Incorporating Maintenance Effectiveness in the Estimation of Dynamic Infrastructure Performance Models

Chih-Yuan Chu

Department of Transportation Technology and Supply Chain Management, Kainan University, Luzhu Shiang, Taoyuan 33857, Taiwan

&

Pablo L. Durango-Cohen*

Department of Civil and Environmental Engineering & Transportation Center, Northwestern University, Evanston, IL, USA

Abstract: We show how intervention analysis can be used in conjunction with dynamic performance modeling to analyze the effect of maintenance activities on the performance of infrastructure facilities. Specifically, we consider state-space specifications of autoregressive moving averages with exogenous inputs models to develop deterioration and inspection models for infrastructure facilities, and intervention analysis to estimate transitory and permanent effects of maintenance, for example, performance jumps or deterioration rate changes. To illustrate the methodology, we analyze the effectiveness of an overlay on a flexible pavement section from the AASHO Road Test. The results show the effect of the overlay on improvements both on surface distress, that is, rutting and slope variance, as well as on the pavement's underlying serviceability. The results also provide evidence that the overlay changes the pavement's response to traffic, that is, the overlay causes a reduction in the rate at which traffic damages the pavement.

1 INTRODUCTION

An important part of making design, construction, maintenance, and rehabilitation (M&R) decisions for infrastructure facilities, for example, pavements and bridges, consists of evaluating the effect of these decisions on the performance of such facilities. The evaluation, in turn, involves assessing and measuring distress and structural properties, for example, cracking, rutting, or elasticity in pavements and forecasting the effect of the aforementioned decisions on future condition. Condition forecasts are generated with performance models, which in this article correspond to statistical expressions that relate condition data to a set of explanatory variables such as design characteristics, traffic loading, environmental factors, and history of maintenance activities. The scale of expenditures associated with the above managerial decisions, as well as the far-reaching and serious negative economic and social impacts of deficient infrastructure have, over the last 40 years, motivated a great deal of research in the development of performance models. Extensive reviews of the literature can be found in references such as McNeil et al. (1992), Hudson et al. (1997), Gendreau and Soriano (1998), and Frangopol et al. (2004).

*To whom correspondence should be addressed. E-mail: pdc@northwestern.edu.
Performance models are estimated using data from panels of facilities. Panel data consist of two components: cross-section data describing the differences between the facilities that comprise the panel, that is, heterogeneity, and time series data describing the evolution of individual facilities over time. Most data sources are unbalanced in that they typically have extensive cross-section data and limited time series data. For example, Madanat et al. (1997) develop a bridge-deck performance model using data collected between 1978 and 1988 for 2,602 bridges in the state of Indiana. More than 80% of the bridges were inspected three or fewer times. No bridge in the data set was inspected more than seven times.1 To exploit this structure, existing performance modeling approaches are static, meaning that condition is represented as a function of contemporaneous explanatory variables and random error terms that describe the cross-sectional differences in the panel. Another important reason that explains the prevalence of static modeling is that commonly used dynamic models, for example, those developed with the Box-Jenkins ARIMA approach to time series modeling, are unattractive to model infrastructure performance because it is cumbersome to include exogenous explanatory variables that represent external factors such as structural characteristics, environmental factors, traffic, and history of M&R activities.

Recent developments that serve as motivation for our work in elaborating a statistical framework to estimate and analyze dynamic infrastructure performance models are:

- Advances in automated data collection technologies, for example, sensors, satellite imaging, video, laser, and radar, that allow for frequent and comprehensive inspections of infrastructure facilities, and that therefore can mitigate the aforementioned (time series) data availability problem. Studies discussing the use of advanced technologies to inspect infrastructure facilities include Maser et al. (1988), Maser (1989), Madanat (2000), and Wang (2000);
- Extensions of the state-space framework to analyze time series (see Harvey (1990), Janacek and Swift (1993), Durbin and Koopman (2001) and the references therein). These developments provide a flexible and rigorous approach to include exogenous variables in the formulation and estimation of dynamic performance models; and
- Increases in computational performance that provide analysts with the capability to manage the abundance and breadth of data generated by advanced inspection technologies, and the flexibility to specify, estimate, and analyze complex statistical models.

In particular, we have conducted two studies in this area. In Chu and Durango-Cohen (2007a) we consider specifications of AutoRegressive Moving Average (ARMA) and structural time series models to estimate inspection and deterioration models for asphalt pavements. The purpose is to illustrate how data generated by different technologies can be processed in the estimation of dynamic performance models.2 We note that the pavements in the data sample (were exposed to traffic and) did not receive maintenance. In Chu and Durango-Cohen (2007b), we consider the estimation of dynamic performance models using data from a panel of facilities. The data in this study consist of bi-weekly Present Serviceability Index (PSI) measurements for a panel of 166 asphalt pavements from the AASHO Road Test. In this study, we consider three classes of multivariate time series models that differ in the assumptions regarding the structure of the underlying mechanisms generating the data sequences corresponding to each pavement. The purpose of this study was to compare the models to assess the poolability of pavement condition data. We note that a subset of the pavements in the data sample failed and were overlaid. This allowed us to estimate the effect of overlays on PSI improvements (for failed pavements). Here, we emphasize that, due to technical and computational complexities that arise when data are pooled across (large) panels of facilities, the scope of the analysis was limited to the estimation of condition improvements.

One of the significant features of developing dynamic statistical performance models for infrastructure facilities is that intervention analysis can be used to rigorously estimate (transitory and permanent) effects of maintenance activities, for example, condition improvements, slope or seasonal changes. Intervention analysis is a procedure to test the nature and magnitude of a known event in time series data (Box and Tiao, 1975). The methodology circumvents difficulties that arise when static modeling approaches are used, for example, the simultaneity between deterioration and maintenance when data come from in-service facilities and that can lead to endogeneity bias. Thus, the aims of this article are to illustrate how intervention analysis can be used to estimate maintenance effectiveness in dynamic infrastructure performance models, and to describe why this approach is attractive. We purposely limit our analysis to the facility level to avoid technical issues that arise when data are pooled across a panel of facilities, as well as to highlight the conceptual differences between static and dynamic performance modeling.

