Measuring the Efficiency of Highway Maintenance Operations: Environmental and Dynamic Considerations

Saeideh Fallah-Fini

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Industrial and Systems Engineering

Konstantinos P. Triantis, Chair

Jesus M. de la Garza

Douglas R. Bish

Hazhir Rahmandad

December 10, 2010

Blacksburg, VA

Keywords: Highway maintenance; Uncontrollable factors; Data envelopment analysis; Meta-frontier; Bootstrapping; System Dynamics; Dynamic Efficiency

Measuring the Efficiency of Highway Maintenance Operations:

Environmental and Dynamic Considerations[1]

Saeideh Fallah-Fini

(ABSTRACT)

Highly deteriorated U.S. road infrastructure, major budgetary restrictions and the significant growth in traffic have led to an emerging need for improving efficiency and effectiveness of highway maintenance practices that preserve the road infrastructure so as to better support society's needs. Effectiveness and efficiency are relative terms in which the performance of a production unit or decision making unit (DMU) is compared with a benchmark (best practice). Constructing the benchmark requires making a choice between an "estimation approach" based on observed best practices (i.e., using data from input and output variables corresponding to observed production units (DMUs) to estimate the benchmark with no elaboration on the details of the production process inside the black box) or an "engineering approach" to find the superior blueprint (i.e., focusing on the transformation process inside the black box for a better understanding of the sources of inefficiencies). This research discusses: (i) the application of the estimation approach (non-parametric approach) for evaluating and comparing the performance of different highway maintenance contracting strategies (performance-based contracting versus traditional contracting) and proposes a five-stage meta-frontier and bootstrapping analytical approach to account for the heterogeneity in the DMUs, the resulting bias in the estimated efficiency scores, and the effect of uncontrollable variables; (ii) the application of the engineering approach by developing a dynamic micro-level simulation model for the highway deterioration and renewal processes and its coupling with

[1] The work described in this dissertation was funded in part by NSF grant # CMMI-0726789. Any opinions, conclusions, and/or findings are those of the authors and do not necessarily reflect the views of NSF and/or VDOT.

calibration and optimization to find optimum maintenance policies that can be used as a benchmark for evaluating performance of road authorities.

This research also recognizes and discusses the fact that utilization of the maintenance budget and treatments that are performed in a road section in a specific year directly affect the road condition and required maintenance operations in consecutive years. Given this dynamic nature of highway maintenance operations, any "static" efficiency measurement framework that ignores the inter-temporal effects of inputs and managerial decisions in future streams of outputs (i.e., future road conditions) is likely to be inaccurate. This research discusses the importance of developing a dynamic performance measurement framework that takes into account the time interdependence between the input utilization and output realization of a road authority in consecutive periods.

Finally, this research provides an overview of the most relevant studies in the literature with respect to evaluating dynamic performance and proposes a classification taxonomy for dynamic performance measurement frameworks according to five issues. These issues account for major sources of the inter-temporal dependence between input and output levels over different time periods and include the following: (i) material and information delays; (ii) inventories; (iii) capital or generally quasi-fixed factors and the related topic of embodied technological change; (iv) adjustment costs; and (v) incremental improvement and learning models (disembodied technological change).

In the long-term, this line of research could contribute to a more efficient use of societal resources, greater level of maintenance services, and a highway and roadway system that is not only safe and reliable, but also efficient.

To My Parents

Acknowledgements

I have come to the end of a long and challenging journey. It started with one of the biggest decisions of my life, leaving my country to pursue higher education, and ended with the biggest achievement of my life thus far. This journey with all of its joyous moments and frustrations will remain as one of the precious memories of my life. I am glad that I have the opportunity to extend my gratitude to all people who had valuable contributions to my academic achievement and to my growing as a person.

First, I want to give special thanks to my advisor and my mentor **Dr. Konstantinos P. Triantis** for being the best advisor one could ever have. He taught me the basics of productivity and performance measurement, introduced me to the productivity research community, gave new directions to my research at the initial stages, facilitated the theoretical rigor in my work, and made sure that I stay on the right path. His thought-provoking questions were able to get the best out of me and his endless support gave me the hope and courage that I needed to reach the destination. He has been a wonderful friend who was always available to provide valuable advice, whether it was a matter of research or my personal life. We will always remain friends and I am extremely fortunate that my life's path has crossed his.

I also have been very privileged by working with a great advisory committee: **Dr. Jesus M. de la Garza**, **Dr. Hazhir Rahmandad**, and **Dr. Douglas R. Bish**. I would like to extend my gratitude to Dr. de la Garza for being extremely generous with his time throughout the course of this research given his busy schedule. It was his challenging questions during our long discussions that helped me better understand the issues that I had overlooked and to address the objectives of this research thoroughly. It was also due to his diligent help that I could overcome one of the most important challenges of this research, obtaining data from the Virginia Department of Transportation. His strong support and valuable advice helped me look at every step of this journey as a valuable experience that can prepare me for my future academic career. I learned from him to be well prepared in advance for everything and to set realistic targets and deadlines. Both of

these helped me release a lot of stress from my life, especially in the past year. I will be forever indebted to him for all I have learned and will learn from him.

Furthermore, I want to thank Dr. Rahmandad for his friendship, guidance, support, and for his significant contribution in this dissertation. His valuable experience in System Dynamics helped me do proper research grounded in theory, make sure that my research is solid from a System Dynamics aspect, and take advantage of various bodies of knowledge. His friendship and advice helped me realize the importance of maintaining a proper balance between workload and personal life both during my PhD studies and in my future career. I am also thankful to him and his beautiful wife, Sara, for being such wonderful hosts during all my trips to the Northern Virginia Campus. Finally, thanks to Dr. Bish for providing valuable advice and suggestions that helped me in completing this dissertation. Although he was not in my committee, I also want to extend my gratitude to **Dr. William Seaver** for his helpful statistical advice and for his contribution to the second chapter of this dissertation. Finishing this dissertation was not possible without the help, support, and contribution of every one of you.

I would like to thank the **Grado Department of Industrial and Systems Engineering at Virginia Tech** for providing the financial support to make this research possible. Many thanks to the ISE staff, especially **Ms. Kim Ooms** and **Ms. Hannah Parks**, for being very supportive and for helping me with all my administrative issues.

I also want to acknowledge the assistance of the **Virginia Department of Transportation** in this research. Especially, I would like to express my sincere appreciation to **Mr. Raja A. Shekharan, Mr. Allen Williams,** and **Mr. Jeff A**. **Wright** for providing the data and for allowing us to tap into their wealth of knowledge.

Last but not least, I want to make a special mention of my family and friends. My mother, **Fatemeh Abolfazli**, and my father, **Mahmoud Fallah Fini,** have been the biggest inspiration of my life and the reason why I have ever accomplished anything. They are very loving, supportive and strong, and have given up so much just so that I can follow my dreams. Everything I do and everything I am is because of my parents and I will be thankful to them for all my life. I also have been very fortunate for having loving,

supportive, and fun sisters and brothers who always sent me their well wishes and prayers and who sacrificed a lot by taking care of everything back home so that I can follow my adventures with peace of mind. I owe everything to your unconditional love.

Finally, many thanks to my wonderful friends, soon to be "Dr." **Evrim Dalkiran** and "Dr." **Ali Hashemi**. Thank you Evrim for believing in me and for giving me the hope and courage that I needed every time that I came to your office, overwhelmed by frustration and confusion. Ali, you never doubted I would be able to accomplish big achievements. Your love and support gave me the energy to overcome all the challenges that I have faced. Most importantly, you are the one who taught me that it is not the final destination that matters, instead I have to enjoy every single moment of the journey in achieving my PhD degree. Thank you Ali for all the sacrifices you have made. Thank you for being my best friend.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Problem Context

Road maintenance (highways and bridges) has received a significant attention in the past two decades (Ozbek et al., 2010a), especially given events such as the collapse of a major bridge in Minnesota in August 2007. Beyond the human tragedy of these unfortunate events lie some daunting numbers. As reported by the American Society of Civil Engineers (ASCE) in 2009, more than one-third of America's major roads are in poor or mediocre condition and 45% of major urban highways are congested. Poor road condition plays a major role in about 33% of traffic fatalities (ASCE, 2009a). Moreover, the deteriorated road system imposes $67 billion a year to U.S. motorists in repairs and operating costs (ASCE, 2009b). The current spending of $70.3 billion per year for improving highway conditions is much less than the estimated $186 billion required annually to make a significant improvement in the road conditions (ASCE, 2009b).

The highly deteriorated road system, major budgetary restrictions, as well as the significant growth in traffic have led to several institutional changes. In particular, road authorities are being challenged to improve performance of the existing highway maintenance policies and practices to preserve a safe, reliable, and efficient road infrastructure that can support society's needs (TRB, 2006). Privatizing some portions of road maintenance operations by state Departments of Transportation (DOTs) under performance-based contracts has been one of the innovative initiatives in response to such a need. A performance-based contract sets the minimum required conditions for the roads and traffic assets without directing the contractor to specific methods to achieve performance targets.

Under these conditions, the implementation of monitoring approaches so as to evaluate the performance of road authorities, such as DOTs, districts, or counties, who are responsible for the maintenance of the road in their administrative area, has significant benefits. As suggested by officials in several national and international highway agencies, maintenance managers should be provided with the mechanisms that allow for the measurement and analysis of maintenance performance, that assure that maximum performance is being achieved (Anastasopoulos et al., 2009; McCullouch and Anastasopoulos, 2009), and that facilitate the realization of improvements, changes, and decisions (TRB, 2006). From the perspective of top management, a

performance evaluation system should differentiate efficient units or decision making units (DOTs, counties, districts) from less efficient ones. Moreover, the performance evaluation system should arrive at a better understanding of the sources of inefficiencies; thus, providing the possibility for improvement for the road authorities that are underperforming. In this setting, performance evaluation can significantly help with budget planning, the design of maintenance policies and the extraction of best practices (Kazakov et al., 1989).

The road maintenance performance measurement systems developed thus far have mainly focused on effectiveness measures (i.e., how good the level-of-service is), with not much elaboration on the efficiency concept (i.e., the amount of resources utilized to achieve such level-of-service), which is also a very essential performance measurement dimension (Ozbek et al., 2010b). State DOTs wish for maintenance operations to be effective, e.g., they should result in safe, travelable, and good quality roads. But it is also reasonable to expect that efficiency, e.g., spending less money and time in performing road maintenance, is also important and thus needs to be accounted for.

The Virginia Department of Transportation (VDOT) is one of the DOTs active in measuring the performance of road maintenance operations. Since 2000, VDOT has been collaborating with Virginia Tech to identify innovative methodologies for evaluating the efficiency and effectiveness of the road maintenance operations and to assess the policies performed by VDOT (e.g., evaluating the performance of private contractors working for VDOT under the terms of performance-based contracts) (Ozbek et al., 2010a).

Within this context, this research was initiated to address the shortcomings of previous approaches by exploring, implementing, and developing new techniques/methodologies to evaluate the performance of road maintenance operations.

1.2 Research Objective and Questions

The main objective of this research relates to measuring the cost *efficiency* of road maintenance operations in relation to their achieved level-of-service (*effectiveness*). Through the use of real data, this research attempts to address the following key questions:

(i) How to evaluate the relative efficiency of different road authorities in performing road maintenance operations?

(ii) How to evaluate and compare the relative efficiency of different road maintenance contracting strategies (e.g., performance-based contracting and traditional contracting)?

(iii) What are the reasons that account for efficiency differences among road authorities?

(iv) What are the benchmarks and best practices of the inefficient road authorities?

Over the long term, this line of research can contribute to a more efficient use of societal resources, greater level of maintenance services, and a highway and roadway system that is safe, reliable, as well as efficient.

1.3 Research Methodological Strategy

In a broad sense, efficiency refers to the performance of a production system in utilizing a set of inputs to produce a set of outputs (Forsund and Hjalmarsson, 1974). Efficiency is a relative term for which the performance of a production system is compared with a benchmark (standard). As pointed by Forsund (2010), constructing the benchmark in any context requires making a choice between an "**engineering approach**" (Triantis, 2004) to find the "**superior blue print**" (Salter, 1960) or using the "**estimation approach**" to find the "**observed best practices**" (Farrell, 1957). Most of the methods in the efficiency measurement literature use the later approach. Meaning that they use data from input and output variables corresponding to observed production units to estimate the benchmark. Therefore, Farrell's measure of efficiency is a comparison between the input/output relations in an observed production unit and the input/output relations in the most efficient (and observable) production units.

The estimation approach looks at the production process as a black box that transforms inputs into outputs, with no elaboration on the details of the production process. However a better understanding of the efficiency differences among production units, especially when the sources of inefficiency are the quality of inputs (e.g., quality of labor) as well as managerial/engineering decisions, needs a focus on the transformation process inside the black box (Forsund, 2010). Thus, there is genuine interest in the literature for formal modeling of the transformation process and constructing the benchmark as the superior blue print. This research uses the ideas of both estimation and engineering approaches to develop appropriate benchmarks for evaluating performance of road maintenance operations.

The body of this research project has been organized into three essays. The next section describes the path that led to the development of the three essays as well as the contribution of each essay to address the core research questions and to achieve the objectives of this research.

1.4 Path Leading to Development of the Three Essays

This research was started by building on the framework developed by Ozbek (2007) for evaluating performance of road authorities that are responsible for the maintaining the road system in their administrative area. Ozbek (2007) uses data on inputs and outputs of road authorities under analysis and then applies a non-parametric approach to estimate a benchmark (based on observed best practices) to evaluate the efficiency of the road authorities. Ozbek (2007) does not elaborate or compare the performance of road authorities that are dissimilar. For example, road authorities may use alternative contracting strategies along with different regulations/limitations to maintain a road system in their administrative area. To address this shortcoming, this research started by developing an analytical framework for evaluating the performance of different highway maintenance contracting strategies. This piece of work led to development of the first essay, which is briefly described herein.

The first essay: "*Measuring the Efficiency of Highway Maintenance Contracting Strategies: A Bootstrapped Non-parametric Meta-frontier Approach"*

This essay uses an estimation approach to construct a benchmark so as to evaluate and compare the performance of "traditional" highway maintenance with "performance-based" highway maintenance contracting. More specifically, recently developed non-parametric performance measurement techniques (non-parametric meta-frontier approach as well as the two-stage bootstrapping technique) are used to evaluate and compare the relative efficiency of the mentioned highway maintenance contracting strategies. The traditional contracting strategy has been applied to 180 miles (within seven counties) of Virginia's Interstate highways maintained by Virginia Department of Transportation (VDOT). The performance-based contracting approach has been applied to 250 miles (within twelve counties) of Virginia's Interstate highways maintained via a PPP (Public Private Partnership). In traditional contracting, the maintenance operations that should be performed as well as the methods that should be used are specified by road authorities in advance. In performance-based contracting, road authorities set the minimum required conditions for the roads and traffic assets without directing the contractor to specific methods to achieve performance targets.

The meta-frontier approach utilized in this essay accounts for the heterogeneity that exists among different road authorities (i.e., counties) due to different limitations and regulations associated with traditional and performance-based highway maintenance contracting strategies.

The two-stage bootstrapping technique accounts for the large set of uncontrollable environmental (e.g., climate condition) and operational (e.g., traffic load) factors that affect the highway deterioration and maintenance processes. It also allows for statistical analysis of the estimated efficiency. A literature review to date has not come across any efficiency measurement analysis that has adopted the meta-frontier framework and the bootstrapping techniques for comparing the efficiency of different types of highway maintenance contracts. The analytical performance measurement approach developed in the first essay allows one to: (i) evaluate the relative efficiency of the two contracting strategies that are used by VDOT; (ii) evaluate the relative efficiency of road authorities (counties) that are using the same type of contract; (iii) explore environmental conditions as the potential sources of efficiency differences among road authorities. Thus, the findings of the first essay addresses the first three key research questions presented in Section 1.2.

The performance measurement approach that is developed in the first essay can be part of any road authority's decision making process that uses various types of contracts for performing maintenance operations. This line of research provides the decision-makers with proper knowledge of the efficiency level of different highway maintenance contracts or projects. This is crucial for guiding future decisions regarding the renewal of contracts, the pricing of these contracts, and the potential efficiency improvement opportunities, since past performance is an important criterion that road authorities can consider when awarding new contracts.

Going through the first essay along with further elaboration of the structure of the input/output data and the transformation process (road deterioration and renewal processes) led to a very important insight and understanding about the nature of the highway maintenance process. More specifically, it was revealed that the utilization of the maintenance budget and treatments that are performed on a road section in a specific year directly affect the road condition and required maintenance operations in consecutive years. Thus, the observed road condition in each year cannot be solely attributed to the latest maintenance policy and operations, but it's the result of a stream of previous maintenance operations/budgets.

For a better understanding, Figure 1-1 illustrates a schematic view of the road condition evolution over time. Assume road section C at period t is affected by a set of deterioration factors such as rainfall, snowfall, traffic load, etc. These factors lead to the deterioration of the road and thus trigger the required maintenance operations. Based on the condition of the road, appropriate

maintenance operations are performed to improve the condition of the road section. Given the realities of limited maintenance budgets, actual maintenance operations are not necessarily equal to the required ones. Due to the interaction between deterioration factors and maintenance operations (renewal process), the road evolves to a new condition at the end of period *t*. Thus, the new road condition is the "output" of the transformation process that happens on road section *C* at period *t*. This new road condition is used as an intermediate input at the start of period *t+1* when road section *C* goes under a similar transformation process due to deterioration factors and maintenance operations that are applied in period *t+1*. Obviously, the maintenance treatments at period *t* affect the road condition at the end of period *t* which is the start point for period *t+1*. Thus, the required maintenance operations at period *t+1* (and consequently the road condition at the end of period *t+1*) depend on the maintenance operations/inputs that have been performed/used in a stream of previous periods.

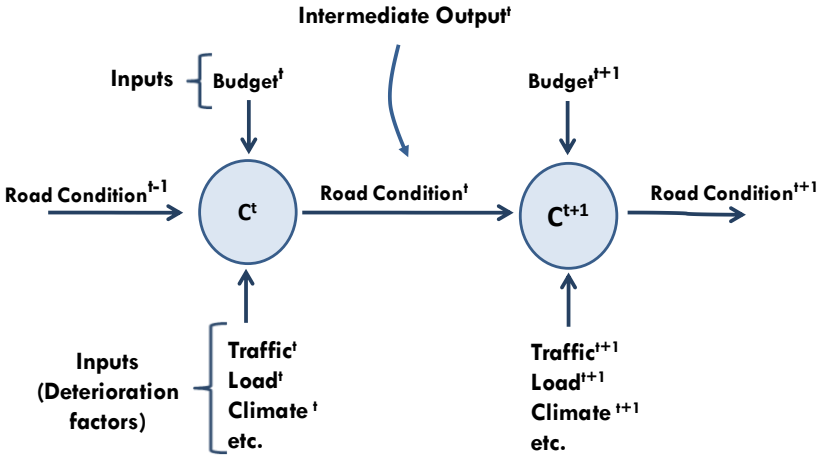


Figure 1-1: Schematic View of Road Condition Evolution over Time

In such a setting, any "static" efficiency measurement framework that ignores the inter-temporal effects of inputs and managerial decisions that affect the future streams of outputs (i.e., future road conditions) is likely to be biased. The fact is that all studies that exist in the literature regarding the performance measurement of highway maintenance activities (e.g., see Kazakov et al., 1989; Cook et al., 1990; Rouse et al., 1997; Ozbek, 2007; Rouse and Chiu, 2008; de la Garza et al., 2009; Fallah-Fini et al., 2009; Ozbek et al., 2010a; Ozbek et al., 2010b) use a static efficiency measurement framework and assume that there is no time interdependence between the input utilization and output realization of a road authority in consecutive periods.

Gaining this important insight and understanding about the dynamic nature of highway maintenance operations coupled with the realization of the shortcomings of previous studies, provided the focus for the rest of this research to explicitly elaborate on the development of a dynamic efficiency measurement framework for evaluating the performance of highway maintenance operations where the inter-temporal dependencies between inputs and outputs are explicitly captured.

A better understanding of the sources of efficiency differences and development of an appropriate benchmark for the highway deterioration and renewal process requires formal modeling of the transformation process. In these situations, engineering approaches, in particular, System Dynamics techniques offer some benefits beyond the standard efficiency measurement approaches that only consider the input and output variables associated with production units under analysis (Triantis, 2004; Vaneman and Triantis, 2007). System Dynamics provides the possibility for modeling the dynamics of the production/transformation process at the required level of aggregation. Especially in the case of highway maintenance, System Dynamics enables us to model delays and, consequently, relate actions (maintenance operations) and payoffs (change in the road condition) that are separated in time. As a result, the second piece of this research uses the System Dynamics approach to develop a micro-level model of the highway deterioration and renewal process that can be used as a stepping stone for constructing a superior blue print for evaluating the performance of road authorities that maintain the road system in their administrative area. This piece of work led to development of the second essay, which is briefly described herein.

The second essay: "*Optimizing Highway Maintenance Operations: Dynamic Considerations*"

System dynamics is used in the second essay to develop a micro-level model of the highway deterioration and renewal processes. More specifically, in the first step, a micro-level simulation model of the highway deterioration process is developed using the principles embodied in the Mechanistic-Empirical models (Huang, 2004) that are among the best available physics-based models for predicting road deterioration in pavement management. This model was then calibrated using empirical data (pavement condition and traffic for approximately 17 miles of Virginia's Interstate highway during the years 2002 to 2007). This data is used to calibrate, adjust and estimate the difficult to measure model parameters. Next, the calibrated model is coupled with the highway maintenance and renewal decision-making process that

allocates the limited budget to different maintenance operations. Moreover, an optimization module is developed and applied to find the best policy for the budget allocation given a set of environmental and operational conditions. In sum, the second essay couples system dynamics modeling with calibration and optimization and builds a framework for designing highway maintenance policies. This analysis offers alternative priority setting schemes (among preventive, corrective and restorative maintenance) that improve current maintenance practices in the highway network under analysis.

The micro-level simulation model constructed in the second essay can be used to develop a superior blueprint for benchmarking efficiency of road authorities in making the best use out of the limited available budget. Thus this essay addresses the second and fourth key research questions presented in Section 1.2 related to evaluating the performance of road authorities and identifying the best practices that pertain to road authorities.

The next logical step after completing the first and second essays would be to construct the appropriate benchmark that accounts for the inter-temporal dependence between utilizing the maintenance expenditure and change in the road condition over future periods. Consequently, this benchmark can be used to develop the dynamic measure of efficiency of the road authorities. Thus, the third essay studies the literature of dynamic efficiency and explores the understanding of this concept in the literature including, the sources of inter-temporal dependence, as well as the models/techniques that have been developed in the literature for evaluating dynamic efficiency. This piece of work led to development of the third essay, which is briefly described herein.

The third essay: "*Dynamic Efficiency Performance: State-of-the-Art"*

The objective of this third essay is to provide a review of the models/approaches in the literature that capture the inter-temporal relation among different periods while developing dynamic measures of performance. In order to systematically review the studies that address relevant aspects of dynamic performance, this essay proposes a classification taxonomy for dynamic performance measurement frameworks according to five issues. These issues account for major sources of the inter-temporal dependence between input and output levels over different time periods and include the following: (i) material and information delays; (ii) inventories; (iii) capital or generally quasi-fixed factors and the related topic of embodied technological change; (iv) adjustment costs; and (v) incremental improvement and learning

models (disembodied technological change). The third essay begins by discussing the comparative static frameworks that have been developed in the literature for quantifying the change in different performance measures over time when no inter-temporal dependence between input and output levels across different time periods is assumed. Thus, this essay points out the studies that potentially belong to the literature of dynamic performance, while no fundamental dynamics have been captured in these studies. Subsequently, the key studies in the literature that capture the inter-temporal relation among different periods with respect to the identified five issues are overviewed. As part of this overview, the strengths, shortcomings, similarities and differences of these key studies as well as the challenges and potential future research areas are discussed.

Exploring the literature through the third essay provided the insight and understanding that would be required to develop an appropriate measure of dynamic performance for road authorities that are in charge of maintaining the road system. It also provided useful insights regarding the likely sources of efficiency differences and the best practices that would pertain to the inefficient road authorities. Thus, third essay contributed to addressing the second, third, and fourth key research questions presented in Section 1.2. The understanding gained from this essay combined with the insights from the first and second essays have set the stage for developing an analytical framework for evaluating the dynamic efficiency of highway maintenance operations. The two important steps for developing this framework are: (i) the development of a benchmark for evaluating the path of development of the road condition over time; and (ii) the development of a measure of dynamic efficiency by benchmarking with respect to the developed frontier. Further discussion regarding the ideas and techniques for developing this analytical framework are presented in the areas for future research presented in Chapter 5 (Conclusion).

1.5 Research Contributions

This research contributes to the body of knowledge with respect to three areas in the literature: (i) highway maintenance; (ii) performance measurement; and (ii) system dynamics.

Contributions to the body of knowledge in the highway maintenance domain: Most of the research in the highway arena has focused on topics such as structural design, selection of materials, safety devices, as well as location of signs, signals, intersections, etc. (Ozbek, 2007). In contrast, not enough research has been done in the area of highway maintenance as well as

highway maintenance performance measurement (TRB, 2006). This research contributes to the body of knowledge in highway maintenance by developing an analytical framework that uses the ex post data to assess and compare (i) the efficiency of various types of highway maintenance contracting strategies; (ii) the efficiency across multiple road authorities that are using the same type of contracts (e.g., state DOTs, districts or counties within a state DOT, etc.). In contrast to most of the road maintenance performance measurement studies that have focused only on the achieved level-of-service (effectiveness) of road authorities, the developed framework also takes into account the efficiency concept (i.e., the amount of resources utilized to achieve such level-of-service) as the other essential dimension of any performance measurement framework. Thus, this research improves the ways that have been used to model and measure the performance of highway maintenance operations, especially in the USA. The impact of the developed framework is believed to be broad and relevant to all agencies that aim to evaluate and compare the performance of units (**homogeneous or heterogeneous**) that are performing under a set of uncontrollable factors that affect their outputs. Prior research that focused on performance measurement of highway maintenance had only considered homogeneous units.

