Bayesian adaptive distributed lag models

Alastair Rushworth

January 23, 2018

Abstract

Distributed lag models (DLMs) express the cumulative and delayed dependence between pairs of time-indexed response and explanatory variables. In practical application, users of DLMs examine the estimated influence of a series of lagged covariates to assess patterns of dependence. Much recent methodological work has sought to develop flexible parameterisations for smoothing the associated lag parameters that avoid overfitting. However, this paper finds that some widely-used DLMs introduce bias in the estimated lag influence, and are sensitive to the maximum lag which is typically chosen in advance of model fitting. Simulations show that bias and misspecification are dramatically reduced by generalising the smoothing model to allow varying penalisation of the lag influence estimates. The resulting model is shown to have substantially fewer effective parameters and lower bias, providing the user with confidence that the estimates are robust to prior model choice.

Keywords: Distributed lag model; penalized spline; adaptive smoothing

1 Introduction

Distributed lag models (DLMs) find application whenever the influence of a time-indexed independent variable is delayed and spread over time. Although DLMs are widely applicable, their development was originally motivated by problems in econometrics (Nerlove et al. (1958), Almon (1965), Zellner and Geisel (1970), Haugh and Box (1977)) and have more recently experienced a surge in popularity in environmental epidemiology (zanobetti, peng, welty, Gasparrini et al. (2010)). Crucially, DLMs enable direct interpretation of the influence of a temporal exposure ensemble, which is particularly useful for characterising the total health impact of persistent environmental exposures such as air pollution or temperature (Gasparrini et al. (2010), Wyzga (1978)). Lag models can help to identify subtle types of time-dependence such as 'mortality displacement', which occurs when exposure related mortality diminishes a vulnerable subpopulation, resulting in lower mortality in subsequent time intervals. Mortality displacement has been widely documented (Schwartz (2000), Braga et al. (2001)) and is conspicuous where lag influence is estimated to have a protective effect at low or moderate lags for what are otherwise harmful exposures (Zanobetti et al. (2000)). Correctly identifying these effects depends on obtaining unbiased estimates of the underlying lag influence, and it is therefore essential that the DLM is correctly specified. Since no previous study has formally quantified the estimation cost of model misspecification, this paper provides a comparison of various existing techniques.

If a time-varying explanatory variables X_t has some influence on a response variable Y_t over a set of lags $0, \ldots, p$, then a simple DLM can be expressed as the regression model $Y_t = \sum_{i=0}^{p} X_{t-i}\beta_i + \epsilon$, where ϵ is an independent error process and the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ are referred to as the 'lag curve' or lag function, as a function of i. $\boldsymbol{\beta}$ can-

not be reliably estimated using ordinary generalised linear models when X_t and X_{t-1} are correlated, and regularisation using shape constraints or smoothness restrictions over the lag parameters $\boldsymbol{\beta}$ are frequently used. Recent studies have implemented various form of flexible parameterisation using polynomial splines and smoothness penalties, for example by combining uniformly spaced B-splines with ridge penalisation (Zanobetti et al. (2000)), logarithmically spaced B-splines (Gasparrini et al. (2010)), penalised B-splines (Rushworth et al. (2013), Obermeier et al. (2015)), and Bayesian penalised splines with varying ridge penalties (Muggeo (2008)). Spline basis representations ensure a minimum level of smoothness, and offer computational relief when there are fewer knots than lags, which reduces the total number of parameters to estimate. Although smoothness restrictions primarily serve to stabilise the model, they can also be exploited to incorporate additional structure over the shape of the estimated lag curve. For example, it is widely assumed that the true lag curve has more curvature at short lags than at large lags, reflecting the intuition that recent events are generally the most influential, and influence should decay rapidly to zero for increasing lags. Further structure can also be used to avoid discontinuity between the influence of the final modelled lag $\beta_p \neq 0$ and the implicitly assumed $\beta_{p+j} = 0$. Some examples include Koyck (1954) who restricted $\boldsymbol{\beta}$ to decay geometrically with increasing p; Muggeo (2008) who combined a B-spline basis with parameters which were shrunk by a lag-varying ridge penalty; Gasparrini et al. (2010) used logarithmic knot placement in the lag dimension to enable varying smoothness; and Welty et al. (2009) constructed a Gaussian process prior distribution that enforces smaller covariances on lag parameters at short lags than longer ones. While each of these approaches achieves a level of lag-varying smoothness, each incorporates a parametric assumption on the type of penalisation that occurs, which may not be appropriate in general.

