Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Part B 42 (2008) 57–81

TRANSPORTATION RESEARCH PART B

www.elsevier.com/locate/trb

Estimation of dynamic performance models for transportation infrastructure using panel data

Chih-Yuan Chu [a,1], Pablo L. Durango-Cohen [b,*]

[a] Department of Transportation Technology and Supply Chain Management, Kainan University, No. 1 Kainan Road, Luchu Shiang, Taoyuan 338, Taiwan

[b] Department of Civil and Environmental Engineering and Transportation Center, Northwestern University, 2145 Sheridan Road, A335, Evanston, IL 60208-3109, USA

Received 25 July 2006; received in revised form 4 June 2007; accepted 7 June 2007

Abstract

We present state-space specifications of time series models as a framework to formulate dynamic performance models for transportation facilities, and to estimate them using panel data sets. The framework provides a flexible and rigorous approach to simultaneously capture the effect of serial dependence and of exogenous factors, while controlling for individual heterogeneity when pooling data across the facilities that comprise the panel. Because the information contained in time series and cross-section data are combined in the estimation, the ensuing performance models capture effects that are not identifiable in either pure time series or pure cross-section data. Also, pooling data across facilities leads to improved estimation results. To illustrate the methodology, we consider three classes of models for a panel of asphalt pavements from the AASHO Road Test. The models differ in the assumptions regarding the structure of the underlying mechanisms generating the data sequences. The results indicate that serial dependence is indeed significant, thereby reinforcing the importance of dynamic modeling. We also compare the specifications to assess the poolability of pavement condition data. The results provide evidence that heterogeneity among the facilities is present in the panel. Finally, we highlight features that elude existing performance models developed with static modeling approaches: the ability to estimate maintenance activities as exogenous variables, and the capability of updating forecasts in response to inspections.

Keywords: Infrastructure performance modeling; Dynamic models; Heterogeneity; Maintenance and rehabilitation; State-space models; Intervention analysis

1. Introduction

The process of allocating resources for the preservation and improvement of transportation infrastructure systems, e.g., pavement and bridge networks, requires evaluating the effect of these decisions on the

* Corresponding author. Tel.: +1 847 491 4008; fax: +1 847 491 4011.
E-mail addresses: jameschu@mail.knu.edu.tw (C.-Y. Chu), pdc@northwestern.edu (P.L. Durango-Cohen).
[1] Tel.: +886 3 341 2500x5011.

0191-2615/$ - see front matter
doi:10.1016/j.trb.2007.06.004

performance of such systems. The evaluation involves assessing and measuring surface distresses and structural properties, e.g., cracking, rutting or elasticity in pavements, and forecasting the effect of decisions on future condition. Condition forecasts are generated with performance models, which in this paper correspond to statistical expressions that relate condition to a set of explanatory variables such as design characteristics, traffic loading, environmental factors, and history of maintenance activities. In developed countries, where much of the transportation infrastructure is mature and portions are nearing the end of their service lives and need to be replaced, the aforementioned managerial decisions are increasingly important. This is due to both the far-reaching and serious negative impacts of deficient infrastructure, as well as the scale of expenditures. In the United States, for example, annual expenditures in preservation of roads and bridges are on the order of tens of billions of dollars (Office of Highway Policy Information, 2006). This, in turn, has over the last 40 years, motivated a great deal of research in the development and analysis of performance models for infrastructure facilities.

Statistical 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, i.e., heterogeneity, and time series data describing the evolution of individual facilities over time, i.e., serial dependence. Most existing data sources 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 2602 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. To exploit this structure, performance modeling approaches (cf. Butler et al., 1985; Paterson, 1987; Shahin, 2005) assume that condition can be fully represented as a function of contemporaneous exogenous variables and random error terms that describe the cross-sectional differences in the panel. These models are static (i.e., they cannot capture serial dependence), which means that condition data at different times, or for different values of the explanatory variables, are assumed to be independent. Studies such as Madanat et al. (1997) provide evidence that this assumption may not be realistic and, in turn, can lead to poor estimation results because critical information contained in the sequence of observations for individual facilities is ignored. This, for example, may explain the difficulties reported in the literature when the effects of deterioration and maintenance are considered simultaneously (refer to the discussion in Ben-Akiva and Ramaswamy (1993)), and in turn, why these factors are often estimated separately, making the models less valuable from a managerial perspective. Other unattractive features of static models are that there is no basis for generating post-sample predictions, and that since forecasting with static models involves modeling the progression of the exogenous variables, as opposed to modeling the transition of dependent variables, forecasts cannot be updated in response to data collected through inspections. The former is a well-known disadvantage of static models. The latter is relevant in the management of transportation facilities, where periodic inspections are (becoming) the norm.

In contrast to static performance modeling approaches, there have been few attempts to develop dynamic performance models for transportation facilities. In addition to the extensive data requirements, i.e., the need for long and detailed historical data for individual facilities, commonly used time series models, e.g., the Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) approach, are unattractive to model infrastructure performance because they do not include exogenous explanatory variables: design characteristics, traffic, environment, maintenance interventions, etc. Furthermore, although the variables that are included (autoregressive or moving average terms) could provide adequate prediction, they can be difficult to interpret physically.

Methodological limitations mentioned in the previous paragraph have been addressed by extending state-space models to analyze time series (see Durbin and Koopman, 2001; Harvey, 1990; Janacek and Swift, 1993 and references therein). The state-space framework provides a unified approach to model dynamic systems. Several classes of time series models (autoregressive moving average (ARMA) or structural time series) can be represented and analyzed within the framework. Among others, the advantages of state-space analysis over traditional time series methodologies include that it is easy to estimate using Kalman filter, exogenous variables and missing values are easy to deal with, and the assumption of stationary data is not required. The general state-space framework and three specifications are described in Section 3. Recently, Chu and Durango-Cohen (2007) adopt the framework to formulate and estimate performance models for individual transportation facilities. In part, the work was motivated by advances in automated data collection technologies, e.g., sensors, that allow for frequent and comprehensive inspections, and that therefore mitigate the data availability problem. Unfortunately, single-facility time series models can lead to poor estimation results (for certain facilities) due to "overfitting" the data (see Section 4.3). Moreover, this approach is unattractive for performance modeling because time series data do not provide information about factors that are constant throughout the series, e.g., structural design, soil characteristics, etc. Also, the estimation of single-facility time series models can be interpreted as the characterization of a stochastic process from a single realization. In turn, this precludes formulating general inferences about the underlying process, and consequently, there is no basis for out-of-sample prediction; that is, using a model estimated with data from one facility to predict the performance of other facilities.

1.1. Overview

We present state-space specifications of multivariate time series models as a flexible and rigorous framework to formulate dynamic performance models for transportation infrastructure, and to estimate them using panel data sets. Two key considerations need to be made when estimating performance models with panel data. The first one has to do with controlling for individual heterogeneity when pooling data across the facilities that comprise the panel. In particular, studies by Archilla and Madanat (2000, 2001), de Solminihac et al. (1999), Madanat et al. (1997) and Prozzi and Madanat (2004) have shown that heterogeneity can be a major source of variability in panel data, and therefore needs to be accounted for in a rigorous fashion. Heterogeneity refers to the assumption that facilities are different or heterogeneous, and that facility-specific factors affecting deterioration are present. The second consideration has to do with capturing the relationships between the facilities that comprise the panel. Recognizing that transportation facilities do not interact with each other, but rather are influenced by similar factors, enables us to relax the requirements of the conventional Vector AutoRegression (VAR) approach to time series modeling, which emphasizes interactions (i.e., cause-and-effect relationships). This, in turn, leads to parsimonious classes of time series models.

The above discussion motivates the development of a general performance modeling framework that fills the gap between static and VAR models as is shown in the conceptual diagram presented in Fig. 1. The framework adopted herein allows for the specification of models under different assumptions regarding the structure/heterogeneity of the underlying dynamic mechanisms that are generating the data sequences, and consequently, in how the data are pooled across the sections for estimation. In particular, we consider three classes of models: individual models (IM), seemingly unrelated time series equations (SUTSE), and single equation models (SE). IM are specified under the assumption that the deterioration of the individual facilities is instances of different and independent stochastic processes. IM models are obtained by removing the cause-and-effect relationships from VAR models. SUTSE are specified under the assumption that the deterioration of the individual facilities is instances of different stochastic processes that exhibit common elements, and that

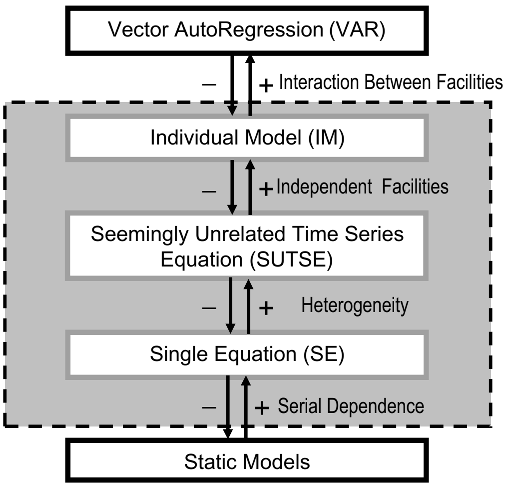

Fig. 1. Conceptual diagram for model assumptions.

are contemporaneously correlated. Heterogeneity among the facilities is captured in SUTSE through the inclusion of individual random error terms for each facility in the panel. Finally, SE models are specified under the assumption that the deterioration of the individual facilities is realizations of the same stochastic process, i.e., homogeneity is assumed and the facilities can be fully explained by the independent variables in the model. SE models differ from static models in that they capture serial dependence.

We illustrate the methodology by estimating performance models using data from a panel of 166 asphalt pavements from the AASHO Road Test (Highway Research Board, 1962). The estimation results indicate that serial dependence is significant in explaining pavement performance. The results also highlight the benefits of estimating aggregate models by pooling data across facilities. Namely, the parameter estimates are highly significant and exhibit correct/intuitive signs. Moreover, the approach allows for the estimation of (constant) factors that describe the characteristics of facilities and are not detectable in time series data. In addition to the estimation, we analyze the data fitting and predictive capabilities of the ensuing models to test for heterogeneity in the panel, which we conclude exist. Finally, we illustrate how the framework can be used to estimate the effect of maintenance activities and to update forecasts in response to condition data generated by inspections. Because the framework is consistent with the latent performance modeling approach of Ben-Akiva and Ramaswamy (1993), the capabilities of the inspection process, i.e., measurement errors and relationships between various indicators/measurements, can be estimated and properly accounted for in the forecasting process.

In addition to advancing the performance modeling literature, the proposed framework is attractive from a managerial perspective because the ensuing models can be used to support the allocation of resources for the preservation of transportation facilities. In particular, the methodology described herein provides a statistically rigorous approach to estimate the parameters that drive maintenance optimization models adopting time series 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 paper is organized as follows: An overview of the statistical performance models that serve as the basis and motivation for our work is presented in Section 2. Section 3 provides a general description of time series models in state-space form. We also present the three classes of specifications that we consider for estimation and analysis. In Section 4 we estimate pavement performance models using data from the AASHO Road Test Highway Research Board, 1962 and use them to conduct various types of analysis. Finally, a summary of the contributions of our work and some of the limitations is presented in Section 5.

2. Related work

As mentioned in the previous section, the main contribution of our work is that we present a methodology that simultaneously and rigorously accounts for serial dependence and heterogeneity. In this section we consider the relationship between the proposed framework and the literature. We also provide an overview of the latent performance modeling approach, which we adopt. A significant feature of the proposed framework is that the effect of maintenance activities can be estimated using intervention analysis. We conclude this section by discussing existing approaches to treat maintenance activities, their limitations, and the caveats that need to be considered in the analysis. Comprehensive surveys of the performance modeling literature are available from Gendreau and Soriano (1998), Hudson et al. (1997), McNeil et al. (1992) and van Noortwijk and Frangopol (2004).