This research also makes practical contributions to the asset management field in the highway maintenance domain by coupling system dynamics and optimization to develop alternative maintenance policies and practices (in terms of priorities of preventive, corrective, and restorative maintenance) that can maintain highway networks in the best possible condition given the limited available budget. Prior research in this field had relied on sensitivity analysis and trial/error for policy analysis and had not utilized an integrated optimization method to find the best maintenance policies.

Finally, this research contributes to the highway maintenance domain by recognizing the presence of inter-temporal dependencies between the utilization of inputs (i.e., the maintenance budget and maintenance treatments) and the realization of outputs (i.e., improvement in the road condition). Performance measurement frameworks that ignore this time dependency between inputs and outputs while measuring performance of highway maintenance operations are biased and/or inaccurate.

Contributions to the body of knowledge in performance measurement: This research contributes to the literature of non-parametric efficiency measurement by introducing highway maintenance performance measurement as a new engineering domain for the application of

recent developments in the performance measurement field (i.e., adapting the meta-frontier framework to the recently developed bootstrapping techniques). In particular, this research has advanced a five-stage meta-frontier and bootstrapping analytical approach to account for the heterogeneity in the production units under analysis, the resulting bias in the efficiency scores, and the effect of uncontrollable variables.

This research also contributes to the literature of dynamic efficiency by (i) providing an overview of the studies that are most relevant to the dynamic efficiency literature as well as their strengths, shortcomings, similarities and differences; (ii) by discussing how an engineering approach can be used to develop a dynamic frontier that can be used as a benchmark for measuring the performance of a production unit or decision making unit (DMU); and (iii) by introducing highway maintenance as an interesting application domain for studies related to dynamic efficiency where there is a time-interdependence between input utilization and output realization over consecutive periods. Prior research has only used estimation approaches (parametric or non-parametric) to develop a benchmark that takes into account the intrinsic temporal dependency between input and output levels.

Contributions to the body of knowledge in system dynamics: By developing one of the first system dynamics simulation models based on the physics of road deterioration and maintenance, this research introduces new concepts and knowledge from the road deterioration literature into the system dynamics literature. Moreover, this research couples the conceptualization steps of the system dynamics method with calibration and optimization to construct a framework for designing effective maintenance policies. In prior SD highway maintenance studies, the model parameters such as deterioration rates are defined either based on synthetic data or using the outputs of other statistical analyses. Therefore the potential to estimate model parameters using original data has not been realized and thus ensuring model reliability has been a challenge. Thus, this research increases the applicability of the system dynamics approach for operational and tactical decision support in the field of highway maintenance. It is hoped that the findings of this research provide a blueprint for the successful introduction of system dynamics methodology in many alternative problem domains.

This research also introduces new concepts from the performance measurement field into the system dynamics literature by introducing system dynamics as an appropriate approach for formal modeling of the transformation process inside a production unit and obtaining a better

understanding of the sources of efficiency differences. In essence this research builds and expands this notion that was first introduced by (Vaneman, 2002). Additionally, coupling system dynamic modeling with optimization leads to the development of an appropriate benchmark for evaluating the performance of the production unit under analysis. Thus, this research introduces system dynamics as an appropriate approach for measuring performance when one chooses to follow an engineering approach.

1.6 Organization of the Dissertation

In Chapters 2, 3, and 4 the first, second, and third essays will be presented, respectively. Chapter 5 concludes and provides directions for future research.

Chapter 2 Measuring the Efficiency of Highway Maintenance Contracting Strategies: A Bootstrapped Non-parametric Meta-frontier Approach

Abstract

Highly deteriorated U.S. road infrastructure, major budgetary restrictions and the significant growth in traffic have led to an emerging need for improving performance of highway maintenance practices. Privatizing some portions of road maintenance operations by state Departments of Transportation (DOTs) under performance-based contracts has been one of the innovative initiatives in response to such a need. Successful implementation of new maintenance policies requires state DOTs to measure the performance of new contracting approaches. This paper adapts the non-parametric meta-frontier framework to the two-stage bootstrapping technique to develop an analytical approach for evaluating the relative efficiency of two highway maintenance contracting strategies. The first strategy pertains to the 180 miles of Virginia's Interstate highways maintained by Virginia DOT using traditional maintenance practices. The second strategy pertains to the 250 miles of Virginia's Interstate highways maintained via a Public Private Partnership using a performance-based maintenance approach. The meta-frontier approach accounts for the heterogeneity that exists among different types of highway maintenance contracts due to different limitations and regulations. The two-stage bootstrapping technique accounts for the large set of uncontrollable factors that affect the highway deterioration/maintenance processes. The preliminary findings, based on the historical data, suggest that (i) road authorities that have used traditional contracting have been more efficient than road authorities that have used performance-based contracting for maintaining the highway network in their administrative area; (ii) climate conditions stood out as significant factors explaining the difference among efficiency scores of Virginia road authorities that are using the same type of contract. This paper recommends that road authorities use hybrid contracting approaches that include best practices of both traditional and performance-based highway maintenance contracting.

Key Words and Phrases: Data Envelopment Analysis; Meta-frontier; Bootstrapping; Highway Maintenance Contracting Strategies; Performance-based Contracting.

2.1 Introduction

2.1.1 Context and Objectives

In the past twenty years, the American Society of Civil Engineers (ASCE) has constantly rated the US road system as being in a poor condition (ASCE, 2009a). As pointed by the ASCE, US road authorities have been facing a huge gap between the level of capital investment and the actual value that is needed to significantly improve the condition of the nation's road system (ASCE, 2009b). In view of this highly deteriorated road system, major budgetary restrictions, and the significant growth in traffic, there is a tremendous interest in the improvement of the efficiency and effectiveness of highway maintenance practices that preserve the road infrastructure so as to better support society's needs.

One of the innovative initiatives undertaken in response to this need was the Virginia Public-Private Transportation Act of 1995 (PPTA). PPTA authorized the Virginia Department of Transportation (VDOT) to establish contracts with private entities for construction, maintenance and improvement of transportation facilities. In the first public-private partnership undertaken under the auspices of the PPTA in 1996, the Virginia Maintenance Services (VMS), a private contractor, took the responsibility for the administration and maintenance of all assets as well as the operation of incident management and snow removal of 250 miles (approximately 25%) of Virginia's Interstate highway system. The very important characteristic of the ten-year contract between VDOT and VMS was its performance-based nature. A performance-based contract (PBC) sets the minimum required conditions for roads, bridges and other assets without directing the contractor to specific maintenance methods that would help achieve the performance targets. Therefore, in principle, this contracting mechanism provides the incentive for contractors to seek innovative maintenance methods that can achieve the pre-determined road performance criteria and targets (Ozbek, 2007). The performance-based approach for contracting highway maintenance projects was introduced as an alternative to the traditional approach, also known as "design-bid-maintain". In the traditional approach, the tasks that are to be performed as well as the methods that should be used are specified in advance.

In addition to state of Virginia, several other states have also tried or are currently experimenting with PBCs over some portions of their Interstate such as Texas, Florida, Maryland, Washington DC, New Mexico, Idaho, and Utah (NCHRP, 2009). The transition to

performance-based contracting has been in response to a nationwide recommendation made by the Federal Highway Administration (FHWA) (FHWA, 2003).

Although there has been an increasing trend toward the application of PBCs for highway maintenance operations, a number of challenges still need to be addressed, in particular, the lack of a system for evaluating performance of PBCs (McCullouch and Anastasopoulos, 2009). Officials in several national and international highway agencies have suggested that there is an emerging need for frequent assessment of PBCs to make sure that the required level of service (LOS) has been met efficiently (Anastasopoulos et al., 2009; McCullouch and Anastasopoulos, 2009). In a performance measurement system, the effectiveness dimension focuses on the achieved level of service and the efficiency dimension focuses on the amount of resources consumed to achieve a given level of service. Most of the performance measurement systems developed and implemented by the State Highway Agencies for road maintenance operations focus mainly on the overall improvement of the road condition (effectiveness) and do not adequately investigate the utilized resources and expenditures (efficiency) that have led to a given level of service (TRB, 2006; Ozbek et al., 2010a; Ozbek et al., 2010b). Providing a highway and roadway system that is safe, reliable, as well as efficient is critical considering current national budgetary restrictions (ASCE, 2009b).

Given this context, an assessment approach that builds on the fundamental relation between the maintenance level of service and budget requirements (the effectiveness and efficiency concepts) should be part of any road authority that is using various types of contracts for performing maintenance operations. Thus, the main objective of this paper is to utilize a combination of analytical non-parametric performance measurement approaches to (i) evaluate and compare the efficiency of various types of highway maintenance contracts at the aggregate level (e.g., the efficiency of performance-based contracts (PBCs) versus the efficiency of traditional contracts) considering their achieved LOS; (ii) evaluate and compare the efficiency of different maintenance projects (contractors) that are under the same type of contract so as to recognize the contractors that are performing better in comparison with others in terms of both efficiency and effectiveness; (iii) identify the potential sources of inefficiency for the inefficient contractors, thus facilitating budget planning. One can use these techniques to find a set of benchmarks for the inefficient contractors and facilitate the design of maintenance policies, but this is beyond the scope of this paper. This research contributes to the literature considering that

there are a limited number of studies (e.g., see Anastasopoulos et al., 2009; de la Garza et al., 2009; McCullouch and Anastasopoulos, 2009) that assess and compare the efficiency of various types of highway maintenance contracts.

2.1.2 Points of Departure

There are two broad types of methods in the efficiency literature for arriving at measures of relative efficiency, i.e., parametric and non-parametric methods (Thanassoulis, 1993). The parametric methods typically assume a functional form for the benchmark (frontier) and use data to estimate the parameters of that function. The estimated function is then used to arrive at estimates of the efficiencies of the units under analysis. The non-parametric methods use data and construct a piecewise-linear function that acts as benchmark for measuring relative efficiency. Data Envelopment Analysis (DEA) (Charnes et al., 1978) is one of the popular non-parametric techniques in the literature and has been widely used in different fields. DEA models are used to examine the relative efficiency of a set of similar decision-making-units (DMUs) (e.g., maintenance projects in a specific year) when a number of factors need to be considered. Several studies exist regarding applications of DEA models for measuring the performance of highway maintenance operations in Ontario, Canada (Kazakov et al., 1989; Cook et al., 1990; Cook et al., 1994), New Zealand (Rouse et al., 1997; Rouse and Chiu, 2008), and Virginia, USA (Ozbek, 2007; de la Garza et al., 2009; Fallah-Fini et al., 2009; Ozbek et al., 2010a; Ozbek et al., 2010b). This paper focuses on a selection of recently developed non-parametric (DEA) performance measurement techniques and assesses their usefulness for measuring the efficiency of highway maintenance strategies.

The first development relates to the difficulty that one encounters when comparing DMUs that may be heterogeneous. For example, highway maintenance projects under different types of contracts have different characteristics in terms of performance targets, methods, and resources. Thus, DEA models that estimate only one frontier for their evaluation may not be applicable. In order to fully capture the heterogeneity of highway maintenance contracts, this paper uses the non-parametric meta-frontier framework developed by a combination of papers by Battese, Rao, and O'Donnell (Battese and Rao, 2002; Battese et al., 2004; O'Donnell et al., 2008). The meta-frontier framework, first evaluates the efficiency of each DMU with respect to its group frontier, where DMUs in each group are assumed to have the same characteristics (e.g., use the same type of contract). In order to make an efficiency comparison across groups, a meta-frontier is then

developed using best practices of all groups. Estimating the gap between each group frontier and the meta-frontier can help decision-makers by identifying performance improvement programs (O'Donnell et al., 2008). A literature review to date has not come across any efficiency measurement analysis that has used the meta-frontier concept for comparing the efficiency of different types of highway maintenance contracts.

The second development relates to the bias as well as the statistical properties of non-parametric efficiency scores. As has been stated by Simar and Wilson (2008), the non-parametric efficiency scores are biased by construction and the bias depends mainly on the sample size (the number of DMUs under analysis) and the dimension of the problem (the number of inputs and outputs). This is one of the potential problems in the meta-frontier literature, since classifying the DMUs into groups based on their characteristics may lead to a limited number of DMUs in each group. Starting with the work of Simar (1992), bootstrapping techniques were introduced into the efficiency measurement literature as attractive approaches for conducting various statistical inferences on the efficiency scores, including the correction for the bias. This paper adapts the meta-frontier framework to the recently developed bootstrapping techniques to correct for the shortcomings that may arise in the meta-frontier analysis. To date, there has been no study in the area of highway asset management that uses bootstrapping techniques for the statistical analysis of efficiency scores.

The third development relates to the integration of environmental and operational conditions into the efficiency analysis. Highway maintenance is a process that is highly affected by uncontrollable environmental factors (e.g., climate condition) and operational conditions (e.g., traffic and load) (Ozbek et al., 2010a). The uncontrollable factors may account for the efficiency differences, thus special attention should be given to these factors. There are many methods for integrating uncontrollable factors into the efficiency analysis, each of which has its own advantages and drawbacks as will be discussed in Section 2.2.4. The two-stage semi-parametric bootstrapping technique by Simar and Wilson (2007) addresses many difficulties that exist in previously developed methodologies. In the first stage, the Simar and Wilson (2007) two-stage bootstrapping technique obtains the non-parametric efficiency scores using only controllable input and output variables, and in the second stage, the observed efficiency patterns are econometrically explained using the set of uncontrollable factors. Adopting this approach constitutes an additional contribution to the literature of highway maintenance performance

measurement, since previous research in this area has not used this approach to integrate uncontrollable factors into the efficiency analysis.

This paper uses a combination of these techniques to develop an analytical approach and applies it on an empirical dataset of pavement condition, traffic, climate condition, and maintenance expenditures. This approach is applied to approximately 180 miles of Virginia's Interstate highways that were maintained by VDOT using traditional maintenance practices over the fiscal years 2002 to 2006 and to 250 miles of Virginia's Interstate highways maintained using a performance-based maintenance strategy over the fiscal years 2002 to 2004.

The remainder of this paper is organized as follows. Section 2.2 describes the proposed analytical approach as well as the detail of the techniques that are used. The results and insights obtained from implementing the approach on the empirical data are discussed in Section 2.3. Conclusion and future directions are provided in Section 2.4.

2.2 Foundations (Methodology)

2.2.1 Non-parametric Estimation of the Production Frontier

A Decision Making Unit (DMU) whose performance is measured is generally regarded as the entity that uses a production process that converts multiple inputs into multiple outputs. The underlying production process is constrained by the "production possibility set" or "technology set" Ψ, which is the set of all physically attainable points ($\boldsymbol{x}, \boldsymbol{y}$), where $x \in \Re_+^N$ is the input vector and $y \in \Re_+^M$ is the output vector.

$$\Psi = \left\{ (x, y) \in \Re_+^{N+M} \mid x \text{ can produce } y \right\} \quad \textbf{(1)}$$

For the purpose of efficiency analysis, the upper boundary of Ψ that is called the efficient boundary or "technology frontier" is of importance. The frontier is defined as the set of best performing DMUs that use the minimum input level to produce a given output level or generate the maximum output given a specific input level (Simar and Wilson, 2008). For a given DMU with input and output variables $(x, y) \in \Re_+^{N+M}$, the measures of technical efficiency can be defined respectively as:

$$\mu(x, y) = \inf \left\{ \mu \mid (\mu x, y) \in \Psi \right\} \quad \textbf{(2)}$$

$$\lambda(x, y) = \inf \left\{ \lambda \mid (x, y / \lambda) \in \Psi \right\} \quad \textbf{(3)}$$

$\mu(x,y)$ and $\lambda(x,y)$ lie between zero and one. $\mu(x,y)$ is the input-oriented measure of efficiency and represents the input reduction required by the DMU to become efficient holding the outputs constant. $\lambda(x,y)$ is the output-oriented measure of efficiency. $1/\lambda(x,y)$ represents the output expansion required by the DMU to become efficient holding the inputs constant. The decision to use an input or output measure of efficiency is usually made based on the objectives of performance measurement. For example, an output-oriented model can help maintenance managers to specify the maximum number of lane-miles they could potentially maintain using the limited available budget.

Data Envelopment Analysis (DEA) uses mathematical programming (i.e., is considered a non-parametric approach) to compute the frontier and the efficiency scores corresponding to all DMUs under analysis. DEA is considered as an appropriate approach for measuring performance of highway maintenance operations since establishing "production standards" and measuring absolute efficiency in this setting is hard, if not impossible (Kazakov et al., 1989; Cook et al., 1994). In addition, DEA allows for the consideration of different non-economic factors, such as traffic, load, climate conditions, etc. where each of these factors plays an important role in the efficiency analysis. Readers are referred to Ozbek, et al. (2009) for a summary on the use of DEA models in the transportation field.

Assuming the information on input and output variables of *n* DMUs is available, the non-parametric estimation of the variable returns to scale technology set Ψ can be written as follows:

$$\hat{\Psi}_{VRS} = \left\{ (x,y) \in \Re^{N+M} \mid y \le \sum_{i=1}^{n} \gamma_i y_i;\ x \ge \sum_{i=1}^{n} \gamma_i x_i \text{ for } \gamma_1,...,\gamma_n \text{ such that } \sum_{i=1}^{n} \gamma_i = 1;\ \gamma_i \ge 0,\ i = 1,...,n \right\} \quad \textbf{(4)}$$

The estimated technology set $\hat{\Psi}_{VRS}$ is a convex set enveloping the data points, thus it is an inner bound approximation of the true (unobserved) and larger technology set Ψ. This means that the distance between each DMU and the estimated frontier, which is a measure of inefficiency of that DMU, is less than the distance between the DMU and the true frontier. Thus, the estimated inefficiency scores are less than the true inefficiency scores, meaning that DEA estimators underestimate the inefficiency scores (or overestimate the efficiency scores). That is why DEA estimators of efficiency, although consistent[2], are biased by construction (Simar and Wilson, 2008). The bias as well as rate of convergence of the DEA estimators to their true (unobserved)

[2] An estimator $\hat{\theta}$ of an unknown parameter θ is consistent if the estimator converges to θ as the sample size *n* increases. Consistency is an asymptotic property for an estimator.

values mainly depends on the properties of the dataset under analysis, namely: (1) the number of observations (DMUs) in the dataset; (2) the number of input and output variables; and (3) the distribution of observations around the frontier (Borger et al., 2008). As the number of observations increases, approximation of the true technology set and corresponding frontier is improved, thus the bias which is the difference between the actual efficiency scores and the estimated ones decreases. Moreover, as the number of input/output variables (dimension of the space) increases, the Euclidean distance between the observations increases. As a result, there will be fewer nearby observations that can convey information about the portions of the efficient frontier which is of interest (Simar and Wilson, 2008). In addition, an increase in the number of input and output variables requires more observations (DMUs) for constructing the efficient frontier, leading to an increase in the bias of the estimated efficiency scores (Simar and Wilson, 2008). In the literature, this situation is referred to as "curse of dimensionality".

Another shortcoming of non-parametric estimators is related to the assumption that DMUs under analysis have the same characteristics and are similar. Thus, a common frontier or benchmark is estimated for evaluating efficiency scores of all DMUs. The reality is that this is not the case in most of the interesting and practical problems and DMUs usually experience some heterogeneity. For example, imagine one is comparing the efficiency of production units that are operating under different regulations or are located in different countries, or have different ownerships (public versus private). Under these circumstances, separate efficiency frontiers need to be estimated for different groups of DMUs (O'Donnell et al., 2008). Sections 2.2.2 and 2.2.3 describe some of the recent remedies that have been suggested in the literature to address the heterogeneity and small sample bias issues, respectively.

2.2.2 Heterogeneity among DMUs

When the DMUs under analysis face different production opportunities, they need to make choices from different sets of feasible input-output combinations. The differences in production opportunities can be attributed to the physical, social, and economic environments in which the production process takes place (e.g., the size and quality of labor force, type of equipment, etc.) (O'Donnell et al., 2008). Under the stated conditions, the DMUs belong to different technology sets (groups), thus using traditional DEA models that estimate one common technology frontier for comparing all DMUs will increase the risk of having unreasonable estimates of the efficiency scores.

This situation is relevant to highway maintenance in that there is a difference among limitations, regulations, and maintenance operations that correspond to the various types of maintenance contracts. For example, traditional contracts are short-term and conservative in nature. Their main focus is on first cost (lowest bidder), thus the corresponding set of techniques, tools and materials to perform the required maintenance operations are chosen accordingly. Performance-based contracts (PBCs) are long-term and tend to optimize the cost over a project's lifecycle. Long-term responsibility of the PBCs motivates the contractors to be innovative in their design (the tasks that should be performed and their specifications), as well as in their selection of tools and materials. Thus, contractors or road authorities under various types of contracts are working in different production environments with different regulations. As a result they may only have access to a restricted part of the production possibility set. The boundaries of these restricted production possibility sets form the group frontiers (O'Donnell et al., 2008). Thus, when comparing the efficiency of the PBCs versus the traditional contracts, the DMUs under each type of contract form separate groups, each of which has its specific operational environment.

The recently developed analytical meta-frontier approach (Battese and Rao, 2002; Battese et al., 2004; O'Donnell et al., 2008) provides an appropriate methodology to evaluate and compare the efficiency of DMUs that belong to different groups. Assuming that there are L groups of DMUs, the meta-frontier framework first pools the observations of all groups and estimates a meta-technology set that contains all input-output combinations that are technologically possible. The boundary of this unrestricted technology set is called the meta-frontier and is used to measure the efficiency of each DMU assuming that technology is freely interchangeable and that the DMUs in all groups have potential access to the same technology. In the second step, the observations of each group are used separately to define group-specific technology sets. The boundaries of the group technology sets are called group frontiers. The distance from an input-output point to its group frontier is a measure of technical efficiency of that DMU, meaning how well a DMU is performing in comparison with the rest of the DMUs in its own group. Each group frontier represents the boundary of a restricted technology set, where the restrictions are defined from resource, regulatory, or other constraints (O'Donnell et al., 2008). Analyzing the gap between a group frontier and the meta-frontier indicates the potential improvements that can be made in the efficiencies of the DMUs of that group when one removes

the restrictions/regulations and uses the best practices that are provided by the meta-technology, which is defined as the technology of all input-output combinations associated with all DMUs in the sample (O'Donnell et al., 2008).

To provide a better understanding, Figure 2-1 provides a graphical illustration of the meta-frontier for a simple example with one input and one output variable. DMUs under analysis belong to two heterogeneous groups, thus two group frontiers represented by XX' and YY' are computed. Consider a specific DMU operating at the input-output combination labeled by *A*. The output-oriented technical efficiency of DMU *A* with respect to its group frontier XX' and meta-frontier MM' are calculated respectively as:

$$TE_{XX'}(A)=\frac{OB}{OC},\ TE_{MM'}(A)=\frac{OB}{OD} \quad \textbf{(5)}$$

To analyze the gap between the group frontier XX' and the meta-frontier MM', the meta-technology ratio (MTR) of DMU *A* is defined as:

$$MTR_{XX'}(A)=\frac{TE_{MM'}(A)}{TE_{XX'}(A)}=\frac{OB/OD}{OB/OC}=\frac{OC}{OD} \quad \textbf{(6)}$$

The meta-technology ratio basically measures how close a group frontier is to the meta-frontier. Moreover, Equation (7) as a reconstruction of equation (6) implies that the technical efficiency of DMU *A* measured with respect to the meta-frontier can be decomposed into the product of technical efficiency with respect to the group frontier (representing the characteristics of the group and its state of knowledge), and the meta-technology ratio for group XX' (representing how close the group frontier is to the meta-frontier) (O'Donnell et al., 2008).

$$TE_{MM'}(A)=TE_{XX'}(A)*MTR_{XX'}(A) \quad \textbf{(7)}$$

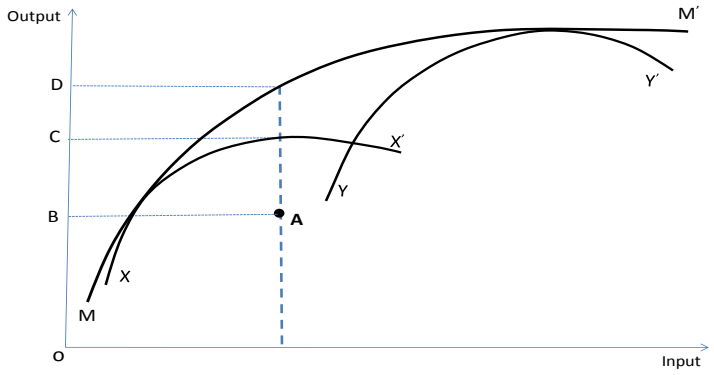


Figure 2-1: Group Frontiers and the Meta-frontier (O'Donnell et al., 2008)

Note that once data on input and output variables of DMUs from all groups are available, the group frontiers and the meta-frontier can be estimated using either parametric or non-parametric methods. As it was mentioned in Section 2.2.1, the non-parametric DEA method is used in this paper for estimating the group frontiers and meta-frontier.

