ESTIMATION OF CONDITION-DEPENDENT MAINTENANCE EFFECTIVENESS FOR TRANSPORTATION INFRASTRUCTURE MANAGEMENT USING EXTENDED KALMAN FILTER

Chih-Yuan Chu (Assistant Professor, Department of Civil Engineering, National Central University, Taiwan, Email address: jameschu@ncu.edu.tw)

ABSTRACT

This paper proposes a procedure for estimating nonlinear dynamic performance models using panel data. Dynamic models that combine performance prediction and maintenance effectiveness are required for state-of-the-art life-cycle cost optimization techniques such as adaptive control. In addition, major maintenance generally takes place only when in-service facilities are near the point of failure. Records of maintenance effectiveness collected from such facilities depend on the pavement condition. To account for this dependence and estimate unbiased maintenance effectiveness, interactions between variables are included using nonlinear models. The relationship between maintenance effectiveness and the current facility condition was found to be polynomial in numerical examples. It was also demonstrated that imposing physical constraints on the maintenance effectiveness based on the actual observations improved the data fit significantly and generated more reasonable models.

Keywords: infrastructure performance modeling, dynamic models, maintenance and rehabilitation, nonlinear state-space models, extended Kalman filter

INTRODUCTION

The most significant reason for adopting a dynamic performance prediction model is that adaptive control using continuous state variables, which is a state-of-the-art optimization framework, requires dynamic models to generate optimal polices and rules for maintenance and rehabilitation (M&R) plans. However, studies such as those of Madanat and Ben-Akiva (1994), Ouyang and Madanat (2006), Durango-Cohen (2007), and Jido et al. (2008) rely on either simplified performance models to reduce the effort in optimization, or more realistic but separate models for prediction of condition and maintenance effectiveness.

As advocated by Lytton (1987), including maintenance in performance models is attractive for supporting maintenance decision-making in transportation infrastructure management. However, this task is not trivial because the maintenance decisions of transportation agencies depend on the facility condition for in-service roads. That is, maintenance is more effective and more likely to be applied to a facility in poor condition. If this dependence is not modeled properly, the model may have incorrect parameter signs and/or a poor data fit (Ben-Akiva and Ramaswamy, 1993).

Dynamic models consider serial dependence, which means that the current condition depends on historical condition. This consideration is particularly useful for estimating the dependence of facility condition on maintenance by analyzing the time series of facility conditions and maintenance schedules. Therefore, they have been used for infrastructure performance modeling in Chu and Durango-Cohen (2008a) and Chu and Durango-Cohen (2008b). Note that those studies rely on linear dynamic models and are incapable of estimating maintenance effectiveness as a function of facility condition. The constant maintenance effectiveness is biased because the maintenance effectiveness is dependent on the pre-maintenance condition. Therefore, nonlinear dynamic models that involve the interactions between independent variables to estimate maintenance effectiveness in conjunction with the pre-maintenance condition are necessary. If the dependence of maintenance on facility condition is estimated, the current facility condition can be predicted by historical facility conditions and in response to maintenance as a function of historical conditions. In such a model specification, facility condition is the dependent variable and maintenance is an independent variable. Note particularly that infrastructure performance models are particularly useful with maintenance as an independent variable. Such models can be used to evaluate various M&R strategies to support maintenance optimization, which is one of the contributions of this research. Motivated by this background, this paper describes a procedure for estimating nonlinear dynamic performance models that capture interactions between variables using panel data.

REVIEW OF PARAMETER ESTIMATION WITH THE KALMAN FILTER

Two popular parameter estimation methods for nonlinear state-space models are joint estimation and dual estimation. The advantage is that the estimation can be conducted on-line and the estimates are updated when new data are observed. However, the disadvantage is that the error covariances are difficult, if not impossible, to estimate as unknown parameters. As a result, the research using joint or dual estimation mentioned above makes various assumptions for the error covariances so that they can be removed the estimation process.

If parameters in the covariance matrices are also unknown and must be estimated, an external optimization routine that maximizes the likelihood function (e.g., Chu and Durango-Cohen (2008a) for linear models and Chow et al. (2008) for nonlinear models) can be used in addition to a nonlinear Kalman filter. This approach is particularly useful when all data are readily available for the estimation, which is often the case in transportation infrastructure management. This paper thus uses the methods described by Chu and Durango-Cohen (2008a) and extends them to consider nonlinear models.

METHODOLOGY

This section describes the nonlinear state-space model formulations and the estimation procedure using panel data.

