987), Anderson & May (1991) and Finkenstädt & 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 $\nu(.)$ and $\mu(.)$ are unknown link functions. See Xia et al (2005). The function $\mu(.)$ 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 & Grenfell (2000) and Xia et al (2005), model (19) can be approximated by

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

Considering 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

$$\log(y_{t_i}) = \sum_{i=1}^{52} \varrho_k D_{k,t_i} + \alpha \log(y_{t_{i-1}}) + \tilde{\mu}(\sum_{\tau=0}^{D} y_{t_{i-\tau}} \kappa(t_i - t_{i-\tau})) + \varepsilon_{t_i}, \tag{20}$$

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

Results based on the weekly notified influenza cases in France are shown in Fig. 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 (19). (II) The estimated seasonal infection rates as shown in Fig.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 *et al.* 1998). (III) The epidemics in France has a decay function of immunity as shown in Fig.8(d). It is noteworthy that the decay pattern of immunity is different from the

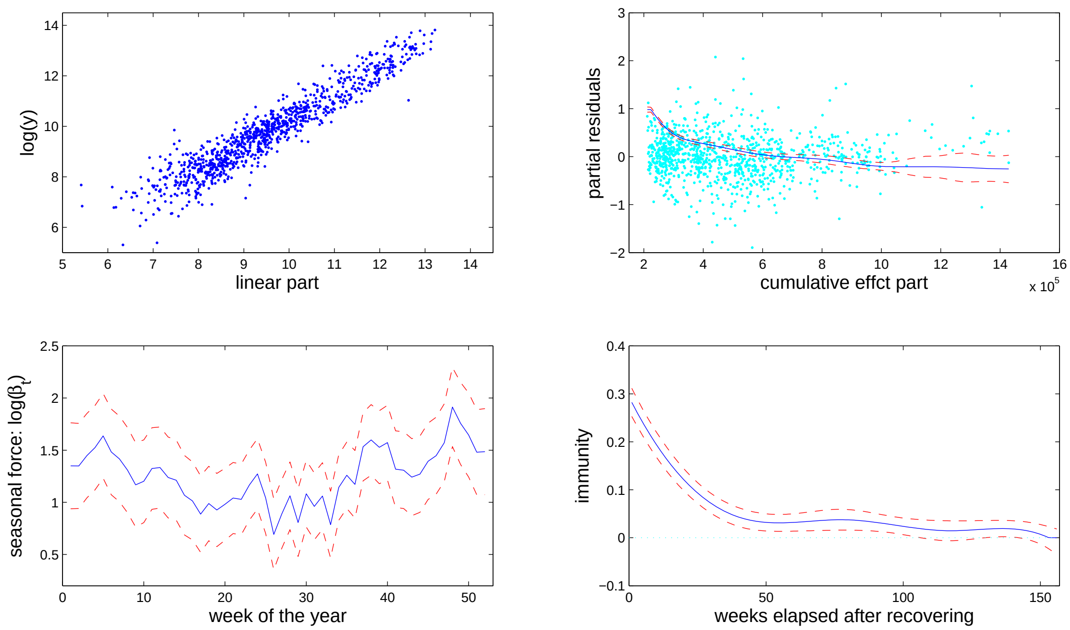

Figure 8: *Results for the influenza data in France. Panel (a) is* $\log(y_{t_i})$ *plotted against the linear part in model (20). In panel (b), the dots are the corresponding partial residuals after removing the linear part in model (20). The x-axis is the expected immunized number* $\sum_{\tau=0}^{D} y_{t_{i-\tau}} \kappa(t_i - t_{i-\tau})$*. The solid line is the estimated link function, the dash lines are the 95% pointwise confidence intervals for* $\tilde{\mu}(.)$*. 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.*

conjecture of Pease (1987). In the first few months, the recovered hosts have high level of immunity. After that, the immunity decreases quite quickly. 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 2 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.* (2003). Our result is also 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 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 & Kasel (1983) and Murphy & Clements (1989).

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 $m(t)$ is continuous deterministic function of t and $B(t)$ 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 certain limit function uniformly in $\xi, \xi^* \in \boldsymbol{\Xi}$, and

$$Q(\xi) = \lim_n \frac{1}{n} \sum_{i=1}^n \{m(\mathbf{v}_i; \xi) - m(\mathbf{v}_i; \xi^0)\}^2 \tag{21}$$