Madanat et al. (1997) were first to simultaneously examine the effects of heterogeneity and state dependence on the performance of infrastructure facilities. State dependence (in discrete-state models) means that the probability that a facility's condition transitions between two states depends on the history of the process. Serial dependence, used in this paper, is the continuous-state analog to state dependence. Madanat et al. (1997) argue that heterogeneity is often mistaken as state dependence. The motivation for their work was to test the validity of the Markovian assumption, i.e., the absence of state dependence, often imposed in infrastructure performance modeling. The argument for capturing the effect of both phenomena simultaneously is that heterogeneity, if properly accounted for, does not invalidate the assumption, whereas state dependence does. Thus, it is critical to separate heterogeneity from state dependence before testing the Markovian assumption. The approach taken in Madanat et al. (1997) and other models that capture heterogeneity (cf. Archilla and Madanat, 2000, 2001; de Solminihac et al., 1999; Prozzi and Madanat, 2004) is to formulate and estimate panel data models, which provide a great deal of flexibility because slope, intercept, and error terms can be controlled for individual facilities. Random-effect specifications are usually adopted to reduce the number of parameters that require estimation. In these specifications, the parameters of individual-specific effects are assumed to be drawn from normal distributions, and thus, the estimation problem is to obtain the moments of the underlying distribution. We do point out, however, that random-effect specifications are inappropriate when random sampling (from a population) is not used, which is the case for some of the studies cited above.

In addition to presenting a modeling approach that rigorously captures the effect of heterogeneity, Madanat et al. (1997) test for state dependence. The test consists of estimating a first model without a lagged condition variable to obtain estimates of the dependent variables. A second model is then estimated with an "instrumental" lagged condition variable (obtained from the first model and treated as an exogenous variable) and the significance of the associated coefficient is examined. The results indicate that, even after controlling for heterogeneity, state dependence is probably present in the deterioration of the facilities in the panel. As noted by the authors, the second model is not statistically satisfactory. This is not surprising given that the test consisted of examining a dynamic effect with a static model. This limitation is well understood by the authors who explain that their scheme merely provides a means to test for state dependence, and should not be used to estimate dynamic performance models. Among the potential problems that arise when lagged dependent variables treated as exogenous variables are used to forecast infrastructure condition is that the prediction mean square error (MSE) is underestimated. This occurs because predictions made for multiple periods do not account for the contribution to MSE arising from the use of predicted values (Harvey, 1990). This is in contrast to the methodology described in this paper, which does provide a rigorous approach to estimate dynamic performance models, while simultaneously accounting for individual heterogeneity. We also point out that, as discussed in Section 1, the data set used in Madanat et al. (1997) has limited time series data, and therefore may be inadequate to model state dependence.

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. The framework described in Section 3 builds on the latent performance modeling approach of Ben-Akiva and Gopinath (1995), Ben-Akiva et al. (1991) and Ben-Akiva and Ramaswamy (1993), who argue that infrastructure condition can be succinctly expressed in terms of a few underlying characteristics, e.g., structural integrity, serviceability, safety, aesthetics, etc. In turn, these characteristics are represented by latent/unobservable variables to capture the ambiguity that exists in defining, and consequently in measuring condition. One or multiple condition indicators/distress measurements are related to the latent condition through a measurement model that accounts for random errors in the data collection process as well as for the relationships between different technologies and measurements. Latent performance models also include a structural model that describes the relationship between a set of factors/explanatory variables and infrastructure condition. Empirical studies by Ben-Akiva and Gopinath (1995) and Ben-Akiva and Ramaswamy (1993) have shown that latent performance models are appropriate to generate condition forecasts for transportation infrastructure, i.e., the goodness-of-fit measures are better than those reported using other statistical approaches, which is why it is used in state-of-the-art performance models, including some of the panel data models described above.

As stated above, one of the significant characteristics of the methodology presented herein is that intervention analysis can be used to estimate the effect of maintenance activities. The inclusion of maintenance in performance models is difficult, which explains why performance modeling approaches often ignore maintenance activities, i.e., observations that follow maintenance activities are excluded from the estimation. Other *ad-hoc* approaches include assuming that maintenance is constant or proportional to traffic. Ben-Akiva and Ramaswamy (1993) provide one of the first rigorous approaches to estimate the effect of maintenance. In addition to capturing the *simultaneity* of deterioration and inspection, they also include maintenance as a third process. The model relates average pavement condition over 5 years to condition indicators, maintenance activities, and other explanatory variables also given as 5 year averages. In the paper, maintenance activities are assumed to be selected and scheduled *in response to facility condition* since the data come from facilities that are in-service. Therefore, historical maintenance activities are treated as endogenous variables, and simultaneous equation models of deterioration and maintenance are estimated. The authors argue that this approach is appropriate when performance models are estimated using cross-section data for in-service facilities. The empirical results corroborate this assertion, as the models appear to be satisfactory. Representing maintenance activities as exogenous variables, i.e., ignoring the simultaneity of deterioration and maintenances, is offered as a possible explanation of why models in the literature (cf. Butler et al. (1985)) exhibit poor fit to data and incorrect coefficient signs. To their credit, Ben-Akiva and Ramaswamy (1993) recognize that when data from a controlled experiment are available, i.e., different traffic loads are run on pavements at different condition levels, and maintenance activities are decided exogenously, then maintenance activities can be represented as exogenous variables. We do point out, however, that even in a controlled experiment, static models are still inappropriate for maintenance modeling because the conditions before and after an intervention are clearly dependent (except for reconstruction), which violates the fundamental assumption of static modeling. In contrast to static performance modeling, simultaneity is not of consideration in dynamic modeling. In dynamic modeling, maintenance-effectiveness is estimated from sequences of observations (time series) that include condition improvements due to maintenance; whereas in static modeling maintenance-effectiveness is estimated by comparing cross-sectional differences in condition due to maintenance. In other words, dynamic models extract the effect of maintenance from the difference in condition for a given facility before and after a treatment, i.e., from observed condition improvements. In static modeling, the effect of maintenance is extracted by comparing facilities that are maintained with those that are not. Simultaneity must be considered to account for the fact that the differences between the facilities are not solely due to maintenance.

The aforementioned difficulties combined with the need to develop models that respond to exogenous interventions in order to support the selection of activities, have motivated various strategies to account for the effect of maintenance. Perhaps the most common approach consists of estimating separate performance and maintenance-effectiveness models. Maintenance-effectiveness models provide estimates of the condition improvement associated with various treatments. The obvious limitation of separate performance and maintenance models is that there is no basis to combine models in order to generate (long-term) forecasts because their predictions can be inconsistent. In contrast the framework presented herein does allow for the simultaneous estimation of performance and maintenance activities. In addition to providing a rigorous approach for forecasting, intervention analysis can be used to estimate maintenance effects other than condition improvements, i.e., slope or seasonal changes Harvey (1990).

In the context of maintenance-effectiveness models, Madanat and Mishalani (1998) argue that when data from in-service facilities 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 a particular (non-random) subset of samples are used to estimate effectiveness of each activity, which can result in selectivity bias. To address this problem, Madanat and Mishalani (1998) propose a multinomial logit (MNL) model to capture the agency's selection process. The results of the study show that selectivity bias can be significant, and that it can be corrected for rigorously using the factors estimated from the MNL model. We note that, although unbiased, a single estimate for the population is still not practical because it is arguable that having estimates for different strata of condition states is more useful for a maintenance decision-making situation where maintenance actions are applied to different condition states. Nevertheless, the significance of this work is to understand that unless an ideal data set that includes a wide range of variability regarding the condition states to which maintenance activities are applied is available, estimates of maintenance-effectiveness for different conditions are not available for maintenance optimization. However, the treatment for biased maintenance-effectiveness without such data set is still an unresolved problem.

3. Methodology

We adopt state-space specifications of time series models to represent the deterioration and inspection of infrastructure facilities. State-space specifications provide a rigorous and flexible framework to model a system's dynamics and to estimate the relevant parameters using panel data sets. We begin this section by presenting the general form of state-space models. We then describe three classes of specifications that we consider for estimation and analysis.

Eqs. (1) and (2) represent a state-space model of the dynamics of a panel of N facilities.

$$\mathbf{X}_{t+1} = \mathbf{g}\mathbf{X}_t + \mathbf{h}\mathbf{A}_t + \mathbf{\Omega}_{t+1} \tag{1}$$

$$\mathbf{Z}_t = \mathbf{\Lambda}\mathbf{X}_t + \mathbf{\Xi}_t \tag{2}$$

where

$X_{i,t}$: d-dimensional *state vector* representing the condition of facility i at the start of period t. Following the latent performance modeling approach of Ben-Akiva and Ramaswamy (1993), the vector components are assumed to be unobservable and can include one or more components to represent characteristics such as functional performance or structural fitness. The state vector can also include lagged dependent variables, structural components (e.g., trend and seasonality), and other elements.

$Z_{i,t}$: k-dimensional *observation vector* representing the set of condition data/indicators collected for facility i in period t. This vector may include measurements of distresses such as cracking, rutting, or raveling. The vector may also include subjective ratings or aggregate condition indices.

$A_{i,t}$: m-dimensional vector of exogenous explanatory variables, e.g., structural design, history of maintenance and rehabilitation activities, environmental factors, and traffic loading, where $A_{i,t} = [a_{i,t}^{(1)} \cdots a_{i,t}^{(m)}]'$.[2]

The above vectors are collected in $\mathbf{X}_t$, Z_t, A_t. That is, $\mathbf{X_t} = [X'_{1,t}, \ldots, X'_{N,t}]'$, $\mathbf{Z_t} = [Z'_{1,t}, \ldots, Z'_{N,t}]'$, and $\mathbf{A_t} = [A'_{1,t}, \ldots, A'_{N,t}]'$.

$\mathbf{g}, \mathbf{h}, \mathbf{\Lambda}$:
The transition matrix $\mathbf{g}$, the parameters associated with the explanatory variables $\mathbf{h}$, and the measurement matrix $\mathbf{\Lambda}$ are the parameters that describe the effects of the above variables. Because we assume that facilities do not interact with each other, the matrices are block diagonal. Depending on the specification, the elements in the matrices can be either facility dependent or independent. This point is explained further when the different specifications are considered later in this section. It is also important to mention that different characterizations of the variables may require different specifications for the parameters embedded $\mathbf{g}$, $\mathbf{h}$, and Λ, and can give rise to different classes of time series models. Chu and Durango-Cohen (2007), for example, describe the formulation of ARMA and structural time series models.

$\mathbf{\Omega}_t, \Xi_t$:
Nd and Nk-dimensional *Gaussian* random vectors that capture disturbances/aleatory uncertainty in the deterioration and inspection processes, respectively. That is, $\Omega_{i,t}$ and $\Xi_{i,t}$ are error terms in deterioration and inspection processes for facility i, and thus $\mathbf{\Omega}_t = [\Omega'_{1,t}, \ldots, \Omega'_{N,t}]'$, and $\mathbf{\Xi}_t = [\Xi'_{1,t}, \ldots, \Xi'_{N,t}]'$. They are assumed to have zero means and finite covariance matrices denoted $\mathbf{\Sigma}_\Omega$ and $\mathbf{\Sigma}_\Xi$. It is conventional to assume that $\mathbf{\Omega}_t$ and $\mathbf{\Xi}_t$ are serially independent and independent of each other for all periods (i.e., $E[\mathbf{\Omega}_t\mathbf{\Omega}_s] = 0, E[\mathbf{\Xi}_t\mathbf{\Xi}_s] = 0\ \forall t, s : t \neq s$, and $E[\mathbf{\Omega}_t\mathbf{\Xi}_s] = 0, \forall t, s$). Commonly used time series specifications, e.g., ARMA or structural time series, assume state vectors with independent components. Thus, the covariances of $\Omega_{i,t}$, Σ_{Ω_i}, are generally diagonal for all i. However, $\mathbf{\Sigma}_\Omega$ is not necessarily diagonal or block diagonal. Its properties will be discussed in the following sections. Similarly, the covariances of $\Xi_{i,t}$, Σ_{Ξ_i}, capture the relationships between the measurements for facility i. To reduce the number of parameters requiring estimation, we assume that $\mathbf{\Sigma}_\Xi$ is block diagonal, which means that the measurement errors for different facilities are uncorrelated. Although the assumption is not necessary to estimate the models, it reduces the computational burden.