2.2.3 The Bias Associated with the Estimated Efficiency Scores

The lack of elaboration on statistical properties of the non-parametric point estimates of efficiency scores early in the literature did not provide the opportunity for performing statistical analysis of the essential properties of the efficiency results (Borger et al., 2008). If probability distributions of the non-parametric estimators are known, then the construction of confidence intervals, the correction for the bias, or conducting statistical inferences on efficiency scores are possible. Considering the very limited number of general analytic derivations for asymptotic sampling distributions of non-parametric frontier estimators, the introduction of the bootstrapping technique in the efficiency measurement literature in the past decade has led to important breakthroughs in approximating the sampling distributions of efficiency estimators. Simar and Wilson (2008) describe in detail the bootstrapping algorithms for estimating the empirical distribution of efficiency scores, correction for their bias, and construction of confidence intervals. The fundamentals of the bootstrapping methodology are as below.

Bootstrapping starts by obtaining a consistent estimation of the data generating process (DGP), the process that has led to the input-output data corresponding to all DMUs in the sample. Repeated simulation of this DGP (e.g., B repetitions) and applying the original non-parametric estimator to the obtained pseudo samples leads to new sets of efficiency scores such that the sampling distribution of the bootstrap efficiency scores mimics the distribution of the original estimator of efficiency scores.

Let X_n and X_n^* represent the original sample and the bootstrap sample, respectively. In addition, θ_A represents the true unknown technical efficiency score of DMU *A*, $\hat{\theta}_A$ represents the original estimate of the efficiency score, and $\hat{\theta}_A^*$ represents the bootstrap estimate of the efficiency score. One of the important results of the bootstrapping approach is that:

$$(\hat{\theta}_A - \theta_A) \mid X_n \overset{approx}{\sim} (\hat{\theta}_A^* - \hat{\theta}_A) \mid X_n^* \qquad \textbf{(8)}$$

The key point in this relation is that within the true world, $\hat{\theta}_A$ is an estimator of the unknown parameter θ based on the original sample X_n generated from the original DGP, but in the bootstrap world $\hat{\theta}_A^*$ is an estimator of $\hat{\theta}_A$ based on the pseudo sample X_n^* generated from the estimated DGP (Simar and Wilson, 2008). By approximating the distribution of the random variable $\hat{\theta}_A - \theta_A$ using the relation between the original estimate $\hat{\theta}_A$ and the bootstrap estimate $\hat{\theta}_A^*$, different statistical inferences on the estimated efficiency score $\hat{\theta}_A$ can be performed. The two analyses that are the focus of this paper are the correction for the bias and construction of confidence intervals for the efficiency scores.

The bias of the DEA estimator for DMU *A*, $bias_A = E(\hat{\theta}_A) - \theta_A$, can be approximated by $\hat{bias}_A = E(\hat{\theta}_A^*) - \hat{\theta}_A$. Thus, estimation of the bias-corrected efficiency score $\tilde{\theta}_A$ is obtained as $\tilde{\theta}_A = \hat{\theta}_A - \hat{bias}_A = 2\hat{\theta}_A - E(\hat{\theta}_A^*) = 2\hat{\theta}_A - B^{-1}\sum_B \hat{\theta}_A^*$, where B is the number of repetitions for simulating the DGP, obtaining the pseudo samples and applying the original non-parametric estimator to these pseudo samples (Simar and Wilson, 2008).

Construction of the confidence intervals of the efficiency scores in this paper will be based on an improved procedure proposed by Simar and Wilson (1999). For the ease of calculations, let $\delta_A = 1/\theta_A$ and $\hat{\delta}_A = 1/\hat{\theta}_A$. As it was discussed before, the DEA efficiency scores overestimate the true efficiency, thus $\hat{\theta}_A \geq \theta_A$ or $\hat{\delta}_A \leq \delta_A$ (Simar and Wilson, 2008). If the distribution of the random variable $\hat{\theta}_A - \theta_A$ or $\hat{\delta}_A - \delta_A$ were known, it would be trivial to find $c_{\alpha/2}$ and $c_{1-\alpha/2}$, the relevant quantiles of the distribution of $(\hat{\delta}_A - \delta_A)$, such that:

$$\text{prob}(c_{\alpha/2} \leq \hat{\delta}_A - \delta_A \leq c_{1-\alpha/2}) = 1-\alpha \quad \textbf{(9)}$$

As a result, (1-α)% confidence interval for efficiency score δ_A would be obtained as:

$$\hat{\delta}_A - c_{1-\alpha/2} \leq \delta_A \leq \hat{\delta}_A - c_{\alpha/2} \quad \textbf{(10)}$$

Considering that the quantiles $c_\alpha, \alpha \in [0,1]$ of the random variable $\hat{\delta}_A - \delta_A$ are unknown, the empirical bootstrap distribution of its counterpart, $\hat{\delta}_A^* - \hat{\delta}_A$, can be used to obtain the estimate $\hat{c}_\alpha$ of $c_\alpha, \alpha \in [0,1]$. Thus, the bootstrap approximation of (1-α)% confidence interval of δ_A would be obtained as follows (Simar and Wilson, 2008):

$$\hat{\delta}_A - \hat{c}_{1-\alpha/2} \le \delta_A \le \hat{\delta}_A - \hat{c}_{\alpha/2} \Rightarrow \frac{1}{\hat{\delta}_A - \hat{c}_{\alpha/2}} \le \theta_A \le \frac{1}{\hat{\delta}_A - \hat{c}_{1-\alpha/2}} \quad \textbf{(11)}$$

This procedure is used to construct confidence interval for the efficiency score of any DMU $(x, y) \in \Re_+^{N+M}$ for which $\hat{\theta}$ exists.

2.2.4 The Consideration of the Effect of Uncontrollable Environmental and Operational Factors

As it was discussed before, the performance of road authorities when making improvements in the road condition does not depend only on the technical efficiency of highway maintenance policies and practices, but also on uncontrollable environmental (such as climate condition) and operational (such as traffic and load, etc.) conditions. Uncontrollable factors may provide important explanation of the underperformance of the DMUs under analysis. These factors are non-discretionary, meaning that they are not under the discretion of the road authorities or maintenance managers. Furthermore, they are contextual factors, meaning that they are characteristics of the environment in which the transformation process (change in the road condition) takes place, thus can directly affect the obtained outputs.

There are various methodologies in the literature that integrate the uncontrollable factors into efficiency measurement. Simar and Wilson (2008) and Triantis, et al. (2010) provide a discussion on various existing approaches and their advantages/disadvantages. The first approach is called one-stage approach, in which uncontrollable variables are included in the linear programming formulation as either an input (if a favorable effect) or an output (if a mitigating effect) variables (Banker and Morey, 1986a). This approach is recommended if uncontrollable variables are considered as part of the transformation process and define the production possibility set. Almost all previous studies that have used DEA for measuring performance of highway maintenance operations have used this approach for considering the effect of uncontrollable factors in measuring performance of highway maintenance operations (Kazakov et al., 1989; Cook et al., 1990; Cook et al., 1994; Rouse et al., 1997; Ozbek, 2007; Rouse and Chiu, 2008; de la Garza et al., 2009; Fallah-Fini et al., 2009; Ozbek et al., 2010a; Ozbek et al., 2010b), since climate condition, load, traffic, and many other uncontrollable factors in this context are part of the transformation process and directly affect the road conditions observed after performing maintenance operations. The main disadvantage of this method is that it requires classification of uncontrollable factors as input or output variables prior to the analysis.

The other approach is the two-stage regression method. In the first stage of this approach, the original non-parametric efficiency scores are obtained. In the second stage, the observed efficiency patterns are explained in a regression model using the set of uncontrollable factors as independent variables. There are several disadvantages to this approach (Simar and Wilson, 2008): namely, (1) the first stage efficiency scores that are used as the dependent variable in the second stage regression are biased as previously discussed, thus it is useful to use the bias-corrected efficiency scores in the second stage regression; (2) the first stage efficiency scores are serially correlated, since they are obtained based on the data of other DMUs on the frontier. In addition, uncontrollable variables are correlated with inputs and outputs, thus, the uncontrollable variables (the independent variables in the second stage regression model) are correlated with the error term of the regression. All these issues lead to inconsistent parameter estimation in the second stage regression; and (3) a restrictive parametric model is used for the second-stage regression.

The three-stage approach is another technique developed by Ruggiero (Ruggiero, 1998) where a single environmental harshness factor that can best represent the effect of uncontrollable variables is developed. The environmental harshness factor is then used for grouping the DMUs such that each DMU is compared with those that are in the similar or worse environmental conditions. When more than one uncontrollable variable impacts output and when differences among the levels of those variables are subtle, defining a single environmental harshness factor to capture the effect of all uncontrollable variables needs further consideration (Triantis et al., 2010). For example, (Fallah-Fini et al., 2009) uses the Analytical Hierarchy Process as well as regression analysis to combine a large set of uncontrollable factors into one variable representing environmental/operational harshness in highway maintenance performance measurement context.

Additionally, in the literature there are other approaches that can be potentially useful when incorporating environmental variables in efficiency analysis, such as the conditional efficiency measures developed by Daraio and Simar (2005; 2007) in which production frontiers are developed conditioned on uncontrollable variables, or the two-stage approach developed by (Triantis et al., 2010) where multivariate methods along with fuzzy clustering (Seaver and Triantis, 1992) are used to group the DMUs based on the values of the environmental variables.

However, the focus of this research is on the innovative application of the bootstrapping techniques to address the difficulties associated with the standard two-stage approach explained before. Simar and Wilson (2007) suggests a bootstrapping algorithm where the bias-corrected efficiency scores are first constructed. The bias-corrected estimates of efficiency scores are then used in the second stage-regression as the dependent variable. Bootstrapping on the second stage regression leads to consistent estimates of the parameters of the regression model.

The algorithm assumes that sample observations (x_j, y_j, z_j) are i.i.d. observations from the random variable (*X*,*Y*,*Z*) with probability density function f(*x*, *y*, *z*) where *x*, *y*, and *z* are the input, output, and uncontrollable variables for DMU j, respectively. In addition, the support of the stated density function is assumed to be $\Psi \times R^r$ where Ψ is the production possibility set constructed by only input-output variables (**x**, **y**) and r represents the dimension of the uncontrollable variables. This is called the separatability assumption between the space of input-output variables and the space of environmental variables (Simar and Wilson, 2008). The relation between efficiency scores δ_j and the environmental variables z_j is assumed to be linear, i.e., $\delta_j = z_j\beta + \varepsilon_j$ where β represents the parameter vector and ε_j, distributed as N(0, σ), represents i.i.d. variables independent of z_j with left truncation at $1 - z_j\beta$, since the dependent variable $\delta_j = \frac{1}{\theta_j} \geq 1$ (based on the Equations (2) and (3), inverse of measures of efficiency are greater than or equal to one). Appendix A describes the details of the employed algorithm.

2.2.5 The Proposed Analytical Approach

After orderly arrangement of the methods and techniques described above, the methodological contribution of this paper can be described as follows:

1. A comprehensive set of key controllable input and output variables pertinent to the transformation process as well as a comprehensive set of external and uncontrollable factors that represent the environmental and operational conditions for road maintenance are developed.
2. DMUs under analysis are classified into different groups based on their characteristics as well as the objectives of the efficiency analysis. For example, in this paper road authorities (DMUs) are classified into two groups based on their contracting strategy (traditional and performance-based).

3. Group frontiers are developed for each group of DMUs using the Simar and Wilson (2007) two-stage bootstrapping technique. Thus the DEA efficiency score of each DMU with respect to its own group frontier is obtained and corrected for the bias.
4. DMUs from all groups are pooled together to develop a meta-frontier using the Simar and Wilson (2007) two-stage bootstrapping technique.
5. Efficiency scores of DMUs with respect to their own group frontiers as well as the meta-frontier are compared via the meta-technology ratio (O'Donnell et al., 2008) to arrive at the measures of performance of different groups of DMUs and potential improvements each group can have to shift to the meta-frontier.

By performing steps 3 and 4 of this methodology, the bias-corrected efficiency scores as well as their corresponding confidence intervals are calculated. Moreover, using the second stage regression, one can also examine if the observed efficiency patterns can be explained based on the environmental and operational conditions. To illustrate the implementation of this methodology as well as the intuition behind each step, an empirical study using real data is presented in the next section.

2.3 Empirical Application: Evaluating the Performance of Road Maintenance Operations

In this section, the proposed methodology is applied to an empirical dataset of pavement condition, traffic, climate condition, and maintenance expenditures for approximately 180 miles of Virginia's Interstate highways that was maintained by VDOT using traditional maintenance practices during the fiscal years 2002 to 2006 and for 250 miles of Virginia's Interstate highways maintained using a performance-based maintenance strategy over the fiscal years 2002 to 2004. The 180 miles maintained by VDOT under the traditional maintenance approach lie within seven counties and the 250 miles maintained using a performance-based maintenance strategy lie within twelve counties. All counties are geographically spread across the state of Virginia. Applying the methodology described in Section 2.2.5 provides the possibility for measuring and comparing performance of the two contracting strategies (traditional versus performance-based). In addition, one can also evaluate and compare the performance of different road authorities (e.g., counties or districts) that are using the same type of contract (traditional or performance-based). The latter type of analysis can capture the managerial inefficiencies that may cause some road authorities not to maintain the roads to their best possible conditions. Detailed discussion on the implementation of each step of the methodology is provided next.

2.3.1 Factors Considered in the Analysis

In order to select the variables that could be utilized in the performance measurement analysis, emphasis was placed on the effects of maintenance activities as well as on a set of explanatory variables that have caused these effects. As a result, a set of input, output, and uncontrollable variables was obtained. Following is the description of these variables, why they have been chosen and how their corresponding numerical values have been calculated (de la Garza et al., 2009; Fallah-Fini et al., 2009).

Lane-miles Served: In the real world, the Interstate inside each county is divided into several road sections and maintenance treatments are defined for each one of these sections. Due to the limited maintenance budget, not all the road sections in a county can be treated each year. The variable "Lane-miles Served" represents only the total lane-miles of the road sections that are maintained within each county. This variable captures the extent of the workload each county has performed, thus is considered as one of the outputs of the maintenance operations.

Change in Pavement Condition: VDOT uses the Critical Condition Index (CCI) to represent the condition of a road section with respect to the load-related and non load-related distresses. The CCI is the main factor to identify deteriorated road sections and the required maintenance operations (JLARC, 2002). The CCI varies between 0 and 100. In this paper, the change in the CCI of the road sections that have been maintained is used to capture the improvement in the road condition due to maintenance operations. In cases in which several road sections in a county are maintained, a weighted average of the change in CCI of those road sections is calculated, where the length of the road sections is used as the weight factor. The change in CCI of the road sections is the other output variable representing the quality of the performed maintenance operations.

Traffic: In order to capture the effect of traffic on pavement deterioration, the Annual Average Daily Traffic (AADT) data estimated and published by VDOT (VDOT, 2010) for each segment of the Interstate within the state of Virginia was used. AADT is the annualized average 24-hour volume of vehicles at a given section of highway (ADOT, 2010). The length of the road sections are used as the weighting factors for calculating the traffic along the entire length of Interstate that lies in a county. Obviously, large values for AADT increase the extent of pavement deterioration and require greater maintenance effort. Traffic is considered as an uncontrollable

factor in this paper and captures the operational conditions under which maintenance activities are performed.

Load: Loads are the vehicle forces exerted on the pavement (e.g., by trucks, heavy equipment, etc.) and obviously cause pavement degradation over time. In order to quantify the traffic load a pavement encounters, the concept of Equivalent Single Axle Load (ESAL) has been used. ESAL is a variable used to convert the pavement loads of various vehicles (passenger cars, light trucks, tracker trailers, etc) into the load of an 18,000 pound single-axle load (Fallah-Fini et al., 2009). VDOT publishes the percentages of AADT for six different types of vehicles each year (VDOT, 2010). By multiplying AADT by the vehicle distribution percentage, the number of vehicles of each type is obtained. The number of vehicles of each type is then multiplied by their corresponding ESAL factors developed by AASHTO (American Association of State Highway and Transportation Officials) and are added together to form the load corresponding to the traffic that affects a specific section of the road. The length of the section is used as a weighting factor for calculating the load present for the entire length of the Interstate that lies within a county. Load is also considered as an uncontrollable factor in this paper and captures the extent of deterioration due to vehicle forces.

Climate Factors: Climate condition factors, as an important set of uncontrollable factors, affect both maintenance efforts and the deterioration of the paved lanes. For example, the amount of snowfall in Southwest Virginia will influence the required winter maintenance workload in comparison with the counties in Southeast Virginia. Climate data was extracted from the National Climate Data Center (NCDC, 2010) and a dataset containing four variables: "Minimum Temperature", "Maximum Temperature", "Total Rainfall", and "Total Snowfall" for all the counties under analysis over the corresponding years was created. Although no clear relationship has been established between the climate factors and the condition of the pavement, it is reasonable to assume that these factors do influence the extent of deterioration (Ruggiero, 1998). Moreover, based on the physical and elevation maps for the state of Virginia, the "Mountainous" dummy variable was created to specify the counties that lie in a mountainous area and thus are experiencing more severe environmental and operational conditions. This variable takes the value of one if a county lies in a mountainous area and zero, otherwise.

Maintenance Expenditure: The cost data corresponding to the 180 miles maintained by VDOT using traditional contracting strategy includes the total costs of sub-contractors to VDOT and the

direct costs of routine maintenance activities for self-performed work (i.e., cost of labor, cost of material, and cost of equipment) at each county. A constant overhead percentage rate provided by VDOT's Controller Office was applied to the direct cost to provide VDOT's total cost. Finally the "BHWA-Highway and Street Construction Cost Index" developed by the Bureau of Labor Statistics was used as an inflation/deflation rate to adjustment the cost data of different years. The cost data corresponding to the 250 miles maintained with a performance-based strategy includes the direct cost of routine maintenance activities, overhead, as well as contractor's profit. Thus, the sum of the stated terms gives the total cost that has been imposed to VDOT for outsourcing highway maintenance operations under a performance-based contract. The performance-based maintenance cost data that has been provided has been lumped in two regions, namely, the total cost related to the counties in the Southeast Virginia region and the cost related to the counties in the Southwest Virginia region. The BHWA cost index was also applied to the contractor's cost data to adjust for inflation/deflation. Maintenance expenditure is one of the most important variables and the only controllable input variable used in this study.

Given the fact that the "maintenance expenditure" data for the traditional maintenance contracting strategy is only available at the county level, the definition of DMUs is limited to the counties of Virginia that encompass the sections of the Interstate system that are maintained using traditional maintenance practices. Thus, as suggested by Ozbek et al. (2010b), each county for each fiscal year is considered as a DMU. This definition for DMUs enables us to analyze trends in the county efficiency scores over time.

However, since the performance-based "maintenance expenditure" is available at an aggregated level for all the counties in the Southeast and Southwest Virginia regions, the definition of contractor's DMUs is potentially limited to the two regions for which the lumped cost data are available. This approach has two main drawbacks. First, treating each region for each year as one DMU leads to only six DMUs (two regions over three years) that are using performance-based contracts. Based on the discussion provided in Section 2.2.1 about the bias and rate of convergence of the estimated efficiency scores, the small number of DMUs accompanied with the large set of input/output variables will lead to a considerable increase in the bias of the efficiency scores and affect the validity of the conventional inferences. Second, defining the counties as the DMUs under the traditional approach and regions as the DMUs under performance-based approach can affect the validity of the comparison between efficiency

results of the two contracting strategies, since the units under analysis do not have the same characteristics. As a result, the lumped performance-based cost data corresponding to the Southwest and Southeast regions were disaggregated among the counties proportional to the total area of the road sections that have been maintained in each county in each year. Based on this assumption, the definition of DMUs can then be changed to the counties that encompass the section of the Interstate system that are maintained using the performance-based strategy.

Finally, the road condition data showed that not all the counties have performed maintenance operations in all years. In the end, 25 DMUs for the traditional approach and 26 DMUs for performance-based approach were used to form the two groups of DMUs (performance-based strategy versus traditional strategy) whose efficiency performance needed to be compared. The results of this comparison provide insights on the type of contracting strategy that has been more efficient when maintaining Virginia's Interstate.

Table 2-1 provides the descriptive statistics for the final dataset corresponding to the two groups of DMUs, i.e., performance-based and traditional maintenance strategies. Several remarks regarding the dataset should be mentioned. First, it is reasonable to assume that all DMUs have access to the same input market, thus are facing the same prices for the inputs that have been defined as part of the maintenance expenditure. Second, there are several weather stations in each county for reporting climate conditions. The precipitation and temperature data that are used in this paper for each county belong to the weather station that is closest to the Interstate under analysis in that county. Third, there were several instances of missing precipitation data for some of the counties under analysis. The missing precipitation data were approximated by the corresponding data from the second closest station in these counties. Fourth, considering the range of the values that have been observed for the minimum temperature or snowfall in each group, it is clear that the counties under analysis are spread in areas with different climatic conditions. Thus, climate conditions are hypothesized to be among the factors that can justify the differences among efficiency scores of DMUs within each group. The developed set of variables and corresponding datasets are used in the next section for evaluating the performance of counties when maintaining their highways.

2.3.2 Estimation of the Group Frontiers

As discussed in the previous section, the DMUs under analysis (i.e., counties) were classified into two groups, traditional and performance-based, according to the contracting

strategy they are using for maintaining the road sections in their administrative area. In this section, group frontiers are constructed so that the performance of each road authority (i.e., county) can be compared with the performance of the rest of road authorities that are using the same type of contract. Applying the Simar and Wilson (2007) bootstrapping algorithm described in Section 2.2.3 provides the possibility for correcting the inherent bias that exists in the non-parametric efficiency scores as well as for constructing the confidence intervals for the efficiency scores. In addition, the potential relations between the efficiency patterns and the set of uncontrollable factors can be evaluated.

Table 2-1: Descriptive Statistics Corresponding to Performance-based and Traditional Maintenance Strategies

Variable	Mean	Std. Dev.	Min	Max
Traditional Maintenance Strategy				
Output				
Change in CCI	37.5	13.74	13.00	74.20
Lane-miles Served (mile)	15.74	8.22	2.14	38.12
Input				
Maintenance Expenditure ($)	1,559,897	1,167,269	334,398	5,001,905
Uncontrollable Factors				
Traffic (Vehicles/Day)	22955	14716	4592	53971
Load (ESAL)	6136	3924	1362	11886
Min Temperature (°F)	23.86	2.99	18.20	29.5
Max Temperature (°F)	86.96	2.55	81.2	93.4
Rainfall (inches)	43.38	6.13	33.59	54.47
Snowfall (inches)	9.35	5.37	1.00	23.00
Mountainous dummy	0.56	0.51	0	1
Performance-Based Maintenance Strategy				
Output				
Change in CCI	29.49	13.23	7.48	64.00
Lane-miles Served (mile)	14.84	10.02	1.5	36.54
Input				
Maintenance Expenditure ($)	1,883,481	1,299,131	217,476	4,890,976
Uncontrollable Factors				
Traffic (Vehicles/Day)	29514	16753	13509	58366
Load (ESAL)	7414	1764	4735	11399
Min Temperature (°F)	22.9	3.10	17.5	29.5
Max Temperature (°F)	84.38	3.64	79.9	92.8
Rainfall (inches)	51.88	8.34	37.24	67.96
Snowfall (inches)	16.27	9.98	0	31.5
Mountainous dummy	0.54	0.51	0	1

To estimate an appropriate non-parametric technology frontier for each group, the orientation (input or output oriented) and returns to scale (variable (VRS) or constant (CRS) return to scale) of the DEA model should be defined. In terms of the returns to scale of the DEA

model, Ozbek (2007) as well as Rouse et al. (1997) discuss a significant presence of scale effects in highway maintenance operations. Moreover, the lower and upper bound of zero and 100 for CCI imposes a lower bound and upper bound for the variable "Change in CCI" as well. In addition, there is a limit for the maximum value that the variable "Lane-miles Served" can take and that is equal to the total area of the Interstate that lie in each county. Thus, a VRS frontier is needed to adjust for the upper and lower bounds of the output variables, since a CRS frontier continues extending linearly without taking any boundary constraint into account (Rouse and Chiu, 2008).

Choosing the orientation of the DEA model mainly depends on the objectives of the analysis. Given the budget limitations that road authorities are facing, maintenance managers mainly need to know the maximum number of lane-miles they can maintain. Thus, the output-oriented BCC model is chosen to analyze if road authorities can improve the output variables "Lane-miles Served" and "Change in CCI", given their limited "Maintenance Expenditure". Table 2-2 shows a summary of the computed efficiency scores in both groups[3]. These results can be used to evaluate how each road authority has performed on average within its own group (i.e., in comparison with other road authorities that are using the same type of contract). Note that these results cannot be used for comparing the groups' average efficiency scores, since they have been obtained with respect to different frontiers. Comparison across groups will be performed in Section 2.3.3.

Table 2-2: Summary of the Technical Efficiency Corresponding to Each Group

Group	Number of DMUs	Mean Efficiency	Std. Dev.	Min	Max	Number of efficient DMUs
Traditional Contracting Strategy	25	0.71	0.18	0.44	1	3
Performance-based Contracting Strategy	26	0.82	0.13	0.64	1	5

Table 2-3 shows the details of the efficiency scores for some of the counties with respect to their group frontier. The shortfalls in the efficiency scores represent the magnitude of improvement that can be achieved in outputs without any additional investment in maintenance expenditure and without the need to new technology. For example, Alleghany County in year 2002 can increase its outputs (the area that has been maintained and the improvement in the road condition) by 21% using the same amount of maintenance expenditure that has already been expended.

[3] We used software R (www.r-project.org) for computing the efficiency scores and running the bootstrapping algorithm.