Single-facility State-space Model

A nonlinear single-facility dynamic model is shown in Eqs. (1) through (2), which constitute a state-space model. Equation (1) is the system equation, which governs the system's dynamics and captures the effect of explanatory variables. Equation (2) is the measurement error equation, which captures random errors in the inspection process. The traditional linear formulation is limited for predicting facility performance because the relationships between facility condition, exogenous variables, and errors are limited to being additive. A nonlinear model is required to consider more realistic relationships between these factors, e.g., the dependence between the maintenance effectiveness and the current facility condition. The functions   and   are nonlinear, and the variables in the function are not necessarily additive but can have any form.

$$X_t = F(g_{t-1}, X_{t-1}, h_{t-1}, A_{t-1}, \omega_{t-1}) \quad (1)$$
$$Z_t = H(\Lambda_t, X_t, \xi_t) \quad (2)$$
$$t = 1, 2, \cdots, T$$

where the variables, parameters, and random error terms in the model are explained briefly below. More details are provided by Chu and Durango-Cohen (2008a).

1. X_t **:** d -dimensional *state vector* representing the unobservable condition of infrastructure facility i at the start of period t, where $X_t = [x_t^{(1)}, \cdots, x_t^{(d)}]'$.

2. A_t **:** s -dimensional vector of exogenous explanatory variables, e.g., structural design, history of maintenance and rehabilitation activities, environmental factors, and traffic loading, where $A_t = [A_t^{(1)}, \cdots, A_t^{(s)}]'$.

3. Z_t **:** k -dimensional *observation vector* representing the set of observable condition measurements or indicators collected for facility i in period t, where $Z_t = [z_t^{(1)}, \cdots, z_t^{(k)}]'$. This vector may include measurements of distress such as

pavement cracking, rutting, or raveling. The vector may also include subjective ratings or aggregate condition indices.

4. g_t, h_t **:** contain the parameters that describe the autoregression of state vectors and the effect of explanatory variables on the state vector, respectively, where $g_t = [g_t^{(1)}, \cdots, g_t^{(d)}]'$ and $h_t = [h_t^{(1)}, \cdots, h_t^{(r)}]'$. The dimensions of the two vectors are d and r, respectively.

5. Λ_t **:** dk -dimensional vector of parameters describing the relationship between measurements and condition, where $\Lambda_t = [\Lambda_t^{(1)}, \cdots, \Lambda_t^{(dk)}]'$.

6. ω_t, ξ_t **:** represent random error terms assumed be normally distributed with finite second moments, where $\omega_t = [\omega_t^{(1)}, \cdots, \omega_t^{(d)}]'$ and $\xi_t = [\xi_t^{(1)}, \cdots, \xi_t^{(k)}]'$. The error terms have dimensions $d \times 1$ and $k \times 1$. Σ_{ω_t} and Σ_{ξ_t} are their covariance matrices with dimensions $d \times d$ and $k \times k$, respectively. They satisfy $\mathrm{E}(\omega_t) = 0$, $\mathrm{Var}(\omega_t) = \Sigma_{\omega_t}$, $\mathrm{E}(\xi_t) = 0$, $\mathrm{Var}(\xi_t) = \Sigma_{\xi_t}$, and $\mathrm{E}(\omega_t \xi_t) = 0$.

State-space Model Using Panel Data

To cover the case of multiple facilities, a new subscript i is added to identify the facility, where $i = 1, \cdots, N$, resulting in $X_{i,t}$, $A_{i,t}$, and $Z_{i,t}$. Many assumptions about heterogeneity among facilities are possible for multiple-facility performance models (Chu and Durango-Cohen, 2008a). The assumption used in this paper is that data from the facilities are instances of the same stochastic process, i.e., different surveys of the same stochastic process. Therefore, the specification has a single set of parameters shared by all facilities.

Equations (3) and (4) are a nonlinear state-space model of the dynamics of a panel of N facilities. The functions $\mathbf{F}$ and $\mathbf{H}$ are collections of the functions for individual facilities. The vectors above are collected in $\mathbf{X_t}, \mathbf{Z_t}, \mathbf{A_t}, \Omega_\mathbf{t},$ and $\Xi_\mathbf{t}$. That is, $\mathbf{X_t} = \left[X_{1,t}, \ldots, X_{N,t}\right]$, $\mathbf{Z_t} = \left[Z_{1,t}, \ldots, Z_{N,t}\right]$, $\mathbf{A_t} = \left[A_{1,t}, \ldots, A_{N,t}\right]$, $\Omega_\mathbf{t} = [\omega_t, \cdots, \omega_t]'$, and $\Xi_\mathbf{t} = [\xi_t, \cdots, \xi_t]'$.