A further problem arises because the maximum lag p must be chosen before model fitting: this is sometimes justified by drawing on previously published studies, for example, in several studies of the delayed effects of temperature on human health, $20 \leq p \leq 30$ is consistently reported (Braga et al. (2002), Armstrong (2006), Gasparrini and Armstrong (2013)). In general, there is a cost to getting p wrong: too large and the model risks overfitting, and too small can bias the resulting lag function estimate. Few studies have investigated sensitivity of different model to choices of p, although several authors recognise that this is an ongoing difficulty for DLMs. Heaton and Peng (2012) and Heaton and Peng (2014) provide notable exceptions, and estimate p as part of a two-stage approach. However, DLMs are often applied in settings where the true lag curve is believed to decay gradually to 0, and in such settings, p is likely to be poorly estimated and will lack a meaningful interpretation.

This paper proposes a pragmatic strategy that ensures flexible and automatic smoothing over the lag structure that avoids the need to either estimate or choose p in advance. By exploiting the property that very large lags have close to 0 influence, an automatic lag-dependent smoother is constructed that ensures strong penalisation at high lags so that p can made arbitrarily large without the risk of overfitting. This follows the tradition of richly specifying a hierarchical model and ensuring parsimony by liberal use of regularisation and smoothing. A new DLM is described in Section 2 and in Section 3 a comprehensive simulation study illustrates the DLMs desirable properties, especially in comparison to existing approaches. This study compares bias resulting from getting p wrong, which is itself a central contribution, as DLMs have not previously been compared in this manner. Section 4 concludes the paper with an overview of the main findings and a discussion of the implications for users of distributed lag models.

2 Adaptive distributed lag models

Define $\mathbf{y} = (y_1, \ldots, y_n)$, a response variable time series of length n, and $\mathbf{x} = (x_1, \ldots, x_n)$, a time series of an exposure or input, whose influence on $\mathbf{y}$ up to some lag p is to be estimated. A Gaussian DLM for these data can be written as $y_i \sim \mathrm{N}\left(\sum_{j=0}^{p} x_{i-j}\beta_j, \ \sigma^2\right)$ where $i = p+1, p+2, \ldots, n$, and σ^2 is the error variance. The lag parameters $(\beta_0, \beta_1, \ldots, \beta_p)$ could be penalised and estimated directly, but for ease of comparison with existing approaches, these will be projected onto a B-spline basis of K functions uniformly spaced over the range of $\mathbf{x}$. The basis representation can be expressed as $\beta_j = \sum_{k=1}^{K} B_k(j) b_k$ where the further set of parameters $\mathbf{b} = (b_1, \ldots, b_K)$ determine the fitted lag curve. The model can be expressed as $\mathbf{y} \sim \mathrm{N}\left(\mathbf{Xb}, \ \sigma^2\right)$, where $\mathbf{X}_{ik} = \sum_{j=0}^{p} x_{i-j} B_k(j)$. To control the roughness of b_k in a way that enables variable smoothing, a similar approach to Brewer and Nolan (2007) and Reich and Hodges (2008) is used, in which the intrinsic autoregressive (ICAR) prior distribution of Besag et al. (1991) is generalised. The conditional prior distribution for $\mathbf{b}$ under the ordinary ICAR is $\mathbf{b} \sim \mathrm{N}(\mathbf{0}, \frac{1}{\lambda}\mathbf{P}^{-1})$ resulting in a conditional prior density of

$$\pi(\mathbf{b}|\lambda) \propto \exp\left(-\frac{\lambda}{2}\mathbf{b}^{\top}\mathbf{Pb}\right) \quad = \quad \exp\left(-\frac{1}{2}\sum_{k=1}^{K-1}\lambda(b_{k+1} - b_k)^2\right). \tag{2.1}$$

The precision parameter λ controls the extent to which the squared differences of consecutive pairs of $\mathbf{b}$ influences the posterior, imparting a smoothing effect over the lag curve derived from $\mathbf{b}$. Since each squared difference $(b_{k+1} - b_k)^2$ contributes equally in the prior in in Equation 2.1, variable smoothing can be achieved naturally by introducing a set of precision parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{K-1})$ that enables pairs of B-spline parameters to be smoothed more or less strongly. This generalised prior can be written