In addition, exploring the patterns of efficiency scores in different counties can have some important policy implications. For example, Spotsylvania County has been 100% efficient over the years 2002 and 2003, but its efficiency score has decreased in the next two years. A similar trend in the efficiency scores has also been observed for Bland County. In contrast, Chesterfield and Henrico Counties show an improvement in their efficiency scores over the years 2002 to 2004. This may require maintenance managers of these counties to explore the changes in their policies and practices over the years under analysis as a potential source for the change in their performance. Further analysis of the results indicated that performance of some of the counties such as Augusta County (traditional approach) and Carroll County (performance-based approach) are of concern since their efficiency scores have been relatively low over all years. By investigating operational and strategic policies of the set of peers corresponding to these two counties, maintenance managers can identify the changes needed to improve the performance of the stated counties.

Table 2-3: Technical Efficiency of Selected Road Authorities with Respect to their Group Frontier

County	2002	2003	2004	2005	2006
Traditional Contracting					
Alleghany	0.79	1.00	1.00	0.63	1.00
Augusta	0.64	0.78	0.44	0.56	NA
Fauquier	0.53	0.71	0.70	0.55	NA
Spotsylvania	1.00	1.00	0.77	0.70	NA
Performance-based Contracting					
Bland	1.00	0.70	0.64	NA	NA
Carroll	0.72	0.74	0.68	NA	NA
Chesterfield	0.77	0.99	1.00	NA	NA
Henrico	0.76	0.86	1.00	NA	NA
Hanover	0.91	0.83	1.00	NA	NA

Note that the efficiency results presented in Table 2-2 and 2-3 do not show any correction for the bias that inherently exists in the non-parametric frontier estimations. To correct and analyze the bias associated with the estimated efficiency scores, the Simar and Wilson (2007) bootstrapping algorithm was used for each group of DMUs. Table 2-4 shows a summary of the results only for a small selection of DMUs from the most efficient to the most inefficient observations in both groups. As it can be seen, Table 2-4 shows the original (uncorrected) efficiency scores, the bias-corrected efficiency scores, and the lower/upper bounds for the 90% confidence intervals of the efficiency scores.

Table 2-4: Bootstrapping Results (the bias-corrected efficiency scores and corresponding confidence intervals)

County	Original Eff. Score	Bias-corrected Eff. Score	Lower Bound (5%)	Upper Bound (95%)
Traditional Contracting				
Spotsylvania(2002)	1.00	0.85	0.78	0.92
Alleghany(2004)	1.00	0.78	0.72	0.88
Albemarle(2006)	0.87	0.69	0.64	0.79
Spotsylvania(2004)	0.77	0.65	0.60	0.71
Augusta(2002)	0.64	0.52	0.48	0.57
Roanoke(2004)	0.56	0.49	0.46	0.52
Augusta(2004)	0.44	0.36	0.33	0.40
Performance-based Contracting				
Chesterfield(2004)	1.00	0.90	0.85	0.97
Dinwiddie(2004)	1.00	0.85	0.77	0.94
Hanover(2002)	0.91	0.83	0.77	0.89
Hanover(2003)	0.83	0.73	0.70	0.78
Chesterfield(2002)	0.77	0.71	0.67	0.75
Washington(2004)	0.68	0.62	0.59	0.65
Bland(2004)	0.64	0.56	0.53	0.60

Analyzing the bias corrected results showed that DMUs that are fully efficient (i.e., the DMUs that define the frontier) have the largest value for bias, such as Alleghany (traditional approach) and Dinwiddie (performance-based approach) counties in year 2004. As was discussed in Section 2.2.1, the density of observations (DMUs) around different sections of the frontier can highly affect the size of the bias. Thus, the large bias of the efficient DMUs can potentially be attributed to the fact that not enough observations in any of the groups are used for constructing the frontiers. Heterogeneity of observations within each group, due to the fact that DMUs (counties) belong to different environmental conditions, can be another important reason for the large bias that has been introduced in the computed efficiency scores (Borger et al., 2008). Overall, the average bias amounted to 13% and 9% for DMUs that are using the traditional and performance-based strategies, respectively. Another important remark about the bootstrapping results is that for many observations, the original (uncorrected) efficiency scores are placed outside the constructed confidence intervals. This kind of behavior underlines the risk associated with using the uncorrected efficiency scores (Borger et al., 2008). Finally, the bootstrapping results suggest that the DMUs whose uncorrected efficiency scores are 1.00 are not potentially 100% efficient. They seem to be efficient only due to the small sample sizes that are available for the analysis in both groups.

The next step is to provide an answer to the question whether any of the uncontrollable (environmental and operational) factors presented in Table 2-1 can significantly justify the difference in efficiency scores among the DMUs within each group. Thus, based on the bootstrapping algorithm described in Section 2.2.4, the bias-corrected efficiency scores in each group are regressed on the set of uncontrollable factors to obtain the bootstrapped second-stage estimates of the parameters of a smooth and continuous function that can capture the relation between the efficiency scores and uncontrollable factors. Exploring different functional forms on different sets of uncontrollable variables in both datasets finally showed that the inverse of the bias-corrected efficiency scores shows a significant relation with the maximum temperature and log of the snowfall in the traditional contracting dataset and with the maximum and minimum temperatures in the performance-based contracting dataset. The details of the regression results and the confidence intervals for the estimated model parameters are depicted in Table 2-5. As it can be seen, the bootstrapped confidence intervals for the estimated parameters do not contain zero, meaning that the parameter estimates are significantly different from zero.

Table 2-5: The results of the Bootstrapped Second-stage Regression on the Uncontrollable Factors

Performance-based Contracting					**Traditional Contracting**			
Variable	Parameter Value	Lower Bound (2.5%)	Upper Bound (97.5%)		Variable	Parameter Value	Lower Bound (2.5%)	Upper Bound (97.5%)
Intercept	4.724	3.569	6.135		Intercept	10.690	5.217	14.102
Max Temp.	-0.030	-0.050	-0.014		Max Temp.	-0.105	-0.147	-0.044
Min Temp.	-0.034	-0.058	-0.012		Log(Snow)	0.118	0.041	0.329

Note that in these models the inverse of the bias-corrected efficiency scores is used as the dependent variable, hence smaller values for the dependent variable mean better efficiency scores. The parameter values for performance-based contracting presented in Table 2-5 indicate that those road authorities whose minimum temperature is lower (representing a worse environmental condition) have worse efficiency scores (i.e., higher values for the inverse of efficiency scores). In terms of the traditional contracting approach, the log of snowfall stood out as a significant variable. Thus higher values for the snowfall indicate a worse environmental condition (i.e., higher values for the inverse of efficiency scores). Thus, the observed positive sign for the coefficient of log(snowfall) is correctly expected.

Moreover, in both models the variable maximum temperature has shown a significant relation and its coefficient has a negative sign. This means that the counties with higher maximum temperature have shown better efficiency scores. This may seem counter-intuitive

initially, since based on the physics of road deterioration, higher frequency and severity of extreme hot days leads to problems related to pavement softening as well as load-related rutting (NRCAN, 2010). The fact is that the stated scenario can potentially happen in states such as Texas or Arizona which experience severe hot days during the Summer with extreme temperatures much higher than 100 °F. As it was shown in Table 2-1, the averages of maximum temperature in both groups are around 80 °F. In addition, there are only three DMUs in the traditional contracting dataset and one DMU in the performance-based dataset that have maximum temperatures around 90 °F. Thus, as our datasets show (and also with respect to the geographical location of the state of Virginia), the variable maximum temperature is not contributing as a deterioration factor. Instead, higher values for the maximum temperature mean that those counties are experiencing a better environmental condition overall and have been able to come up with better results after performing maintenance operations. No other uncontrollable factor in any of the datasets could significantly explain the dispersion of the efficiency scores.

2.3.3 Estimation of the Meta-Frontier

In Section 2.3.2, the relative performance of each road authority within its group (traditional contracting or performance-based contracting) was evaluated. However, considering the main objective of the paper, the issue of considerable interest is to measure the relative performance of road authorities that are using the traditional approach with the efficiency levels of road authorities that are using the performance-based approach. As a general rule, road authorities' efficiency levels measured with respect to one frontier (e.g., traditional contracting frontier) cannot be directly compared with other road authorities' efficiency levels measured with respect to another frontier (e.g., performance-based contracting frontier).

In this section, based on the meta-frontier approach, a common meta-frontier for measuring the relative efficiency of road authorities across groups is computed. As it was described in Section 2.2.2, the meta-frontier envelopes the two group frontiers that were computed in Section 2.3.2. So, each road authority's efficiency score measured relative to the meta-frontier can be decomposed into two components: the first component measures the distance of the road authority to its group frontier and the second component measures the gap between the group frontier and meta-frontier. Evaluation of this gap using the meta-technology ratio helps with analyzing the effects of the physical and economic characteristics of contract types (e.g.,

regulations, limitations, quality of labor force, type of machinery, etc.) on the performance of road authorities (i.e., more efficient use of available resources).

To construct the meta-frontier, the datasets corresponding to both traditional and performance-based contracting strategies were pooled. Table 2-6 shows the descriptive statistics for the pooled dataset. Next, the output-oriented BCC model with the same structure as before was applied to the pooled dataset to measure the efficiency of the road authorities relative to the estimated meta-frontier, irrespective of their contract type. As Table 2-6 shows, the uncontrollable factors such as rainfall and snowfall have a wide range in the pooled dataset. This again confirms the fact that DMUs (counties) are geographically spread across the state of Virginia and are experiencing different environmental conditions. Thus, it is postulated that the environmental factors play an important role in explaining the differences among efficiency scores of DMUs in the pooled dataset.

Table 2-6: Descriptive Statistics Corresponding to the Pooled Dataset

Variable	Mean	Std. Dev.	Min	Max
Output				
Change in CCI	30.74	12.99	7.48	64.00
Lane-miles Served (miles)	13.76	9.32	1.5	33.94
Input				
Maintenance Expenditure ($)	1,721,944	1,175,222	217,476	4,104,299
Uncontrollable Factors				
Traffic (Vehicles/Day)	30847	16762	14035	58366
Load (ESAL)	7620	1676	5099	11399
Min Temperature (°F)	22.18	3.14	17.50	29.50
Max Temperature (°F)	84.59	3.71	79.90	92.80
Rainfall (inches)	52.14	8.54	37.24	67.96
Snowfall (inches)	16.02	9.87	0	31.50
Mountainous dummy	0.5	0.51	0	1

Table 2-7 reports the summary statistics of the efficiency scores with respect to the group frontiers, meta-frontier, as well as the meta-technology ratios for each county and for each group. For example, Table 2-7 shows that technical efficiency score for Hanover County with respect to its group frontier was estimated to vary between 0.83 and 1 during the three-year period with an average of 0.91. In comparison with the meta-frontier, Hanover County's efficiency score is estimated to vary between 0.64 and 0.83 with an average of 0.70. Obviously, the efficiency scores of road authorities with respect to the meta-frontier are less than or equal to the efficiency scores obtained with respect to the group frontiers. The average meta-technology ratio of 0.77 for

Table 2-7: Summary of the Technical Efficiency Corresponding to Each Road Authority in the Pooled Dataset

County		Technical Efficiency wrt Group frontier				Technical Efficiency wrt Meta-frontier				Meta-technology Ratio			
		Mean	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
Traditional Contracting	Albemarle	0.71	0.14	0.58	0.87	0.66	0.18	0.52	0.87	0.93	0.12	0.78	1.00
	Alleghany	0.88	0.16	0.63	1.00	0.88	0.16	0.63	1.00	1.00	0	1.00	1.00
	Augusta	0.60	0.14	0.44	0.78	0.60	0.14	0.44	0.78	1.00	0	1.00	1.00
	Fauquier	0.62	0.09	0.53	0.71	0.62	0.09	0.53	0.71	1.00	0	1.00	1.00
	Roanoke	0.53	0.05	0.46	0.59	0.50	0.06	0.44	0.56	0.94	0.13	0.75	1.00
	Rockbridge	0.71	0	0.71	0.71	0.71	0	0.71	0.71	1.00	0	1.00	1.00
	Spotsylvania	0.87	0.15	0.70	1.00	0.87	0.15	0.70	1.00	1.00	0	1.00	1.00
Performance-based Contracting	Bland	0.78	0.19	0.64	1.00	0.63	0.28	0.44	0.96	0.79	0.14	0.69	0.96
	Carroll	0.71	0.03	0.68	0.74	0.55	0.04	0.52	0.60	0.77	0.04	0.73	0.82
	Chesterfield	0.92	0.13	0.77	1.00	0.58	0.21	0.41	0.82	0.64	0.21	0.41	0.83
	Dinwiddie	1.00	0	1.00	1.00	1.00	0	1.00	1.00	0.77	0.13	0.64	0.91
	Hanover	0.91	0.08	0.83	1.00	0.70	0.10	0.64	0.83	0.77	0.13	0.64	0.91
	Henrico	0.87	0.12	0.76	1.00	0.53	0.13	0.39	0.67	0.61	0.14	0.52	0.78
	Prince George	0.1	0	1.00	1.00	0.86	0.12	0.78	0.94	0.86	0.12	0.78	0.94
	Smyth	0.73	0.04	0.70	0.76	0.56	0.07	0.50	0.61	0.76	0.06	0.71	0.80
	Washington	0.75	0.08	0.68	0.85	0.53	0.23	0.39	0.80	0.69	0.21	0.56	0.94
	Wythe	0.70	0.04	0.64	0.73	0.45	0.10	0.38	0.57	0.65	0.13	0.51	0.79
Traditional Contracting Group		0.71	0.18	0.44	1.00	0.70	0.19	0.44	1.00	0.98	0.07	0.75	1.00
Perf-based Contracting Group		0.82	0.13	0.64	1.00	0.61	0.19	0.38	1.00	0.73	0.15	0.41	1.00

Note: Table 2-7 reports the original efficiency scores. Using the bias-corrected efficiency scores leads to the same conclusions. However, differences in magnitude of the bias for each DMU with respect to the group and meta-frontiers occasionally results in meta-technology ratios greater than one.

Hanover County shows that the improvement in the road condition achieved by this county using the performance-based contracting approach is 77% of the improvements that could potentially be achieved using the unrestricted meta-technology. Thus, Hanover County can improve its outputs by 23% if it is not limited to regulations as well as physical and economic characteristics of performance-based contracting.

Looking at the group level efficiency scores in Table 2-7 shows that the technical efficiency of the performance-based contracting group is 0.82 with respect to its group frontier. This means that, on average, road authorities that are using performance-based contracting can improve their outputs by 18% using the same amount of maintenance expenditure they have already expended. In addition, the technical efficiency of the performance-based contracting group with respect to the meta-frontier is 0.61. This means that based on the unrestricted meta-technology (with no limitations and regulations), the road authorities that are using performance-based contracting could improve their output by 39% (more than twice as the potential improvement they could achieve based on the performance-based contracting). The average meta-technology ratio of 0.73 for performance-based contracting group shows that on average, the maximum output that can be achieved by this group is 73% of the maximum output that can potentially be achieved using the unrestricted meta-technology (by removing limitations and regulations).

As Table 2-7 shows, a large number of counties in the traditional contracting group have shown the meta-technology ratio of 1. This means that the counties that are using the traditional maintenance contracting approach are playing a more important role in constructing the meta-frontier than the counties that are using the performance-based contracting approach. In other words, the tangency of the meta-frontier and the traditional contracting group frontier is much more than the tangency between the meta-frontier and the performance-based contracting group frontier. This can also be seen by comparing the average meta-technology ratios of 0.98 and 0.73 for the traditional and performance based contracting groups, respectively. This preliminary finding suggests that in the road authorities under analysis over the years 2002 to 2006, traditional contracts for performing pavement maintenance operations have been more efficient than performance-based contracts. Note that traditional contracting focuses on the lowest-bid combined with method-based specifications. In contrast, performance-based contracting, in its purest form, focuses on life-cycle cost without detailing on how, when, and where the work

should be performed (NCHRP, 2009). Our finding may suggest that VDOT should not rely only on LOS specifications in performance-based contracting, instead VDOT may want to use some hybrid approaches by bringing some of the features of traditional highway maintenance contracting into performance-based maintenance. This is in fact the expectation created by the meta-frontier in that all DMUs should have access to the best practices associated with each group.

Another plausible explanation for our finding is that traditional contracting has been used and tested many times with different road authorities, but this is VDOT's first experience with performance-based contracting. As a result, there are many reasons that can affect the successful implementation of PBCs, such as contractor's quality/capacity, the acquisition/award process, managing the cultural change inside the organization, the methods used for monitoring and evaluating the contractors, risk management process, etc. (NCHRP, 2009). It is very possible that VDOT's shift to the needed PBC culture had not been fully developed during the execution of this first performance-based maintenance pilot project; in fact the culture needed to support PBC is perhaps 180 degrees apart from the Traditional one. Furthermore, the contractor industry is not used to think in terms of "life-cycle" costing models which are radically different from "first cost" models. These and other dimensions should be considered while analyzing the numerical results from Table 2-7.

Just like in Section 2.3.2, the bootstrapping algorithm is used to see which of the uncontrollable (environmental and operational) factors can significantly justify the difference in DMUs' efficiency scores evaluated with respect to the meta-frontier. Note that in the pooled datasets the observations (DMUs) belong to two different types of contracts. Thus, a dummy variable representing the contract type should be added to the set of independent variables. After exploring various functional forms, the inverse of the bias-corrected efficiency scores showed a significant relation with the dummy variables "Mountainous" and "Contract type". The details of the regression results and the confidence intervals for the estimated model parameters can be observed in Table 2-8. As it can be seen, the 90% bootstrapped confidence intervals for the estimated parameters do not contain zero, meaning that the parameter estimates are significantly different from zero.

Note that the coefficient of the "Mountainous" dummy variable shows a positive relation with the inverse of the efficiency scores (dependant variable), meaning the counties that are

located in a mountainous area with more severe climate condition have maintained lower efficiency scores. Moreover, significance of the "Contract Type" dummy variable in this regression again confirms that the type of maintenance contracts has a significant role in the difference among efficiency scores of different road authorities and they do not belong to the same technology (frontier). Note that this dummy variable takes the value of one if a DMU has used the traditional contracting approach and zero, otherwise. Thus, the negative sign of the dummy variable implies that those road authorities that have used the traditional contracting approach have lower dependant variable (i.e., higher efficiency scores).

The Kruskal-Wallis test was also run to test if there is any significant statistical difference between the efficiency scores of road authorities that have used traditional contracting and those that have used performance-based contracting. The test results showed that the efficiency scores of road authorities under different types of contracting strategies are significantly different at 5% confidence interval.

Table 2-8: The Results of the Bootstrapped Second-stage Regression on the Uncontrollable Factors

Variable	Parameter Value	Lower Bound (5%)	Upper Bound (95%)
Intercept	1.82	1.60	2.02
Mountainous dummy variable	0.437	0.208	0.666
Contract Type dummy variable	-0.473	-0.699	-0.244

2.4 Conclusions

This paper utilizes recent developments on non-parametric (DEA) frontier estimations to develop a framework for evaluating and comparing the performance of highway maintenance projects that are using different types of contracts. The developed approach is applied to an empirical dataset of pavement condition, traffic, climate condition, and maintenance expenditures for approximately 180 miles of Virginia's Interstate highways maintained by VDOT using traditional maintenance practices and 250 miles of Virginia's Interstate highways maintained using a performance-based maintenance strategy.

The non-parametric meta-frontier framework is exploited in this paper to study the differences in efficiency across groups of heterogeneous DMUs (i.e., road authorities that are using traditional maintenance contracting and road authorities that are using performance-based maintenance contracting). The meta-technology ratio that is computed for each group depicts how close a group frontier is to the meta-frontier. Evaluation of this gap helps road authorities and decision makers by analyzing the effects of the economic and physical characteristics of

contract types (e.g., regulations, limitations, size and quality of labor force, type of machinery, etc.) on the performance of road authorities. Thus, potential improvements in performance of road authorities resulting from any change in regulations and limitations corresponding to contract types can be assessed.

Comparing the meta-technology ratio of the traditional maintenance contracting group (98%) with that of the performance-based maintenance contracting group (73%) in the dataset under analysis showed that, based on our historical data, road authorities that have used traditional contracting for transforming the societal resources for the improvement of the road conditions are more efficient than road authorities that have used the performance-based contracting approach. As it was discussed, since this is VDOT's first experience in using performance-based contracting, many reasons can potentially justify the under-performance of PBCs. For example, difficulty in implementing the needed PBC organizational culture, challenges associated with "life-cycle" costing models needed in PBCs versus "first cost" models needed in traditional contracts, quality/capacity of the contractors, etc. Moreover, using some hybrid contracting approaches by bringing some of the features/best practices of traditional highway maintenance contracting into performance-based maintenance can potentially be very helpful with improving the efficiency of highway maintenance contracting. Additional research using a more comprehensive dataset that includes more road authorities across the state of Virginia is recommended for evaluating the validity of these preliminary results.

As it has been discussed in the literature, the non-parametric measures of efficiency are biased by construction. Thus, the Simar and Wilson (2007) bootstrapping technique was applied on the non-parametric (DEA) group frontiers as well as on the meta-frontier to correct for the bias in the estimated efficiency scores, and also to construct confidence intervals for efficiency scores. For many road authorities, the original (uncorrected) efficiency scores are located outside the constructed confidence intervals. This finding underlines the risk involved with using uncorrected efficiency scores. Based on this observation, road authorities whose uncorrected efficiency scores are 1.0 are not in reality 100% efficient. They seem to be efficient only due to the relatively small sample sizes that were available for the analysis in both groups.

The Simar and Wilson (2007) bootstrapping technique also revealed that the patterns in the estimated efficiency scores with respect to group frontiers and the meta-frontier can significantly be explained by environmental factors, in particular minimum/maximum temperatures, snowfall,

and characteristics such as being located in a mountainous area. This finding suggests that road authorities that are in worse environmental conditions have been less efficient in improving road conditions. Introducing the environmental variables as an important potential source of inefficiency for road authorities can considerably help decision makers with their budget planning. Considering the important role played by the uncontrollable factors in the deterioration process of the Interstate, more extensive research needs to be done on finding alternative methods for incorporating these uncontrollable factors in performance measurement.

Despite the data limitations this paper has, this line of research provides the decision-makers with proper knowledge of the efficiency level of different highway maintenance contracts or projects. This is crucial for guiding future decisions regarding the renewal of contracts, the pricing of these contracts, and the potential efficiency improvement opportunities, since past performance is one of the very important criteria that road authorities can consider when awarding new contracts.

Acknowledgment: We would like to acknowledge the assistance of the Virginia Department of Transportation for providing the data in this research. This research is funded by the National Science Foundation, Award # CMMI-0726789. Any opinions and/or findings are those of the authors and do not necessarily represent the views of the sponsors.

References

ADOT. 2010. Traffic Data by Arizona Department of Transportation, http://tpd.az.gov/data/aadt.php.

Anastasopoulos PC, McCullouch BG, Gkritza K, Mannering FL, Kumares SC. 2009. A Cost Saving Analysis for Performance-based Contracts For Highway Maintenance Operations. ASCE Journal of Infrastructure Systems, http://dx.doi.org/10.1061/(ASCE)IS.1943-555X.0000012:

ASCE. 2009a. American Society of Civil Engineers: Reportcard for America's Infrastructure, http://www.infrastructurereportcard.org/fact-sheet/roads. Retrieved October 2009,

ASCE. 2009b. American Society of Civil Engineers: Facts About Roads, www.asce/reportcard. Retrieved October 2009,

Banker RD, Morey RC. 1986a. Efficiency Analysis for Exogenously Fixed Inputs and Outputs. Operations Research **34** (4): 513-521.

Battese GE, Rao D. 2002. Technology Gap, Efficiency, and a Stochastic Metafrontier Function. International Journal of Business and Economics **1** (2): 87-93.

Battese GE, Rao D, O'Donnell CJ. 2004. A Metafrontier Production Function for Estimation of Technical Efficiency and Technology gaps for Firms Operating Under Different Technologies. Journal of Productivity Analysis **21** (1): 91-103.

Borger B, Kerstens K, Staat M. 2008. Transit Cost and Cost Efficiency: Bootstrapping Non-parametric Frontiers. Research in Transportation Economics **23**: 53-64.

Charnes A, Cooper WW, Rhodes E. 1978. Measuring the efficiency of decision making units. European Journal of Operational Research **2** (4): 429-444.

Cook WD, Kazakov A, Roll Y. 1990. A DEA Model for Measuring the Relative Efficiency of Highway Maintenance Patrols. INFOR **28** (2): 131-124.

Cook WD, Kazakov A, Roll Y. 1994. "Chapter 10: On the Measurement and Monitoring of Relative Efficiency of Highway Maintenance Patrols". Data Envelopment Analysis: Theory, Methodology and Applications. A. Charnes, W. Cooper, A. Y. Lewin and L. M. Seiford, . Boston, Kluwer Academic Publishers. 195-210.

Daraio C, Simar L. 2005. Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach. Journal of Productivity Analysis **24** (1): 93–121.

Daraio C, Simar L. 2007. Conditional Nonparametric Frontier Models for Convex and Nonconvex Technologies: A Unifying Approach. Journal of Productivity Analysis **28**: 13–32.

de la Garza JM, Fallah-Fini S, Triantis K. 2009. Efficiency Measurement of Highway Maintenance Strategies Using Data Envelopment Analysis. In Proceedings of the Proceedings of 2009 NSF Engineering Research and Innovation Conference. Hawaii, USA.

Fallah-Fini S, Triantis K, de la Garza JM. 2009. Performance measurement of highway maintenance operation using data envelopment analysis: Environmental considerations. In Proceedings of the IIE Annual Conference. Miami, FL.

FHWA. 2003. Special Experimental Project Task Force 14. Retrieved October 2009,

JLARC. 2002. Adequacy and Management of VDOT's Highway Maintenance Program. Richmond.