$$\mathbf{X_t} = \mathbf{F}(g_{t-1}, \mathbf{X_{t-1}}, h_{t-1}, \mathbf{A_{t-1}}, \Omega_\mathbf{t-1}) \tag{3}$$

$$\mathbf{Z_t} = \mathbf{H}(\Lambda_t, \mathbf{X_t}, \Xi_\mathbf{t}) \tag{4}$$

EMPIRICAL STUDY

This section describes the data source for the empirical study and the estimation results of the nonlinear dynamic performance model.

Data Source

The empirical study used the AASHO Road Test dataset (Highway Research Board, 1962). The models described in the previous section require the specification of variables representing distress measurements and explanatory variables. The distress measurements as well as other factors such as the number of traffic loadings and temperature were recorded biweekly. Therefore, 56 records were available for each variable and were indexed by $t = 1, \cdots, 56$ in the model specifications. For example, $t = 1$ indicates the data collected on November 3, 1958 and $t = 56$ indicates the data collected on December 5, 1960. The actual dates of the data collection can be found in the source (Highway Research Board, 1962). To eliminate the influence of missing values, the pavement sections selected for analysis were those whose data were recorded during the whole period. The treatment of missing values in state-space models can be found in Chu and Durango-Cohen (2007) and Durbin and Koopman (2001). The selected pavement sections were indexed by i, where $i = 1, \cdots, 188$ and the pavement condition in the data was represented by the present serviceability index (PSI). The PSI is a widely-accepted indicator of a pavement's serviceability or functional performance in the range 0.0–5.0. The average PSI values for the inner and outer wheelpaths of each pavement section were used in this paper.

The exogenous explanatory variables and their notations considered in this study were as follows:

1. SN_i**:** Structural number of section i. A pavement's structural number serves as a proxy for its structural design and is a function of the pavement surface, base, and subbase thicknesses. The range of SN in the dataset was 0.44–5.66.

2. $TRF_{i,t}$**:** Seasonal weighted traffic loading applied to section i during time period t measured in 10^5 equivalent single-axle loads (ESALs). Traffic and weathering are the critical factors in pavement deterioration. Load applications were transformed to ESALs according to the axle loads and axle configurations. The seasonal weighting function that accounts for the ambient temperature and frost depth at the time of loading was established by the Highway Research Board (1962), and is not presented here due to space limitation. The motivation for using the seasonal weighted traffic load was to account for environmental factors that make pavement more or less vulnerable to deterioration due to traffic.

3. $OVR_{i,t}$**:** Indicator variable where $OVR_{i,t} = 1$ when an overlay is applied on section i between time t and $t+1$, and 0 otherwise.

Two key observations reported by the Highway Research Board (1962) are used later in the model estimation. The first observation is that the average PSI of newly constructed pavements was 4.2. This value is the upper limit of the pavement condition in the analysis that follows. The second key observation is that the application of overlays was difficult due to the cold weather and short pavement sections, and the average PSI value of the pavement immediately after overlay application was only 3.4.

Estimation Results and Discussion

Various model specifications were tested and the preferred model is listed in Eqs. (5) and (6) as an example. All parameters have intuitive signs; for example, $g^{(1)} = 0.991 < 1$ indicates that condition deteriorates over time. Note that the inspection interval is relatively small and the deterioration between two inspections is minimal, which explains why the parameter values for $g^{(1)}$ are very close to 1.

Two nonlinear phenomena were considered in the system equation for this example. The first is the effect of structural design on traffic impact ($SN - TRF$). With a nonlinear model, structural design can have a nonlinear relationship with traffic impact and thus also have an indirect effect on the pavement condition. In other words, the impact of a unit of traffic depends on the structural design, which is a more reasonable and relaxed assumption. Figure 1 shows the relationship between the structural number and the corresponding traffic impact. As expected, pavement with high SN (i.e., stronger) experiences small traffic impact and vice versa. This specification is more meaningful than including SN as an additive variable. A negative parameter for TRF in Eq. (5) indicates that the traffic causes damage to the pavement ($-0.035 + 0.004 SN_i < 0$ due to the range of SN).