$\pi(\mathbf{b}|\lambda) \propto \exp\left(-\frac{1}{2}\sum_{k=1}^{K-1}\lambda_k(b_{k+1}-b_k)^2\right) = N(\mathbf{0},\mathbf{Q}^{-1})$ where $\mathbf{Q}$ is the precision matrix

$$\mathbf{Q} = \begin{pmatrix} \lambda_1 & -\lambda_1 & 0 & 0 & \cdots & 0 \\ -\lambda_1 & \lambda_1+\lambda_2 & -\lambda_2 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2+\lambda_3 & -\lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & -\lambda_{K-1} & \lambda_{K-1}+\rho \end{pmatrix}.$$

The additional parameter $\rho > 0$ added to the $(K,K)^{th}$ element of $\mathbf{Q}$ serves the dual purpose of ensuring $\mathbf{Q}$ is non-singular, and in providing an additional penalty that encourages the last element of $\mathbf{b}$ to approach zero. This joint specification for $\mathbf{b}$ results in the conditional distributions

$$b_k \mid \mathbf{b}_{-k}, \boldsymbol{\lambda} \sim \begin{cases} N\left(b_{k+1}, \frac{1}{\lambda_i}\right) & \text{if } k = 1 \\ N\left(\frac{\lambda_{k-1}b_{k-1}+\lambda_i b_{k+1}}{\lambda_k+\lambda_{k+1}}, \frac{1}{\lambda_k+\lambda_{k+1}}\right) & \text{if } k = 2,\ldots,K-1 \\ N\left(b_{k-1}, \frac{1}{\lambda_{k-1}}\right) & \text{if } k = p. \end{cases}$$

When $\lambda_1 = \lambda_2 = \ldots = \lambda_{K-1} = \lambda$ and $\rho = 0$, then $\mathbf{Q} = \lambda\mathbf{P}$, which shows that the ordinary ICAR prior in Equation 2.1 is a special case of the adaptive prior. A further smoothing prior distribution is placed over λ to restrict their flexibility - this is particularly attractive as it also reflects our intuition that the curvature of the lag function should not change too rapidly. For this purpose, an ordinary ICAR prior is used to smooth λ such that $\boldsymbol{\tau}|\zeta^2 \sim N(\mathbf{0}, \zeta^2\mathbf{K}^{-1})$

where $\mathbf{K}$ is the $(K-1) \times (K-1)$ precision matrix

$$\mathbf{K} = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Therefore, for small values of the variance ζ^2, pairs of consecutive elements of $\boldsymbol{\lambda}$ are more strongly smoothed towards each other. Finally, the variance parameters were assigned weakly informative priors $\sigma^2, \zeta^2 \sim \text{Inverse-Gamma}\,(1, 1/2)$. The combination of adaptive smoothing with smoothed variances as described, results in a very flexible model that can accommodate complex shapes using a relatively simple specification. The model described above assumes a Gaussian errors for simplicity, but the ideas easily generalise to accommodate other response distributions.

Inference for models fitted in the simulation study in Section 3 was achieved using Markov-chain Monte Carlo (MCMC) simulation using a mixture of Gibbs sampling and Metropolis-Hastings steps, using a mixture of optimised R (R Core Team (2016)) and C++ functions implemented using the package Rcpp (Eddelbuettel (2013)). However, the models can be implemented using standard Bayesian model fitting software such as STAN (Carpenter et al. (2016)).

3 Simulation study

The adaptive model detailed in Section 2 was compared with several currently available distributed lag models, labelled $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, $\mathcal{M}_4$ and $\mathcal{M}_5$. $\mathcal{M}_1$ is a non-adaptive version of the DL described in Section 2 (achieved by fixing $\lambda_1 = \lambda_2 = \ldots = \lambda_{K-1} = \lambda$), $\mathcal{M}_2$ is a Bayesian implementation of the DLM with linearly increasing ridge penalty proposed by Muggeo (2008), $\mathcal{M}_3$ is the new adaptive model described in Section 2, $\mathcal{M}_4$ is the P-spline model of Obermeier et al. (2015) and $\mathcal{M}_5$ is the DLM with logarithmically spaced knots of Gasparrini et al. (2010).