Kazakov A, Cook WD, Roll Y. 1989. Measurement of Highway Maintenance Patrol Efficiency: Model and Factors. Transportation Research Record **1216**:

McCullouch BG, Anastasopoulos PC. 2009. Performance-based Contracting-Yes or No: An in Depth Analysis. 12th AASHTO-TRB Maintenance Management Conference.

NCDC. 2010. Natinal Climate Data Center Archive. Retrieved May 2009,

NCHRP. 2009. NCHRP Synthesis 389 : Performance-based Contracting for Maintenance, Transportation Research Board of National Academies.

NRCAN. 2010. Climate change impacts and adaptation: A Canadian perspective (impacts on transportation infrastructure). Retrieved March 2010,

O'Donnell CJ, Prasada Rao DS, Battese GE. 2008. Metafrontier Frameworks for the Study of Firm-level Efficiencies and Technology Ratios. Empirical Economics **34**: 231-255.

Ozbek EM. 2007. Development of a comprehensive framework for the efficiency measurement of road maintenance strategies using data envelopment analysis. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. Blacksburg.

Ozbek M, de la Garza JM, Triantis K. 2010a. Data and Modeling Issues Faced during the Efficiency Measurement of Road Maintenance using Data Envelopment Analysis. ASCE, Journal of Infrastructure Systems **16** (1): 21-30.

Ozbek ME, de la Garza JM, and Triantis K. 2010b. Efficiency Measurement of Bridge Maintenance using Data Envelopment Analysis. ASCE, Journal of Infrastructure Systems **16** (1): 31-39.

Ozbek ME, de la Garza JM, Triantis K. 2009. Data Envelopment Analysis as a Decision Making Tool for the Transportation Professionals. Journal of Transportation Engineering **135** (11): 822-831.

Rouse P, Chiu T. 2008. Towards Optimal Life Cycle Management in a Road Maintenance Setting Using DEA. European Journal of Operational Research doi: 10.1016/j.ejor.2008.02.041:

Rouse P, Putterill M, Ryan D. 1997. Towards a General Managerial Framework for Performance Measurement: A Comprehensive Highway Maintenance Application. Journal of Productivity Analysis **8**: 127–149.

Ruggiero J. 1998. Non-discretionary Inputs in Data Envelopment Analysis. European Journal of Operational Research **111** (461-469):

Seaver WL, Triantis K. 1992. A Fuzzy Clustering Approach for Measuring Technical Efficiency in Manufacturing. Journal of Productivity Analysis **3** (337-363):

Simar L. 1992. Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-parametric, and Semi-parametric Methods with Bootstrapping. Journal of Productivity Analysis **3**: 167-203.

Simar L, Wilson P. 2007. Estimation and Inference in Two-stage Semi-parametric Models of Production Processes. Journal of Econometrics **136** (1): 31-64.

Simar L, Wilson P. 2008. Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. The Measurement of Productive Efficiency and Productivity Chaneg. H Fried, Lovell CAK, Schmidt S. New York, Oxford University Press. 421-521.

Thanassoulis E. 1993. A Comparison of Regression Analysis and Data Envelopment Analysis as Alternative Methods for Performance Assessments. Journal of Operational Research Society **44** (I1): I 29-1144.

TRB. 2006. Maintenance and Operations of Transportation Facilities 2005 Strategic Vision. E-C092.

Triantis K, Seaver WL, Sarayia D. 2010. Using Multivariate Methods to Incorporate Environmental Variables for Local and Global Efficiency Performance Analysis. forthcoming, Informational Systems and Operational Research:

VDOT. 2010. Virginia traffic data publications, http://www.virginiadot.org/info/ct-TrafficCounts.asp. Retrieved March 2010,

Appendix A: Bootstrapping Algorithm for Construction of Bootstrapped Efficiency Scores as well as the Second Stage Regression (Simar and Wilson, 2007)

1. Using the original data in X_n (set of controllable input and output variables) to estimate the original input oriented efficiency scores $\hat{\delta}_i = \hat{\delta}(x_i, y_i \mid \hat{\Psi})$ based on a non-parametric DEA model (where $\hat{\delta}_i = \frac{1}{\hat{\theta}_i}$).
2. Use the maximum likelihood method to estimate the truncated regression of $\hat{\delta}_i$ on the set of uncontrollable variables z_i ($\hat{\delta}_i = z_i \beta + \varepsilon_i$) to obtain the estimates $\hat{\beta}$ and $\hat{\sigma}_\varepsilon$.
3. Loop over the steps 3.1 to 3.4 L_1 times to obtain L_1 bootstrap estimates $\hat{\delta}_i^*$ of $\hat{\delta}_i$ for each DMU *i*.
   3.1. For each DMU i, draw a sample ε_i^* from the $N(0, \hat{\sigma}_\varepsilon^2)$ distribution truncated at $1 - z_i \hat{\beta}$.
   3.2. For each DMU i, compute $\delta_i^* = z_i \hat{\beta} + \varepsilon_i^*$.
   3.3. Construct the pseudo sample (x_i^*, y_i^*) by setting $y_i^* = y_i, x_i^* = x_i \frac{\delta_i^*}{\hat{\delta}_i}$ for all DMUs.
   3.4. Calculate the bootstrapped efficiency estimates $\hat{\delta}_i^* = \delta(x_i, y_i \mid \hat{\Psi}^*)$ for each DMU where $\hat{\Psi}^*$ is the production possibility set estimated based on the pseudo sample (x_i^*, y_i^*).
4. For each DMU i, compute the bias-corrected efficiency score $\tilde{\delta}_i = 2\hat{\delta}_i - L_1^{-1} \sum_{L_1} \hat{\delta}_i^*$.
5. Use the maximum likelihood method to estimate the truncated regression of $\tilde{\delta}_i$ on the set of uncontrollable variables z_i to obtain the estimates $\hat{\hat{\beta}}$ and $\hat{\hat{\sigma}}$.
6. Use the steps 6.1 to 6.3 L_2 times to find a set of bootstrap estimates $\hat{\beta}^*$ and $\hat{\sigma}^*$.
   6.1. For each DMU i, draw a sample ε_i^* from the $N(0, \hat{\hat{\sigma}})$ distribution truncated at $1 - z_i \hat{\hat{\beta}}$.
   6.2. For each i compute $\delta_i^{**} = z_i \hat{\hat{\beta}} + \varepsilon_i^*$.
   6.3. Use the maximum likelihood method to estimate the truncated regression of δ_i^{**} on the set of uncontrollable variables z_i to obtain the estimates $\hat{\beta}^*$ and $\hat{\sigma}^*$.
7. Use the bootstrap estimates $(\hat{\beta}^*, \hat{\sigma}^*)$ and the original estimates $(\hat{\hat{\beta}}, \hat{\hat{\sigma}})$ to construct confidence intervals for β and σ.

Chapter 3 Optimizing Highway Maintenance Operations: Dynamic Considerations[4]

Abstract

Effective highway maintenance depends on several activities, including the understanding of current and the prediction of future pavement conditions and deciding how to best allocate limited resources for maintenance operations. In this paper, a dynamic micro-level simulation model of highway deterioration and renewal processes is presented. This model is calibrated with data from eight road sections in Virginia and is coupled with an optimization module that optimizes maintenance operations. The analysis offers alternative priority setting schemes that improve current maintenance practices at the project and network levels. This approach provides a blueprint for designing optimal highway maintenance practices.

Key Words and Phrases: Highway maintenance; pavement management systems; micro simulation; optimal maintenance budget allocations.

[4] Used with Permission of John Wiley and Sons, 2010.

Fallah-Fini, S., Rahmandad, H., Triantis, K., de la Garza, J.M., 2010, **Optimizing Highway Maintenance Operations: Dynamic Considerations**, System Dynamics Review, Special Issue: System Dynamics and Transportation, Volume 26, Issue 3, pages 216–238.

3.1 Introduction

Gradual deterioration of the U.S. road infrastructure in the past two decades (ASCE, 2009a) along with major budgetary restrictions for making significant improvements in road conditions (ASCE, 2009b) have raised the importance of improving highway maintenance practices. In fact, the Federal Highway Administration (FHWA) has endorsed "asset management" as the future approach for road maintenance in all state Departments of Transportation (state DOTs) (JLARC, 2002). Asset management calls for the utilization of engineering, management, and economics principles to help state DOTs with allocating limited resources to preserving, operating, and managing the nation's road infrastructure (Ozbek, 2007).

This focus, over the past two decades, has lead to the development of Pavement Management Systems (PMSs) (Gendreau and Soriano, 1998) that draw on systematic methods to best use available budgets and to design cost-effective maintenance policies. One of the key components of a PMS is planning methods to decide which maintenance operations should be performed when considering the current and predicted condition of the pavement (Gendreau and Soriano, 1998). The current research contributes to this component of a PMS by using the System Dynamics approach to find the optimal policy for allocating maintenance budget given a set of environmental and operational conditions.

Figure 3-1 specifies the major steps of this chapter. First (Section 3.3.1), we develop a simulation model of the highway deterioration process at a micro (road section) level using the principles embodied in the Mechanistic-Empirical models (Huang, 2004) that are among the best available physics-based models for predicting road deterioration in pavement management. Second (Section 3.3.2), we use empirical data (pavement condition and traffic for approximately 17 miles of Virginia's Interstate highway during the fiscal years 2002 to 2007) to estimate the unknown parameters of this model. Finally (Section 3.3.3), we couple the calibrated pavement deterioration model with the highway maintenance and renewal decision-making process to find the best policy for allocating maintenance budget to different maintenance operations. Reliable policy recommendations require optimization to be performed in robust and calibrated models. Thus, this paper couples system dynamics modeling with calibration and optimization and builds a concrete framework for designing maintenance policies across different settings.

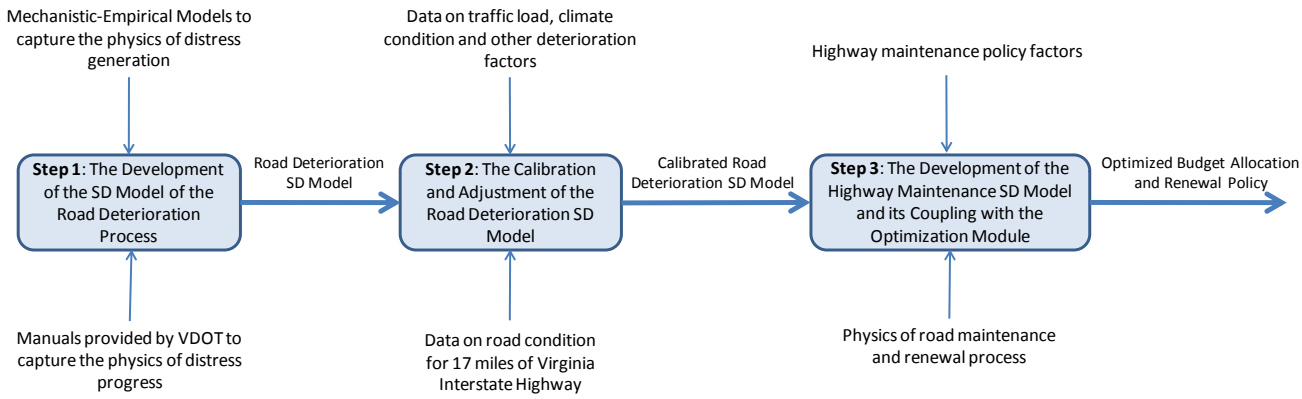


Figure 3-1: The Three Main Steps of this Chapter toward Optimizing Highway Maintenance Policies

The remainder of this paper is organized as follows. The approaches that exist in the literature for the planning of maintenance operations are described in Section 3.2. Section 3.3 describes the methods used for modeling the dynamics of road deterioration and maintenance as well as the calibration and optimization of the model. The results and insights are discussed in Section 0. Conclusions and future directions are provided in Section 3.5.

3.2 Approaches for Planning of Road Maintenance Operations

In the past two decades, several approaches such as, expert systems, analytical models, and system dynamics have been utilized to formulate highway maintenance planning and scheduling. With the advances in expert knowledge elicitation and analysis methods, "Expert Systems" have become popular for pavement management in both highway and airport networks (Ismail et al., 2009). One of the advantages of expert systems is their ability to involve expert knowledge and subjective human reasoning (Chang Albitres et al., 2005) that typically are complicated to incorporate in mathematical models. However, lack of agreement among experts in the field and inadequate approaches for representing domain knowledge remain as major shortcomings (Ismail et al., 2009).

The second approach for highway maintenance scheduling and budget planning is analytical modeling. These approaches formulate the decision-making problem as a model where the objective function (e.g., minimizing overall costs) and the constraints (e.g., minimum acceptable condition) are represented by mathematical expressions of the decision variables (Gendreau and Soriano, 1998). These mathematical modeling techniques are either deterministic (Lytton, 1985; Feighan et al., 1987; Fwa et al., 1988; Smadi, 1994; Fwa et al., 2000; Wang et al., 2003) or stochastic (Butt et al., 1994; Gao et al., 2007; Gao and Zhang, 2008). They often can

handle problems with major detail complexity, yet they also suffer from (i) extensive computational requirements; (ii) lack of prior knowledge of the data generating process that is required for stochastic modeling of the pavement deterioration process; and (iii) the gap between theory and practice that has led to the challenge of interpreting the complex analytical models by maintenance managers (Dekker, 1996; Gao and Zhang, 2008).

The third approach is the application of System Dynamics (SD). The road deterioration and maintenance is a dynamic phenomenon embedded in a nonlinear feedback system where the uncontrollable factors of the environment (e.g., climate condition) and operational conditions (e.g., traffic load) as well as the controllable factors (e.g., maintenance policy) affect the road condition. SD is an appropriate modeling approach for this context because it captures the dynamics of road conditions and accounts for feedback loops that determine the physical road deterioration processes. There are a few applications of SD to transportation systems (Ogunlana et al., 1998; Ogunlana et al., 2003; Ibbs and Liu, 2005; Lee and Pena-Mora, 2005; Lee et al., 2005; Thompson and Bank, 2010) and highway maintenance policy (Chasey, 1995; Chasey et al., 1997; de la Garza et al., 1998; Kim, 1998; Bjornsson et al., 2000; Chasey et al., 2002; de la Garza and Krueger, 2007). For example, the impact of deferred maintenance on the Interstate's level of availability (i.e., in terms of capacity and allowable traffic volume) and on the level of operation (i.e., the road physical condition) has been modeled by a number of researchers (Chasey, 1995; Chasey et al., 1997; de la Garza et al., 1998; Chasey et al., 2002). In addition, at the macro-level, the effects of the deteriorated roads on user benefits (i.e., vehicle operating cost and travel time) and non-user benefits (e.g., increase in business opportunities) have been considered.

All of the prior SD highway maintenance studies are at the network-level meaning that the unit of analysis is the overall highway network under the control of the road authority. In the real world, the whole highway network is divided into several road sections and maintenance treatments are defined for each one of these sections. Thus, the road section is the smallest unit for which the maintenance decisions are made. Therefore network-level models are sometimes too aggregate to provide practical decision support at the road-section level. In fact maintenance managers would like decision support both at the road-section and the overall network-levels.

In addition, in previous SD highway maintenance studies, the model parameters such as deterioration rates are defined either based on synthetic data or using the outputs of other

statistical analyses. Therefore the potential to estimate model parameters using original data has not been realized and thus ensuring model reliability has been a challenge. Furthermore, previous studies have relied on sensitivity analysis and trial and error for policy analysis and have not used an integrated optimization method to find the best maintenance policies. This paper advances the application of SD to highway maintenance by addressing these shortcomings.

3.3 Methods

At an aggregate level, the deterioration and maintenance dynamics can be summarized in two major feedback loops (Figure 3-2). The pavement condition is deteriorated by traffic load, climate condition, and other deterioration factors. The balancing loop B1 (Maintenance Fix) describes how maintenance operations performed by road authorities attempt to bring the road condition towards desired conditions by reducing the area under distress. On the other hand, the reinforcing loop R1 (Accelerated Deterioration) illustrates the effect of the budget shortfall on the delayed maintenance and further deterioration of pavement conditions. Note that although the physics of road deterioration and maintenance is complicated, the corresponding feedback structure is relatively simple. The real challenge lies in building a reliable model that quantifies the road deterioration process and corresponds to empirical findings. This is the challenge we tackle in the next sections.

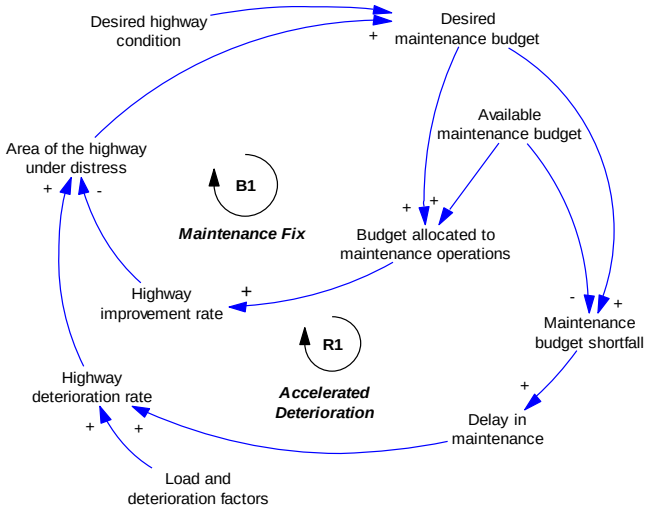


Figure 3-2: The Highway Deterioration and Maintenance Causal Loop Diagram

3.3.1 Modeling the Dynamics of Road Deterioration

Given that the goal of this study is to use the model for policy analysis through optimization, a modeling framework that is robust to extreme conditions and one that is understood by policy makers is required. To this end, pavement engineering literature was explored to represent the physics of the pavement deterioration (Huang, 2004). Within the System Dynamics (SD) framework (Forrester, 1971; Sterman, 2000), these physically based dynamics were studied in conjunction with macro-level maintenance operations. This combination allows for building a simulation model that is grounded in the physics of road operations that can be validated (Fallah-Fini and Triantis, 2009), and that considers the social and managerial factors. We seek a broad model boundary that is robust to extreme input levels and that allows for reliable policy analysis (Sterman, 2000).

The pavement deterioration module of the SD model captures distress generation and propagation on road sections. It is a critical component of the model, since it forms the main input to the maintenance decision-making process and can significantly affect the effectiveness of the recommended maintenance policies. This module relies on Mechanistic-Empirical (ME) models (Huang, 2004) to capture the road deterioration process. ME models use mechanical analysis to design the pavement structure. Pavement structure determines the "allowable number of load cycles". This parameter represents the number of load cycles a pavement can bear before a specific percentage of the pavement area experiences distress.

The damage ratio, which is the ratio of the "actual" to the "allowable" number of load cycles, shows the damage corresponding to a specific condition (e.g., traffic load). Damage to the pavement accumulates as the actual number of load cycles increases. In practice, it is common to consider that when the ratio reaches one, 20 percent of the lane area experiences fatigue cracking (Priest and Timm, 2006). In this study, it is assumed that the damage that gradually appears on a road section is proportional to the damage ratio.

The allowable number of load cycles is specified for each type of distress. For flexible pavements, such as those in highways, the main load related distresses include fatigue cracking and rutting. Fatigue cracking is a series of interconnected cracks caused by fatigue failure of the pavement surface due to repeated traffic loading and is considered as the major structural distress on the road surface (Priest and Timm, 2006). Thus, in this study we focus on fatigue cracking.

The actual number of load cycles is calculated by considering the availability of traffic and vehicle classification data for each road section. See (Fallah-Fini et al., 2009) for details about calculating the actual number of load cycles. The allowable number of load cycles is calculated using the data related to pavement structure at each road section. This data requires detailed knowledge of road construction and material, not typically available for older roads. Thus, in this SD model, the "allowable" number of load cycles is an unknown parameter that is estimated through calibration.

In the VDOT, fatigue cracking levels are specified by two parameters, namely, severity and density of the cracks (VDOT, 2002). Severity, which represents depth and openness of a crack, is defined at three levels: Not-Severe (NS), Severe (S), and Very-Severe (VS). Density of the cracks represents the percentage of the right lane area that is covered with cracks. The three density levels that are used in practice are Rare (i.e., less than 10% of the right lane area is filled with the cracks), Occasional (i.e., between 10% and 50%), and Frequent (i.e., more than 50%).

When a road section is evaluated, the density levels corresponding to all three fatigue cracking severity levels (i.e., NS, S, and VS) are defined. For example, a road section may contain 5% of NS cracks, 10% of S cracks, and 5% of VS cracks. The overall condition of a road section at any point in time is defined by: (1) Choosing the severity level with the maximum density (i.e., S in the example); (2) Adding up the density of the three severity levels to determine the overall density level (i.e., 20% in the example). Thus, the overall condition of the example road section is reported as "Severe & Occasional". In addition, if a road section is in a perfect condition, it is represented as "Not Severe & None", meaning that no distress has been observed in the road section. These described steps are referred to as the "road condition evaluation procedure", hereafter. This procedure is performed annually for all road sections across a highway network. Table 3-1 shows the possible combinations of severity and density levels in a three-by-three matrix.

In this study, the damage ratio is used to model the generation of fatigue cracking through the life cycle of a road section. It is assumed that a crack is in a Not-Severe condition when it is initially generated. Deferred maintenance operations due to lack of maintenance budget, sub-optimal maintenance policies, and repeated traffic loads cause the continuous transition of lane-miles under distress from Not-Severe to Severe and from Severe to Very-Severe conditions,

respectively. Thus, three stocks are defined for the three severity levels of fatigue cracking (See Figure 3-3).

Table 3-1: Possible Combinations of Severity and Density Levels

		Density Levels		
		Rare (R)	**Occasional (O)**	**Frequent (F)**
Severity Levels	**Not-Severe (NS)**	NS-R	NS-O	NS-F
	Severe (S)	S-R	S-O	S-F
	Very-Severe (VS)	VS-R	VS-O	VS-F

There is a delay between applying maintenance treatments on each road section and recurrence of distress on that road section. Thus the road sections that are maintained are reassigned to the "Not-Severe & None" condition and remain in that state for some time. The "Maintained Lane-miles" stock is defined to accumulate the maintained lane-miles that are depleted from the three stocks of distresses. The outflow of the "Maintained Lane-miles" stock releases the maintained lane-miles into the deterioration process.

The time constants representing the speed of transfer across different stocks depend largely on the structure of the road sections and environmental conditions. Thus, theses parameters are estimated in the calibration process. Figure 3-3 shows the simplified structure of the pavement deterioration module, its main variables, inputs and outputs. The four parameters that are computed by calibration process are highlighted in bold and italic fonts and the exogenous inputs coming from exogenous time series data are underlined and highlighted in bold.

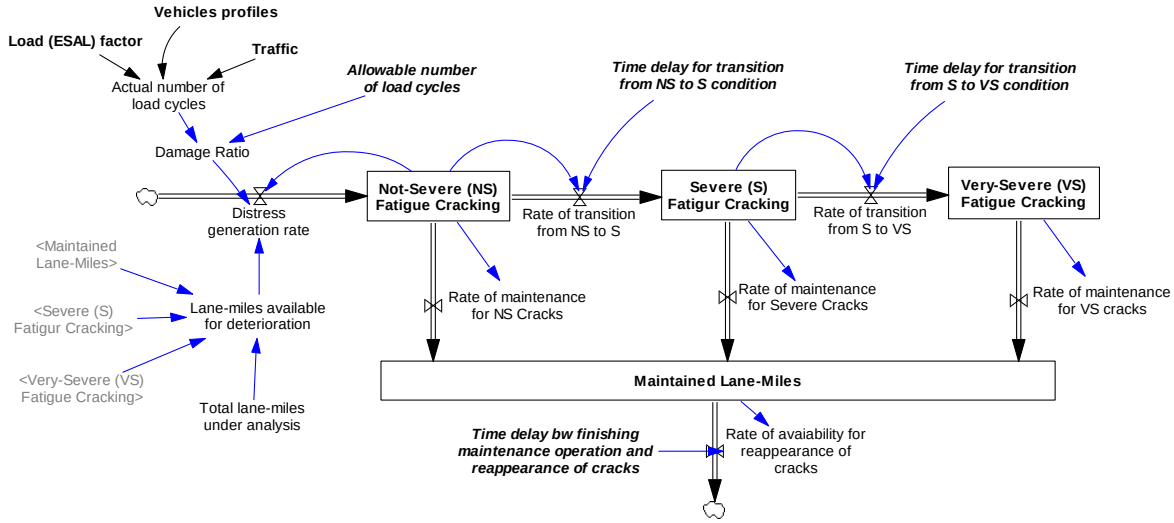


Figure 3-3: Simplified Structure of the Pavement Deterioration Module

3.3.2 Model Calibration

In this section, a model-based calibration method for parameter estimation (Oliva, 2003) is used to find appropriate values for unknown parameters. To analyze road maintenance operations on a project-level, the units under analysis are defined as road sections. By modeling multiple road sections, analysis of network-level metrics are also possible. The highway network, in this study, is represented by eight road sections that lie in a county in the state of Virginia.

VDOT uses the Load Related Distress (LDR) index to reflect the condition of the road with respect to distresses that are mainly caused by the traffic load, such as fatigue cracking and rutting. LDR values range between 0 and 100, where LDR of 100 represents a road section in perfect condition. To obtain the LDR for each road section, one subtracts from 100 the deduct value corresponding to each one of the nine possible road conditions presented in Table 3-1. Table 3-2 shows the standard deduct values for fatigue cracking developed by VDOT's Asset Management Division (Chowdhury, 2007). Based on Table 3-2, the LDR value of a road section that is in "Severe & Rare" condition is reported as 87=100-13.