From a managerial perspective, an attractive feature of the proposed framework is that the ensuing performance models can be used to support the allocation of resources for the preservation of infrastructure facilities. In
particular, the methodology described herein provides a statistically rigorous approach to estimate the parameters that drive maintenance optimization models to represent deterioration such as Durango-Cohen (2007). This means that the work described herein can be compared to the estimation of Markovian transition probabilities that are used to drive maintenance optimization models formulated as (latent) Markov decision processes. For a review of this literature the reader is referred to Mishalani and Madanat (2002) and the references therein.

The remainder of the article is organized as follows. Section 2 provides an overview of the approaches that have been proposed to estimate and capture maintenance effectiveness in performance models for infrastructure facilities. In Section 3, we describe the methodology used in this article, which consists of the formulation and estimation of the infrastructure performance models as ARMAX (AutoRegressive Moving Average with eXogenous variables) time series models. In Section 4, we consider an empirical example to illustrate the methodology. Specifically, we analyze the deterioration of pavement section 476, a flexible pavement, from the AASHO Road Test, and show how intervention analysis can be used to estimate the effect of an overlay both on surface distress, that is, rutting and slope variance, as well as on its serviceability. A summary of the contributions of the article and directions for future work are presented in Section 5.

2 LITERATURE REVIEW

The aims of the article are to illustrate how intervention analysis can be used to estimate and incorporate maintenance effectiveness in dynamic infrastructure performance models, and to describe why this approach is attractive. In this section, we consider the relationship between the proposed approach and other modeling approaches appearing in the literature, paying special attention to the caveats that need to be considered in the analysis. To facilitate the discussion in the following sections, the relationships between the different approaches are displayed in Figure 1. In addition, we provide an overview of the latent performance approach of Ben-Akiva and Ramaswamy (1993), which we adopt to rigorously capture the characteristics of the inspection/data-collection process.

Researchers such as Lytton (1987) have long recognized the need to "incorporate the effect of maintenance directly into pavement performance models," and to develop models that respond to exogenous interventions. Both characteristics are necessary to evaluate the effect of different managerial policies on the performance of infrastructure facilities, and thus to support the selection of M&R activities. There are, however, significant technical problems that make inclusion of maintenance in performance models difficult, and in turn, explain why the most common approach is to estimate separate deterioration and maintenance-effectiveness models. The obvious limitation is, of course, that there is no basis to combine separate models to generate (long-term) forecasts. We proceed to discuss these issues in detail.

The vast majority of existing performance models predict pavement deterioration without maintenance (cf. Highway Research Board, 1962; Paterson, 1987;
American Association of State Highway and Transportation Officials, 1993; Shahin, 2005). As depicted in Figure 1a, the approach is to model condition as a function of cumulative exogenous variables representing factors related to traffic, weather, age, and so on. These types of models are useful to predict time/traffic-loading to failure for new or reconstructed facilities, but cannot be used to support M&R decision-making. As depicted in Figure 1b, representing (cumulative) maintenance activities with exogenous variables and including them in the aforementioned performance models would seem to be a natural approach to include maintenance effectiveness in performance models. Butler et al. (1985) is one of the few published documents that employ this approach even though other studies (cf. Highway Research Board, 1962) report having tried it. This observation can probably be attributed to difficulties in the estimation. The models in Butler et al. (1985), for example, exhibit poor fit-to-data ($R^2$ values ranging from 0.225 to 0.384), and coefficients associated with traffic that are either statistically insignificant or that exhibit incorrect signs.

A common approach to deal with the aforementioned problems in the context of performance modeling is to assume (rather than estimate) the effect of various maintenance actions. Livneh (1998) and Archilla and Madanat (2001), for example, assume that overlays restore pavements to their original condition. It follows that estimation of maintenance effectiveness is not required because it is given by the difference between new condition and the condition before maintenance. Another approach involves separate estimation of maintenance-effectiveness models (cf. Paterson, 1987; Watanabe et al., 1987; Al-Mansour et al., 1994; Madanat and Mishalani, 1998). These models provide estimates of the condition improvement associated with various interventions as a function of cumulative exogenous factors, as well as type, extent, and quality of the actions. Condition before maintenance is sometimes included in the formulation of maintenance-effectiveness models. As stated above, the drawback of having separate maintenance and deterioration models is that there is no basis to combine them to generate (long-term) condition forecasts (under different M&R policies). One of the sources of inconsistencies is that (static) performance models explain aggregate deterioration (from new to failed), whereas maintenance-effectiveness models predict incremental condition changes. Among the possible problems that might arise, we note that unless specific treatments are considered, for example, reconstruction, it is difficult/impossible to sustain the dependency between condition and maintenance if they are forecasted with two or three separate models. This dependency dictates that condition after maintenance is determined by the condition before maintenance and the effectiveness of the maintenance activity.

There are other technical difficulties in the estimation of maintenance-effectiveness models. Significantly, Madanat and Mishalani (1998) argue that when data from in-service pavements are used, it is important to correct for selectivity bias. They explain that when maintenance activities are selected in response to condition, it is critical to recognize that only activities that are deemed effective are applied to the facilities. Thus, only self-selected (as opposed to random) samples are used to estimate the effectiveness of each activity, which can result in selectivity bias. The implication is that the effectiveness estimates may not be valid when other policies are adopted. To address the problem, as illustrated in Figure 1a, they propose a discrete choice model that captures the decision-maker’s selection process. The empirical results presented in the article show that selectivity bias can be significant. We note that the need to correct for sampling/selectivity bias in maintenance-effectiveness models is unclear because it is arguable that having estimates for different strata within a population of facilities is more desirable than having unbiased (average) estimates for the population. Thus, the significance of this work is to understand a given model’s limitations when generating forecasts, for example, if the effect of a given treatment is estimated with a sample of failed facilities, then the model should not be used to forecast the effect of the treatment on new facilities.