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

$$\Omega(\xi) := \lim_n \frac{1}{n} \sum_{i=1}^n m^{(1)}(\mathbf{v}_i; \xi) m^{(1)}(\mathbf{v}_i; \xi)^\top \tag{22}$$

and $n^{-1} \sum_{i=1}^n \frac{\partial^2 m(\mathbf{v}_i; \xi)}{\partial \xi_{j_1} \partial \xi_{j_2}}$, $j_1, j_2 = 1, \cdots, 2qr + p$, 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 et al (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 $i = 1, \cdots, n$ (Lemma 6.1). Such requirement is commonly made in dealing with continuous-time model but with only discretized data available, especially in finance and biology, see e.g. Fan & Zhang (2003) and Fan & Jiang (2005). Assumptions (A3)-(A7) are similar to those in Yu & 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 $\tilde{\mathbf{Z}}(t) = E\{\mathbf{Z}^\top(t)|\mathbf{X}(s),\, s \le t\}$. We have

$$E\{Y_t|\mathbf{X}(s), s \le t\} \quad = \quad \tilde{\mathbf{Z}}^\top(t)\beta^0 + \sum_{k=1}^{q} g_k\left(\int_0^\Delta X_k(t-\tau)\theta_k(\tau)dt\right). \tag{23}$$

Subtracting (2) by (23), we have

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

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

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

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

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

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

$$\lim_{h\to 0} \Delta_2(\tilde{x}_2^a(t), x_2(t)) \quad = \quad \theta_2(0)g_2'(\int_0^\Delta x_2(t-\tau)\theta_2(\tau)d\tau). \tag{24}$$

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

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

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 $m(.)$, which together with $\int \theta_2(v)dv = 1$ establishes the identifiability of the function $\theta_2(.)$. Similarly, the other weight functions are identifiable.

By (24) and the identifiability of $\theta_2(.)$, the derivative of $g_2(.)$ is identifiable. By the first assumption in (3), $g_2(.)$ is identifiable. Similar arguments can be applied to $g_1(.)$. □

Lemma 6.1 *Under (A1) and (A2), we have*

$$X_{t_i}^{n,k} - X_{t_i}^k = O((n/\log n)^{-1/2}) \quad a.s. \tag{25}$$

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

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

$$|B(t) - B(t+h)| \le c|h/\log h|^{-1/2} \text{ for any } t \in [0, T) \text{ and } h < \delta$$

Substitute n^{-1} for h and we have (25) as $m(t)$ is uniform 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 6.2 *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* $j_1, j_2 = 1, \cdots, 2qr + p$.

Proof The result follows directly from Lemma 6.1 and an application of the Cauchy-Schwartz Inequality.

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

$$
\begin{aligned}
Q_{n,\lambda}(\xi) &= \frac{1}{n}\sum_{i=D+1}^{n}\Big\{Y_{t_i} - m_n(\mathbf{v}_i;\xi)\Big\}^2 + \lambda_n\delta^\top\Sigma\delta \\
&= \frac{1}{n}\sum_{i=D+1}^{n}\Big\{Y_{t_i} - m(\mathbf{v}_i;\xi^0) + m(\mathbf{v}_i;\xi^0) - m(\mathbf{v}_i;\xi) + m(\mathbf{v}_i;\xi) - m_n(\mathbf{v}_i;\xi)\Big\}^2 + \lambda_n\delta^\top\Sigma\delta \\
&= \frac{1}{n}\sum_{i=D+1}^{n}\varepsilon_{t_i}^2 + \frac{2}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi^0) - m(\mathbf{v}_i;\xi)\}\varepsilon_{t_i} + \frac{1}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi^0) - m(\mathbf{v}_i;\xi)\}^2 \\
&\quad + \frac{2}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi) - m_n(\mathbf{v}_i;\xi)\}\varepsilon_{t_i} + \frac{1}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi) - m_n(\mathbf{v}_i;\xi)\}^2 \\
&\quad + \frac{2}{n}\sum_{i=D+1}^{n}\{m(\mathbf{v}_i;\xi) - m_n(\mathbf{v}_i;\xi)\}\{m(\mathbf{v}_i;\xi^0) - m(\mathbf{v}_i;\xi)\} + \lambda_n\delta^\top\Sigma\delta \\
&= \frac{1}{n}\sum_{i=D+1}^{n}\varepsilon_{t_i}^2 + T_1 + T_2 + T_3 + T_4 + T_5 + T_6. \qquad (26)
\end{aligned}
$$

All the 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 6.1, 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,

$$Q_{n,\lambda}(\xi) \to Q(\xi) + \sigma^2 \text{ uniformly for all } \xi \in \boldsymbol{\Xi}. \qquad (27)$$

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

As $\hat{\xi}_{n,\lambda}$ is 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 the following two results:

$$\sqrt{n}\frac{\partial Q_{n,\lambda}}{\partial \xi}\Big|_{\xi^0} \xrightarrow{D} N(0, 4\sigma^2\Omega(\xi^0)) \tag{28}$$

and

$$\frac{\partial^2 Q_{n,\lambda}}{\partial \xi \partial \xi^\top}\Big|_{\tilde{\xi}} \xrightarrow{P} 2\Omega(\xi^0). \tag{29}$$