The parameters to be estimated are the transition matrix $\mathbf{g}$; the parameters for the explanatory variables $\mathbf{h}$; the measurement matrix $\mathbf{\Lambda}$; and the covariance matrices $\mathbf{\Sigma}_\Omega$ and $\mathbf{\Sigma}_\Xi$. Note that as presented in Eqs. (1) and (2) and assumed hereafter the model is *time-homogeneous/time-invariant*, i.e., the parameters are not indexed by time. This assumption reduces the computational effort required for model estimation. We also note that

[2] The notation Y' denotes the transpose of vector Y. To simplify the presentation, we assume that d, k, m do not depend on the individual facilities, and that the corresponding vector components represent the same characteristics.

linear models with lagged-dependent variables are capable of capturing non-linear deterioration patterns, as is shown in Chu and Durango-Cohen (2007). An overview of the estimation process and our implementation is presented in Appendix A.

Eq. (1) is the *system equation* and Eq. (2) is the *measurement equation*. The system equation governs the evolution/deterioration of the system of facilities and describes effect of the explanatory variables. The measurement equation describes the inspection process, i.e., it relates the state-vector, $\mathbf{X}_t$, to the measurements, $\mathbf{Z}_t$, and captures random errors in the data collection process. The presence of systematic measurement errors does not invalidate the methodology since a prior estimate obtained from other measurement error analysis methods such as Humplick (1992) can be used to correct for these errors. Often distress measurements corresponding to different physical characteristics are collected, and, therefore, the number of measurements is larger than the number of latent condition variables (i.e. $k > d$). In such situations, the measurement model captures the relationships between the measurements and accounts for the fact that they are imperfect surrogates of the latent state variables.

The structure of the general state-space model depends on assumptions used to specify the parameters and random error terms. The case of single-facility time series ($N = 1$) is considered in Chu and Durango-Cohen (2007), where specifications of ARMA and structural time series models are used to estimate performance models for individual facilities. In the following subsections we present three specifications for the case of $N > 1$. As stated in Section 1 the specifications we consider differ in terms of the assumptions regarding the structure of the underlying mechanisms generating the data sequences, and consequently, in how the data are pooled across the sections for estimation.

3.1. Individual model (IM)

In this specification, facilities are assumed to be completely heterogeneous with independent deterioration processes. The specification can be represented as an instance of the general state-space model as shown below:

$$\begin{bmatrix} X_{1,t+1} \\ \vdots \\ X_{N,t+1} \end{bmatrix} = \begin{bmatrix} g_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g_N \end{bmatrix} \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{N,t} \end{bmatrix} + \begin{bmatrix} h_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & h_N \end{bmatrix} \begin{bmatrix} A_{1,t} \\ \vdots \\ A_{N,t} \end{bmatrix} + \begin{bmatrix} \Omega_{1,t} \\ \vdots \\ \Omega_{N,t} \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} Z_{1,t} \\ \vdots \\ Z_{N,t} \end{bmatrix} = \begin{bmatrix} \Lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Lambda_N \end{bmatrix} \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{N,t} \end{bmatrix} + \begin{bmatrix} \Xi_{1,t} \\ \vdots \\ \Xi_{N,t} \end{bmatrix} \tag{4}$$

where

$$\mathbf{\Sigma}_{\Omega_t} = \mathrm{Var}(\mathbf{\Omega}_t) = \mathrm{Var}\left(\begin{bmatrix} \Omega_{1,t} \\ \vdots \\ \Omega_{N,t} \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{\Omega_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Sigma_{\Omega_N} \end{bmatrix} \tag{5}$$

$$\mathbf{\Sigma}_{\Xi_t} = \mathrm{Var}(\Xi_{\mathbf{t}}) = \mathrm{Var}\left(\begin{bmatrix} \Xi_{1,t} \\ \vdots \\ \Xi_{N,t} \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{\Xi_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Sigma_{\Xi_N} \end{bmatrix} \tag{6}$$

If we expand Eqs. (3) and (4), this specification yields N single-facility time series models, each with its own set of parameters $g_i, h_i, \Lambda_i, \Sigma_{\Omega_i}, \Sigma_{\Xi_i}$. That is, $X_{i,t+1} = g_i X_{i,t} + h_i A_{i,t} + \Omega_{i,t+1}$ and $Z_{i,t} = \Lambda_i X_{i,t} + \Xi_{i,t}\ \forall i$. Since facilities are assumed independent, it follows that the covariance matrices in Eqs. (5) and (6) are block diagonal. Furthermore, the models can be estimated either simultaneously as shown in Appendix A, or individually as is done in Chu and Durango-Cohen (2007) because parameters for each facility are independent from those of other facilities. The large number of parameters and the flexibility in the estimation means that this specification can be expected provide excellent fit to data. Indeed, we use this specification to benchmark the data fitting capabilities of the pooled models (described in the following subsections). However, because data are not pooled across sections, this specification suffers from the same limitations as single-facility time series models. Significantly,

1. The estimation procedure corresponds to identifying a stochastic process from a single realization, and therefore it is not appropriate to formulate general inferences about the underlying process. For example, it is possible that the estimation yields incorrect or insignificant parameter estimates for certain facilities.
2. Constant explanatory variables, whose effect does not change over time, e.g., structural design, location, and materials, cannot be included because their effect cannot be identified from time series data.

3.2. Seemingly unrelated time series equations (SUTSE)

In seemingly unrelated time series equations (SUTSE), originally proposed by Harvey (1990), the deterioration of individual facilities are assumed to be instances of different stochastic processes that exhibit common elements, and that are contemporaneously correlated. Several SUTSE formulations can be obtained by imposing different assumptions or constraints. In this section, we present three classes of SUTSE models: SUTSE-1 a general class, SUTSE-2 consistent with the assumption that measurement errors are attributed to inspection technologies, and SUTSE-3 where a non-linear data transformation is used to capture the effect of constant explanatory variables.

SUTSE-1 is presented in Eqs. (7)–(11). In this formulation, the model parameters are facility independent, i.e., the transition matrix, g, the set of parameters for the explanatory variables, h, and the measurement matrix, Λ, are common to all facilities. Only part of the facility deterioration is captured by explanatory variables, $A_{i,t}$. The rest of the deterioration is explained by the individual heterogeneity among the facilities through individual error terms, $\Omega_{i,t}$ and $\Xi_{i,t}$. This property is called heteroscedasticity or heterogeneity of variances in statistics. As stated in Greene (2000), for panel data with large number of time periods as the case in the empirical study, time series analysis can be emphasized and the disturbance terms in the context of stochastic processes can be studies. In other words, considering heterogeneity of variances in multivariate time series models is meaningful in this setting. It is in contrast to considering heterogeneity of intercepts and slopes in panel data models (e.g. fixed-effects and random-effect models) for a large panel with small number of time periods. The measurement equation of SUTSE-1 follows Eq. (8) or Eq. (11), and it has the most general assumptions for measurement errors. In SUTSE-1, measurement errors are attributed to both inspection technologies and facilities, which means the elements on the diagonal of $\mathbf{\Sigma}_{\Xi}$ are different (as in Eq. (6)). It follows that the number of parameters that require estimation to describe measurement errors increases with the number of facilities and inspection technologies. As a result, this specification can be unattractive when a large panel is available.

$$\begin{bmatrix} X_{1,t+1} \\ \vdots \\ X_{N,t+1} \end{bmatrix} = \begin{bmatrix} g & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g \end{bmatrix} \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{N,t} \end{bmatrix} + \begin{bmatrix} h & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & h \end{bmatrix} \begin{bmatrix} A_{1,t} \\ \vdots \\ A_{N,t} \end{bmatrix} + \begin{bmatrix} \Omega_{1,t+1} \\ \vdots \\ \Omega_{N,t+1} \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} Z_{1,t} \\ \vdots \\ Z_{N,t} \end{bmatrix} = \begin{bmatrix} \Lambda & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Lambda \end{bmatrix} \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{N,t} \end{bmatrix} + \begin{bmatrix} \Xi_{1,t} \\ \vdots \\ \Xi_{N,t} \end{bmatrix} \tag{8}$$

$$\mathbf{\Sigma}_{\Omega_t} = \mathrm{Var}(\mathbf{\Omega}_t) = \mathrm{Var}\left(\begin{bmatrix} \Omega_{1,t} \\ \vdots \\ \Omega_{N,t} \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{\Omega_1} & \cdots & \Sigma_{\Omega_1\Omega_N} \\ \vdots & \ddots & \vdots \\ \Sigma_{\Omega_N\Omega_1} & \cdots & \Sigma_{\Omega_N} \end{bmatrix} \tag{9}$$

where $\Sigma_{\Omega_i\Omega_j}$ means the covariance of deterioration process errors of facilities i and j.

Alternatively,

$$\mathbf{X}_{t+1} = (I_N \otimes g)\mathbf{X}_t + (I_N \otimes h)\mathbf{A_t} + \mathbf{\Omega}_{t+1} \tag{10}$$
$$\mathbf{Z}_t = (I_N \otimes \Lambda)\mathbf{X}_t + \Xi_t \tag{11}$$

where $\otimes$ denotes a Kronecker product and I_N is an $N \times N$ identitiy matrix.

Because facilities are assumed not to interact with each other, but to deteriorate under the influence of similar exogenous factors, the off-diagonal elements in the transition matrix $\mathbf{g}$ in (Eq. (7) or Eq. (10)) have no physical meanings and are set to zero, i.e., $\mathbf{g}$ is block diagonal. Also, the structure of $\mathbf{\Sigma}_\Omega$ has two implications. If $\mathbf{\Sigma}_\Omega$ has the same diagonal property as in Eq. (5), only cross-sectional heteroscedasticity is considered in the specification. That is, heterogeneity is captured in the disturbance terms of each facility. When $\mathbf{\Sigma}_\Omega$ is not block diagonal, both cross-sectional correlation and heteroscedasticity are considered as in Eq. (9). In other words, in addition to heterogeneity, the equations are further linked by the disturbance terms. As stated in Zellner (1962), estimating regression equations together by linking them with disturbance matrix provides more efficient estimates than estimating them equation-by-equation. Note, however, that considering cross-sectional correlation is only practical when the number of facilities is small.

The assumption of facility-dependent measurement errors in SUTSE-1 seems strong when the same type of facilities are considered. To reduce the number of parameters requiring estimation, the measurement equation of SUTSE-2 is specified under the assumption that measurement errors are attributed to the inspection technologies generating the condition data as Eq. (12) shows. This restriction is equivalent to assuming that the vectors $\Xi_{i,t}\ \forall i,t$ are independent and identically distributed, meaning that the covariance matrix follows Eq. (13). The system equation of SUTSE-2 is identical to that of SUTSE-1

$$\begin{bmatrix} Z_{1,t} \\ \vdots \\ Z_{N,t} \end{bmatrix} = \begin{bmatrix} \Lambda & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Lambda \end{bmatrix} \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{N,t} \end{bmatrix} + \begin{bmatrix} \Xi_t \\ \vdots \\ \Xi_t \end{bmatrix} \tag{12}$$

where

$$\mathbf{\Sigma}_\Xi = \mathrm{Var}(\Xi_\mathbf{t}) = \mathrm{Var}\left(\begin{bmatrix} \Xi_t \\ \vdots \\ \Xi_t \end{bmatrix}\right) = \begin{bmatrix} \Sigma_\Xi & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Sigma_\Xi \end{bmatrix} = I_N \otimes \Sigma_\Xi \tag{13}$$

Unfortunately, neither SUTSE-1 nor SUTSE-2 can capture the effect of constant explanatory variables. The reason here is different than the one discussed in the context of IM specifications. Although variables such as structural design are constant over time, they are supposed to provide valuable information when common parameters of explanatory variables, h, are used and the samples from different facilities are pooled together. However, the effect of design (and other constant explanatory variables) overlaps with the (deterioration and measurement) error terms, $\Omega_{i,t}$ and $\Xi_{i,t}$ (or Ξ_t) when linear, time-homogeneous state-space models are adopted. As a result, the identification of the relevant parameters would be difficult, and the parameter estimates would be biased. To deal with this problem, the effect of structural design has to be removed (from SUTSE-1 or SUTSE-2) and attributed to the error terms. In this paper, we consider a more appealing approach (than discarding the constant variables) to deal with the collinearity while preserving the linearity of the model. The approach consists of combining/pairing the effect of a constant explanatory variable with that of another time-dependent variable. The logic is that constant explanatory variables generally represent physical characteristics of facilities, and thus it is likely that they contribute/determine/influence the effect of other explanatory variables.