3.1 Comparison under different lag structures

To compare estimation performance of models $\mathcal{M}_1 - \mathcal{M}_5$, time series response and covariate data were simulated assuming each of five different-shaped lag relationships, each with maximum lag $p = 50$. The curves are shown in Figure 1 and included a delayed peak scenario (DELAYED PEAK), an exponential decay curve (DECAY CURVE), a horizontal line at 0 equivalent to a null effect (NULL CURVE), a curve exhibiting a shape similar to a mortality displacement effect (DISPLACEMENT) and a sharply peaked function that declines rapidly to 0 (SHARP PEAK).

A covariate time series, x_t, was generated by assuming $x_t \sim \mathrm{N}(0.5x_{t-1}, 0.1^2)$, and conditional on each lag function, a response series y_t was generated using $y_t = \sum_{j=0}^{50} \beta_j x_{t-j} + \epsilon_t$ for $t = 51, \ldots, 500$. The error term ϵ_t assumed autocorrelation such that $\epsilon_t \sim \mathrm{N}(0.2\epsilon_{t-1}, 0.1^2)$, whose parameters were selected for an acceptable signal-to-noise ratio. The autocorrelation in both x_t and ϵ_t reflects the temporal dependence in covariates and errors typically observed in real data.
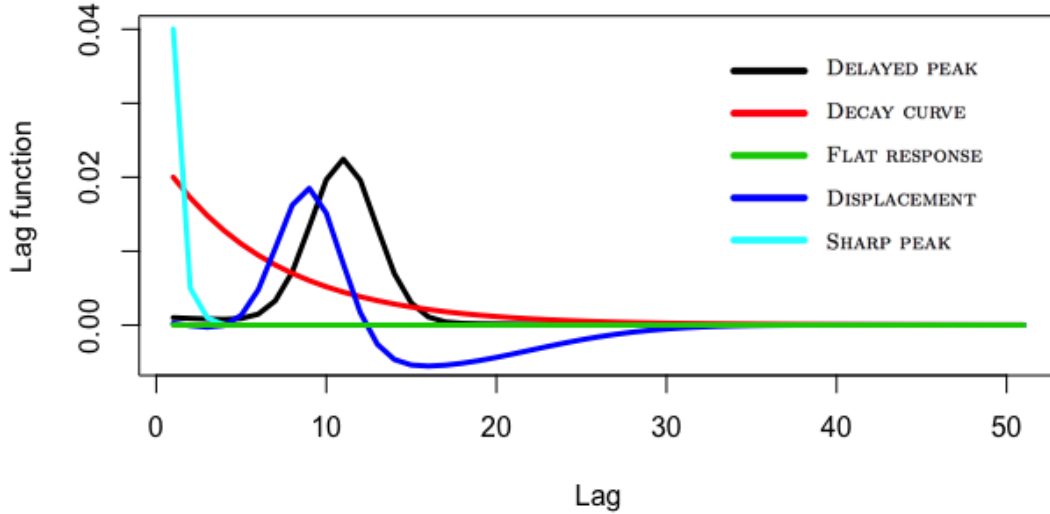
Figure 1: The five lag curves used to simulate response and covariate data with different lag response.