Ben-Akiva and Ramaswamy (1993) were first to shed light on the difficulties associated with combining deterioration and maintenance-effectiveness modeling. They point out that when dealing with cross-section data from in-service pavements (or in other situations), maintenance activities are selected and scheduled in response to condition, that is, cost and other considerations dictate that only pavements in poor or failed condition are repaired. The technical implication is that maintenance activities are scheduled endogenously instead of exogenously. To address the problem rigorously, they propose simultaneous equation models, where condition and maintenance are represented as endogenous variables. As shown in Figure 1c, these variables depend both on cumulative exogenous factors, as well as on each other. Empirical results reported in this and other studies (cf. Mohamad et al., 1997) have been used to validate this modeling approach. The drawback is that the simultaneous dependence between deterioration and maintenance means that it is not possible to forecast conditions under various M&R policies, and therefore, the ensuing models are not useful to support M&R decision-making.

To their credit, Ben-Akiva and Ramaswamy (1993) suggest dynamic modeling as an approach to
circumvent the problems associated with the simultaneity of deterioration and maintenance. They state that when detailed time series data are available, it is possible to formulate dynamic performance models where maintenance activities are represented as exogenous variables. This approach is radically different. In static models, condition (at a given time) is represented as a function of cumulative exogenous variables. Dynamic models, in contrast, are models of serial dependence. That is, facility condition (at a given time) is represented as a function of lagged-dependent variables, that in turn, make condition dependent on the entire prior sequence of condition measurements for that particular facility. By using state-space specifications of time series models and intervention analysis, as is done herein, it is also possible to estimate the effect of (incremental or cumulative) exogenous factors, including maintenance, on the ensuing sequence of condition measurements. These relationships are depicted in Figure 1d where historical condition is included as one of the inputs. This is more powerful than (static) maintenance-effectiveness models that only capture the effect of maintenance on incremental condition improvements, a transitory effect, and therefore permits the identification of both transitory and permanent effects of maintenance actions. We adopt the terminology of Labi and Sinha (2003), where condition improvements are labeled performance jumps and permanent effects are referred to as deterioration rate changes. Before describing the methodology in detail, we point out that:

- Because dynamic models include lagged-dependent variables as explanatory factors, it is possible to capture the effect of historical condition trends on the effect of maintenance; and
- Researchers such as de Solminihac et al. (1999) and Prozzi and Madanat (2003) have proposed incremental models as an alternative approach. These models also address the difficulties caused by simultaneity between deterioration and maintenance in static models. However, serial-dependence is not treated rigorously in this modeling approach.

2.1 Latent performance modeling approach

The characteristics of the inspection process constitute another significant source of variability in infrastructure condition data that needs to be accounted for in a rigorous fashion. A common approach to process and use condition data to forecast infrastructure condition involves computing indices that aggregate condition data into a single quantity/rating. Examples include the Concrete Bridge Deck Condition Ratings, the Present Serviceability Index (PSI), the Pavement Condition Index (PCI), the Pavement Quality Index (PQI), and the International Road Roughness Index (IRI) developed by Office of Engineering Bridge Division (1979), Highway Research Board (1962), Shahin (2005), Karan et al. (1983), and Paterson (1986), respectively. Numerous statistical performance models have been developed by regressing these indices on explanatory variables. As explained by Ben-Akiva and Ramaswamy (1993), although intuitive and simple, the aforementioned indices lack rigorous justification and rely on predetermined sets of distress measurements (which precludes incorporating new ones). The aforementioned characteristics motivated Ben-Akiva and Ramaswamy (1993) to propose the latent performance modeling approach as a flexible and rigorous framework to forecast infrastructure condition when multiple distress measurements are collected simultaneously (using multiple technologies). This approach is similar to the indices described above in that it reflects the notion that infrastructure condition can be succinctly expressed in terms of a few underlying characteristics, for example, structural integrity, serviceability, safety, aesthetics, and so on. The fundamental difference is that condition of a facility is represented by latent/unobservable variables that capture the ambiguity that exists in defining, and consequently in measuring, condition. Empirical studies by Ben-Akiva and Ramaswamy (1993) and by Ben-Akiva and Gopinath (1995) have shown that latent performance models are appropriate to generate condition forecasts of transportation infrastructure, that is, the goodness-of-fit measures are better than those reported using other statistical methods. We build on this approach and use it in our empirical analysis to estimate the effect of maintenance both on surface distress, that is, rutting and slope variance, as well as on serviceability.

3 METHODOLOGY

In this section, we describe ARMAX models in state-space form to formulate the deterioration and inspection of infrastructure facilities. Such models can be represented as follows:

\[ x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + h_t^{(1)} a_{t-1}^{(1)} + \cdots + h_t^{(m)} a_{t-1}^{(m)} + \theta_1 e_{t-1} + \theta_2 e_t + \cdots + \theta_q e_{t-q}, t = 1, 2, \ldots, T \]  

(1)

\[ z_t^{(k)} = \lambda_{k,t} x_t + \xi_t^{(k)}, k = 1, 2, \ldots, K, t = 1, 2, \ldots, T \]  

(2)

Equation (1) is referred to as the structural equation. It is used to represent a facility’s deterioration.
The parameters and variables in the equation are as follows: \( x_i \) is a random variable used to represent facility condition at the start of period \( t \). Following the latent performance modeling approach of Ben-Akiva and Ramaswamy (1993), it is assumed to be unobservable, and may correspond to characteristics such as serviceability or structural fitness. \( \varepsilon_i \) is a variable that captures the random error in the deterioration process at time \( t \). \( \phi_{i,t} \) is the autoregressive parameter of order \( r \) at time \( t \). Collectively, the autoregressive terms capture the historical condition trends. \( \theta_{j,t} \) is the moving average parameter of order \( j \) at time \( t \). These terms capture the history of the prediction errors. The orders of the autoregressive and moving average terms, \( p \) and \( q \), indicate the time lags associated with their effects. The variables \( a_{pr}^{(l)} \) and their associated coefficients \( h_{pr}^{(l)} \) are used to capture the effect of exogenous factors such as structural design, environmental factors, traffic loading or (historical) M&R activities.