The combined effect of two variables is obtained by applying a non-linear data transformation that captures the interactions between the variables. The system equation for SUTSE-3 specifications is shown in Eq. (14). The m components of $A_{i,t}$, representing the exogenous variables for facility i, are classified as either constant or time-dependent and grouped in $\widetilde{A}_i$ and $\widehat{A}_{i,t}$, respectively. $A_{i,t}$ can thus be rewritten as $A_{i,t} = [\widetilde{A}'_i, \widehat{A}'_{i,t}]'$. To obtain SUTSE-3 we let the reduced coefficient matrices that includes only the effects of the time-dependent variables, $h(\widetilde{A}_i)$, be functions of the constant variables, $\widetilde{A}_i$. For example, when we combine

r pairs of variables, the row in $h(\widetilde{A}_i)$ that corresponds to the lth component in the state vector would be $[\mathscr{F}_1(a_i^{(1)}, h^{(1+(l-1)\cdot(m-r))}), \ldots, \mathscr{F}_r(a_i^{(r)}, h^{(r+(l-1)\cdot(m-r))}),\ h^{(r+1+(l-1)\cdot(m-r))}, \ldots, h^{(m-r+(l-1)\cdot(m-r))}]$, where $\mathscr{F}_j$ is a non-linear function of the jth constant variable, $a_i^{(j)}$, and the coefficient $h^{(j+(l-1)\cdot(m-r))}$. The superindex $(j+(l-1)\cdot(m-r))$ means that the coefficient can, in general, depend on the component of the state vector, l, and on the pair of variables that are combined. Note that there are $m-r$ time-dependent variables, and that $m \geqslant 2r$ to enable the pairing. When it is assumed that a given variable has no effect on the corresponding component of the state vector, the relevant coefficient is set to zero. Different transformations can be made without violating the structure of the general state-space model or the estimation procedure. Pairing variables and specifying transformations depend on an analyst's judgment. The approach is illustrated in the empirical study presented in the following section. The specification of $\mathbf{\Sigma}_\Omega$ in SUTSE-3 is as in SUTSE-1. The measurement equation can be as in either SUTSE-1 or SUTSE-2.

$$\begin{bmatrix} X_{1,t+1} \\ \vdots \\ X_{N,t+1} \end{bmatrix} = \begin{bmatrix} g & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g \end{bmatrix} \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{N,t} \end{bmatrix} + \begin{bmatrix} h(\widetilde{A}_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & h(\widetilde{A}_N) \end{bmatrix} \begin{bmatrix} \widehat{A}_{1,t} \\ \vdots \\ \widehat{A}_{N,t} \end{bmatrix} + \begin{bmatrix} \Omega_{1,t+1} \\ \vdots \\ \Omega_{N,t+1} \end{bmatrix} \tag{14}$$

Finally, readers should keep in mind that to keep the computational effort in the estimation reasonable, analysts have to consider trade-off between capturing general behavior and capturing heterogeneity and correlation in SUTSE specifications depending on the size of the panel. When the size of panel is small, which is often the case in studies appearing in the multivariate time series analysis literature, more considerations are allowed for factors such as cross-sectional correlation in the covariance matrices as well as facility-dependent measurement errors. On the other hand, when the size of the panel is large, as is the case for the panel of pavements considered in the empirical study, capturing cross-sectional correlation and/or facility-dependent measurement errors becomes costly, and thus, capturing only cross-sectional heteroscedasticity is more practical. The advantage of large panels is that they lead to robust estimates of the common elements in g, h, Λ.

3.3. Single equation (SE)

The assumption in this specification is that data from the facilities in the panel are the realizations of the same stochastic process, i.e., different surveys of the same statistical unit. Homogeneity is assumed and the deterioration is entirely captured by the exogenous variables. The system equation for SE follows Eq. (15) or Eq. (17) while the measurement equation follows Eq. (16) or Eq. (18). Σ_Ω and Σ_Ξ follow Eqs. (19) and (13), respectively. The former implies that heterogeneity does not exist and the latter indicates that measurement errors are attributed to only inspection technologies. This specification yields a single set of parameters g, h, Λ, Σ_Ω, Σ_Ξ shared by all facilities. The most important advantages of this specification are that it supports formulating general inferences about the effect of explanatory variables on the performance of facilities, and that it is parsimonious. Constant variables can be estimated in this specification because data from facilities are pooled together, and their effects can be identified by comparing cross-sectional differences. Although SE estimates a model in pooled form, our specification is different than estimating an aggregate model that uses average data from individual facilities. In SE models, panel data remain unchanged, and goodness-of-fit is measured as the minimum overall prediction error for all data. This, in turn reduces aggregation bias.

$$\begin{bmatrix} X_{1,t+1} \\ \vdots \\ X_{N,t+1} \end{bmatrix} = \begin{bmatrix} g & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g \end{bmatrix} \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{N,t} \end{bmatrix} + \begin{bmatrix} h & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & h \end{bmatrix} \begin{bmatrix} A_{1,t} \\ \vdots \\ A_{N,t} \end{bmatrix} + \begin{bmatrix} \Omega_{t+1} \\ \vdots \\ \Omega_{t+1} \end{bmatrix} \tag{15}$$

$$\begin{bmatrix} Z_{1,t} \\ \vdots \\ Z_{N,t} \end{bmatrix} = \begin{bmatrix} \Lambda & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Lambda \end{bmatrix} \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{N,t} \end{bmatrix} + \begin{bmatrix} \Xi_t \\ \vdots \\ \Xi_t \end{bmatrix} \tag{16}$$

or,

$$\mathbf{X_{t+1}} = (I_N \otimes g)\mathbf{X}_t + (I_N \otimes h)\mathbf{A}_t + \mathbf{\Omega}_{t+1} = (I_N \otimes g)\mathbf{X}_t + (I_N \otimes h)\mathbf{A}_t + \mathbf{1}_{N\times 1} \otimes \Omega_{t+1} \tag{17}$$

$$\mathbf{Z}_t = (I_N \otimes \Lambda)\mathbf{X}_t + \mathbf{\Xi}_t = (I_N \otimes \Lambda)\mathbf{X}_t + \mathbf{1}_{N\times 1} \otimes \Xi_t \tag{18}$$

where $\mathbf{1}_{N\times 1}$ is an N-by-1 matrix of ones.

$$\mathbf{\Sigma}_\Omega = \mathrm{Var}(\mathbf{\Omega}_t) = \mathrm{Var}\left(\begin{bmatrix} \Omega_t \\ \vdots \\ \Omega_t \end{bmatrix}\right) = \begin{bmatrix} \Sigma_\Omega & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Sigma_\Omega \end{bmatrix} = I_N \otimes \Sigma_\Omega \tag{19}$$

4. Empirical study

In this section, we exploit the flexibility of the state-space model to formulate and estimate pavement performance models under the various assumptions discussed in the preceding section. The models were estimated using a panel of 166 facilities from the AASHO road test. The objectives of the empirical study are:

1. To evaluate the significance of serial dependence as a factor to explain pavement performance.
2. To present a comprehensive comparison of model specifications and, equivalently, the appropriateness of assumptions under which pavement condition data for the given data set can be pooled. The comparison is done via model fitting, in-sample prediction, and out-of-sample prediction, and is used to examine the presence of heterogeneity in the panel.
3. To highlight key advantages of dynamic performance modeling approaches over existing static approaches. In particular, we use our models estimate the exogenous effect of maintenance interventions on the performance of infrastructure facilities. We also discuss the benefits of updating forecasts in response to inspection data.
4. To illustrate the process of estimating the parameters that drive maintenance optimization models using time series models to describe deterioration (e.g. Durango-Cohen, 2007).

4.1. Data: 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 is uniform and representative of the soils in large area of the country and because the climate is typical of those in the northern United States (Highway Research Board, 1962). The study set out to evaluate the impact of structural design (surface, base and sub-base thicknesses) and traffic loading (axle applications, load and configuration) on the performance of flexible pavements. The care with which factors such as construction quality and inspection errors were controlled and the comprehensive test of important factors, including traffic loads and pavement structure configurations, explain why the data collected during this experiment is still one of the most widely used sources in the development of pavement performance models and pavement design criteria. The condition data collected through the duration of the study focused on serviceability/functional performance, and included distresses such as rut depth, slope variance, cracking, and patching.

Highlights of experimental design and relevant descriptive statistics are presented in Table 1. The test tracks consisted of six loops with two lanes each. The two lanes of loop 1 were not subjected to traffic for the purpose of comparison. The other ten lanes had approximately the same number of axle applications but different axle loading and configurations during the two-year period. In general, the traffic loading and pavement thickness increase as the loop number increases.

A total of 332 main factorial design sections were tested. Data for 308 sections are available for the analysis because the raw data from lane 2 of loop 1 are missing. Overall, the sections exhibit three failure patterns. Note that sections with average PSI (taken over the inner and outer wheelpaths) below 1.5 were classified as failed. The first pattern is for sections that do not fail during the experiment. The second pattern is for sections that failed early during the experiment. These sections were reconstructed after the failure and taken out

Table 1
AASHO road test: experiment design and failure patterns

Factor		Loop					
		1	2	3	4	5	6
Factorial test sections		48	44	60	60	60	60
Traffic loading (lb)	Lane 1	0	2000S	12000S	18000S	22400S	30000S
(S: single, T: tandem)	Lane 2	0	6000S	24000T	32000T	40000T	48000T
Surface thicknesses (in.)		1, 3, 5	1, 2, 3	2, 3, 4	3, 4, 5	3, 4, 5	4, 5, 6
Base thicknesses (in.)		0, 6	0, 3, 6	0, 3, 6	0, 3, 6	3, 6, 9	3, 6, 9
Sub-base thicknesses (in.)		0, 8, 16	0, 4	0, 4, 8	4, 8, 12	4, 8, 12	8, 12, 16
Pattern 1 (no failure)[a]		18/6	14/4	4/0	6/0	6/0	8/1
Pattern 2 (early failure)[a]	Lane 1	0/0	4/0	19/3	14/0	10/1	5/1
Pattern 3 (late failure)[a]		0/0	0/0	4/0	8/2	11/2	14/1
Pattern 1 (no failure)[a]		18/6	8/0	1/0	7/0	6/0	11/1
Pattern 2 (early failure)[a]	Lane 2	0/0	8/1	21/3	12/0	11/1	6/0
Pattern 3 (late failure)[a]		0/0	5/0	5/0	9/2	11/1	10/2

[a] The entries in the last six rows are for the number of available sections for estimation/validation.

of the test. That is no distress measurements were collected after the reconstruction. Finally, the third pattern is for sections failed toward the end of the experiment.

To obtain the panel of 166 pavements, we excluded the 120 sections following the second pattern, as well as the 22 replicate sections,[3] from the data set for estimation. The exclusion of the sections following the second pattern is somewhat arbitrary and potentially causes sampling bias; however, it is justified because the time series data collected for these sections is limited (18 observations per section on average, compared to 56 observations for the other sections). The fact that the missing values are at the end of the data set means that the (post-sample) predictive capabilities of these sections differs substantially from that of the other sections, and therefore, it may not be reasonable to pool them in the estimation. In addition, we note that the experimental design for these sections was inconsistent with the objective of causing failures toward the end of the experiment. Finally, we note that while the methodology presented in Section 3 and the estimation procedure presented in Appendix A can be modified to account for missing values as is done in Chu and Durango-Cohen (2007), dealing with a balanced panel simplifies the procedure. The sections following the third pattern were overlaid after failure. These sections are of great importance because they are useful for maintenance activity modeling. After excluding the sections following the second pattern, the ratio of number of the third pattern to the first pattern increases as the loop number increases in general, which indicate traffic loading is heavier relative to pavement thickness in higher loops.