Models $\mathcal{M}_1 - \mathcal{M}_5$ were fitted to each of 200 realisations of the processes $\{x_t, y_t\}$ described above. The models were assessed in terms of their ability to recover the true lag curve, as measured by root-mean squared error (RMSE) and squared bias (Bias^2) of the estimates, defined for the jth simulation as $\left(\frac{1}{51} \sum_{i=0}^{50} (\hat{\beta}_i^{(j)} - \beta_i^{(j)})^2 \right)^{1/2}$ and $\left(\frac{1}{512} \sum_{i=0}^{51} \hat{\beta}_i^{(j)} - \beta_i^{(j)} \right)^2$, respectively. Following Hodges and Sargent (2001), model complexity was recorded for fitted model $\mathcal{M}_i^{(j)}$ using the effective degrees of freedom $\text{ED} = \text{tr}[\mathbf{X}(\mathbf{X}^\top \mathbf{X} + \hat{\mathbf{S}}_{\mathcal{M}_i^{(j)}})^{-1} \mathbf{X}^\top]$ where $\hat{\mathbf{S}}_{\mathcal{M}_i^{(j)}}$ is the estimated penalty component for the ith model fitted to the jth simulation. In particular, $\hat{\mathbf{S}}_{\mathcal{M}_1^{(j)}} = \hat{\lambda}^{(j)} \text{diag}(1, 2, \ldots, p+1)$; $\hat{\mathbf{S}}_{\mathcal{M}_2^{(j)}} = \hat{\lambda}^{(j)} \mathbf{P}$; $\hat{\mathbf{S}}_{\mathcal{M}_3^{(j)}} = \hat{\mathbf{Q}}^{(j)}$ and $\hat{\mathbf{S}}_{\mathcal{M}_4^{(j)}} = \hat{\lambda}^{(j)} \mathbf{P} + \hat{\rho}^{(j)} \mathbf{I}$. For models $\mathcal{M}_1 - \mathcal{M}_4$ the model matrix $\mathbf{X}$ was identical, resulting from assuming $(2/3) \times (p+1) = 34$ uniformly spaced B-spline basis functions to represent the lag curve, such that the $(i, j)^{\text{th}}$ element, $\mathbf{X}_{i,k} = \sum_{j=0}^{p} x_{i-j} B_k(j)$. Since lag curve smoothness for $\mathcal{M}_5$ is decided by finding the number of logarithmically spaced knots prior that minimises AIC, the effective dimension in this case is the number of knots, $\text{ED} = n_z^{(j)}$ for

9

the jth simulation.

Table 3.1 shows the average RMSE, Bias^2 and ED for each model and lag curve scenario combination. The adaptive model, $\mathcal{M}_3$, outperformed all of the models in terms of RMSE across all lag curve scenarios. Large separations were observed in model complexity across each scenario, with the adaptive smoothing $\mathcal{M}_3$ resulting in the lowest ED across all non-null scenarios, ranging between 3.86 to 7.97 compared to 6.4 and 19.09 for the other models. A particularly concerning result, is that only $\mathcal{M}_3$ and $\mathcal{M}_4$ correctly result in degree of freedom ≈ 1 under the FLAT RESPONSE scenario compared to a range of 7.7 and 18 for the others, which suggests that other approaches may falsely identify structure when none exists at all. Bias was generally low across all models and scenarios, although higher values were noted across $\mathcal{M}_1$-$\mathcal{M}_4$ under the SHARP PEAK scenario. As might be expected, the log-spaced knots of $\mathcal{M}_5$ performed favourably when lag influence declined rapidly from 0 (SHARP PEAK), and much less so when the strongest lag effect was larger than 0 (DELAYED PEAK, DISPLACEMENT). The comparison clearly highlighted the need for care in choosing a lag smoother in practical application, and that the quality of fit strongly depends on the underlying lag function. Furthermore, the simulations provide encouraging evidence that the newly proposed adaptive model is far more robust to these underlying differences, and might be a suitable choice when little is known about the structure being estimated.

3.2 DLM misspecification with maximum lag p

Next, simulation was used to explore robustness of $\mathcal{M}_1$-$\mathcal{M}_5$ to different choices of maximum lag p. The series x_t and y_t were generated exactly as in Section 3.1, using only the DISPLACEMENT lag function with 50 lags. The models $\mathcal{M}_1$-$\mathcal{M}_5$ were fitted to the simulated x_t and y_t, assuming maximum lags of $p = 50, 75, 100, 125$. To ensure the fairest comparison, the

Lag curve shape	Summary	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
	RMSE	17.31	14.53	10.32	15.51	22.68
DELAYED PEAK	Bias^2	3.31	3.24	3.32	9.59	3.83
	ED	18.53	14.99	7.12	12.14	15.81
	RMSE	16.57	11.59	8.45	9.46	14.15
DECAY CURVE	Bias^2	1.45	1.65	2.32	4.30	0.79
	ED	18.15	11.54	5.42	6.40	7.88
	RMSE	16.60	8.80	1.59	2.00	13.68
FLAT RESPONSE	Bias^2	1.14	0.82	0.28	0.22	1.51
	ED	18.07	8.08	0.77	0.68	7.74
	RMSE	17.32	14.13	11.24	15.29	20.34
DISPLACEMENT	Bias^2	2.79	3.30	3.36	8.79	2.51
	ED	18.55	13.97	7.97	12.78	13.37
	RMSE	23.98	16.65	10.97	27.72	14.07
SHARP PEAK	Bias^2	16.01	11.63	6.81	24.45	0.86
	ED	19.09	11.05	3.86	10.85	7.94