Equation (2) is the measurement model. The random variable \( \xi_t^{(k)} \) represents one of \( K \) measurements collected during \( t \). Following Ben-Akiva and Ramaswamy (1993) the measurements are modeled as manifestations of the underlying condition during the period, and may include surface distress (e.g., cracking or rutting), structural properties (e.g., elasticity), and subjective ratings (e.g., PSI). Linear measurement models for infrastructure facilities were proposed by Hudson et al. (1987) and Humphrick (1992). Such models capture both random errors/noise variables \( \xi_t^{(k)} \), as well as multiplicative biases through the coefficients \( \lambda_k \). We follow the aforementioned studies in that we attribute measurement errors to the precision and accuracy of the inspection technologies, even though they may also depend on factors such as location, time period, equipment operator, or the true value of the underlying condition. When the measurements correspond to different physical characteristics, \( \xi_t^{(k)} \) and \( \lambda_k \) also account for the fact that the measurements are imperfect condition proxies.

Two assumptions in the estimation and diagnosis of time series models in state-space form are that:

1. The random errors in Equations (1) and (2) follow Gaussian Distributions with zero means and finite covariance matrices. They are also assumed to be serially independent and independent of each other for all periods (i.e., \( E[\varepsilon_t | \varepsilon_s] = 0, E[\xi_t | \xi_s] = 0, \forall (t,s): t \neq s \), and \( E[\varepsilon_t \xi_s] = 0, \forall t, s \)). These assumptions are not overly restrictive, and are analogous to the assumption of normal, independent and identically distributed residuals in regression analysis. In dynamic modeling, the errors are associated with incremental (one-step-ahead) predictions. In regression analysis, a static modeling approach, the residuals are for contemporaneous prediction errors.

2. The true condition is generally associated with one of the measurements, a reference measurement. This involves setting one of the \( \lambda \)'s to 1 in Equation (2).

Thus, the parameters to be estimated are \( \phi_{1,i}, \ldots, \phi_{p,i}, h_{1,i}, \ldots, h_{m,i}, \theta_{1,i}, \ldots, \theta_{q,i}, \sigma_{\varepsilon}^2, \lambda_{1,i}, \ldots, \lambda_{K,i}, \Sigma_\xi \), where \( \sigma_{\varepsilon}^2 \) is the variance of \( \varepsilon_i \) and \( \Sigma_\xi \) is the covariance matrices for \( \xi_i \), respectively. As presented in Equations (1) and (2), the model is time-varying/time-dependent, that is, the parameters are indexed by time. However, to reduce the number of parameters, and thus, make the estimation meaningful and practical, time-dependence is usually restricted to certain periods. For example, in the empirical study presented in Section 4, we pay special attention to changes induced by maintenance. Thus, we consider up to two phases with associated parameters: pre-maintenance and post-maintenance. Capturing seasonal effects, for example, freeze-thaw cycles in the spring, is an example of another situation where having the flexibility afforded by time-dependent parameters can be useful.

4 EMPIRICAL STUDY

In this section, we estimate and analyze performance models for section 476, a flexible pavement section, from the AASHO Road Test. The objectives are:

1. To illustrate how the statistical framework described in this article can be used to estimate and analyze the effects of maintenance activities, that is, overlays, on the performance of infrastructure facilities. Specifically, we consider performance jumps and deterioration rate changes both on surfaces distress, that is, rutting and slope variance, as well as on the serviceability of section 476; and

2. To highlight differences between static and dynamic performance modeling approaches, as well as to illustrate how dynamic models can circumvent the technical difficulties that are inherent in the development of static performance models.

Prior to presenting estimation results and analysis, we briefly provide background about our data source, the AASHO Road Test, and the Present Serviceability Index developed as part of the study. We then describe in detail the performance models that were considered for estimation. The results and analysis are presented in Section 4.3.
4.1 Data source: The AASHO road test

The AASHO Road Test was conducted between October 1958 and November 1960 near Ottawa, Illinois (about 120 km southwest of Chicago). The site was chosen because the soil and climate in the area are representative of soils and climates in large areas of the United States. Details of the experimental design are available from Highway Research Board (1962). Overall, the test tracks consisted of six loops with two lanes each. The pavements in the first loop were not subjected to traffic for the purpose of comparison. The other ten lanes had approximately the same number of axle applications, load, and configuration on the performance of flexible pavements. The care with which other factors, for example, construction quality, inspection errors, and so on, were controlled explains why the data collected during this experiment are still among the most widely used sources in the development of pavement performance models and pavement design criteria. The condition data collected throughout the study focused on serviceability/functional performance, and included distress such as rut depth, slope variance, roughness, cracking, and patching.

One of the important contributions of the AASHO Road Test was the development of the Present Serviceability Index (PSI) as a means to quantify pavement serviceability/functional performance. In our analysis we consider the effect of overlays on improving pavement serviceability/functional performance, and included distress such as rut depth, slope variance, roughness, cracking, and patching.

The above equation was calibrated by having a panel of experts rate the serviceability of the pavements (from 0 to 5) to generate samples of the dependent variable.