4.2. Variables

The models described in the previous section require the specification of the variables representing the distress measurements and explanatory variables. The distress measurements as well as other factors such as traffic applications and temperature are recorded biweekly; as a result, 56 records are available for each variable and they are indexed by $t = 1, \ldots, 56$ in the model specifications. As examples, $t = 1$ indicates the data collected on 3 November 1958 and $t = 56$ indicates the data collected on 5 December 1960. The actual dates of the data collection can be found in Highway Research Board (1962). Also, the pavements selected for analysis are indexed by i, where $i = 1, \ldots, 166$. The variables are as follows:

$z_{i,t}$: Average PSI values for the inner and outer wheelpaths of section i at time t, measured in 10 PSI. The PSI is a widely accepted indicator of a pavement's serviceability, i.e., its functional performance. The index is computed as follows:

[3] The replicates have identical structural design and external conditions (traffic and environmental factors) to 22 sections included in the data set used for estimation. These sections were excluded for model validation.

$$\text{PSI} = 5.03 - 1.91 \cdot \log_{10}(1 + \text{SV}) - 1.38 \cdot \text{RD}^2 - 0.03 \cdot \sqrt{C + P} \tag{20}$$

where SV is slope variance measured in square inches, RD is the rut depth in inches, and C and P correspond to cracking and patching in square feet.Because the objective of the present study is to illustrate how panel data can be pooled, for simplicity, we restrict the number of condition indicators to one ($k = 1$). The structure of the measurement model implies that the average PSI for facility i in period t is a manifestation single-dimensional ($d = 1$) latent variable $x_{i,t}$, which may be interpreted as a pavement's true underlying serviceability. That is, in this case the model is only capturing errors in the PSI measurements. The more general case where different distress measurements are related to a single latent variable is considered in Chu and Durango-Cohen (2007).

$x_{i,t}$: State variable representing section i's serviceability at t. In the empirical study, we consider characterizations of the state variable that lead to AutoRegressive(p) or AR(p) models, which means that $x_{i,t}$ is a function of p-order lagged-dependent variables. In specific, the state vector for facility i in period t, $X_{i,t}$, is obtained by collecting the variables as follows: $X_{i,t} = [x_{i,t}, x_{i,t-1}, \ldots, x_{i,t-p+1}]'$ where the dimension of the state vector $d = p$. It follows that the structure of the transition matrix for facility i, g_i, is

$$g_i = \begin{bmatrix} \phi_{1,i} & \cdots & \phi_{p-1,i} & \phi_{p,i} \\ & & & 0 \\ & I_{p-1} & & \vdots \\ & & & 0 \end{bmatrix} \tag{21}$$

where $\phi_{j,i}$ is the coefficient associated with the j-order lagged-dependent variable. The coefficient is referred to as the ARj parameter.

The exogenous explanatory variables that we consider in our models are as follows:

SN_i: Structural number of section i. A pavement's structural number serves as a proxy for its structural design. It is given as a function of the pavement's surface, base, and sub-base thicknesses. The formula calibrated by Highway Research Board (1962) and used here is

$$\text{SN} = 0.44D_1 + 0.14D_2 + 0.11D_3 \tag{22}$$

where D_1, D_2, and D_3 represent the thickness of surface, base, and sub-base, respectively. It can be seen, that the thickness of the surface layer provides more strength than the thicknesses of the other two layers.

$\text{TRF}_{i,t}$: 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 critical factors that cause pavement deterioration. Load applications are transformed to ESALs according to the axle loads and axle configurations. 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. The motivation is to account for the fact that environmental factors can make pavements more or less vulnerable to deterioration due to traffic.

$\text{OVR}_{i,t}$: Indicator variable where: $\text{OVR}_{i,t} = \{1$, when overlay is applied on section i between time t and $t + 1$; 0, otherwise$\}$. Maintenance activities can be analyzed using *intervention analysis* in time series modeling. Interventions can be classified by their effects as transitory, slope change, or change in seasonal pattern Harvey (1990). Facility-specific factors such as structural design or other factors such as overlay thickness can also be included easily as factors explaining maintenance-effectiveness. The representation here captures an overlay's ability to improve condition (a transitory effect). However, other effects mentioned above could also be used to capture changes due to maintenance in the underlying deterioration mechanism.

4.3. Model specifications and estimation results

We consider five different model instances for estimation: IM, SUTSE-2, SUTSE-3, and two instances of SE. Because the main objective of the study is to illustrate how panel data can be pooled to estimate pavement performance models, and in order to provide a fair basis for comparison, we predetermine a common set of variables for all specifications. In particular, through preliminary experiments, we found that the autoregressive parameters of order two and above raise problems because they tend to be insignificant or redundant. See for example the models labeled "eliminated" in Table 2, where the AR2 parameter for the SE model is insignificant at the 95% confidence level. Also, SUTSE models with AR2 parameters (not listed) lead to non-invertible Hessian matrices, which is an indication of a misspecified model. Thus, we restrict our final models to AR(1) specifications.

The other principle we used in specifying and estimating models is to include as many exogenous variables as allowed by each specification (without regard to parameter signs or level of significance). Thus, the SE specification includes TRF, OVR, and SN. SUTSE-2 and SUTSE-3 include TRF and OVR. In addition, the effect of SN was included in the SUTSE-3 specification by applying a non-linear data transformation. It is believed that thicker pavements suffer smaller damage for the same traffic loads. Therefore, we use the constant variable, SN_i, to create a non-linear function for the parameter of the time-dependent variable, $\mathrm{TRF}_{i,t-1}$. Examples of the transformation using exponent and power functions are shown in Eqs. (23) and (24). In each equation $h^{(1)}$ and $h^{(2)}$ are the parameters to estimate. We found that Eq. (24) is the most appropriate transformation for this empirical example. The results will be presented and the choice will also be validated later in this section. Moreover, to reduce the number of parameters, the measurement model in SUTSE-3 specification follows that of SUTSE-2 as in Eq. (12), which assumes facility-independent measurement error

$$h(\widetilde{A}_i)\widehat{A}_{i,t-1} = [\mathscr{F}_1(\mathrm{SN}_i, h^{(1)}),\ h^{(2)}]\begin{bmatrix}\mathrm{TRF}_{i,t-1}\\ \mathrm{OVR}_{i,t-1}\end{bmatrix}$$

$$h(\widetilde{A}_i)\widehat{A}_{i,t-1} = e^{h^{(1)}\mathrm{SN}_i}\mathrm{TRF}_{i,t-1} + h^{(2)}\mathrm{OVR}_{i,t-1} \tag{23}$$

$$h(\widetilde{A}_i)\widehat{A}_{i,t-1} = \mathrm{SN}_i^{h^{(1)}}\mathrm{TRF}_{i,t-1} + h^{(2)}\mathrm{OVR}_{i,t-1} \tag{24}$$

Table 2
Estimation results

Parameter	Benchmark	Final		Eliminated	
	IM estimate[a]	SUTSE-3 Estimate[a]	SE (AR1) Estimate[a]	SE (AR2) Estimate[a]	SUTSE-2 Estimate[a]
σ_ϵ	(88%)[e]	(100%)[b]	1.852 (64.3)	1.832 (48.8)	(100%)[b]
σ_ξ	(72%)[b]	1.071 (48.9)	1.112 (39.3)	1.148 (34.4)	1.067 (48.5)
AR1, ϕ_1	(100%)[b]	0.995 (2250.4)	0.984 (552.8)	0.999 (63.0)	0.995 (2223.5)
AR2, ϕ_2	–	–	–	−0.014 (−0.9)[d]	–
SN	–	–	0.067 (3.6)	0.061 (3.3)	–
TRF	(27%)[b]	$-\mathrm{SN}^{-1.062}$ (−16.8)	−0.207 (−8.3)	−0.208 (−8.3)	−0.198 (−10.0)
OVR	(86%)[b]	15.349 (46.3)	15.739 (56.7)	15.675 (55.6)	15.278 (46.1)
No. of parameters	723	170	6	7	170
LL[b]	−19085.4	−19997.8	−20823.8	–	−20004.1
AIC[e]	713.372	726.206	749.851	–	726.433
SSE[e]	41072.1	51228.3	51141.3	–	51282.1
Generalized PEV[e]	1.850E+96	1.177E+111	1.609E+124	–	1.501E+111
Q-test[c]	158/166 (95%)	150/166 (90%)	138/166 (83%)	–	150/166 (90%)

[a] The dependent variables, $x_{i,t}$, are measured in 10 PSI. The values in the parentheses are t-statistics.

[b] LL: log likelihood, AIC: Akaike information criterion, SSE: One-step prediction sum of square errors, generalized PEV: generalized prediction error variance. SSE is measured in PSI^2 and generalized PEV are measured in PSI^{332}. The definitions and formulas to compute these statistics can be found in Harvey (1990).

[c] Q-test/Portmanteau test: The single-facility version is used and the percentages of facilities that passed the test are reported.

[d] Entries in bold-face type indicate that a parameter is insignificant at the 95% confidence level.

[e] When facility-specific parameters are used, the estimates are not reported due to space limitation. Instead, the percentages of parameters that are significant at 95% level are listed in the parentheses.

Finally, the IM specification includes only the time-dependent exogenous variables: OVR for sections that were overlaid, and TRF for sections that were loaded with traffic. A summary of the estimation results is presented in Table 2.

Parameter estimates for IM specification, and estimates of the deterioration process error for the SUTSE specifications are omitted from Table 2 due to space limitations. To make the results and discussion tangible we present the final specifications below:

$$\text{IM}:$$
$$x_{i,t} = \phi_{1,i} x_{i,t-1} + h_{2,i}\text{TRF}_{i,t-1} + h_{3,i}\text{OVR}_{i,t-1} + \omega_{i,t} \tag{25}$$
$$z_{i,t} = x_{i,t} + \xi_{i,t} \tag{26}$$

$$\text{SUTSE-2}:$$
$$x_{i,t} = 0.995 x_{i,t-1} - 0.198\text{TRF}_{i,t-1} + 15.278\text{OVR}_{i,t-1} + \omega_{i,t} \tag{27}$$
$$z_{i,t} = x_{i,t} + \xi_t, \;\; \xi_t \sim N(0, 1.067^2) \tag{28}$$

$$\text{SUTSE-3}:$$
$$x_{i,t} = 0.995 x_{i,t-1} - \text{SN}_i^{-1.062}\text{TRF}_{i,t-1} + 15.349\text{OVR}_{i,t-1} + \omega_{i,t} \tag{29}$$
$$z_{i,t} = x_{i,t} + \xi_t, \;\; \xi_t \sim N(0, 1.071^2) \tag{30}$$

$$\text{SE}:$$
$$x_{i,t} = 0.984 x_{i,t-1} + 0.067\text{SN}_i - 0.207\text{TRF}_{i,t-1} + 15.739\text{OVR}_{i,t-1} + \omega_t,$$
$$\omega_t \sim N(0, 1.852^2) \tag{31}$$
$$z_{i,t} = x_{i,t} + \xi_t, \xi_t \sim N(0, 1.112^2) \tag{32}$$

The results highlight the advantages of each of the specifications. In particular, the benchmark model, IM specification, provides significantly better fit to data than the pooled models (as measured by LL, SSE and Generalized PEV). Somewhat surprisingly, we observe that IM specification is also superior to the pooled models based on the AIC, which considers trade-off between goodness-of-fit and number of parameters. Among the pooled models, the SUTSE specifications dominate SE model, which suggests that heterogeneity is present in the panel. The advantages of the pooled models are also apparent as they permit the estimation of constant exogenous factors. Moreover, we observe that the parameter estimates are highly significant (at the 95% confidence level) and have correct signs. Another indication that the pooled models adequately capture the effect of the explanatory factors is that the different specifications only lead to modest changes in the parameter estimates. Significantly, we point out that the AR1 parameters are highly significant for all specifications, which means that serial dependence plays an important role in explaining pavement performance. This, in turn, reinforces the importance of developing dynamic performance models.