Table 3-2: Standard Deduct Values for Fatigue Cracking Developed by VDOT

	Rare (R)	**Occasional (O)**	**Frequent (F)**
Not-Severe (NS)	10	20	30
Severe (S)	13	30	40
Very-Severe (VS)	16	38	52

Table 3-3 summarizes the road condition data for the eight road sections in the highway network under analysis after some data cleaning. These data points are used for the calibration process. The highlighted cells show the years in which maintenance operations were performed. As the data shows, when a maintenance operation is performed on a road section, the condition of the road goes back to Not-Severe & None state in the next year. Thus, when a maintenance operation is performed in the calibration, the three stocks of distress are fully depleted. Maintenance happens exogenously in the calibration runs, so as to be consistent with the data.

The objective of the calibration process is to determine the values of the four unknown parameters presented in Table 3-4 so that the simulated road sections follow the data provided in Table 3-3. The payoff function that is minimized in the calibration is the weighted sum of eight error terms for different road sections. These error terms are zero if the road section condition in the model equals what is observed in the data and one otherwise. Error terms are weighted

proportional to the inverse of the standard deviation of the error terms, which leads to maximum likelihood estimates for the estimated parameters (Greene, 2002). Table 3-4 shows the calibration results.

Table 3-3: Road Condition Data over the Fiscal Years 2002-2007 for Road Sections in the Network under Analysis

Road Section ID	Length (mile)	2002			2003			2004			2005			2006			2007		
		Severity	Density	LDR	Severity	Density	LDR	Severity	Density	LDR	Severity	Density	LDR	Severity	Density	LDR	Severity	Density	LDR
1	0.85	S	R	87	S	R	87	S	O	70	NS	None	100	NS	None	100	S	R	87
2	3.49	S	R	87	S	R	87	S	O	70	S	F	60	S	F	60	S	F	60
3	0.77	NS	None	100	NS	None	100	S	R	87	NS	R	90	S	R	87	S	R	87
4	2.12	S	O	70	NS	None	100	NS	None	100	NS	R	90	S	R	87	S	R	87
5	2.16	S	R	87	S	R	87	S	O	70	NS	None	100	NS	Non	100	S	R	87
6	0.75	S	O	70	S	O	70	NS	None	100	NS	None	100	VS	R	84	S	R	87
7	1.07	S	R	87	S	R	87	S	O	70	S	O	70	NS	None	100	S	R	87
8	5.65	S	R	87	NS	None	100	NS	None	100	NS	R	90	NS	R	90	NS	R	90

The calibration results are in a reasonable range considering the road condition data in Table 3-3. For example, Table 3-3 shows that most of the road sections remain in Not-Severe & None condition for two years before recurrence of distresses. As it can be seen, the estimated value for the corresponding parameter is 23.7 months. Figure 3-4, as an example, shows the comparison of data and the model estimates for LDR in the road section 4 and road section 6. See Appendix A for the corresponding graphs for the rest of the road sections. Figure 3-5 shows the behavior of the LDR variable at the network-level and its comparison with the network-level LDR variable obtained from the data. The network-level LDR was obtained as the weighted average of the LDR of the road sections, where the lengths of the road sections were used as the weight factor.

Table 3-4: Estimated Values for the Parameters of Interest Obtained from the Calibration Process

Parameter	Estimated Value
Allowable number of load cycles	1e+8 Load cycles
Time delay for transition from not-severe cracks to severe cracks	10.9 Months
Time delay for transition from severe cracks to very-severe cracks	20.1 Months
Time delay between finishing maintenance operations and reappearance of cracks	23.7 Months

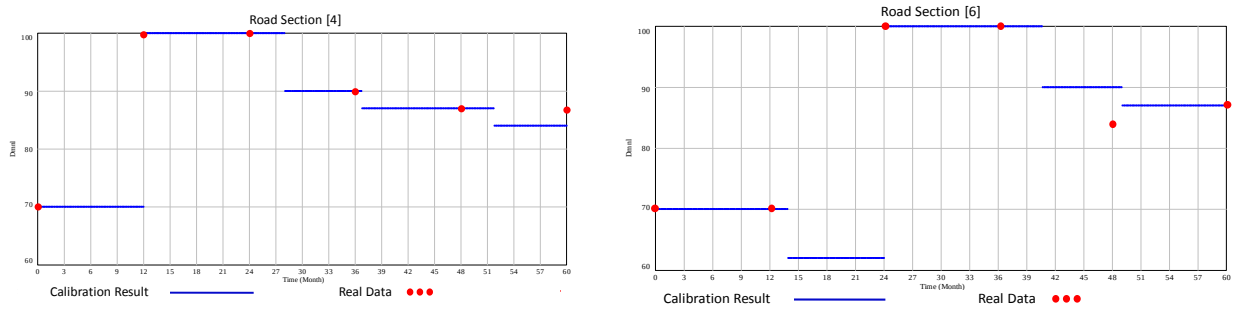


Figure 3-4: Comparing the LDR Index Obtained from the Real Data and the Model after Calibration for the Road Sections 4 and 6

Note that data points only record the road condition at the beginning of each year, whereas the model behavior is continuous. For example, in Figure 3-5 data (dots) at month 48 and month 60 show that network-level LDR has changed from 93 at year 2006 (month 48) to 88 at year 2007 (month 60), while the model (line) shows that the network-level road condition has continuously deteriorated through the year 2006 to the start of the year 2007. Also note that LDR takes different constant values over different periods of time because of the calculation method used by VDOT (See Table 3-2). As a result, LDR values change discretely even though their driving variables are continuous.

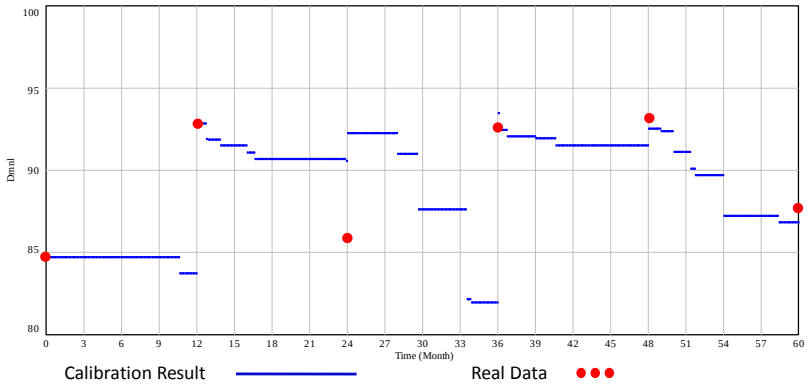


Figure 3-5: Network-level LDR Obtained from the Data as well as the Model after Calibration

3.3.3 Modeling and Optimization of the Maintenance Budget Allocation

In this section, the road deterioration module is coupled with a "Maintenance Budget Allocation" module to capture the dynamics of the budget allocation for various maintenance operations for each road section. The dynamics that are described in this section mostly relate to loop B1 (see Figure 3-2) where deteriorations generated in the road sections (as discussed and

modeled in section 3.3.1) are evaluated to define the required maintenance operations for each road section. Then the limited available budget is allocated to the required maintenance operations based on the policy factors that are discussed and optimized using the optimization module. Thus, the optimization module finds an optimum balance between performing maintenance operations (loop B1) to avoid accelerated deterioration (loop R1) and to save on the costs of maintenance. In what follows, the modeling and optimization of the budget allocation processes are discussed.

The required maintenance activities (preventive, corrective, and restorative maintenance) are defined based on the types of distresses on a road section, as well as the distress's severity and density levels. Table 3-5 shows the decision matrix that is used by VDOT (Chowdhury, 2007) to determine the maintenance activity suitable for a road section in presence of fatigue cracking. For example, the maintenance strategy for the "Severe & Occasional" condition is preventive maintenance. Preventive Maintenance (PM) refers to the treatments that are performed to reduce the rate of deterioration and preserve the existing pavement integrity. Corrective Maintenance (CM) refers to the treatments that maintain the characteristics and structural integrity of an existing pavement for continued serviceability. Restorative Maintenance (RM) refers to new surface layers that restore the pavement structure to a level similar to the original condition of the pavement (Smith and Nazarian, 1992; Kim, 1998). Rehabilitation and reconstruction activities are not considered in this study. Costs for different maintenance activities, in dollars per lane-mile of pavement, were estimated by investigating the literature (Mahoney et al., 2010) and consulting with practitioners. When an area within a road section needs to be maintained, the entire road section is maintained to restore the pavement to the "Not-Severe & None" condition.

Table 3-5: Decision Matrix for Fatigue Cracking (DN: Do Nothing, PM: Preventive Maintenance, CM: Corrective Maintenance, RM: Restorative Maintenance)

	Rare (R)	**Occasional (O)**	**Frequent (F)**
Not-Severe (NS)	DN	DN	PM
Severe (S)	DN	PM	CM
Very-Severe (VS)	CM	CM	RM

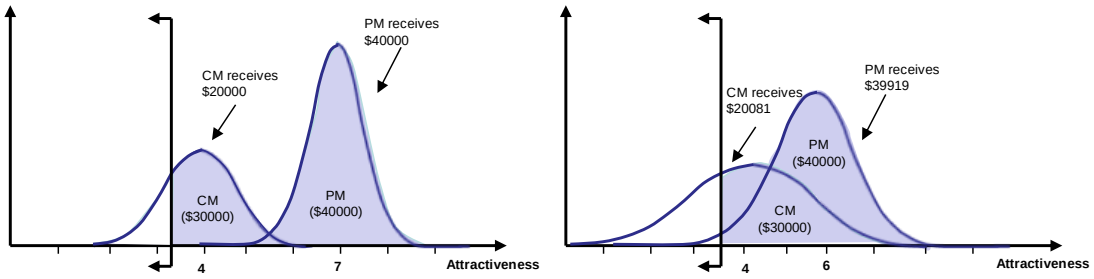
Highway maintenance budgets are limited, thus not all road sections will be maintained every year. What follows represents the real-world allocation of a limited network-level maintenance budget among the road sections. First, Table 3-1 and Table 3-2 are used to define the LDR of each road section. The road section's LDR is then compared with a "LDR threshold"

set by road authorities so that only those road sections with LDRs below the threshold are eligible to receive maintenance this year. For example, with the LDR threshold of 80, a road section that has LDR of 90 is not eligible for maintenance. Next, Table 3-5 is used to determine the appropriate maintenance operations (i.e., PM, CM, and RM) for road sections that are eligible for maintenance. As a result, the total required budgets for PM, CM, and RM in the highway network are determined. The next step is to allocate the limited network-level budget among competing maintenance operations, i.e., PM, CM, and RM, based on their priorities.

The "*Allocate Available*" function of VENSIM was used to model the allocation process. This function takes as its argument the network-level available budget, the total budget required by the PM, CM, and RM operations, and the priority profiles corresponding to each of these maintenance operations. The priority *profile* is represented by a normal distribution curve with a given mean (priority) and standard deviation (width), where the area under the curve represents the demand (required budget to perform the specific type of maintenance). The priority and width defined for the three different maintenance operations determine the budget allocation policy.

The following illustration shows how for given priority and width levels, the "*Allocate Available*" function allocates the budget among multiple demands. Consider a case in which a limited maintenance budget of $60,000 needs to be allocated between PM and CM, where PM requires $40,000 and CM requires $30,000. Two scenarios are depicted in Figure 3-6: a) PM and CM have priorities of 7 and 4 and widths of 0.5 and 1, respectively; b) priorities and widths are 6 and 1 for PM and 4 and 2 for CM. Once the normal curves with these profiles are portrayed graphically, the allocation function works by distributing the resource (budget in this case) to the areas under the curves from right to left, until all the resource (budget) is allocated or all demand curves are fully satisfied. In these graphs the resulting budget allocations are shown by dark colors under the curves for PM and CM, and the exact values are reported in Figure 3-6.

In the cases similar to Figure 3-6(a) when the allocated budget for a maintenance operation (i.e., allocated budget for CM in this example) is not enough for fixing all road sections in need of that operation, the road sections that are in worse conditions (i.e., have lower LDRs) have a higher priority for receiving the limited available budget.



(a) PM priority=7, PM width=0.5, CM priority=4, CM width=1 (b) PM priority=6, PM width=1, CM priority=4, CM width=2

Figure 3-6: Allocation of the Limited Available Budget between PM and CM Based on their Priorities

There are several decision parameters in the procedure described above that can potentially affect the outcome of budget allocation and the final condition of the road sections. The base case values of these parameters are used to represent the current practice and, an optimization module is used to find improvements to the base case policy that can help road authorities make the best use of limited available resources.

The first set of parameters is related to the priority profiles of PM, CM, and RM. As it is shown in Figure 3-6, assigning different values to the mean (priority) and standard deviation (width) of the demand curves can lead to different scenarios for allocating the limited available budget among the maintenance operations and the corresponding road sections. Based on the discussion with maintenance managers, CM usually receives a higher priority. Finding the optimal set of priorities for PM, CM, and RM and their sensitivity with respect to the available network-level budget are analyzed in the Section 0.

The second set of parameters is related to the density thresholds (used in Table 3-1) for defining Rare vs. Occasional (i.e., when 10% of the road section is covered with distress in practice) and Occasional vs. Frequent (i.e., when 50% of the road section is covered with distress in practice). By changing these thresholds, the conditions under which a road section goes from one density category to the other, and thus the required maintenance activities, are changed. However, one cannot change these thresholds arbitrarily since very high density threshold values may make it infeasible to perform the necessary maintenance operations. Therefore, as will be discussed in the next section, some feasible ranges for these parameters are set based on expert input and the optimum values are determined through optimization.

The third set of parameters that can be evaluated through optimization is the type of maintenance operation recommended for each cell in Table 3-5. Again, not all alternative treatments are feasible for each cell. However, when the recommended treatment is "Do

Nothing", one can always apply PM instead and test if this can improve the overall performance of the road network.

Finally, the fourth parameter is the LDR threshold. In practice, this parameter is defined based on the availability of the maintenance budget. Road authorities that are facing a shortage in their budget choose lower LDR thresholds so as to be able to assign the limited budget to those road sections that are in critical need for maintenance, i.e., have very low LDRs. After setting a feasible range for this parameter based on expert input, optimization is used to find its best value.

In general, an "optimized" policy is compared with a "base case" scenario. The decision rules that are currently used by road authorities are utilized to construct the base case scenario (See Table 3-6). It is postulated that by using the calibrated model from Section 3.3.2 and an optimization method that varies the parameters discussed in this section, an optimal use of the maintenance budget is possible. Specifically, the optimization module of VENSIM is used in this study to find the optimum values of the parameters such that a desired utility function is minimized. The utility function in this paper is defined as the weighted sum of the area of the highway network which is under distress with different severity levels[5]. The optimal maintenance budget allocations are defined such that sum of the utility function over the analysis horizon is minimized.

Table 3-6: The Parameter Values Representing the Base Case Scenario

Parameters Considered in Optimization	**Values in the Base Case Scenario**
Priorities of the PM, CM, and RM	First priority: CM, Second priority: PM, Third priority: RM
The density threshold used in Table 3-1	10% and 50% for Rare and Occasional, respectively
Variations of Table 3-5	Table 3-5 is used as it is in the base case
The LDR threshold	85

Another complicating factor concerns the initial conditions of the road sections. The historical initial conditions that were observed for this network may lead the optimization algorithm to a budget allocation policy that is not generally reasonable. To account for this possibility, the model is subscripted for 50 replications of the same road network, which differ only in their initial values of stocks. To be realistic, the initial values for the stocks are randomly drawn from the distribution of road conditions over the calibrated base case simulations. For the optimization, the value of the utility function is summed over all 50 road networks

[5] The weights associated with the Very-Severe, Severe, and Not-Severe distresses are assumed to have the values of 3, 2, and 1, respectively.

configurations, rather than focusing on a single network configuration with only one set of initial conditions. Therefore, the resulting solution is assumed to be robust to the initial road conditions of the three stocks. The optimization results and corresponding discussion are presented next.

3.4 Optimization Results and Discussion

State DOTs are interested in identifying the best maintenance resource allocation alternatives given different budgetary scenarios. Therefore, in order to find these alternatives, four budget scenarios are considered, namely, "Low", "Most likely", "High", and "Extreme". The values associated with these scenarios (see Table 3-7) were found by investigating the literature (Mahoney et al., 2010) and consulting with practitioners. For each budget scenario, the optimization results are presented in Table 3-7.

A few interesting observations emerge. First, RM is rarely needed for any of these scenarios since the optimally allocated budget is sufficient to avoid the deterioration of the road sections to "Very-Severe & Frequent" where road sections would need RM. As a result, the RM profiles (with their respective priorities and widths) are not reported in Table 3-7.

Second, overall, PM gets a higher priority when compared with CM in the optimized solutions. This result is robust when considering the four different budget levels. In essence, by applying PM, the road authorities can effectively decrease the need for future CM while spending less overall. However, the overlap in the PM and CM profiles increases as the total budget goes down. This shows the need for sharing the budget between PM and CM, rather than satisfying PM first and then allocating leftover resources to CM. This overlap is justified when one considers that with low maintenance budgets not all PM work can be completed, which logically leads to an increase in the required CM work in subsequent years.

Table 3-7: Priority Profiles Corresponding to the PM, CM, and RM Obtained from the Optimization

	PM priority	PM std	CM priority	CM std	Rare UB Threshold	Occasional UB Threshold	LDR Threshold	Optimal Utility Function	Base case Utility Function
Base case scenario	7	1	12	1	0.10	0.5	85		
Low Budget (200000$/year)	7	1	1.92	1.16	0.107	0.5	86.31	7.38e+8	7.39e+8
Most Likely Budget (340000$/year)	7	1	1.44	1.02	0.103	0.5	86.31	7.00e+8	7.05e+8
High Budget (400000$/year)	7	1	1.41	1.20	0.100	0.5	86.31	6.85e+8	6.91e+8
Extreme Budget (600000$/year)	7	1	0.57	1.00	0.093	0.5	86.31	6.38e+8	6.46e+8

It is interesting to note that the current resource allocation practice favors CM over PM, which is not optimum according to this analysis. In practice, however, the preference of CM over PM likely reflects a short-term perspective vis-à-vis a long-term one. That is, road maintenance engineers struggle with the difficult notion of allowing current road sections requiring CM to further deteriorate, even if they know that conducting PM first is better over the long run.

The impact of the distress density thresholds for defining Rare vs. Occasional (Rare Upper Bound (UB)) and Occasional vs. Frequent (Occasional Upper Bound (UB)) is more subtle, even though these thresholds have a noticeable impact on the utility function because they influence how different road sections are categorized. Whether a road section is seen as highly deteriorated and in need of maintenance or not depends on these thresholds. For example, consider road sections A and B in which "Very Severe" cracks have the highest density with 7% and 35%, respectively. If the Rare UB is set to 5%, both road sections A and B lie in the "Very Severe & Occasional" cell of Table 3-2 and receive the same deduct value of 38. In this case, both road sections have the same priority in terms of receiving limited funds for performing CM. This leads to a proportional allocation of the limited available budget among road sections A and B. On the other hand, if Rare UB is set to 10%, then the deduct values corresponding to road sections A and B will be equal to 16 and 38, respectively, meaning that road section B will have higher priority for receiving the limited maintenance funds. Furthermore, sensitivity analysis showed that the utility function is more sensitive to the Rare UB in comparison with the Occasional UB. As discussed earlier in this section, most road sections in all scenarios are repaired before they fall into a condition where they have major problems. Therefore, the Occasional UB remains relatively unimportant. Interestingly, the optimal Occasional UB threshold (see Table 3-7) found in this analysis closely corresponds to the threshold used by VDOT in practice, suggesting the wisdom of current practice.

Another insight is that LDR levels may be too coarse a metric to prioritize among different road sections in need of the same maintenance type (e.g., PM or CM). In essence, road sections with very different distress density levels (e.g., 11% and 45%) lie in the same cell of Table 3-2 and thus have the same LDR value. In order to give a higher priority to the road sections that are in a worse condition, the exact distress density levels of the road sections could be a more useful metric for budget allocation than LDR. The alternative metric will ensure that road sections with

a higher distress density receive a higher priority for their required maintenance given the limited budget.

The optimization analysis shows little benefit in doing additional PM where currently "Do-Nothing" is advised. Two mechanisms explain this observation. First, this policy increases maintenance demand, but does not add much value because it requires maintenance for road sections that are otherwise acceptable (i.e., road sections are in the "Not-Severe & Rare" and "Severe & Rare" conditions that have an LDR greater than the LDR threshold). This leads to a potential waste of resources even if enough resources are available to complete this maintenance operation. Second, the condition "Not-Severe & Occasional" where the impact of this policy change is mostly observed does not happen very often. The transition rate from the Not-Severe cracks to Severe cracks obtained through calibration (see Table 3-4) in most instances prevents the Not-Severe cracks to accumulate to the point where they become "Not-Severe & Occasional". Instead, in most cases, the Not-Severe cracks become Severe cracks.

The analysis of the utility function with respect to the LDR threshold also shows little sensitivity to this parameter. In essence, it was found that this threshold is somewhat redundant. The maintenance operation for the cells that lead to LDRs greater than 85 (base case threshold is 85) is Do-Nothing (see Table 3-2 and Table 3-5). Alternative priorities for road sections could be assigned with more precision when using the priority profiles and the Rare UB threshold. The LDR threshold therefore remains of limited use and its optimum value is relatively close to the current decision rules.

Finally, in comparing the base case with the optimized utility function results, it can be observed that current VDOT practices, while not optimum, are not terribly inferior to the suggested practices that were found through optimization. The major departure, i.e., the importance of PM vs. CM, only modestly impacts the utility function values. Overall, the base case rules of thumb that have evolved over the years in practice are fairly effective. This may not come as a surprise once we consider the length of time over which these rules have evolved and have been fine-tuned to the local conditions.

3.5 Conclusions

The contributions of this paper to the body of knowledge in system dynamics are two-fold. First, this paper develops one of the first system dynamics simulation models of the physics of road deterioration and maintenance. As such, it introduces new concepts and knowledge from the

road deterioration into the SD literature. Second, this paper couples the conceptualization steps of system dynamics method with calibration and optimization to construct a concrete framework for designing effective maintenance policies. This framework increases the applicability of the system dynamics approach for operational and tactical decision support in the field of maintenance. In fact, this combination of modeling, calibration, and policy optimization leverages the unique strengths of system dynamics in conceptualizing and modeling managerially relevant problems while at the same time effectively using available numerical data. This combination has been useful and attractive for the clients of this research project who are not knowledgeable in system dynamics. We therefore hope that our analysis provides a blueprint for successful introduction of system dynamics methodology in many alternative problem domains.

The study also makes practical contributions relevant to highway maintenance. The policies that were found through optimization pointed to alternative priorities where preventive maintenance is preferred over corrective maintenance. The extent of this contribution is however limited by the small improvements that are feasible. In essence, the results suggest that if one follows the current decision making strategy, there is only limited room for improvement.

On the other hand, there might be alternative decision making structures that can provide improvement opportunities. For example the current "road condition evaluation procedure" described in Section 3.3.1 translates a density function distribution of road section severity levels into a single point in a 3*3 matrix (see Table 3-1). This simplification is valuable given its cognitive simplicity and drives much of the decision making and analysis conducted in practice. On the other hand, with increasingly computerized analysis and decision-making tools, there is room for using all the information available in the severity distribution of road to drive the optimum maintenance policy. Future research can further explore this area of work.

Additionally, this model can be augmented by incorporating a user-benefit module that could include vehicle operating costs, travel time costs, and safety to change the specification of the optimized utility function. This would provide the opportunity to analyze the effects of different maintenance alternatives on societal costs and benefits. More work to establish a comprehensive utility function is left for the future extensions of the model.

The results presented in this paper are also limited by the number of road sections analyzed and the number of years for which the data were available. Calibrating the model using a more

comprehensive dataset of road sections, road conditions and operational environments can add to the validity of the model and the usefulness of the results. Moreover, to be able to compare the optimization results with the real dataset, the optimization horizon was limited to the number of years for which the data were available (i.e., five years). Considering that the developed model can potentially be run over a longer horizon, using a data set that contains longer periods of time can further reinforce the "long-term perspective" in road maintenance planning. Finally, capturing the distress generation and propagation process for other types of distresses (e.g., rutting) is an important future extension. Despite these limitations, in the long-term, this line of research could contribute to a more efficient use of societal resources, greater level of maintenance services, and a highway and roadway system that is safer and more reliable.

References

ASCE. 2009a. American Society of Civil Engineers: Reportcard for America's Infrastructure, http://www.infrastructurereportcard.org/fact-sheet/roads. Retrieved October 2009,

ASCE. 2009b. American Society of Civil Engineers: Facts About Roads, www.asce/reportcard. Retrieved October 2009,

Bjornsson HC, de la Garza JM, Nasir MJ. 2000. A decision support system for road maintenance budget allocation. In Proceedings of the 8th International Conference on Computing in Civil and Building Engineering. Palo Alto, CA.

Butt AA, Shahin MY, Carpenter SH, Carnahan JV. 1994. Application of Markov process to pavement management systems at network level. In Proceedings of the 3rd International Conference on Managing Pavements. San Antonio, Texas.

Chang Albitres C, Krugler P, Smith R. 2005. A knowledge approach oriented to improved strategic decisions in pavement management practices. First Annual Inter-university Symposium of Infrastructure Management. Waterloo, Canada.

Chasey AD. 1995. A framework for determining the impact of deferred maintenance and/or obsolescence of a highway system. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. Blacksburg.

Chasey AD, de la Garza JM, Drew DR. 1997. Comprehensive level of service: Needed approach for civil infrastructure systems. Journal of Infrastructure Systems **3** (4): 143-153.

Chasey AD, de la Garza JM, Drew DR. 2002. Using simulation to understand the impact of deferred maintenance. Computer-Aided Civil and Infrastructure Engineering **17**: 269–279.