4.2 Variables in the empirical study

Equations (1) and (2) require the specification of the variables representing distress measurements and explanatory factors. The measurements in our models are:

- \( z^{(1)}_t \): Transformed Rut Depth (RD) during period \( t \):
  \[ z^{(1)}_t = (0.0393 \cdot RD + 1)^c \] (c is a constant described below); and

- \( z^{(2)}_t \): Transformed Slope Variance (SV) during period \( t \):
  \[ z^{(2)}_t = \log(10SV + 1). \]

The above transformations are consistent with those used in the development of the PSI and facilitate the comparison of our results with those obtained in the original study. However, a further step is taken in this study. That is, a multiplication of 10 is used for \( SV \) and 1 is added for \( RD \) to make \( z^{(1)}_t \) and \( z^{(2)}_t \) have similar magnitude, which is a common technique to stabilize and expedite the model estimation. The constant, 0.0393, in the transformation of \( RD \) converts the raw data measured in inches into millimeters (mm). The original data can be easily obtained and analyzed by reversing the transformations. In our models, we set \( z^{(2)}_t \) as the reference measurement. The contributions of cracking and patching, \( CR + PA \), to the PSI for the pavements in the AASHO Road Test happens to be almost negligible (Huang, 1993), and therefore we exclude these measurements from our models.

The exogenous explanatory factors are:

- \( a^{(1)}_t \): Adjusted traffic loading during period \( t \) (seasonal-weighted Equivalent Standard Axel Loads in \( 10^5 \) ESALs). Traffic and weathering are critical factors that cause pavement deterioration. In the study, we use seasonal-weighted ESALs to represent traffic loading. The seasonal weighting function accounts for the ambient temperature and frost depth at the time of loading and was established by Highway Research Board (1962). We do not present the function due to space limitations; and

- \( a^{(2)}_t \): Dummy variable defined as follows: \( a^{(2)}_t = \{ 1, \text{if an overlay is applied during period } t; 0, \text{otherwise} \} \).

M&R activities in the AASHO Road Test included skin patch, deep patch, landing mat, remove and replace surface, spot seal, fog seal, overlay, and
reconstruction. Unfortunately, due to deficiencies/unavailability of records for other treatments, in this study we focus on modeling the effectiveness of overlays. We do note, however, that as discussed in American Association of State Highway and Transportation Officials (1993), few studies in the literature (none that we are aware of) have reported adequate estimation results when external interventions are included as explanatory variables for data from the AASHO Road Test.

4.3 Results and analysis

The objective of the study is to illustrate how intervention analysis can be used to estimate the effect of overlays on distress measurements (rutting and slope variance), as well as on the serviceability of a pavement. We reiterate that the analysis is restricted to the facility level. This is done to keep the analysis intuitive, as well as to avoid difficulties that arise when data are pooled across sections. One such difficulty is caused by the presence of unobserved heterogeneities between the sections in the data set (Prozzi and Madanat, 2003; Chu and Durango-Cohen, 2007b). Unobserved heterogeneities refer to the presence of persistent, facility-specific, but unobserved factors, for example, construction quality, that cause seemingly identical facilities to perform differently. To a certain extent, unobserved heterogeneities motivate model validation at the facility-level as is presented below (even when data from different sections are pooled in the estimation). The results and analysis presented in the remainder of the section is in the spirit of traditional time series analysis, where only a single realization of a stochastic process is available. Specifically, we estimate and analyze deterioration and measurement models for section 476 from the AASHO Road Test. This section is representative of sections that failed and were overlaid toward the end of the test (12 May 1960). The analysis of all sections in the experiment that exhibited this failure pattern is included in the first author’s doctoral dissertation. We chose section 476 from a total of 25 pavements in Loop 5 that exhibited the same failure pattern. The design of the section was such that it had 127mm., 228.6mm., and 101.6mm. of surface, base, and sub-base thicknesses, respectively. The section was subjected to 40,000 lb (177.9 kN) tandem axle loads, which correspond to 2.06 ESALs. A total of 1,114,000 load applications were made during the test.

In the remainder of this section we first analyze and compare several instances of the ARMAX model that were considered for estimation. The details of the estimation procedure appear in Chu and Durango-Cohen (2007a). We then use our preferred model to analyze the effect of the overlay on section 476’s performance.

4.3.1 Estimation results, model validation, and selection

The original RD and SV measurements for section 476 are shown in Figures 2 and 3. The distress measurements were transformed as follows: \( z_t^{(1)} = (0.0393RD + 1)^2 \) and \( z_t^{(2)} = \log(10SV + 1) \). Preliminary specifications, not presented in the article, resulted in statistically insignificant MA terms (at the 95% confidence level). Thus, we restricted our analysis on autoregressive models with exogenous inputs, ARX for short. The initial set of candidate models and estimation results for section 476 are presented in Table 1. Models ARX(1) and ARX(2) are time-homogeneous, whereas ARX(1)-\( \phi \) and ARX(1)-\( h^{(1)} \) are time-varying models. Because the second order autoregressive parameter, \( \phi_2 \), in ARX(2) is
insignificant, we limit the set of candidate deterioration models to first-order autoregressive models. Corroborating this choice, the Akaike Information Criterion (AIC) suggests that ARX(1) provides better fit-to-data than ARX(2).7

To select between the three remaining candidate models, we compare their predictive capabilities. The predictions obtained for the ARX(1) model are shown in the first row of Figure 4, along with the transformed measurements. Noticing that the deterioration rate has significant change after overlay, and the model fits worse for the measurements that follow the overlay, motivates consideration of the time-varying models.

Model ARX(1)- $\phi$ includes an (auto-regressive) parameter, $\phi_{1,PT}$, prior to the treatment (overlay), and a second (auto-regressive) parameter, $\phi_{1,AT}$, after the treatment. ARX(1)-$h^{(1)}$ uses one parameter, $h_{PT}^{(1)}$, for traffic prior to the overlay, and a second parameter, $h_{AT}^{(1)}$, for traffic applied after the overlay. The predictions obtained with the time-varying models are shown in the second and third rows of Figure 4. We also note that the two time-varying models are preferred to the time-homogeneous model based on the AIC. The engineering explanation is that in addition to a condition improvement, the overlay induces a change in the underlying deterioration process/rate/mechanism. We proceed to explore this issue further.