The estimation results also point to the benefits of intervention analysis (and dynamic modeling) to represent and estimate the (transitory) effect of maintenance activities, overlays in this case. The effect of overlay is estimated approximately 1.5 PSI for SUTSE and SE models, which is a reasonable value based on the actual observation that actual increases of PSI average 2.2 PSI and over 90% of them range from 1.1 to 3.3. The fact that the results are correct/intuitive is significant given that previous research (using the same data set) either has ignored the effect of maintenance or has reported difficulties in the estimation (see Highway Research Board (1962) and numerous subsequent studies). To connect the results to the discussion on *selectivity bias* presented in Section 2, we note that the maintenance-effectiveness estimates apply to failed pavements (PSI $\leqslant$ 1.5). Thus, strictly speaking, the estimates may be biased. This, however, is not important in the empirical analysis because we only use the estimates to predict the condition improvement for failed pavements, and in turn why it is not necessary to correct for selectivity bias, in this particular case. In the case of maintenance optimization, where the estimates of improvements in different pavement conditions are required, a different data set and more modeling considerations are necessary. When overlays are applied to pavements in different conditions, we might, for example, consider multiple maintenance variables to reflect condition-dependent condition improvements. This extension would be straightforward in dynamic modeling because current condition is one of the inputs for predicting future condition.

The estimation results for IM specification are different than those for the pooled models. The results are expected and consistent with identifying stochastic processes from a single realization (i.e. unreliable parameter estimates). In particular, we observe that the absence of (significant) traffic, and hence deterioration, in loops 1 and 2, leads to insignificant parameter estimates, which means that the sections' performance is best represented as a random noise process. Loops 3 and 4 are only subjected to light traffic loading, and thus the parameter estimates (for the exogenous variables) are often insignificant or have incorrect signs. In Table 2, the fact that only 27% of TRF parameters are statistically significant suggests that it is naive to expect that these factors alone can fully explain the deterioration process of all pavements (a premise of static modeling approaches). To corroborate this assertion, we note that the AR1 variable is capable of producing satisfactory predictions without the exogenous variables for these sections. As a result of the heavy traffic loading in loops 5 and 6, the effects of the explanatory variables tend to be captured accurately.

In Table 2, we also report the results of the Q-test, which examines the serial independence of the residuals, and hence, is an indication of satisfactory models. The multivariate Q-test is not applicable to our specifications because total number of measurements (166) is larger than the number of time periods (56). This structure leads to sample covariance matrices that are singular. Instead, we test facilities individually and report the percentage that exhibit serially independent residuals. Overall, the models perform satisfactorily. As expected, the models with the greatest flexibility have the highest percentage of facilities passing the Q-test.

4.4. Model validation and comparison

In this section, we compare the models presented in the previous subsection. The comparison, in turn, provides empirical evidence to assess the poolability of pavement performance data from a panel data set. In addition to measures of goodness-of-fit that are often the sole selection criteria appearing in the performance modeling literature, we evaluate models based on their in-sample predictive capabilities (i.e., reproducing data and predicting failures). We validate our models by conducting out-of-sample testing using the replicate sections in the data set. Out-of-sample testing is attractive in this situation because in addition to testing a model's forecasting capabilities, it is consistent with the objective of formulating general inferences about deterioration. We begin this section by justifying our selection of SUTSE-3 as the final SUTSE specification, and hence as the representative model capturing heterogeneity in the panel.

The selection of SUTSE-3 over SUTSE-2 was based on both goodness-of-fit and practical appeal. Both specifications have the same number of parameters and are estimated with the same data. The statistics reported in Table 2 suggest that the SUTSE-3 is better from a goodness-of-fit standpoint, although the difference is not overwhelming. The benefit of adding SN as an explanatory variable can be analyzed by examining the two specifications (see Eqs. (27) and (29)). The difference between the specifications is that in SUTSE-2 the effect of TRF on condition does not vary with SN (the coefficient is constant and equal to -0.198); whereas in SUTSE-3, it does (TRF's coefficient is equal to $-\mathrm{SN}^{-1.062}$). In Fig. 2, we graph the coefficients for the two specifications as a function of SN. We observe that the coefficient for SUTSE-3 is higher when SN is small (i.e. weak pavement), and lower when SN is high (i.e., strong pavement). This relationship is consistent with the physical characteristics of asphalt pavements, and therefore, makes SUTSE-3 appealing from a practical perspective. Table 2 shows that overall prediction SSE decreases as a result of introducing SN variables.

In the remainder of this section we test the predictive capabilities of the final specifications: IM, SUTSE-3 and SE. The objective is to provide empirical evidence of the poolability of panel data to model infrastructure performance. In-sample testing is used to assess the specification's ability to reproduce, i.e., capture the variability in, the data set used for estimation. Out-of-sample testing is used to assess "overfitting", as well as the specifications' ability to provide information about the underlying process. The validation data (not used in the estimation) consist of the 22 replicate sections described in Section 4.1. The replicates include 6, 4, 0, 4, 3, and 5 sections, respectively, from loops 1 through 6. All of the sections from loop 4 and 5 were overlaid during the experiment. Three out of five sections from loop 6 were overlaid, and no sections from loop 1 or 2 were overlaid.

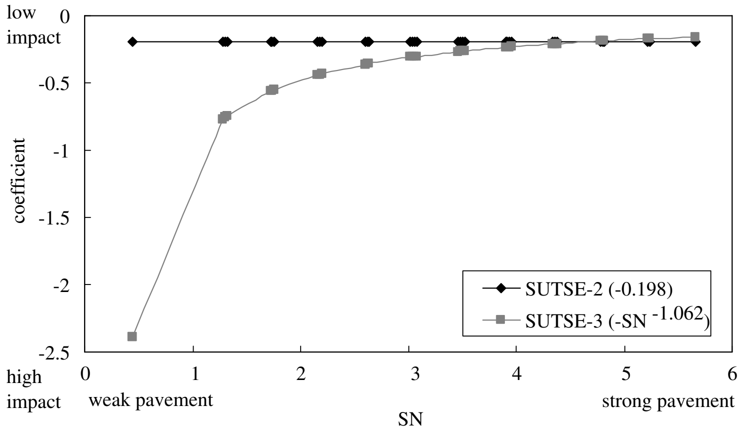

Fig. 2. Effect of non-linear structural number component.

The test we conduct consists of assessing the ability of each specification to predict deterioration over the second year of the experiment. This choice is somewhat arbitrary, although, we note that performance over the second year is more variable than in the first, which makes the test more interesting. Also, the majority of the failures in the panel occurred during the second year and one of the criteria that we use to test the specifications is their ability to predict failures. A one-year forecast horizon is sufficient from a managerial perspective because agencies are likely to allocate resources for pavement preservation on a yearly basis. Also, a one-year forecast horizon is the minimum required to assess seasonal effects on forecasting capabilities. When analyzing the results in absolute terms, it is important to recall that pavements in the AASHO road test were subjected to accelerated traffic loading. Therefore, the one-year horizons in our tests correspond to more than one calendar year. The test results are presented in Table 3 and in Fig. 3.

Root mean square error (RMSE) and correct failure prediction are criteria used to evaluate predictive capability. Recall that in the AASHO road test sections with average PSI of inner and outer wheelpath below 1.5 were classified as failed. Some of the sections have severe distress in one of the wheelpaths and they were considered failed and overlaid before average PSI reaches 1.5. Therefore, the trigger of overlay is set to 2.0 in this analysis. Note that some of the sections failed were not overlaid immediately, which is unpredictable for the models and can only be attributed to error terms. Each entry in the correct failure prediction row of Table 3 corresponds to the percentages of sections where a specification correctly predicts failures or lack thereof within the year. Incorrect failure predictions (i.e. the complement of the percentages reported in Table 3) correspond exclusively to situations where unanticipated failures were observed, as there are no cases where predicted failures did not realize.

The pooled models exhibit noticeably worse in-sample predictive capability than the IM specification: SUTSE-3 and SE, respectively, exhibit 15% and 21% more RMSE error than IM. This is explained by IM's additional flexibility. This does not seem to indicate that the (pooled) models do not adequately capture variability in the data set because the magnitude of the RMSE is small. Also, the pooled models are comparable to IM in yielding correct failure predictions. To further assess the reliability of the predictions, we consider the RMSE evolution over the forecast horizon. We also calculate the percentage of sections with

Table 3
Second-year (54-week) predictions

Measure	In-sample prediction			Out-of-sample prediction		
	IM	SUTSE-3	SE	IM	SUTSE-3	SE
Root mean square error (PSI)	0.53	0.61	0.64	0.64	0.56	0.67
Correct failure prediction (%)	91.0	86.1	84.9	100	86.4	86.4

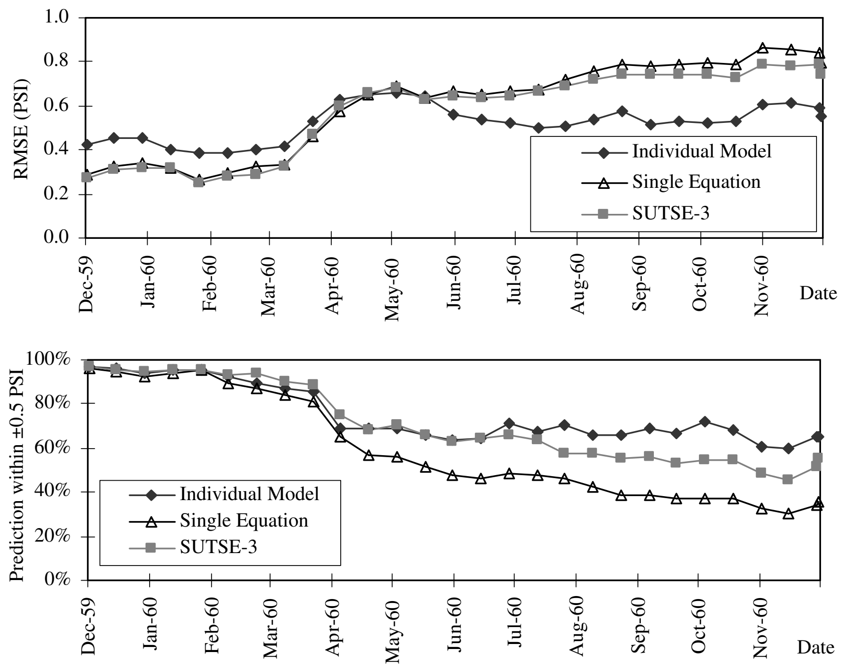

Fig. 3. Upper: Section-average of prediction RMSE (PSI). Lower: Percentage of sections with prediction error within ±0.5 PSI. (estimation data).

prediction errors within ±0.5. The latter measure is important in practice because pavement PSI (ranging from 0 to 5) is often categorized into 5 levels, and thus predictions within ±0.5 would seem to be acceptable. The results are presented in Fig. 3. There are four major sources of error in the prediction: transition process, measurement process, the effect of overlay, and the timing of failure. Since IM captures measurement error and effect of overlay individually and thus more accurately, these two sources of error are less likely to occur in IM. However, the overlay after failure improves the condition drastically. When the prediction of timing of failure and overlay is not exact, the prediction errors between the predicted date and observed date of failure would be enormous comparing to other sources of error. Furthermore, since this error is proportional to the effect of overlay, the prediction error of IM is often overestimated, especially for the sections with large PSI improvement due to the overlay. Since the intercepts of the models in the figure are not attributed to the deterioration process error, the above arguments also explain the counterintuitive fact that the intercept of RMSE of IM is larger than those of SUTSE-3 and SE in Fig. 3. On the other hand, as the prediction continues over time, the transition of performance contributes more error than the measurement process and maintenance prediction for SUTSE-3 and SE, which explains the fact that the prediction error is significantly large for SUTSE-3 and SE in the end of prediction.