Chowdhury T. 2007. Supporting document for pavement models and decision matrices development process used in the needs-based budget. Richmond, VA, Virginia Department of Transportation Asset Management Division.

de la Garza JM, Drew DR, Chasey AD. 1998. Simulating highway infrastructure management policies. Journal of Management in Engineering **14** (5): 64-72.

de la Garza JM, Krueger DA. 2007. Simulation of highway renewal asset management strategies. In Proceedings of the International Workshop on Computing in Civil Engineering, American Society of Civil Engineers. Carnegie Mellon University, Pittsburgh.

Dekker R. 1996. Applications of maintenance optimization models: a review and analysis. Reliability Engineering and System Safety **51**: 229-240.

Fallah-Fini S, Triantis K. 2009. Evaluating the productive efficiency of highway maintenance operations: environmental and dynamic considerations. The XI European Workshop on Efficiency and Productivity Analysis. Pisa, Italy.

Fallah-Fini S, Triantis K, de la Garza JM. 2009. Performance measurement of highway maintenance operation using data envelopment analysis: Environmental considerations. In Proceedings of the IIE Annual Conference. Miami, FL.

Feighan KJ, Shahin MY, Sinha KC. 1987. A dynamic programming approach to optimization for pavement management systems. In Proceedings of the Second North American Conference on Managing Pavement. Toronto, Ontario.

Forrester JW. 1971. World Dynamics. in: (Ed.), Wright-Allen Press: Cambridge, Mass.

Fwa TF, Chan WT, Hoque KZ. 2000. Multiobjective optimization for pavement maintenance programming. Journal of Transportation Engineering **126** (5): 367–374.

Fwa TF, Sinha KC, Riverson JDN. 1988. Highway routine maintenance programming at network level. Journal of Transportation Engineering **114** (5): 539–554.

Gao L, Tighe SL, Zhang Z. 2007. Using markov process and method of moments for optimizing management strategies of pavement infrastructure. 86th Annual Meeting of the Transportation Research Board. Washington, D.C.

Gao L, Zhang Z. 2008. Robust optimization for managing pavement maintenance and rehabilitation. Transportation Research Record: Journal of the Transportation Research Board **2084**: 55–61.

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

Greene WH. 2002. Econometric Analysis. in: (Ed.), Fifth edition, Prentice Hall: Upper Saddle River, NJ.

Huang YH. 2004. Pavement Analysis and Design. in: (Ed.), Pearson/Prentice Hall: Upper Saddle River, NJ.

Ibbs W, Liu M. 2005. System dynamic modeling of delay and disruption claims. AACE Construction Engineering **47** (6): 12-15.

Ismail N, Ismail A, Atiq R. 2009. An overview of expert systems in pavement management. European Journal of Scientific Research **30** (1): 99-111.

JLARC. 2002. Adequacy and management of VDOT's highway maintenance program. Richmond, VA, http://jlarc.state.va.us/reports/rpt273.pdf.

Kim K. 1998. A transportation planning model for state highway management: A decision support system methodology to achieve sustainable development. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. Blacksburg.

Lee S, Pena-Mora F. 2005. System dynamics approach for error and change management in concurrent design and construction. In Proceedings of the Winter Simulation Conference. Orlando, FL.

Lee S, Pena-Mora F, Park M. 2005. Quality and change management model for large scale concurrent design and construction projects. Journal of Construction Engineering and Management **131** (8): 890–902.

Lytton RL. 1985. From ranking to true optimization. In Proceedings of the North American Pavement Management Conference. Toronto, Ontario.

Mahoney JP, Uhlemeyer J, Morin P, Luhr D, Willoughby D, Mouench ST, banker T. 2010. Pavement preservation funding and performance in Washington State. Transportation Research Board Annual Meeting. Washington, D.C.

Ogunlana S, Li H, Sukhera F. 2003. System dynamics approach to exploring performance enhancement in a construction organization. Journal of Construction Engineering and Management **129** (5): 528–536.

Ogunlana SO, Lim J, Saeed K. 1998. DESMAN: A dynamic model for managing civil engineering design projects. Computers & Structures **67** (5): 401-419.

Oliva R. 2003. Model calibration as a testing strategy for system dynamics models. European Journal of Operational Research **151**: 552-568.

Ozbek EM. 2007. Development of a comprehensive framework for the efficiency measurement of road maintenance strategies using data envelopment analysis. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. Blacksburg.

Priest AL, Timm DH. 2006. Methodology and calibration of fatigue transfer functions for mechanistic-empirical flexible pavement design. National Center for Asphalt Technology. Auburn University, AL.

Smadi OG. 1994. Network pavement management system using dynamic programming: Application to Iowa interstate network. In Proceedings of the 3rd International Conference on Managing Pavements. San Antonio, Texas.

Smith RE, Nazarian S. 1992. Defining pavement maintenance and distress precursors for pavement maintenance measurement, maintenance of pavements, lane markings, and road sides. Transportation Research Record: Journal of the Transportation Research Board **1334**: 16-18.

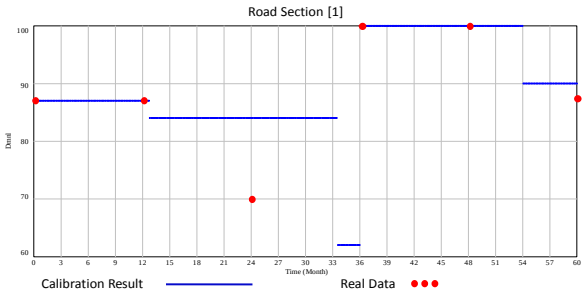
Sterman J. 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World. in: (Ed.), Irwin/Mc-Graw Hill: Homewood, IL.

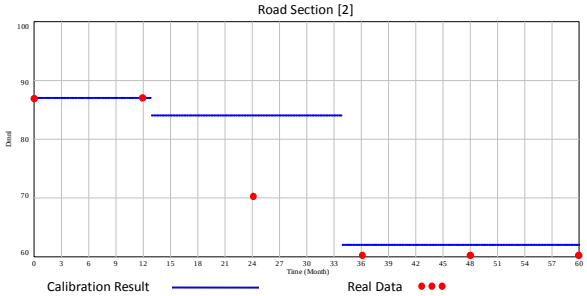
Thompson BP, Bank LC. 2010. Use of system dynamics as a decision-making tool in building design and operation. Building and Environment **45** (4): 1006-1015.

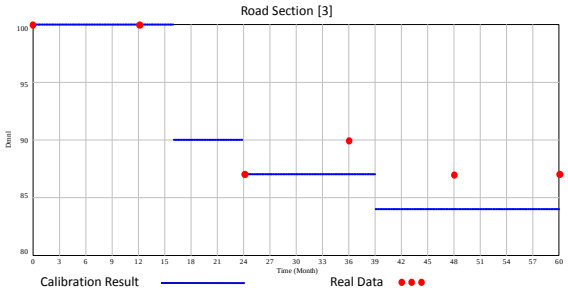
VDOT. 2002. 2002 Windshield Survey Program User Manual. Virginia, Pavement Management Program & Virginia Department of Transportation Asset Management.

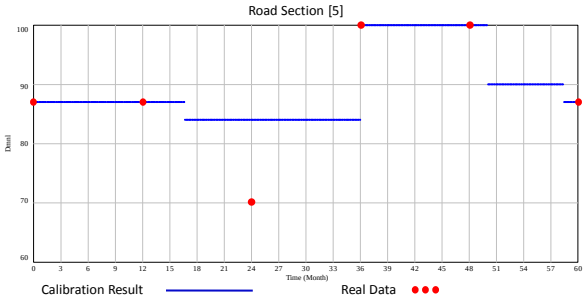
Wang F, Zhang Z, Machemehl RB. 2003. Decision-making problem for managing pavement maintenance and rehabilitation projects. Transportation Research Record: Journal of the Transportation Research Board **1853**: 21–28.

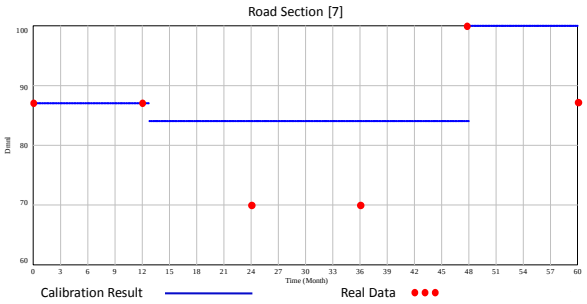
Appendix A: Comparing the LDR Index Obtained from the Real Data as well as the Model after the Calibration for the Rest of the Road Sections under Analysis

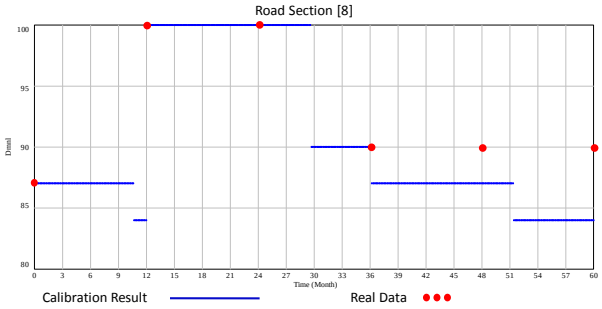












Chapter 4 Dynamic Efficiency Performance: State-of-the-Art

Abstract

Most of the studies in the efficiency measurement literature have focused on static frameworks for measuring efficiency performance where (i) there is no time interdependence between the input utilization and output realization of a production unit in consecutive periods; (ii) inputs or outputs can be adjusted to their optimal levels instantaneously without imposing any adjustment cost/time; and (iii) the focus is on the short run or steady state production in which quasi-fixed inputs do not change. The restrictive assumptions of static framework can lead to unrealistic efficiency estimates justifying the rising attention to dynamic performance measurement frameworks where the inter-temporal nature of the decision making process in a production unit is explicitly captured. In order to systematically review the studies that address relevant aspects of dynamic efficiency performance, this paper classifies dynamic performance measurement frameworks according to five issues. These issues account for major sources of the inter-temporal dependence between input and output levels over different time periods and include the following: (i) material and information delays; (ii) inventories; (iii) capital or generally quasi-fixed factors and the related topic of embodied technological change; (iv) adjustment costs; and (v) incremental improvement and learning models (disembodied technological change). We begin by discussing the comparative static frameworks that have been developed in the literature for quantifying the change in different performance measures over time when no inter-temporal dependence between input and output levels across different time periods is assumed. Thus, we point out the studies that potentially belong to the literature of dynamic performance, while no fundamental dynamics have been captured in these studies. Subsequently, the key studies in the literature that capture the inter-temporal relation among different periods with respect to the identified five issues are overviewed. As part of this overview we identify the strengths, shortcomings, similarities and differences of these key studies and focus on how they relate to each of the previously identified five issues. Based on this discussion we conclude with challenges and potential future research areas.

Key Words and Phrases: Dynamic efficiency; dynamic performance; inter-temporal dependence.

4.1 Introduction

Broadly speaking, efficiency refers to the performance of a production unit when utilizing a set of inputs to produce a set of outputs (Forsund and Hjalmarsson, 1974). Efficiency is a relative term since the performance of a production system is compared with a benchmark (frontier). Most of the studies in the efficiency measurement literature have focused on static frameworks. The static efficiency measurement frameworks assume that the inputs in a specific period are fully used for producing outputs in the same period. Thus, there is no time interdependence between the input utilization/output realization for a production unit[6] in consecutive time periods (Silva and Stefanou, 2007). In addition, it is assumed that all time periods are equivalent, thus the data collected at period t can be compared with the data collected at period t+1 (Vaneman and Triantis, 2007).

One of the main drawbacks of static representations is that they ignore the effects of input consumption as well as the managerial/engineering decisions in one period on output levels over several consecutive periods. Another important weakness of static representations lie in their assumption that firms are able to adjust instantaneously, thus these static models are unable to explain how some inputs or outputs are gradually adjusted to their optimal level (Silva and Stefanou, 2003). Moreover, these frameworks usually focus on the short run or steady state production in which capital or other quasi-fixed inputs are fixed (Sengupta, 1994b). But in the medium and long run, quasi-fixed factors can change through capacity expansion. This introduces the inter-temporal aspect of input utilization into the production framework. To address these shortcomings, researchers have focused on developing dynamic frameworks for measuring efficiency.

In a dynamic production framework there is a time interdependence among different periods (Färe and Grosskopf, 1996), meaning that input consumption or production decisions in one period not only depend on input consumption or production decisions in previous periods, but also affect the outputs in future periods. In other words, there can be a time lag between input consumption and output production. Moreover, in a dynamic production framework, firms usually need some time periods to adjust the levels of their inputs or outputs to target levels (Sengupta, 1996). Reviewing the literature reveals that this time interdependence among different periods can be attributed to one or a combination of issues associated with the dynamic

[6] The terms production unit and firm are used interchangeably in this paper.

aspects of production, namely, (i) material and information delays; (ii) inventory (inventories of exogenous inputs, inventories of intermediate and final products, etc.); (iii) capital or generally quasi-fixed factors (embodied technological change, vintage specific capital); (iv) adjustment costs; and (v) incremental improvement and learning models (disembodied technological change). The following discussion presents why each one of these five issues affects the inter-temporal nature of decision making for a production unit and why they contribute to the dynamic nature of production processes.

Delays (Lagged output): In production, there are significant time delays between input utilization and output realization. In these cases the efficient use of some inputs is relative to the outputs produced several periods later. The lag between the efforts of the sales team and an increase in the actual recorded sale is an example of presence of lagged output in a production process (Emrouznejad and Thanassoulis, 2005). Thus a standard cross-sectional model would mis-specify a relationship between inputs and outputs in the same period when the relationship is actually between inputs and outputs several periods apart. In addition to the material delays one typically encounters information delays that affect how inputs are transformed into outputs. For example, new best practice innovations are not communicated instantaneously but are diffused within the production unit.

Inventories: Inventories can cause dynamics in production because production from earlier periods can be held and used to meet demand in future periods. The idea is that producing in advance of demand can potentially help with leveling resource consumption and fulfilling demand that is occasionally greater than production capacity (Hackman, 1990). This can lead to inventories of exogenous inputs, intermediate products, final outputs, and other materials over time. When one allows for inventories, the periods in which exogenous/intermediate inputs or outputs appear are not potentially the same as the periods in which they are used. This leads to inter-temporal resource/inventory transfer in a production unit and contributes directly to the dynamic nature of production.

Capital: An important characteristic of capital is that the level of this variable at any point in time can be defined as a function of this same variable at one or more previous time periods. Moreover, the production capacity that is created by capital expansion is utilized over future time periods (Sengupta, 1994b). Thus, capital investment decisions depend on the previous decision concerning capital expansion, which also affect the outputs over periods beyond the one in which

the investment has taken place. Given these conditions, efficiency measurement studies ought to explicitly consider the inter-temporal nature of capital inputs and their effects on future streams of production outputs. As it was discussed earlier, capital (quasi-fixed inputs) changes over the long run. Thus, studies that focus on productivity growth due to capital expansion require a long term perspective.

Furthermore, a capital asset can be distinguished by its vintage, i.e., the year in which the capital good was built/installed (Frenger, 1992). In such a setting, adoption of new vintage technologies and scrapping of old vintages represents an explicit source of technical change (Färe and Grosskopf, 1996). Adopting a new vintage technology is referred to as "embodied" technical change (Johansen, 1959), where the state of the technology is reflected by the time at which the new capital goods were installed.

Note that once a new vintage technology is introduced (i.e., a new piece of capital good is built and is in operation), then the substitution possibilities between capital and variable inputs are limited. In other words, there are different substitution possibilities (between capital and variable inputs) before and after investments in new production technologies (embodied in capital equipment) (Forsund and Hjalmarsson, 1974). This has led to development of ex ante and ex post concepts of production initially introduced by Johansen (1959). Assuming variable inputs and capital are the factors of production, ex ante substitution possibilities between capital and variable inputs exists when increments in output can be obtained by increments in capital and variable inputs. Once a piece of capital good is built and is in operation, it imposes restrictions on possible combinations of the production factors (capital and labor inputs) through its life time. Thus, the vintage model implies that substitution possibilities between capital and variable inputs exist ex ante but are limited ex post. Due to these characteristics, an ex ante decision regarding the introduction of a new vintage can be defined as an inter-temporal optimization problem where an optimal decision requires information on future expected prices, demand, and other factors (Frenger, 1992).

Adjustment cost: Typically, a production unit cannot instantly change the levels of production variables (such as labor, capital, etc.) to their optimal values without incurring some costs of adjustment. Adjustment costs are usually in the form of foregone outputs or resources for a production unit due to the investment in new capital. For example, installation of a new machine usually requires that a portion of the work force to stop working on the production line for some

time. There is also a need to be trained to work with the new machine. Thus, installing new capital goods is a trade-off between current production (giving up some of the output) and current expansion and future production (Silva and Stefanou, 2007). Adjustment costs can also be imposed in the form of deviation of inputs or outputs from target levels during a planning horizon (Sengupta, 1996). Both approaches for capturing adjustment costs require introducing an inter-temporal planning horizon into the decision making process.

Incremental improvement and learning models: Although the vintage of a production technology may not change, the methods of production using the current capital may improve. In fact, firms tend to learn and adjust their production processes once they obtain the required information about the sources of technical inefficiency (Sengupta, 1994b). These types of efficiency changes along with the changes in productivity through learning by doing, coming up with better managerial practices, etc. are referred to as "disembodied" technical change. The efficiency frontier requires the consideration of disembodied technical changes as well.

The objective of this paper is to provide a review of the models/approaches in the literature that capture the inter-temporal relation among different periods by focusing on dynamic aspects of production process (i.e., production delays, inventories, capital, adjustment costs, and learning effects). This paper provides a taxonomy of existing studies, their strengths, shortcomings, similarities and differences. It should be noted that even though the coverage of studies in this paper is intended to be as comprehensive as possible, there are studies that have not been included due to space limitations. We apologize to the authors of these studies.

In Section 4.2 we provide a brief review on the static frameworks that have been developed in the literature for quantifying the change in different performance measures over time when there is no inter-temporal dependence between input and output levels across different time periods. The purpose of Section 4.2 is to point out the studies that potentially belong to the literature of dynamic performance, while no fundamental dynamics are captured in these studies in the sense that datasets do not show or support any inter-temporal relation between inputs, outputs, or technologies across different periods. Section 4.3 then provides the main focus of the paper by presenting the studies that capture the inter-temporal relation among different periods by focusing on dynamic aspects of production. Section 4.4 concludes and discusses challenges and potential future research areas.

4.2 Static Frameworks that Consider Time but not Inter-temporal Relations

This section provides a brief review on models that look at the snapshots of the system under analysis over time using a panel data or time series and try to obtain measures of "productivity change" (by evaluating how productivity indexes change over time), "technical change" (by evaluating how a frontier changes over time), or "technical efficiency change" (by evaluating how the position of a production unit with respect to the frontier changes over time). Thus, this section focuses on quantifying the change for different performance measures that are obtained independently at different instances of time while no fundamental dynamics are involved in the respective models.

4.2.1 Measuring Productivity Change

There are several methods in the literature that use a comparative static framework for evaluating the change in relative total factor productivity[7] (TFP) over time. Total factor productivity is commonly defined as the ratio of an aggregate output to an aggregate input (O'Donnell, 2008). Figure 4-1 represents the schematic view of the comparative static framework where both inputs and technology in any period *t* are assumed to be exogenous, but technical change can happen over time (Färe and Grosskopf, 1996). This is called a static framework because it compares a series of static measures of productivity over time where there is no dependence among inputs, outputs or technology across different periods.

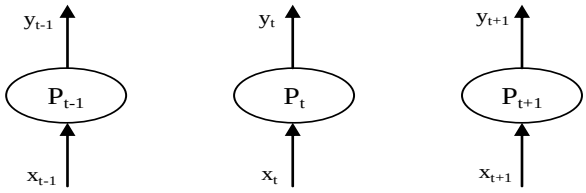


Figure 4-1: The Comparative Static Technology (Färe and Grosskopf, 1996)

The index number methods are the most promising class of methods for estimating productivity change. Diewert (1992a) discusses different index numbers developed in the literature for measuring total factor productivity change in complex technologies, where production units produce multiple outputs using multiple inputs. These indexes are commonly functions of aggregate indexes that measure the change in input/output quantities and/or prices in a production unit going from period *s* to period *t*. Diewert (1992a) discusses the Laspeyres,

[7] The terms total factor productivity and productivity are used interchangeably in this paper.

Paache, Fisher ideal, and Tornqvist aggregate indexes as the four commonly used functional forms for defining aggregate input (output) quantity and price indexes. The advantages of index number methods are that (i) they can handle large set of input/output variables; and (ii) no assumption regarding the functional form of the technology that transforms inputs into outputs is required. The disadvantage of index number approach is that (i) it needs the assumption of competitive profit maximizing behavior; and (ii) it requires data on prices for all inputs and outputs which may not easily be available.

The Malmquist productivity index, proposed by Caves et al. (1982), is another index for measuring productivity change following a static comparative framework. However, this method was then popularized by Färe et al. (1994a). The Malmquist index, defines the measure of change in productivity between consecutive periods as the ratio of distance functions. Distance functions measure how the outputs (inputs) of a production unit can be scaled so that the production unit lies on the boundary of the output (input) production possibility set. Caves et al. (1982) discuss that there are two ways for measuring productivity change over adjacent periods, namely, an input-based productivity index (constructed by input distance functions) and output-based productivity index (constructed by output distance functions). In each case, the Malmquist productivity index is developed as the ratio of distance functions of a production unit at period $t+1$ and t, respectively, where the distance functions can be developed with respect to the frontier for period t or $t+1$. To estimate the Malmquist index, Caves et al. (1982) need to assume a specific functional form for the distance function.

To prevent choosing an arbitrary reference technology (frontier), Färe et al. (1994a) define productivity change as the geometric mean of the two alternative Malmquist productivity indices developed by Caves et al. (1982). Färe et al. (1994a) use the non-parametric technical efficiency measures developed by Farrell (1957) to estimate the distance function component of the Malmquist index. Thus, no specific assumption about the functional form of technology is needed. This also enabled Färe et al. (1994) to decompose the measure of the productivity change into two components capturing "change in technical efficiency" and "shifts in frontier". Färe et al. (1994b) further decomposes the efficiency change component of Malmquist productivity index into "change in scale efficiency" and "change in pure technical efficiency". Färe and Grosskopf (1996) impose the constant returns to scale assumption and formulate the Malmquist productivity indexes in terms of distance functions. Färe (1992) also applies a

nonparametric approach for calculating the input-oriented version of the Malmquist index and calculates productivity change in Swedish pharmacies.

O'Donnell (2008) also develops a conceptual static framework for measuring productivity change with a focus on "complete total factor productivity indexes"[8]. O'Donnell (2008) shows that changes in the TFP index can be decomposed into technical change, change in pure technical efficiency, change in scale efficiency, and change in mix efficiency. Mix efficiency measures the change in productivity if restrictions on input and output mixes are removed and is closely related to allocative efficiency (O'Donnell, 2008; O'Donnell, 2010a). Mix and scale efficiency terms capture the changes in economies of scope and scale of the production units, respectively (O'Donnell, 2008). The class of multiplicatively complete TFP indexes includes Fisher, Tornqvist, and Hickes-Moorsteen indexes (O'Donnell, 2008). But, the Malmquist productivity index is complete only if technology is inversely homothetic[9] and exhibits constant returns to scale (Färe et al., 1996; O'Donnell, 2008). If these conditions do not hold, then the Malmquist productivity index represents a systematically biased measure of change in productivity (O'Donnell, 2008). O'Donnell (2010b) also builds on the framework developed by O'Donnell (2008) and shows how a multiplicatively complete TFP index can be decomposed into measures of technical change, technical efficiency change, and a combination of scale and mix efficiency change.

4.2.2 Measuring Technical Change

Most of the developed approaches for measuring technical change identify this concept with shifts in production function or cost function over time. Diewert (1980;1981) focuses on four approaches to the measurement of technical change, namely, (i) econometric estimation of the production and cost functions; (ii) the divisia; (iii) the exact index number; and (iv) nonparametric approaches.

The econometric approach assumes a convenient functional form for the production or cost functions and uses regression equations to estimate the unknown parameters of the production or cost functions. This approach is based on the assumption that inputs and outputs corresponding to period t are on the firm's period t production/cost function. Thus they assume there in no

[8] TFP indexes that can be expressed as the ratio of an output quantity index to an input quantity index are referred to as multiplicatively complete (O'Donnell, 2008).

[9] Inverse homotheticity essentially means that the output distance function can be defined as multiplication of an output and an input component function (Färe et al., 1996).

technical/allocative inefficiency. Derivatives of the production or cost functions with respect to time estimate the technical change. The divisia approach uses time series data of inputs, output, and prices and assumes that production or cost functions exist and they are differentiable at any moment of time. Then the continuous derivative of the production function or cost function with respect to time is used as a measure of shifts in productivity over time. The continuous time differences of the divisia approach can be approximated by discrete differences. The exact index number approach assumes that the firm's production or cost function has a specific functional form (e.g., translog function that is quadratic in its arguments). Then it is shown that change in the production or cost function can be written as an implicit quantity index in inputs divided by an implicit quantity index in outputs. Finally, the non-parametric approach uses a time-series data on inputs and outputs and assumes that technology sets are convex and accommodate the free disposal assumption. The non-parametric techniques are then used for estimating the maximum possible output for a firm at any point of time and evaluating its changes over time.

Note that no inter-temporal relation is assumed between input and output levels in consecutive periods in any of these methods. Moreover, the discussed methods basically follow a sequential approach since they construct the production reference set by adding new observations into the reference set period by period. In other words, as new observations are obtained, firms never forget how they have performed before. Finally, these models enforce technical progress and do not allow for technical regress.