In Figure 4, we examine the (24-week-ahead) post-sample predictive capabilities of the two time-varying models. A problem associated with ARX(1)- $\phi$ is that the model yields improving (decreasing) distress measurements after the overlay. This observation is inconsistent with deterioration, and renders the model useless. This problem does not occur with ARX(1)-$h^{(1)}$ where the parameters for traffic and overlay, as well as the ensuing forecasts (see Figure 4) have signs that are consistent with deterioration. Therefore, we conclude that attributing the change of deterioration rate to the reduction of traffic damage due to the newly applied overlay is more realistic than assuming a new deterioration rate. In addition, ARX(1)-$h^{(1)}$ has a superior AIC. The ARX(1)-$h^{(1)}$ model is presented in Equations (4)–(6). Note that $a^{(2)}_t = 0$ after the overlay and thus it is excluded from Equation (6).

$$x_t = \phi_1 x_{t-1} + h_{PT}^{(1)} a^{(1)}_t + h^{(2)} a^{(2)}_t + \epsilon_t,$$

$\epsilon_t \sim N(0, 0.003)$ (before and at overlay)

$$x_t = 0.999 x_{t-1} + 0.106 a^{(1)}_t - 1.329 a^{(2)}_t + \epsilon_t,$$  \hspace{1cm} (4)

$$\epsilon_t \sim N(0, 0.003) \text{ (after overlay)}$$

$$x_t = \phi_1 x_{t-1} + h_{PT}^{(1)} a^{(1)}_t + \epsilon_t$$

$$x_t = 0.999 x_{t-1} + 0.018 a^{(1)}_t + \epsilon_t,$$  \hspace{1cm} (5)

$$\epsilon_t \sim N(0, 0.003) \text{ (after overlay)}$$

$$\begin{bmatrix} z^{(1)}_t \\ z^{(2)}_t \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ 1.000 \end{bmatrix} x_t + \begin{bmatrix} \xi^{(1)}_t \\ \xi^{(2)}_t \end{bmatrix} = \begin{bmatrix} 0.966 \\ 1.000 \end{bmatrix} x_t + \begin{bmatrix} \xi^{(1)}_t \\ \xi^{(2)}_t \end{bmatrix},$$

$$\xi_t \sim N(0, \Sigma_\xi), \Sigma_\xi = \begin{bmatrix} 0.054 & -0.017 \\ -0.017 & 0.024 \end{bmatrix} \hspace{1cm} (6)$$

To generate additional evidence that section 476’s deterioration is time-dependent, we conduct (one-year-ahead) in-sample predictive testing for the second year of the experiment. The results are shown in Figure 5. We note that the sum of the square prediction errors for the time-varying model, ARX(1)-$h^{(1)}$, is 54% less for rut depth and 64% less for slope variance than that of the time-homogeneous model, ARX(1). Finally,
Table 1

Estimation results for section 476 (with overlay)

<table>
<thead>
<tr>
<th>Parameter&lt;sup&gt;1,3&lt;/sup&gt;</th>
<th>Estimate</th>
<th>t-statistic</th>
<th>Estimate</th>
<th>t-statistic</th>
<th>Parameter&lt;sup&gt;1,2,3&lt;/sup&gt;</th>
<th>Estimate</th>
<th>t-statistic</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma^2_{\xi}^{(1)})</td>
<td>0.006</td>
<td>3.855</td>
<td>0.006</td>
<td>4.100</td>
<td>(\sigma^2_{\xi}^{(2)})</td>
<td>0.002</td>
<td>2.877</td>
<td>0.003</td>
<td>2.926</td>
</tr>
<tr>
<td>(\sigma^2_{\xi}^{(0)})</td>
<td>0.244</td>
<td>7.069</td>
<td>-0.240</td>
<td>-7.089</td>
<td>(\sigma^2_{\xi}^{(1)})</td>
<td>0.235</td>
<td>7.518</td>
<td>-0.233</td>
<td>-7.465</td>
</tr>
<tr>
<td>(\sigma^2_{\xi}^{(v)})</td>
<td>-0.072</td>
<td>-2.412</td>
<td>0.074</td>
<td>2.516</td>
<td>(\sigma^2_{\xi}^{(v)})</td>
<td>-0.071</td>
<td>-2.693</td>
<td>0.073</td>
<td>2.773</td>
</tr>
<tr>
<td>(\sigma^2_{\xi}^{(m)})</td>
<td>0.119</td>
<td>5.245</td>
<td>0.116</td>
<td>5.031</td>
<td>(\sigma^2_{\xi}^{(m)})</td>
<td>0.137</td>
<td>5.668</td>
<td>0.137</td>
<td>5.583</td>
</tr>
<tr>
<td>(\Sigma_{\xi})</td>
<td>a</td>
<td>-</td>
<td>b</td>
<td>-</td>
<td>(\Sigma_{\xi})</td>
<td>c</td>
<td>-</td>
<td>d</td>
<td>-</td>
</tr>
<tr>
<td>(\phi_1)</td>
<td>1.024</td>
<td>9.304</td>
<td>0.999</td>
<td>113.979</td>
<td>(\phi_{1,PT})</td>
<td>1.002</td>
<td>163.593</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(\phi_2)</td>
<td>-0.025</td>
<td>-0.228&lt;sup&gt;4&lt;/sup&gt;</td>
<td>-</td>
<td>-</td>
<td>(\phi_{1,AT})</td>
<td>-</td>
<td>-</td>
<td>-0.999</td>
<td>159.248</td>
</tr>
<tr>
<td>(h^{(1)})</td>
<td>0.071</td>
<td>2.528</td>
<td>0.071</td>
<td>2.484</td>
<td>(h_{PT}^{(1)})</td>
<td>0.093</td>
<td>3.706</td>
<td>0.106</td>
<td>3.892</td>
</tr>
<tr>
<td>(h^{(2)})</td>
<td>-1.210</td>
<td>-5.983</td>
<td>-1.227</td>
<td>-6.179</td>
<td>(h_{AT}^{(1)})</td>
<td>-</td>
<td>-</td>
<td>-0.018</td>
<td>0.694&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
<tr>
<td>(\lambda_1)</td>
<td>0.966</td>
<td>40.577</td>
<td>0.965</td>
<td>40.957</td>
<td>(\lambda_1)</td>
<td>0.966</td>
<td>40.978</td>
<td>0.966</td>
<td>40.978</td>
</tr>
</tbody>
</table>

Loglikelihood | 14.685 | 15.540 | Loglikelihood | 18.869 | 18.965
AIC | -0.132 | -0.234 | AIC | -0.317 | -0.320(preferred)

1. \(h^{(1)}\) is the parameter of traffic (weighted ESAL) and \(h^{(2)}\) is the parameters of overlay. The unit of traffic is \(10^5\) ESAL. \(a^{(2)}_1 = 1\) when overlay is applied and 0 otherwise.