The second nonlinearity considered in the model specification was the condition-dependent maintenance effectiveness ($x - OVR$). As explained in Section 1, maintenance effectiveness depends on the pavement condition and can be formulated as a function of historical pavement condition to predict the current condition. As previously mentioned, the Highway Research Board (1962) observed that the condition of newly constructed pavements was PSI=4.2 on average. Therefore, improving the condition to a PSI greater than 4.2 is unreasonable. In addition, it was reported that the average PSI value immediately after overlay application was only 3.4 due to cold weather and short pavement sections. To make sure the predictions to be consistent with the actual observations, constraints based on the physical conditions of the test would be necessary to improve the data fit and obtain more reasonable models. The first constraint comes from the observation that the overlay will bring the pavement condition to PSI=3.4. Because overlays were applied to pavement in poor condition, the extreme situation where the maintenance effectiveness for the worst possible pavement (PSI=0) would be 3.4 is used to represent these pavements. However, overlays were never applied to pavement with PSI$>$2.2 in the test, and the effectiveness of maintenance on other pavement conditions is unknown. Therefore, a second constraint must be assumed for this range of conditions. Because of the difficulty in applying overlay reported by the Highway Research Board (1962), the assumption is that the condition after overlay cannot exceed PSI=3.4 and that applying overlay on very new pavement will worsen its condition (i.e., negative effectiveness). In summary, the specifications satisfy that the effectiveness is 3.4 for $x_{i,t-1}$=0 (*Constraint 1*) and the effectiveness is 0 for $x_{i,t-1} = 3.4$ (*Constraint 2*). Figure 2 shows the relationship between the current condition and the maintenance effectiveness for the two constraints. *Constraint 1* forces all specifications to pass the location of the circle, and *Constraint 2* requires the functions to pass the location of the square. Clearly, the different specifications are almost identical after imposing the two constraints and, more important, they all produce very reasonable results. The results also show that the methodology is flexible for imposing constraints and these constraints could be beneficial for capturing the underlying pavement mechanism and improving the data fit.

$$x_{i,t}^{(1)} = 0.991x_{i,t-1}^{(1)} + (-0.035 + 0.004SN_i)TRF_{i,t-1}$$
$$+(3.400 - 0.998x_{i,t-1} - 0.0005(x_{i,t-1})^2)OVR_{i,t-1} + \omega_{t-1}^{(1)},\ \omega_t^{(1)} \sim N(0, -0.181^2) \quad (5)$$
$$z_{i,t}^{(1)} = x_{i,t}^{(1)} + \xi_t^{(1)}, \xi_t^{(1)} \sim N(0, 0.108^2) \quad (6)$$