To prove (28), notice that

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

and

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

As $\lambda_n = o(n^{-1/2})$, the last term can be ignored. Under (A2) and (A3), it follows

$$\mathbf{A}(\eta_k^{0\top}X_{t_i}^k) - \mathbf{A}(\eta_k^{0\top}X_{t_i}^{n,k}) = o(n^{-1/2}), \text{ uniformly in } i \text{ and } k,$$

which together with Lemma 6.2 leads to

$$\frac{2}{\sqrt{n}}\sum_{i=D+1}^{n}\sum_{k=1}^{q}\eta_k^{0\top}[\mathbf{A}(\eta_k^{0\top}X_{t_i}^k) - \mathbf{A}(\eta_k^{0\top}X_{t_i}^{n,k})] m_n^{(1)}(\mathbf{v}_i;\xi^0) \to 0. \tag{30}$$

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

$$\frac{1}{\sqrt{n}}\sum_{i=D+1}^{n}\varepsilon_{t_i} m_n^{(1)}(\mathbf{v}_i;\xi^0) \xrightarrow{D} N(0, \sigma^2\Omega(\xi^0)). \tag{31}$$

Combination of (30) and (31) yields (28).

For (29), we have

$$\begin{aligned}\frac{\partial^2 Q_{n,\lambda}}{\partial \xi \partial \xi^\top}|_{\tilde{\xi}} =& \frac{2}{n}\sum_{i=D+1}^{n} m_n^{(1)}(\mathbf{v}_i;\xi) m_n^{(1)}(\mathbf{v}_i;\xi)^\top|_{\tilde{\xi}} \\ & -\frac{2}{n}\sum_{i=D+1}^{n}\Big\{Y_{t_i} - m_n(\mathbf{v}_i;\xi)\Big\}\frac{\partial^2 m_n(\mathbf{v}_i;\xi)}{\partial \xi_{j_1}\partial \xi_{j_2}}|_{\tilde{\xi}} + 2\lambda_n\Sigma \\ =& 2\Omega_n(\tilde{\xi}) - \frac{2}{n}\sum_{i=D+1}^{n}\Big\{Y_{t_i} - m(\mathbf{v}_i;\tilde{\xi})\Big\}\frac{\partial^2 m_n(\mathbf{v}_i;\xi)}{\partial \xi_{j_1}\partial \xi_{j_2}}|_{\tilde{\xi}} \\ & -\frac{2}{n}\sum_{i=D+1}^{n}\Big\{m(\mathbf{v}_i;\tilde{\xi}) - m_n(\mathbf{v}_i;\tilde{\xi})\Big\}\frac{\partial^2 m_n(\mathbf{v}_i;\xi)}{\partial \xi_{j_1}\partial \xi_{j_2}}|_{\tilde{\xi}} + 2\lambda_n\Sigma.\end{aligned}$$

The first term on the right hand side goes to $2\Omega(\xi^0)$ by Lemma 6.2, (A7) and the fact that $\tilde{\xi}\to\xi^0$ almost surely. The second term is $o_p(1)$ as argued in Yu and Ruppert (2002). By Lemma 6.1 and (A3), the last two term are also to $o_p(1)$. This completes the proof of (29). $\square$

References

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

American Thoracic Society (2000) What constitutes an adverse health effect of air pollution? *Amer. J. Resp. Crit. Care Medi.*, **161**, 665-673.

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

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

Cai, Z., Fan, J. and Yao, Q. (2000). 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. Ass.,* **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. *Am. J. Epidemiol.*, **156**, 193-203.

Dubé, M, et al (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., Fan, Y. and Jiang, J. (2007) Dynamic integration of time- and state-domain methods for volatility estimation. *J. Amer. Statist. Ass.,* (to appear).

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. Ass.*, **98**, 118-134.

Fan, J. and Zhang, W.Y. (1999) Statistical estimation in varying coefficient models. *Ann. of 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. *Proceedings of the National Academy of Sciences*, **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. Ass.*, **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. Ass.*, **100**, 565-575.

Jennrich, R. I. (1969) Asymptotic properties of nonlinear least estimators, *Annals of Mathematical Statistics*, **40**, 633-643.

Liew, C. K. (1976) Inequality constrained least-squares estimation. *J. Amer. Statist. Ass.*, **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 Re- gressions: A Semiparametric Approach. *Journal of Applied 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. and 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. *Statistical Science*, **3**, 425-461.

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

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. Ass.*, **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

U.S. Environmental Protection Agency, Report No. 40 CFR Part 50 (1996, available at http://es.epa.gov/ncer/publications/terms/tscs99.html).

World Health organization, Reports on a WHO/HEI working group, Bonn, Germany, (2003, 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. and Tong, H. (2006) Cumulative effects of air pollution on public health. *Statistics in Medicine*, **25**, 3548-3559.

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. & 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. Ass.*, **97**, 1042-1054.