2. Parameters are denoted by additional symbols in time-varying models. For example, ARX(1)-\(h^{(1)}\) model has two parameters for traffic variable. Therefore, \(h_{PT}^{(1)}\) is the traffic parameter Prior to Treatment (overlay) and \(h_{AT}^{(1)}\) is the traffic parameter After Treatment (overlay). Similarly, time-varying autoregressive parameters are denoted by \(\phi_{1,PT}\) and \(\phi_{1,AT}\).

3. The covariance matrix of measurement errors are defined by

\[
\Sigma_{\xi} = \begin{bmatrix}
(\sigma^2_{\xi}^{(1)}) & (\sigma_{\xi}^{(1)}\phi_{1}^{(1)}) \\
(\sigma_{\xi}^{(0)}\phi_{1}^{(0)}) & (\sigma^2_{\xi}^{(0)})
\end{bmatrix} = \Sigma_{\xi}^{T} \Sigma_{\xi}^{T'} = \begin{bmatrix}
(\sigma^2_{\xi}^{(1)}) & (\sigma_{\xi}^{(1)}\phi_{1}^{(1)}) \\
(\sigma_{\xi}^{(0)}\phi_{1}^{(0)}) & (\sigma^2_{\xi}^{(0)}) + (\sigma^2_{\xi}^{(m)})^2
\end{bmatrix}
\]

and

\[
\Sigma_{\xi}^{T} = \begin{bmatrix}
\sigma^2_{\xi}^{(1)} & 0 \\
\sigma_{\xi}^{(0)} & \sigma^2_{\xi}^{(m)}
\end{bmatrix}.
\]

As defined throughout the empirical study, the first measurement is rut depth and the second is slope variance. The use of \(\Sigma_{\xi}^{T} \Sigma_{\xi}^{T'}\) assures positive definite covariance matrices of measurement error. The values of covariance matrices areas follows.

\[
a = \begin{bmatrix}
0.059 & -0.018 \\
-0.018 & 0.019
\end{bmatrix}, \quad b = \begin{bmatrix}
0.059 & -0.018 \\
-0.018 & 0.019
\end{bmatrix}, \quad c = \begin{bmatrix}
0.055 & -0.017 \\
-0.017 & 0.024
\end{bmatrix},
\]

and

\[
d = \begin{bmatrix}
0.054 & -0.017 \\
-0.017 & 0.024
\end{bmatrix}.
\]

4. Bold entries indicate the parameters are insignificant at 95% level of confidence.

Comparing the forecasts for the two models we observe that they show identical trends before the overlay is applied; however, after the overlay is applied, deterioration rate of ARX(1)-\(h^{(1)}\) is noticeably smaller than that of the ARX(1). Also, the 95% confidence interval of ARX(1)-\(h^{(1)}\) is narrower than that of the ARX(1) model, meaning that there is higher confidence in ARX(1)-\(h^{(1)}\) forecasts.

The model fitting (i.e., one-step-ahead or two-week-ahead predictions) obtained with the “preferred” model, ARX(1)-\(h^{(1)}\), are shown in Figures 2 and 3. An important observation is that the residuals/errors do not satisfy the independence assumption. One possible explanation may be given by the fact that the measurement error model imposes a constraint on the relationship between the two measurements and the latent variable, that is,
it assumes that the measurements are manifestations of the latent variable and that the relationship is linear. In spite of this constraint, the fits for the two measurements appear adequate.

An attractive feature of the structure used to specify the measurement error model is that it allows us to extract an index representing the pavement’s serviceability. The index corresponds to the expected value of the latent variable, \( x_t \). In Figure 6, we present a plot of the negative expected value of \( x_t \), normalized over the range of the section’s PSI, \([0.9, 4.3]\), and compare it to the PSI index for the section as given by Equation (3). Two advantages of using the proposed approach to develop and compute such indices are that it does not rely on predetermined distress measurements, and that the estimation does not rely on experts’ judgment.

4.3.2 Maintenance effectiveness. Here, we use the preferred model, ARX(1)-\( h^{(1)} \), to analyze the effect of the overlay on section 476’s performance. From Table 1, we see that the parameter of traffic after overlay is not significantly different from zero. A plausible explanation for the test result is that the overlay was applied near the end of the test, and therefore the number of observations is not sufficient to identify the parameter. If more data were available, we would expect the parameter to be statistically significant. Another explanation is that the effect of traffic is negligible when the overlay is newly applied. More importantly, the parameter of traffic after overlay is significantly different from 0.106, the parameter of traffic before overlay. Additional evidence that the estimate is adequate is that the parameter of traffic (\( h^{(1)} = 0.071 \)) in ARX(1) falls between those of the time-varying model (0.106 and 0.018), which is a reasonable result and shows the potential parameter bias when time-homogeneity is assumed. Therefore, it is conclusive that time-homogeneity does not exist for this example. Moreover, the impact of traffic is (instantaneously) reduced by 83%. (The coefficient of traffic changes from 0.106 to 0.018 as listed in Table 1). The
values seem reasonable for the short-term due to the newly applied overlay; however, there is no basis to validate this result.