Fig. 3 corroborates that the models adequately capture the variability in the data set. At first glance, the results in Fig. 3 do not appear to be stellar for a short, one-year forecast horizon (35–65% of predictions within ±0.5 PSI). We emphasize, however, that pavements in the AASHO road test were subjected to accelerated traffic loading. Overall, the results indicate that the pooled models adequately reproduce the original data. SUTSE-3 reproduces samples as well as IM for the first seven months. SUTSE-3's performance degrades gradually thereafter. SE performs adequately for five months and then the performance worsens rapidly. At the end of the forecast horizon, SUTSE-3 and SE, respectively, are 10% and 30% worse than IM. An interesting observation is that the prediction capabilities degrade rapidly in the spring. This, of course, is consistent with the fact that freeze-thaw cycles deteriorate pavements, and therefore, cause variability in pavement performance. This seasonal effect is captured in the seasonal weighted ESAL and purposely not included in the model formulations presented herein. A discussion of how such seasonal effects can be incorporated in state-space models is presented in Chu and Durango-Cohen (2007).

The results for the out-of-sample testing with validation data are shown in Table 3, and in Fig. 4. The tests are the same as those for in-sample prediction, described above. Generally, there is no basis for out-of-sample prediction with IM specifications since each facility is an independent stochastic process. However, the fact that each replicate section is identical to a section in the original data set allows us to assess the out-of-sample predictive capabilities with validation data of the IM specification. From the figure the deterioration patterns of out-of-sample testing are similar to those for in-sample testing. The first key different between the two tests is that SUTSE-3 outperforms the IM in most of the time periods, which is consistent with the results in Table 3 that SUTSE-3 has the lowest RMSE. The second observation is that both IM and SE's performances decline substantially relative to SUTSE-3. IM's diminished performance may be an indication of "overfitting", although it may also be attributed to the presence of heterogeneity leading to replicates that are not identical to the original sections. This reinforces the argument that out-of-sample forecasts with the IM specification are not rigorous. SE's performance invalidates the assumption of a universal pavement deterioration process. SUTSE-3 captures both general behavior and heterogeneity of transportation infrastructure and avoids the problem of "overfitting", which is the explanation of superior performance in out-of-sample testing.

4.5. Updating forecasts through inspections

As discussed earlier, one of the important features of dynamic models is that because they capture serial dependence, it is possible to update forecasts in response to data collected through inspections. This is particularly appealing in situations where (periodic) inspections are the norm, such as in the management of transportation infrastructure. In contrast, static models are specified under the assumption that condition can be fully explained by exogenous explanatory variables, and thus there is no basis to update forecasts in response to inspection data. Fig. 5 illustrates how inspection data impact SUTSE-3's forecasting capability. The setup for the test was similar to the in-sample prediction test described in the previous subsection. The only difference is that the condition data collected on 1 June 1960, 28 weeks after the start of one-year prediction, was used to update the forecasts. In the figure, the inspection data improves RMSE by 0.3 PSI and the percentage of sections with prediction errors within 0.5 PSI increases by approximately 20%.

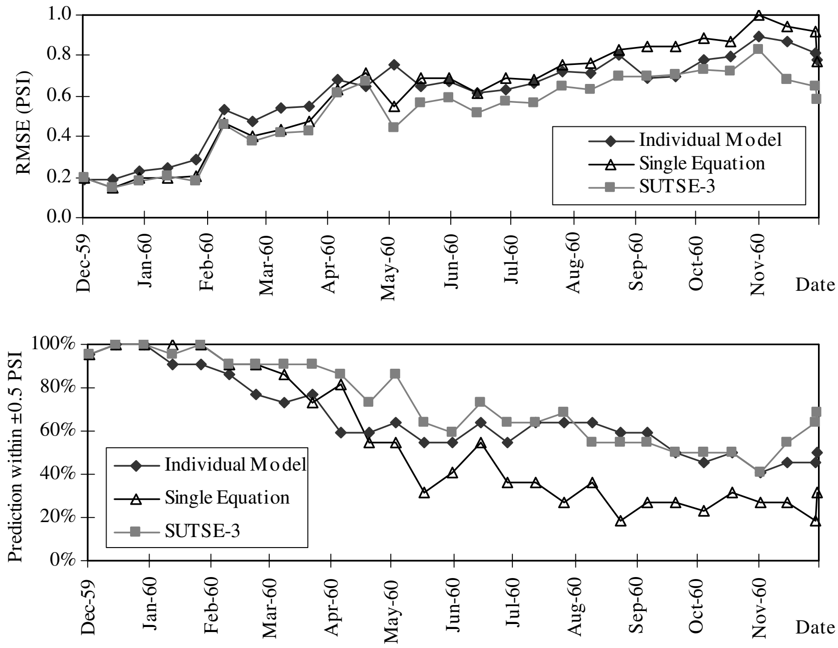

Fig. 4. Upper: Section-average of RMSE (PSI). Lower: Percentage of sections with prediction error within ±0.5 PSI (validation data).

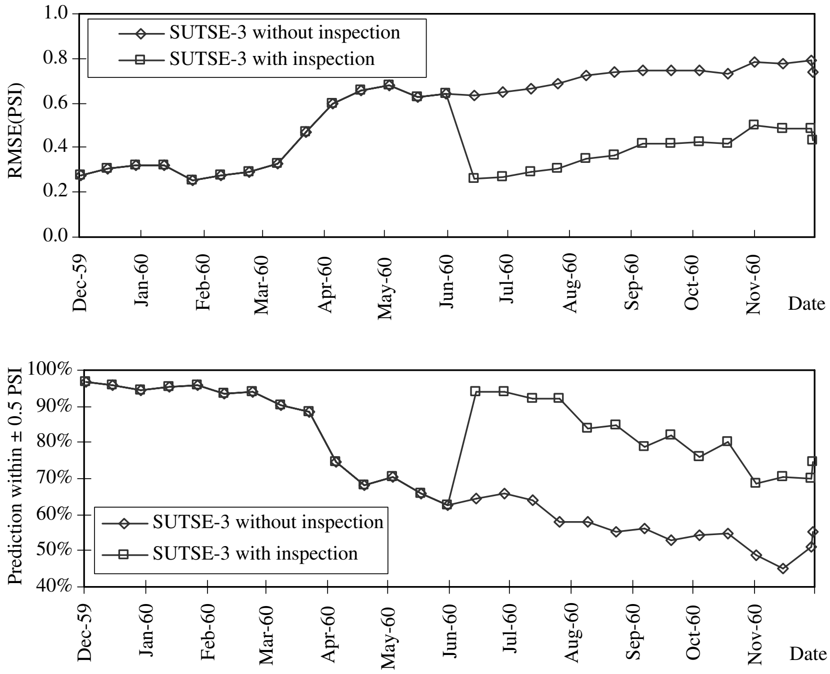

Fig. 5. Upper: Effect of inspection on in-sample prediction RMSE. Lower: Percentage of prediction error within ±0.5 PSI.

5. Conclusions and discussion

In this paper, we present state-space specifications of multivariate time series models as a general framework to formulate dynamic performance models for transportation facilities, and to estimate them using panel data sets. The framework simultaneously captures the effect of serial dependence and of exogenous factors while controlling for individual heterogeneity when pooling data across the facilities that comprise the panel. The latter is achieved by formulating performance models under various assumptions regarding the structure of the underlying mechanisms generating the data sequences. The proposed methodology has several advantages over the existing performance modeling approaches. In particular, because the information of time series and cross-section data are combined in the estimation, the ensuing models capture effects that cannot be identified otherwise, e.g., serial dependence in pure time series data or constant exogenous factors in cross-section data. The fact that serial dependence is shown to be statistically significant reinforces the need for dynamic performance modeling.

In addition to describing the methodology in detail, we estimate instances of three classes of performance models for a panel of 166 asphalt pavements from the AASHO road test. The results highlight the benefits of pooling data across facilities in the estimation as the parameter estimates are highly significant and exhibit correct/intuitive signs. In addition, we conduct a comprehensive comparison of the data fitting and predictive capabilities of the models. The results provide empirical evidence of the presence of heterogeneity in the panel, and in turn invalidate the SE specification's assumption of a universal deterioration process. Heterogeneity is captured as individual error terms for each of the facilities in SUTSE specifications. Due to the linearity and time-homogeneity of the state-space model, the effect of these error terms overlaps with constant explanatory variables, representing (physical) characteristics such as structural design for SUTSE-1 and SUTSE-2 specification. This problem can be addressed by employing non-linear data transformations combining the effect of different variables as SUTSE-3 specification. We illustrate the technique in the paper by estimating a model that combines the effect of traffic and structural design. In addition to addressing a technical problem, the data transformation leads to a more appealing model. Finally, our out-of-sample test indicates that SUTSE-3 performs better than IM, which suggests that "overfitting" can be a problem when model evaluation and selection rely exclusively in-sample performance measures.

Throughout the paper, we highlight features that elude existing performance models developed with static modeling approaches. One of the attractive characteristics of the methodology described herein is that intervention analysis to estimate performance models that capture the effect of maintenance activities. We illustrate the technique by estimating the (transitory) effect of overlays (on failed facilities). We also consider the benefits of updating condition forecasts in response to inspection data.

Among the limitations of the present study, we note the exclusion of a large number of sections exhibiting early failures. This exclusion was justified in order to maintain a balanced panel data set, and thus, to simplify the implementation of the estimation procedure. The exclusion is not unreasonable because the number of observations for the excluded sections differs significantly from the sections in the sample. Strictly speaking, however, the impact of the exclusion on the results should be tested rigorously. Another important limitation in the analysis is that the raw data collected in the AASHO road test are not utilized. First, as mentioned in Section 3, using an aggregate rating, PSI, instead of rut depth and slope variance does not fully exploit the latent performance framework. In this case, the model only captures errors in the PSI measurements. The advantage of selecting the PSI as the only dependent variable is that the ensuing models can be compared with numerous studies that use the same data set and dependent variable. Second, thickness, traffic loads, axle configuration, and environmental factors are aggregated into the SN and TRF variables that were estimated in Highway Research Board (1962). This leads to biases because the original estimation was not rigorous. Finally, the number of facilities in the out-of-sample testing (22) is much smaller than the number in the in-sample testing (166). This is an outcome of the original experimental design; however, as a result the conclusions drawn from this type of testing are not as strong as the results obtained from the in-sample testing.

In terms of future work, it would be interesting to analyze a richer data set in order to develop improved performance models. In particular, having data for more than two years would lead to models capturing seasonal or time-dependent effects. Having reliable and variable maintenance records would allow us to consider maintenance-effectiveness for different facility conditions, which is critical for maintenance decision-making, and effects beyond condition improvements, e.g., treatment combinations, (slope or seasonal) changes induced on the underlying deterioration mechanism, etc.

To conclude, we note that the work presented herein is an interesting and relevant application of state-space modeling techniques for the formulation and estimation of time series models. The approach is not widely used in traditional time series analysis where, often, only a single data sequence is available. The framework can be particularly useful in the analysis of panel of experimental data sets, which are common in engineering and other disciplines.

Acknowledgements

This work was partially supported by the Northwestern University Transportation Center through a Dissertation Year Fellowship award to the first author, and by the National Science Foundation through Grant 0547471 awarded to the second author. The work was done while the first author was a graduate student at Northwestern University.

Appendix A. Overview of estimation procedure

This section provides an overview of the procedure used to estimate the parameters of the models described in Section 3. We also describe how critical issues are addressed in our implementation. In particular, we describe the evaluation of likelihood function, the Kalman filter, and the evaluation of the score vector.

A.1. Likelihood function

We adopt maximum likelihood estimation (MLE) due to the asymptotic properties of the ensuing estimates. In time series the observations are assumed to follow a joint distribution, $f(\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_{T-1}}, \mathbf{X_T})$. The assumption required for MLE to estimate time series models is that $\mathbf{\Xi}_t$ and $\mathbf{\Omega}_t$ follow multivariate normal distributions. It follows that the conditional distribution of $\mathbf{X}_t$ given data up to $t-1$ is also normal. Based on the above assumption, the loglikelihood function of a state-space model can be formulated as Eq. (A.1). A detailed derivation can be found in Harvey (1990).

$$\log \mathscr{L} = -\frac{N(T-d)}{2}\log 2\pi - \frac{1}{2}\sum_{t=1+d}^{T}\log\det(\mathbf{F}_t) - \frac{1}{2}\sum_{t=1+d}^{T}\mathbf{V}_t'\mathbf{F}_t^{-1}\mathbf{V}_t \tag{A.1}$$

where

$\mathbf{V}_t$ is one step prediction error at time t
$\mathbf{F}_t$ is the prediction error variance (PEV) at time$t$. det($\mathbf{F}_T$) is called generalized PEV and is a single measure of goodness-of-fit.
N is the total number of facilities.
T is the number of time periods.
d is the cardinality of the state vector of individual facility.