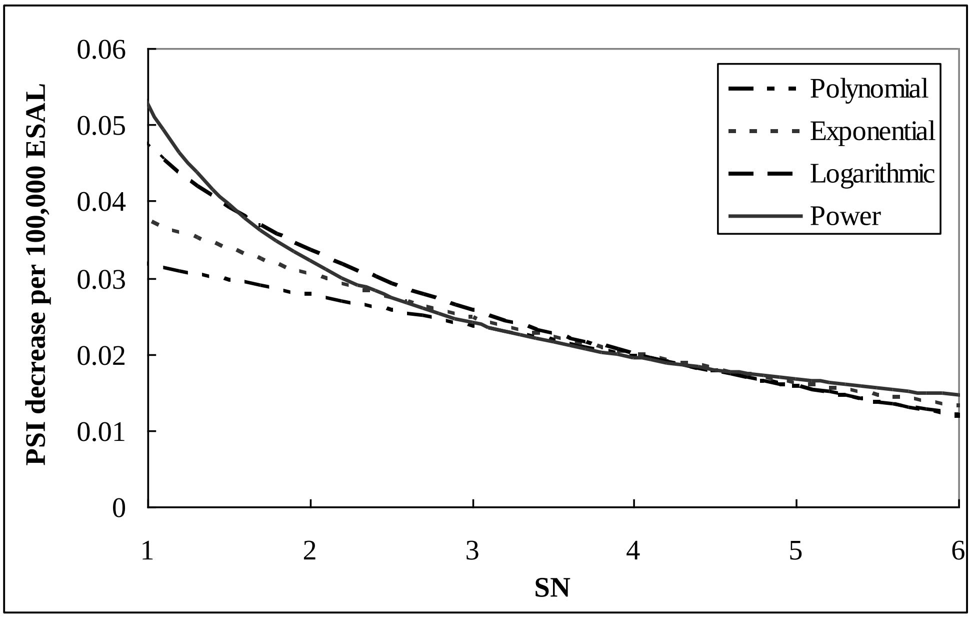

Figure 1 - Relationship between structural number and traffic impact

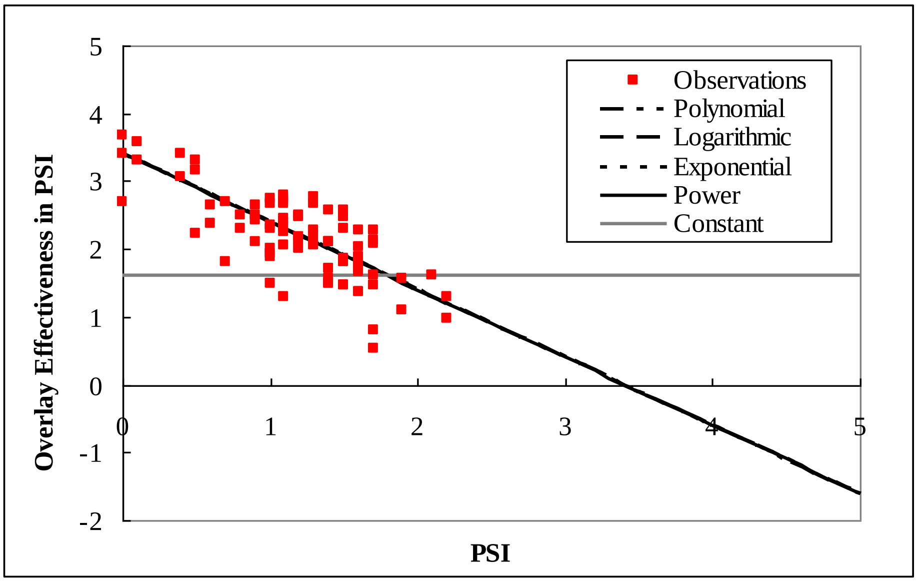

Figure 2 - Relationship between pavement condition and maintenance effectiveness (nonlinear models with Constraints 1 and 2)

CONCLUSION AND FUTURE RESEARCH

A methodology for estimating nonlinear dynamic performance models for transportation infrastructure management has been proposed. The dynamic models combine panel data, performance prediction, maintenance effectiveness, and variable interactions, all of which have not been previously studied. As a result, a state-of-the-art life-cycle cost optimization framework—adaptive control—can be supported with models estimated using empirical maintenance data.

ACKNOWLEDGMENTS

This work was supported by the National Science Council of Taiwan through research grant NSC 97-2218-E-008-013.

REFERENCES

Ben-Akiva, M. and Ramaswamy, R. (1993). An approach for predicting latent infrastructure facility deterioration. Transportation Science, 27(2), 174–193.

Chow, S.-M., Ferrer, E., and Nesselroade, J. R. (2008). An unscented kalman filter approach to the estimation of nonlinear dynamical systems models. Multivariate Behavioral Research, 42(2), 283–321.

Chu, C.-Y. and Durango-Cohen, P. L. (2007). Estimation of infrastructure performance models using state-space specifications of time series models. Transportation Research Part C, 15(1), 17–32.

Chu, C.-Y. and Durango-Cohen, P. L. (2008a). Estimation of dynamic performance models for transportation infrastructure using panel data. Transportation Research Part B, 42(1), 57–81.

Chu, C.-Y. and Durango-Cohen, P. L. (2008b). Incorporating maintenance effectiveness in the estimation of dynamic infrastructure performance models. Computer-Aided Civil and Infrastructure Engineering, 23(3), 174–188.

Durango-Cohen, P. L. (2007). A time series analysis framework for transportation infrastructure management. Transportation Research Part B, 41(5), 493–505.

Durbin, J. and Koopman, S. J. (2001). Time Series Analysis by State Space Methods, Oxford University Press Inc., New York, NY, USA.

Highway Research Board (1962). The AASHO Road Test, Special Reports No. 61A-E, National Academy of Science, National Research Council, Washington, DC, USA.

Jido, M., Otazawa, T., and Kobayashi, K. (2008). Optimal repair and inspection rules under uncertainty. Journal of Infrastructure Systems, 14(2), 150–158.

Lytton, R. L. (1987). Concepts of pavement performance prediction and modelling. Second North American Conference on Managing Pavements Proceedings, Vol. 2, 2.3–2.19.

Madanat, S. and Ben-Akiva, M. (1994). Optimal inspection and repair policies for infrastructure facilities. Transportation Science, 28(1), 55–62.

Ouyang, Y. and Madanat, S. (2006). An analytical solution for the finite-horizon pavement resurfacing planning problem. Transportation Research Part B, 40(9), 767–778.