Table 1: Estimated lag function RMSE and effective degrees of freedom for four different types of lag curve shape, averaged across 200 simulated data sets. RMSE and Bias have each been scaled by 10^3.

range of flexibility for each model was restricted so that $0 < ED < 2p/3$, and also ensured that lag curve flexibility could grew proportionally with p. For each of 200 fits, the RMSE of the lag curve estimate, ED and squared bias were recorded and the results are summarised in Figure 2.

Figure 2 clearly illustrates, for $\mathcal{M}_1$ and $\mathcal{M}_5$ that ED increases steeply with p, even although the underlying true lag function is identical. Furthermore, both the RMSE and ED are substantially smaller for the adaptive models $\mathcal{M}_3$ compared to any of the others in the comparison, and indicate much better overall performance. The new adaptive approach $\mathcal{M}_3$ is more robust to user misspecification of 'large' p, and the simulations raise concerns about the potential for overfitting using existing approaches. However, it is noted that the effective degrees of freedom for $\mathcal{M}_3$ are not constant and do increase modestly with p, from
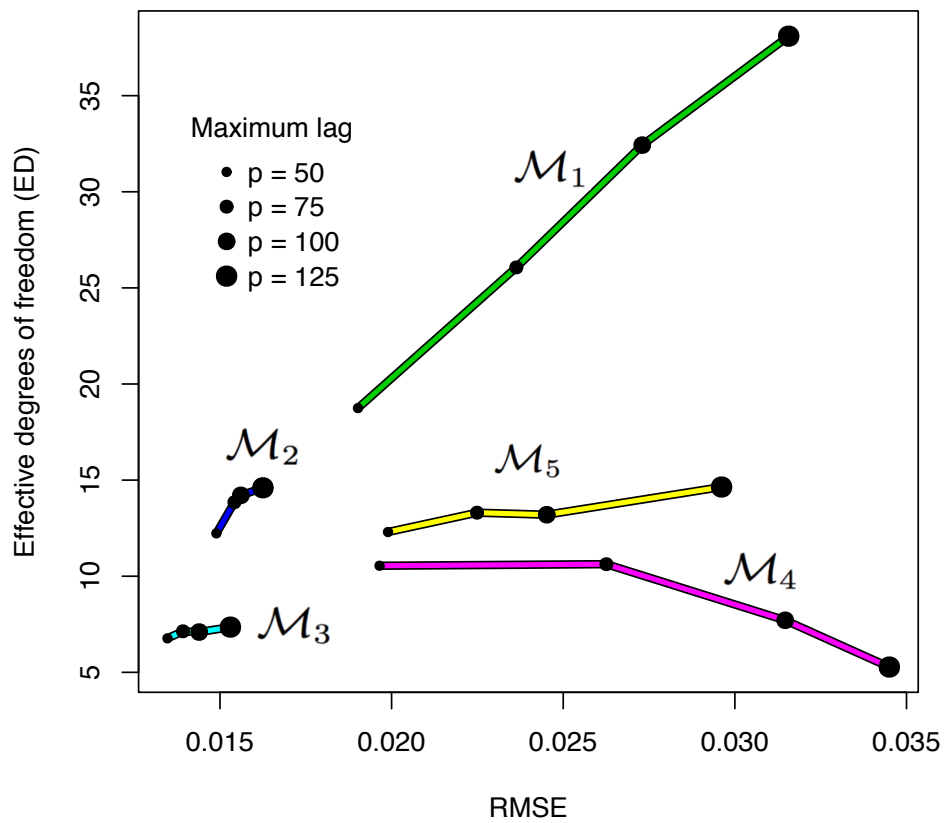
Figure 2: Effective number of parameters and RMSE for each of $\mathcal{M}_1$ - $\mathcal{M}_5$. Increasing diameter of the plotting symbols indicates larger numbers of maximum lags assumed in advance of model fitting.

6.7 ($p = 50$) and 7.4 ($p = 125$). Model $\mathcal{M}_2$ resulted in a comparatively favourable balance between RMSE and degrees of freedom, reinforcing the suggestion that adaptive models are preferable choices.