The evaluation of the above function, depends on values of $\mathbf{V}_t$ and $\mathbf{F}_t$. These values can be generated by the Kalman filter described in the next subsection.

A.2. Kalman filter algorithm

The Kalman filter is an algorithm that updates the first two moments of the conditional distribution of $\mathbf{X}_t$ given data up to $t-1$. Recall that under the assumptions discussed in the previous subsection, the conditional distribution is multivariate normal. $\mathbf{P}_t$ denotes the covariance matrix of state vector at time t. In the estimation, it is necessary to use the Kalman filter to compute $\mathbf{V}_t$ and $\mathbf{F}_t$, which appear in the log-likelihood function.

In terms of the implementation, $\mathbf{X}_1$ and $\mathbf{P}_1$ are required to initialize the filter. We adopt the diffuse initial values approach; that is, $\mathbf{X}_1 = \mathbf{0}$ and $\mathbf{P}_1 = kI_{Nd}$ where k is a large number. This implies that the knowledge of initial values is unavailable and they are calculated using the first d period(s) of observations (Nd observations total). It also explains why the same time periods of the prediction errors are not counted in Eq. (A.1).

Kalman filter algorithm

For $t = 1,\ldots,T$:
$\widehat{\mathbf{Z}}_t = \mathbf{\Lambda X_t}$
$\mathbf{V}_t = \mathbf{Z_t} - \widehat{\mathbf{Z}}_t$
$\mathbf{F}_t = \mathbf{\Lambda P_t \Lambda}' + \mathbf{\Sigma_\Xi}$
$\mathbf{K}_t = \mathbf{gP_t\Lambda' F}_t^{-1}$
$\mathbf{L}_t = \mathbf{g} - \mathbf{K}_t\mathbf{\Lambda}$
$\mathbf{X_{t+1}} = \mathbf{gX}_t + \mathbf{hA_t} + \mathbf{K}_t\mathbf{V_t}$
$\mathbf{P_{t+1}} = \mathbf{gP}_t\mathbf{L_t'} + \mathbf{\Sigma_\Omega}$

A.3. Optimization routine

The estimation problem consists of identifying parameters that maximize the log-likelihood function, i.e., Eq. (A.1). We solve the estimation problem with the non-linear unconstrained optimization routine in MATLAB. The implementation, described below, involves a hybrid analytical and numerical approach to evaluate the score vector.

Unconstrained optimization is sufficient for maximizing the log-likelihood function for the following reasons. First, one major advantage of the specifications using Kronecker product in Section 3 is that no constraints fixing parameter values are required. Second, we use transformations to avoid imposing constraints that guarantee positive-definite covariance matrices. The transformations are based on the matrix factorizations as shown in Eqs. (A.2) and (A.3), where $\mathbf{\Sigma_\Xi^L}$ and $\mathbf{\Sigma_\Omega^L}$ are lower-triangular matrices.

$$\mathbf{\Sigma}_{\Xi} = \mathbf{\Sigma}_{\Xi}^{\mathbf{L}'}\mathbf{\Sigma}_{\Xi}^{\mathbf{L}} \tag{A.2}$$

$$\mathbf{\Sigma}_{\Omega} = \mathbf{\Sigma}_{\Omega}^{\mathbf{L}'}\mathbf{\Sigma}_{\Omega}^{\mathbf{L}} \tag{A.3}$$

The gradient or score vector of the loglikelihood function is critical when estimating the models using optimization routines. In general, to accelerate the procedure, computing the score vector analytically is preferred to numerically differentiating the likelihood function such as finite difference approximation. The analytical approach to compute the score vector is proposed by Engle and Watson (1981) and Koopman and Shephard (1992). However, for state-space model, computing the score vector analytically involves the computation of a considerable number of matrix multiplications, which could be more time-consuming than numerical approach if the state vector is large. Therefore, Koopman and Shephard (1992) concludes that when unknown parameters exist in the matrix **g**, **h**, and **Λ**, it is better to evaluate the score vector numerically than analytically.

If the parameters in $\mathbf{\Sigma}_{\Omega}$ and $\mathbf{\Sigma}_{\Xi}$, denoted by the vector μ, are independent from the parameters in **g**, **h**, and **Λ**, denoted by the vector θ, a rapid way to determine the score vector for the parameters in $\mathbf{\Sigma}_{\Omega}$ and $\mathbf{\Sigma}_{\Xi}$ is proposed by Koopman and Shephard (1992). This case is true for almost all the state-space specifications, including all the specifications in this paper. The score vector for the parameters in **g**, **h**, and **Λ** can then be determined using standard numerical approaches. Eqs. (A.4)–(A.7) are the results of the disturbance smoother by Koopman (1993). The smoother evaluates the minimum mean squared error estimator of the disturbance vector and the smoothed disturbances are used to evaluate the score vector for the parameters in the covariance matrices, $\mathbf{\Sigma}_{\Omega}$ and $\mathbf{\Sigma}_{\Xi}$ via Eq. (A.8).

$$\mathbf{e}_t = \mathbf{F}_t^{-1} - \mathbf{K}_t'\mathbf{r}_t \tag{A.4}$$

$$\mathbf{r}_{t-1} = \mathbf{\Lambda}'\mathbf{F}_t^{-1} + \mathbf{L}_t'\mathbf{r}_t \tag{A.5}$$

$$\mathbf{D}_t = \mathbf{F}_t^{-\mathbf{1}} + \mathbf{K}_t'\mathbf{U}_t\mathbf{K}_t \tag{A.6}$$

$$\mathbf{U}_{t-1} = \mathbf{\Lambda}'\mathbf{F}_t^{-\mathbf{1}}\mathbf{\Lambda} + \mathbf{L}_t'\mathbf{U}_t\mathbf{L}_t \tag{A.7}$$

$$t = 1, \ldots, T$$

$$\frac{\partial \log \mathscr{L}(\mathbf{Z_1}, \ldots, \mathbf{Z_T}|\theta)}{\partial \mu_i} = \frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}\left[(\mathbf{e}_t\mathbf{e}_t' - \mathbf{D}_t)\frac{\partial \mathbf{\Sigma}_{\Xi}(\boldsymbol{\mu})}{\partial \mu_i}\right] + \frac{1}{2}\sum_{t=2}^{T} \mathrm{tr}\left[(\mathbf{r_{t-1}}\mathbf{r_{t-1}'} - \mathbf{U_{t-1}})\frac{\partial \mathbf{\Sigma_\Omega}(\boldsymbol{\mu})}{\partial \mu_i}\right] \tag{A.8}$$

where tr denotes the trace operator and λ_i is the ith element in the parameter vector $\boldsymbol{\lambda}$.

This mixed approach is highly attractive to SUTSE specifications since the facilities are assigned individual error terms and the number of parameters is significant in $\boldsymbol{\lambda}$ (i.e. $\mathbf{\Sigma}_{\Omega}$ and $\mathbf{\Sigma}_{\Xi}$). On the other hand, the facilities have common parameters in $\boldsymbol{\theta}$ (i.e. **g** and **h**) and the number of parameters required for numerical evaluation is small. Thus, for each iteration in the optimization procedure, the first pass of Kalman filter and the disturbance smoother generates a large portion of information for computing the score vector. A very small number of passes of Kalman filter are required for the rest of the parameters. Comparing to the numerical approach that has to run Kalman filter at least once for each parameters, the saving of this mixed approach is expected. In our implementation, the speed of the mixed approach was considerably faster than that of the numerical approach for SUTSE-3 models. Furthermore, the solution obtained by the mixed approach was better than the one obtained by the numerical approach in terms of objective function values, although the difference, as expected, was small. It is also clear that the benefits of the mixed approach are not high for IM and SE since the number of parameters that are estimated simultaneously and the computational effort for numerical evaluation are not high.

References

Archilla, A.R., Madanat, S., 2000. Development of a pavement rutting model from experimental data. Journal of Transportation Engineering 126 (4), 291–299.

Archilla, A.R., Madanat, S.M., 2001. Estimation of rutting models by combining data from different sources. Journal of Transportation Engineering 127 (5), 379–389.

Ben-Akiva, M., Gopinath, D., 1995. Modeling infrastructure performance and user costs. Journal of Infrastructure Systems 1 (1), 33–43.

Ben-Akiva, M., Humplick, F., Madanat, S., Ramaswamy, R., 1991. Infrastructure Management Under Uncertainty: the Latent Performance Approach. Department of Civil Engineering, Massachusetts Institute iof Technology, Cambridge, MA.

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

Butler, B.C., Carmichael III, R.F., Flanagan, P.R., 1985. Impact of Pavement Maintenance on Damage Rate. Federal Highway Administration FHWA, Washington, DC.

Chu, C.-Y., 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.

de Solminihac, H., Salsilli, R., Covarrubias, J.P., Vidal, M., 1999. Rehabilitation performance prediction models for concrete pavements. Transportation Research Record 1684, 137–146.

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

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

Engle, R.F., Watson, M.W., 1981. A one-factor multivariate time series model of metropolitan wage rates. Journal of the American Statistical Association 76, 774–781.

Gendreau, M., Soriano, P., 1998. Airport pavement management systems: an appraisal of existing methodologies. Transportation Research Part A 32 (3), 197–214.

Greene, W.H., 2000. Econometric Analysis, fourth ed. Prentice Hall, Upper Saddle River, NJ.

Harvey, A.C., 1990. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, New York, NY.

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

Hudson, W.R., Haas, R., Uddin, W., 1997. Infrastructure Management: Integrating Design, Construction, Maintenance, Rehabilitation, and Renovation. McGraw-Hill Companies, New York, NY.

Humplick, F., 1992. Highway pavement distress evaluation: modeling measurement error. Transportation Research Part B 26 (2), 135–154.

Janacek, G., Swift, L., 1993. Time Series Forecasting, Simulation, Applications. Ellis Horwood Limited, New York, NY.

Koopman, S.J., 1993. Disturbance smoother for state space models. Biometrika 80 (1), 117–126.

Koopman, S.J., Shephard, N., 1992. Exact score for time series models in state space form. Biometrika 79 (4), 823–826.

Madanat, S., Mishalani, R., 1998. Selectivity bias in modeling highway pavement maintenance effectiveness. Journal of Infrastructure Systems 4 (3), 134–137.

Madanat, S.M., Karlaftis, M.G., McCarthy, P.S., 1997. Probabilistic infrastructure deterioration models with panel data. Journal of Infrastructure Systems 3 (1), 4–9.

McNeil, S., Markow, M., Neumann, L., Ordway, J., Uzarski, D., 1992. Emerging issues in transportation facilities management. Journal of Transportation Engineering 118 (4), 477–495.

Mishalani, R.G., Madanat, S.M., 2002. Computation of infrastructure transition probabilities using stochastic duration models. Journal of Infrastructure Systems 8 (4), 139–141.

Office of Highway Policy Information, 2006. Highway Statistics 2004. Federal Highway Administration, Washington, DC.

Paterson, W.D.O., 1987. Road Deterioration and Maintenance Effects: Models for Planning and Management. The Johns Hopkins University Press, Baltimore, MD.

Prozzi, J.A., Madanat, S.M., 2004. Development of pavement performance models by combining experimental and field data. Journal of Infrastructure Systems 10 (1), 9–22.

Shahin, M.Y., 2005. Pavement Management for Airports, Roads, and Parking Lots. Springer, New York, NY.

van Noortwijk, J.M., Frangopol, D.M., 2004. Deterioration and maintenance models for insuring safety of civil infrastructures at lowest life-cycle cost. In: Frangopol, D.M., Bruhwiler, E., Faber, M.H., Adey, B. (Eds.), Life-Cycle Performance of Deteriorating Structures: Assessment, Design and Management. American Society of Civil Engineers, Reston, VA, pp. 384–391.

Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions, and tests for aggregation bias. Journal of American Statistical Association 57 (298), 348–368.