Regarding the effect of the overlay on the pavement’s condition improvement, the actual effectiveness of the overlay (performance jump) is 20.3 (mm) for rut depth and 30.0 for slope variance or 92% of reduction of rut depth and 81% of slope variance. We also note that the PSI reduction is 2.4. From the results of in-sample prediction (or \( ARX(1) - h^{(1)} \) in Figure 5), the predicted reduction of rut depth is 10.5 (mm) and that of slope variance is 59.5 or 64% and 94% of reduction, respectively. The predicted improvement in the expected value of \( x_t \) is 2.3. These results show that the model correctly captures the direction and magnitude of maintenance effectiveness. The errors in predicting the improvement on the distress measurements, however, are quite significant. The observation may be related to assumptions, such as that the distress measurements are imperfect manifestations of the latent condition variable, to the linear structure of the measurement model, or to the data transformations that were applied to the raw distress measurements. The result is not unexpected because the latent condition variable has to explain both measurements.

To be precise and consistent with the discussion about selectivity bias in Section 2, we should note that the effectiveness of the overlay was estimated for section 476 when the pavement failed. In other words, the application of the overlay might have resulted in a different effect had it been applied when section 476 was in a different condition.

In summary, the results show that the proposed approach adequately captures the effect of the overlay on improving section 476’s condition, as well as in capturing the induced change on the effect of traffic that affects the deterioration rate after the overlay is applied. The facts that one intervention is sufficient to estimate the condition improvement, and that the associated parameter estimate, \( h^{(2)} \), is statistically significant is somewhat surprising and highlights a fundamental difference between static and dynamic modeling. When estimating maintenance-effectiveness models (such as the ones alluded to in Section 2), only the condition improvement induced by the overlay is included in the data sample (for a total of one observation). In contrast, the estimation of \( h^{(2)} \) in time series modeling relies on comparing the condition improvement induced by the overlay to the incremental condition changes in the other periods (for a total of 56 observations). As shown above, other changes induced by the overlay on the ensuing sequence can also be identified.

5 CONCLUSIONS

In this article, we consider state-space specifications of time series models, that is, ARMAX models, as a framework to develop deterioration and inspection models for transportation infrastructure facilities. The proposed models can be combined with intervention analysis to estimate permanent and transitory effects of maintenance activities, for example, performance jumps and deterioration rate changes. The approach can be used in
situations where there are extensive time series data for infrastructure facilities, and is attractive because it circumvents technical problems, for example, the simultaneity between deterioration and maintenance, that arise in the estimation of maintenance effectiveness with static performance modeling approaches (that rely on cumulative exogenous variables to predict aggregate condition changes).

To illustrate the approach, we present an empirical example where we develop performance models for a flexible pavement section from the AASHO Road Test. Specifically, we use the statistical framework to estimate and analyze performance jumps and deterioration rate changes induced by an overlay on both surface distress, that is, rutting and slope variance, as well as on the serviceability of the pavement section. In addition to obtaining adequate estimates of the performance jumps, the empirical results provide evidence that the overlay changes the pavement’s response to traffic, that is, the overlay results in an (instantaneous) 83% reduction in the rate at which traffic damages the pavement.

In the article, and as is done in traditional time series analysis, we restrict our analysis to the facility level. This is done to highlight the differences between static and dynamic performance modeling, to facilitate the interpretation of the results, and to avoid technical and computational problems that arise when data are pooled across multiple sections. One such difficulty is caused by the presence of unobserved heterogeneities between the sections in the data set. These heterogeneities can cause seemingly identical facilities to perform differently, and in turn, motivate model validation at the facility-level as is conducted in this article (even when data from different sections are pooled in the model estimation).

The drawbacks of restricting the estimation to the facility level are that there is little or no basis for out-of-sample predictions (i.e., using a model estimated with data from one facility to predict the performance of other facilities), and that constant exogenous factors, for example, structural design, soil or material properties, and so on, cannot be estimated (because they do not exhibit variability within the data sequence). These issues are explored in Chu and Durango-Cohen (2007b) where the methodology is extended to deal with data from multiple facilities. Significantly, Chu and Durango-Cohen (2007b) generalize the methodology described herein to develop models that are transferable across facilities.

In terms of future work, it would be interesting to analyze a richer data set in order to develop improved, practical, and portable performance models. In particular, having data for more than two years may lead to models capturing seasonal or other time-dependent effects. Having reliable maintenance records would allow us to consider the effects of other maintenance actions, as well as other situations such as repeated treatments, treatment combinations, (slope vs. seasonal) changes induced on the underlying deterioration mechanism, and so on. Having access to a data set reflecting modern traffic loads and pavement design standards would lead to models that are meaningful in current practice. Unfortunately, we are not aware of the existence of such a data set.
NOTES

1. Seminal data collection efforts for (in-service) pavements are described in Paterson (1987). Overall, their structure is similar to that of the bridge data set in Indiana.

2. The data consist of deflection and pressure measurements generated by a falling weight deflectometer and by pressure sensors, respectively. The data are for asphalt pavements located on a closed-loop test track run by the Minnesota Road Research Program (MnROAD).

3. Empirical studies such as Madanat et al. (1997) and Chu and Durango-Cohen (2007b) provide evidence that serial dependence, that is, auto-correlation, can be significant in explaining condition, thereby reinforcing the importance of dynamic modeling.

4. We use the notation $E[· | ]$ to represent the expectation operator.

5. Sections with average PSI (taken over the inner and outer wheelpaths) below 1.5 were classified as failed.

6. This approach to model refinement where statistically insignificant factors are successively removed from the specification is referred to as backward elimination. For additional details, the reader is referred to Janacek and Swift (1993).

7. We use the Akaike Information Criterion (AIC) to choose between models that are statistically satisfactory. The AIC is a measure that trades-off goodness-of-fit with number of model parameters. It is widely used to assess “overfitting.” The reader is referred to Chu and Durango-Cohen (2007a) to find the formula to compute the AIC. Plots of autocorrelation and partial autocorrelation functions are commonly used to assess and compare ARMA models. This analysis, however, does not apply to ARMA models in state-space form where stationary data are not required.

8. We use the Q-test to check the independence of the residuals (Harvey, 1990).

REFERENCES


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