4   Discussion

This paper makes several important contributions. It was shown by simulation in Section 3 that estimation of lag structure can strongly depend on the type of smoothing model that is assumed, and that some existing approaches. Several existing DLM models were shown to be non-robust to the choice of maximum lag p, even when the underlying lag function is identical, which suggests that the interpretation of lag estimates should be made with caution. A new model has been proposed that combines automatic adaptive smoothing with a pragmatically large choice of p to ensure simple and flexible smoothing of the lag curve that avoids sensitivity to the choice of p. The new approach provides users of DLMs with a new way to explore their data with confidence that the estimates are not contaminated by artefacts that result from particular model choices.

References

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, 178–196.

Armstrong, B. (2006). Models for the relationship between ambient temperature and daily mortality. *Epidemiology 17*(6), 624–631.

Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics 43*(1), 1–20.

Braga, A. L., A. Zanobetti, and J. Schwartz (2002). The effect of weather on respiratory and cardiovascular deaths in 12 us cities. *Environmental health perspectives 110*(9), 859.

Braga, A. L. F., A. Zanobetti, and J. Schwartz (2001). The time course of weather-related deaths. *Epidemiology 12*(6), 662–667.

Brewer, M. J. and A. J. Nolan (2007). Variable smoothing in bayesian intrinsic autoregressions. *Environmetrics 18*(8), 841–857.

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell (2016). Stan: A probabilistic programming language. *J Stat Softw*.

Eddelbuettel, D. (2013). *Seamless R and C++ integration with Rcpp*. Springer.

Gasparrini, A. and B. Armstrong (2013). Reducing and meta-analysing estimates from distributed lag non-linear models. *BMC Medical Research Methodology 13*(1), 1.

Gasparrini, A., B. Armstrong, and M. Kenward (2010). Distributed lag non-linear models. *Statistics in medicine 29*(21), 2224.

Haugh, L. D. and G. E. Box (1977). Identification of dynamic regression (distributed lag) models connecting two time series. *Journal of the American Statistical Association 72*(357), 121–130.

Heaton, M. J. and R. D. Peng (2012). Flexible distributed lag models using random functions with application to estimating mortality displacement from heat-related deaths. *Journal of agricultural, biological, and environmental statistics 17*(3), 313–331.

Heaton, M. J. and R. D. Peng (2014). Extending distributed lag models to higher degrees. *Biostatistics 15*(2), 398–412.

Hodges, J. S. and D. J. Sargent (2001). Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika 88*(2), 367–379.

Koyck, L. M. (1954). Distributed lags and investment analysis.

Muggeo, V. M. (2008). Modeling temperature effects on mortality: multiple segmented relationships with common break points. *Biostatistics 9*(4), 613–620.

Nerlove, M. et al. (1958). Distributed lags and demand analysis for agricultural and other commodities. *Distributed lags and demand analysis for agricultural and other commodities.*.

Obermeier, V., F. Scheipl, C. Heumann, J. Wassermann, and H. Küchenhoff (2015). Flexible distributed lags for modelling earthquake data. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 64*(2), 395–412.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Reich, B. J. and J. S. Hodges (2008). Modeling longitudinal spatial periodontal data: A spatially adaptive model with tools for specifying priors and checking fit. *Biometrics 64*(3), 790–799.

Rushworth, A. M., A. W. Bowman, M. J. Brewer, and S. J. Langan (2013). Distributed lag models for hydrological data. *Biometrics 69*(2), 537–544.

Schwartz, J. (2000). Harvesting and long term exposure effects in the relation between air pollution and mortality. *American journal of epidemiology 151*(5), 440–448.

Welty, L. J., R. Peng, S. Zeger, and F. Dominici (2009). Bayesian distributed lag models: estimating effects of particulate matter air pollution on daily mortality. *Biometrics 65*(1), 282–291.

Wyzga, R. E. (1978). The effect of air pollution upon mortality: a consideration of distributed lag models. *Journal of the American Statistical Association 73*(363), 463–472.

Zanobetti, A., M. Wand, J. Schwartz, and L. Ryan (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics 1*(3), 279–292.

Zellner, A. and M. S. Geisel (1970). Analysis of distributed lag models with applications to consumption function estimation. *Econometrica: Journal of the Econometric Society*, 865–888.