Baltagi and Griffin (1988) also develop an econometric method for estimating a general index of nonneutral[10] technical change. This method aims to address the shortcomings of the time trend parametric models for technical change where time *T* and its interaction with both input prices and output are defined as regressors of the cost function. One of the shortcoming of this approach is that the estimated pure technical change stays constant or is increased or is decreased at a constant rate (Baltagi and Griffin, 1988). To address these shortcomings, Baltagi & Griffin (1988) follow Robert Solow and replace the time trend *T* in a time trend parametric model with a purely general index of technical change developed by Solow (1957). To allow for the estimation of the cost function, time-specific dummy variables are defined. Baltagi & Griffin

[10] Generally speaking, technical change is "neutral" if it raises productivity of production factors (e.g. capital and labor) by the same proportion. Hick's definition of "neutral" technology requires no change in the ratio of factor marginal products when factor ratios are held constant (Drugge SE. 1988).

(1988) then show that the estimated measure of pure technical change in this approach can capture any pattern of technical change.

4.2.3 Measuring Technical Efficiency Change (Panel Data Models)

The focus of this section is on parametric (stochastic frontier analysis (SFA)) or non-parametric panel data models for estimating technical efficiency change. In SFA, the estimated production frontier includes a production function of the regular regression type and an error term composed of two components representing the usual statistical noise (distributed as a normal distribution) and the non-positive technical inefficiency term. Schmidt and Sickles (1984) overcome the problem of assuming a distributional form for the inefficiency component by extending the SFA model using a fixed effects panel data model. If one assumes that all variation other than inefficiency is controlled for by observable variables, then the individual specific fixed effect is composed of each individual's inefficiency. The main drawback to this approach is that it assumes time invariant efficiency. However, this assumption may hold true for many panel data sets that include only a few years of data at frequent increments such as monthly or weekly. Pitt and Lee (1981) proposed a random effects panel data model, similar to the fixed effects model, which can also be used to estimate efficiency. However, it requires that the individual specific component to be uncorrelated with all regressors and the error component. It is difficult to justify these assumptions in most real world situations.

Cornwell et al. (1990) also relax the assumption that efficiency is time invariant and let both cross-sectional and temporal variations to exist in efficiency levels. The technical inefficiency term is defined as a quadratic function of time whose parameters depend on firms. Thus, the technical inefficiency is allowed to be both firm-specific and time-variant. However, this approach has the drawbacks that were raised by Baltagi and Griffin (1988), meaning that the firm-specific technical inefficiencies either stay constant or are increased/decreased with a constant rate. Sickles et al. (1986) evaluate efficiency growth over time by modeling a profit function that captures allocative efficiency using a flexible function of time, but the efficiency component is assumed to be firm-invariant. Allocative inefficiency happens when firms adjust their inputs or outputs based on wrong price ratios. In both models developed by Sickles et al. (1986) and Cornwell et al. (1990) no distributional assumptions about technical or allocative inefficiencies are needed. Kumbhakar (1990) also models firm-specific and time-varying technical and allocative inefficiencies using a cost minimization framework.

Note that in contrast to the models discussed in the Section 4.2.2, none of the parametric models presented in this section assume that inputs and outputs corresponding to the production units at time *t* lie on the production function. In fact, these models do not estimate a dynamic production function. Instead by estimating a time-invariant production function and allowing for technical inefficiency, these models use a panel data structure and measure time-varying technical inefficiencies of production units.

The stochastic frontier production function using panel datasets on firms has also been extended (e.g., see Battese and Coelli, 1992; Battese and Coelli, 1995) to capture technical change in stochastic frontiers by incorporating an independent time variable representing the time period of the observations.

Among non-parametric approaches, "window analysis" in data envelopment analysis (DEA) that was initially proposed by Charnes et al. (1985) also uses a static framework for capturing the technical efficiency variations over time. Given a panel dataset, window analysis treats each unit in each time period as a different production unit. By defining a window size *W,* the production units in the first *W* time periods are first evaluated using a relevant DEA model. Next, a new period is added to the window and the earliest period is dropped and the DEA model is run again over the new set of production units. This process is repeated till the end of the time horizon. Analysis of the firm's efficiency scores over time leads to a study of trends of efficiency scores. One of the important drawbacks of window analysis is that it drops any observation that is older than the size of the window and forgets about how firms have performed in the past. Window analysis has been developed to provide a larger comparison set for performing a non-parametric estimation of the technical efficiency but there is no theoretical underpinning of why firms would forget after a specific time window. This is in contrast to the sequential models by Diewert (1980) where new observations are added to the comparison dataset over time and make the dataset larger.

As it was stated before, all the key studies discussed in this section focus on quantifying the change in different performance measures obtained independently at different instances of time without assuming any inter-temporal dependence between input and output levels across different time periods. In contrast, the next section provides an extensive discussion on the models/approaches in the literature that capture the inter-temporal relation among different periods by focusing on dynamic issues of the production processes, namely, production delays,

inventories, capital, adjustment costs, and learning effects. Appendix A briefly describes the methods discussed in Section 4.2 as well as their strengths/limitations.

4.3 Modeling Dynamics

4.3.1 Delays in Production (Lagged output)

This section focuses on the approaches developed in the literature for capturing the time delays between input utilization and output realization when measuring technical efficiency of production units. One potential approach to account for the production delays is to add lagged variables to any of the parametric methods (e.g., Pitt and Lee, 1981; Schmidt and Sickles, 1984; Sickles et al., 1986; Cornwell et al., 1990; Kumbhakar, 1990; Battese and Coelli, 1992; Battese and Coelli, 1995) that use panel data to measure time-variant or time-invariant technical efficiency. This is actually the definition of an autoregressive model or a dynamic model for a parametric approach. For example, in the following model, the dependent variable y_{it} is defined as a function of current values of explanatory variables and lagged values of the dependent variable.

$$y_{it} = \alpha + X'_{it}\beta + \gamma y_{i\,t-\tau} + v_{it} + u_{it},\ \tau \in \{1,...,t-1\} \quad \textbf{(12)}$$

Chen and Dalen (2010) develop a non-parametric model for incorporating lagged productive effects (or lagged effects) of input consumption when analyzing technical efficiency. In their model inputs at period *t* incorporate outputs of the current period *t* and future *m* periods (i.e., periods $t, t+1, ..., t+m$). The productive effects of each specific input *p* utilized at period t over outputs in current and future periods are represented by a collection of lag parameters *D* that capture the degree of output augmentation at any of the future periods associated with that specific input *p*. For simplicity, it is assumed that the lag parameters are invariant with respect to starting time period *t*. It is also assumed that the technology constant return to scale. Thus, it is reasonable to assume that lag parameters will not change with respect to the scale of the production.

Let $k=1,...,K$ represents production units over $t=1,...,T$ time periods, each with inputs $x_k^t \in \Re_+^I$ and outputs $y_k^t \in \Re_+^J$. Chen and Dalen (2010) define $\tilde{y}_k^t$ as the actual output produced from input x_k^t as a fraction (g_m) of outputs produced over periods $t, t+1, \ldots, t+m$ (i.e., $[y_k^{t+r}]_{r=0}^m$) where the relative intensity of each output $[y_k^{t+r}]_{r=0}^m$ is defined by lag parameters *D*. $\tilde{y}_k^t$ is called

the dynamic output associated with input x_k^t .Then the output-oriented dynamic technical efficiency $\tilde{\theta}_k^t$ of a production unit *k* at period *t* is obtained from the following non-parametric optimization model (Chen and Dalen, 2010):

$$
\begin{aligned}
&\underset{\tilde{\theta}_k^t,\lambda_i}{Max} \quad \tilde{\theta}_k^t \\
&s.t. \quad \sum_{i=1}^{K} \lambda_i \tilde{y}_{iq}^t \geq \tilde{\theta}_k^t \tilde{y}_{kq}^t, \quad q = 1,...,J, \\
&\quad \sum_{i=1}^{K} \lambda_i x_{ip}^t \leq x_{kp}^t, \quad p = 1,...,I, \\
&\quad \lambda_i \geq 0, \quad i = 1,...,k \\
&\quad \tilde{y}_{iq}^t = g_m([y_{iq}^{t+r}]_{r=0}^m, D), \quad i = 1,...,K
\end{aligned}
\qquad \textbf{(13)}
$$

The dynamic or inter-temporal input-output dependence in this model is captured through the last constraint that relates the technical efficiency at time *t* to a panel of input-output data for periods $t, t+1,...,t+m$ and the lag parameters. The function g_m is assumed to be a linear function of output vectors $[y_k^{t+r}]_{r=0}^m$ weighted by lag parameters. To be able to estimate the dynamic efficiencies $\tilde{\theta}_k^t, k = 1,...,K$, (Chen and Dalen, 2010) estimates the lag parameters by applying panel vector autoregressive method to the empirical input-output panel data. Providing an approach for estimating the lag parameters is an important strength of this study. However, an important shortcoming is that lag parameters are assumed to be invariant over all time periods and for all production units. If one allows for variable returns to scale, it is more realistic for lag parameters to change with respect to the scale of the production.

4.3.2 Inventories

This section focuses on models that introduce a dynamic relationship in the production process by allowing for inter-temporal inventory transfer in a production unit. Hackman (1990) develops an axiomatic framework for modeling a dynamic production process while allowing for inventory of exogenous inputs, as well as inventory of intermediate and final outputs. Hackman's framework defines a production process as a collection of interrelated production activities operating together to produce final outputs. These activities consume two types of inputs: (i) exogenous inputs such as non-storable inputs (labors, machines, etc.) and storable inputs (raw materials, purchased parts, etc.); and (ii) intermediate inputs transferred from other activities. The set of axioms allow one to handle inventory in a production process and are appropriate for

production networks in which inventories are of significant concern. In contrast to traditional production theory, Hackman's framework does not treat a production process as a black box. Instead, by explicit modeling of the production activities, Hackman establishes an appropriate framework for analyzing dynamic production processes.

Hackman (1990) first develops a set of assumptions/constraints that should be satisfied by feasible flows (i.e., a collection of activity input consumptions, transfers of intermediate products from one activity to the others, as well as activity final outputs) that can produce a specific output level over time. Next, the set of axioms that can characterize the space of feasible flows are developed. These axioms are developed to generate the production static axioms such as no free lunch, closeness and convexity of production possibility set, etc. In Hackman's framework, the production function corresponding to each activity is represented by a dynamic production function that transforms time-varying inputs into time-varying outputs in a given period. This relation can change as the time period changes, thus Hackman's framework allows for technical change to happen through time. One of the shortcomings of this framework is that capital is not considered as one of the time-varying aspects of the production process. Instead, capital is treated as an input that changes in a longer term, thus it is fixed during the time horizon under analysis. Another minor shortcoming is that each activity is assumed to produce only one output. Thus, more nodes are necessary to model multiple outputs.

Consistent with the axioms of Hackman's framework, Färe and Grosskopf (1996) present a set of Network DEA (NDEA) models that explicitly capture the network of activities (production nodes) and intermediate products inside the transformation process. By allowing production nodes to represent production in different time periods, the network formulation (dynamic NDEA) can also be used for modeling dynamic production processes (Färe and Grosskopf, 2000). In dynamic NDEA models each period is treated as an activity with its own technology or production process, as well as time-specific inputs/outputs. The relation between periods is captured by time-intermediate products (Färe and Grosskopf, 2000), meaning that some outputs at period t are used as inputs in the next period $t+1$. Figure 4-2 represents the schematic view of a dynamic network model where there is a linkage between consecutive periods. In this model, the total output at any period t is formed by both final output fy_t and intermediate output iy_t. Agriculture industry is a good example of where the dynamic NDEA model can be applied when

part of the produced crops at each period is consumed as the final product and part of it is used as an intermediate input for the next period.

Dynamic NDEA model allows for formulating storable exogenous inputs, where the appearance of inputs and use of the inputs do not necessarily happen at the same time period. Assume there are *K* production units at each period *t* with storable exogenous inputs $x_{tn}, n = 1,...,N^S$ and non-storable inputs $x_{tn}, n = N^S + 1,...,N$. Then, some of the storable inputs at each period *t* can be stored (x_{tn}^S) to be used in future periods. The dynamic network model presented in Figure 4-2 simply allows for exogenous inputs to be stored only for one period. Thus, in this model, there are two sources of interdependence among consecutive periods *t* and *t+1*: (i) intermediate inputs iy_t, (ii) storable inputs x_t^S.

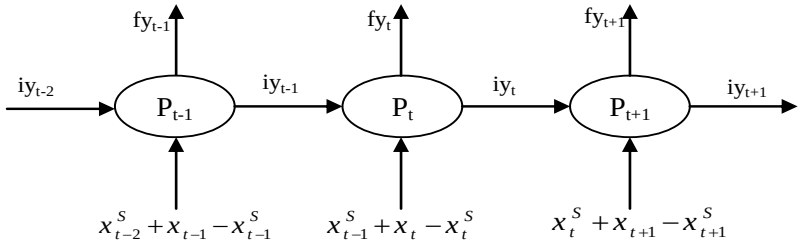


Figure 4-2: The Dynamic Network Model With Storable Inputs (Färe and Grosskopf, 1996)

Then the dynamic efficiency corresponding to Figure 4-2 can be obtained by solving the following model (Färe and Grosskopf, 1996):

$$
\begin{aligned}
&\underset{\{\theta_\tau,\lambda_\tau,iy_\tau,x_\tau^S\}}{Max} \sum_{\tau=t-1}^{t+1}\theta_\tau \\
&s.t.\ \ \theta_\tau fy_{\tau m} + iy_{\tau m} \le \sum_{k=1}^{K}(fy_{\tau m}^k + iy_{\tau m}^k)\lambda_\tau^k \quad m = 1,...,M,\quad \tau = t-1,t,t+1, \\
&\quad x_{\tau-1n}^S + x_{\tau n} - x_{\tau n}^S \ge \sum_{k=1}^{K}(x_{\tau-1n,k}^S + x_{\tau n,k} - x_{\tau n,k}^S)\lambda_\tau^k \quad n = 1,...,N^S,\quad \tau = t-1,t,t+1, \\
&\quad x_{\tau n} \ge \sum_{k=1}^{K} x_{\tau n,k}\lambda_\tau^k \qquad n = N^S+1,...,N,\quad \tau = t-1,t,t+1, \\
&\quad iy_{\tau m} \ge \sum_{k=1}^{K} iy_{\tau m,k}\lambda_\tau^k \quad m = 1,...,M,\ \tau = t-1,t,t+1, \\
&\quad \lambda_\tau^k \ge 0,\quad k = 1,...,K,\ \tau = t-1,t,t+1.
\end{aligned}
\qquad \textbf{(14)}
$$

This model maximizes the output efficiency of a production unit over time by calculating the intermediate inputs iy_t and storable inputs x_t^S endogenously. By allowing for different

technologies at different periods (through time-varying inputs, outputs, and intensity variable λ_t), dynamic NDEA models provide the possibility for technological change throughout the time horizon under analysis. Note that Färe and Grosskopf (1996) limit the scope of their analysis to a single production process connected over time periods using the intermediate inputs. Moreover, their model only allows for storability of exogenous inputs and does not allow the stored inputs to perish over time.

Chen (2009) addresses some of these shortcomings by adopting a broader network perspective and allowing for the production process at each period to be composed of sub-production processes that use exogenous or intermediate inputs to produce final or intermediate inputs that can be used by other sub-production processes. The intermediate inputs produced by sub-production processes are allowed to be stored and used in future periods. Moreover, the stored intermediate inputs are allowed to perish over time, meaning that their effectiveness on future outputs decreases over time. Figure 4-3 illustrates the structure of the dynamic production network formulated by Chen (2009). $\alpha_{IJ}^{t_m t_m}$ represents the percentage of the intermediate input $iy_{IJ}^{t_m}$ produced at time t_m by sub-production process I to be used at time t_m by sub-production process J. $\alpha_{IJ}^{t_m t_n}$,on the other hand, represents the percentage of the intermediate input $iy_{IJ}^{t_m}$ produced at time t_m by sub-production process I to be used by sub-production process J at time t_n ($n>m$). $\beta_{IJ}^{t_m t_n} \geq 0$ represents the effectiveness (or perishability) of unconsumed intermediate input $iy_{IJ}^{t_m}$ that is produced at time t_m to be used at time t_n.

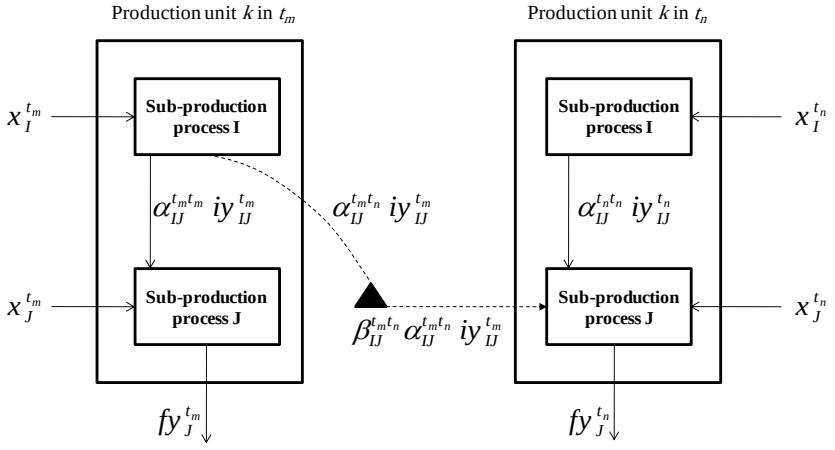


Figure 4-3: Structure of the Dynamic Production Network Formulated by Chen (2009)

For each production unit *k*, Chen (2009) first develops efficiency measures of each sub-production process *I* (*J*) benchmarked with similar sub-production process *I* (*J*) of other production units in the same time period. These efficiency measures are then used to develop dynamic measures of efficiency of sub-productions *I* and *J* of production unit *k* while the network structure and inter-temporal effects are taken into account. For example, the possible input reduction in sub-production *I* of production unit *k* in period t_m not only depends on the technical efficiency of sub-production *I*, but also on the input reduction that is imposed by sub-production *J* at time period t_m as well as time period t_n due to presence of storable intermediate inputs. The technical efficiency of sub-production *I* (or *J*) of production unit *k* is finally defined as the ratio of the minimum input that can be used by sub-production *I* (or *J*) over the time horizon to the observed total input that has been used by the sub-productions. Moreover, the dynamic efficiency of production unit *k* is defined as the multiplication of the dynamic efficiency of its sub-production processes. By explicit modeling of the internal structure of a production unit, this approach allows for detecting the potential sources of inefficiency. One of the shortcomings of this study is that it does not discuss how the dynamic parameters $\beta_{IJ}^{t_m t_n}$ can be estimated. Moreover, it does not consider storability of exogenous inputs.

Färe (1986) also use a non-parametric approach to develop a measure of output efficiency for the whole planning horizon *T* assuming that two types of inputs are available: one that is given at any time *t* and one whose total volume is given for the whole time horizon *T*. Assume data of *n* inputs $x_i, i = 1,...,n$ and one output *y* corresponding to *K* production units are available. Let *p* be a subset of inputs that are given at any time *t* over the planning horizon $(\text{i.e., } x_i^t,\ i \in p)$, and *q* be the subset of inputs for which the total amount is given for the whole planning horizon $(\text{i.e., } x_i,\ i \in q)$. A production unit then decides on the allocation of the second type of inputs over finitely many time periods. To develop an output measure of efficiency for production unit *k* over the whole planning horizon *T*, Färe (1986) compares the sum of observed outputs $Y = \sum_{t=1}^{T} y^t$ to the sum of largest potential outputs that can be obtained by optimal allocation of input *i*, $i \in q$ over time, as follows:

$$
\begin{aligned}
&\underset{x_{ki}^{t},\, i \in q}{Max} \sum_{t=1}^{T}\sum_{j=1}^{K} y_{j}\lambda_{j}^{t} \\
&s.t. \quad \sum_{j=1}^{K} x_{ji}\lambda_{j}^{t} \le x_{ki}^{t}, \quad i \in n,\ t = 1,...,T, \\
&\qquad \sum_{t=1}^{T} x_{ki}^{t} \le x_{ki}, \quad i \in q, \\
&\qquad \sum_{j=1}^{K} \lambda_{j}^{t} = 1, \qquad t = 1,...,T, \\
&\qquad \lambda_{j}^{t} \ge 0, \quad \forall j,t.
\end{aligned}
\tag{15}
$$

Inefficient allocation of input *i*, $i \in q$ over time leads to the observed output in some periods to be less than the maximum potential outputs in those periods. In this model, allocation of the second type of inputs over time adds a dynamic aspect to the general non-parametric models. This is because the amount of input *i*, $i \in q$ allocated for production in each period *t* depends on the total amount of input *i* that has been allocated to previous periods and the total amount of input *i* that has been left for future periods. Finally, the inputs and outputs that have formed the technology set in this model are time invariant. Thus, this model does not allow for technical change over time.

4.3.3 Capital (embodied technical change)

4.3.3.1 The Vintage Concept

Most of the studies in literature do not treat capital as a vintage specific factor. In these studies usually replacement cost or book value of the capital is used to come up with a measure of capital that can capture the quality of different pieces of capital that potentially have different vintages. But these measures cannot appropriately account for the fact that different pieces of capital bought/installed at different points of time have different levels of productivity. Vintage models can address this issue to a large extent.

Taking into account the vintage of capital can help with constructing homogeneous measures of capital when, for example, one accounts for the age of machines in a production unit. Several studies in the literature (e.g., see Forsund and Hjalmarsson, 1974; Forsund and Eilev, 1983; Forsund, 2010) have analyzed the dynamic aspects of efficiency where technology difference is the main reason for explaining inefficiencies of production units. In these studies technology is embodied in the capital equipment and capital equipment is essential for production. Moreover, progress in the technology or structural change happens due to investment

in new capital (embodying new technology) and scrapping the old capital (that is no longer beneficial to keep). This leads to a constant dynamic development of a firm and consequently affects its inputs and outputs. In such a setting, the vintage approach, initially developed by Johansen (1972), can explain efficiency differences. The idea is that inefficiency in production units that have different vintages of capital equipment can potentially be explained by the heterogeneity of their capital equipment (Forsund, 2010).

One of the key characteristic of the vintage models is that a micro unit (e.g., a firm or a plant) needs to distinguish between substitution possibilities between capital and variable inputs before the investment takes place (micro unit ex ante production function) and substitution possibilities after the time of the investment (micro unit ex post production function) (Johansen, 1972; Forsund and Eilev, 1983). For a micro unit, let y be the output, $x_i, i = 1,...,n$ stands for the inputs and *K* represent the capital. Then a micro unit ex ante production function at current time *t=0* is represented as (Forsund, 2010):

$$y = f_0(x_1,...,x_n,K), f'_{0i}, f'_{0K} > 0, i = 1,...,n \quad \textbf{(16)}$$

At the time of investment, a micro unit decides on expanding its capital based on the ex ante production function. After making a choice of capital, then the volume of capital is fixed and there are no substitution possibilities between capital and variable inputs. This is equivalent with choosing a point on the isoquant corresponding to the ex ante production function (Forsund, 2010). In other words, after the investment time ν the input coefficients $\xi_i(\upsilon) = \frac{\bar{x}_i(\upsilon)}{\bar{y}(\upsilon)}, i = 1,...,n$ are fixed, where $\bar{y}(\upsilon)$ represent the output capacity based on the investment in capital at time ν and $\bar{x}_i(\upsilon)$ represents the required input *i* if the production unit produces at full capacity. Then the ex post micro unit production function capturing the actual production at any point of time $t \geq \upsilon$ can be represented as (Forsund, 2010):

$$y(t,\upsilon) = \min\left[\frac{x_1(t,\upsilon)}{\xi_1(\upsilon)},...,\frac{x_n(t,\upsilon)}{\xi_n(\upsilon)},\bar{y}(\upsilon)\right], \quad \textbf{(17)}$$

where $x_i(t,\upsilon)$ represent the amount of input *i* used in production unit at time *t*. Note that there is no explicit representation of real capital *K* in the ex post production function. Considering these two production functions and their differences when analyzing investment decisions can lead to different understanding of productivity changes and their driving factors (Forsund, 2010).

Having set the ex ante and ex post production functions and assuming that firms are producing at full capacity over the economic life time *T* of capital, Forsund (2010) develops an optimization problem (Equation (18)) to decide on the initial capital level (investment decision) and the initial level of variable inputs for a micro unit as:

$$\begin{aligned} &\underset{K(0),x_i(0,0)\forall i}{Max}\ \pi(0)=\int_{t=0}^{T} e^{-rt}[p(t)y(t,0)-\sum_{i=1}^{n} q_i(t)x_i(t,0)]dt-q_k(0)K(0) \\ &s.t. \qquad y(0,0)=f_o(x_1(0,0),...,x_n(0,0),K(0)) \quad \text{for } t=0 \\ &\qquad\quad y(t,0)=\bar{y}(0),\ x_i(t,0)=\xi_i(0)\bar{y}(0) \quad \text{for } t\in[0,T] \end{aligned} \qquad \textbf{(18)}$$

In this model q_i is the price for variable input *i*, q_K is the price per unit of capital, *p* is the output price, and *r* is the discount factor. Input/output prices in future periods are assumed to be known with certainty. This model maximizes the present value of the net profit of investment in capital over the economic lifetime of capital. In this model, the initial decision (at time zero) on the level of capital and the input coefficients are made based on the ex ante production function (first constraint). The actual output during the planning horizon *T* is then defined based on the ex post production function (second constraint).

To analyze the structural change at the industry level, Forsund and Eilev (1983) aggregate the ex post production functions of micro units and develop an industry level short-run production function by maximizing the output of the industry given the limited available industry-level inputs. Assuming that an industry consists of *N* micro units with homogeneous inputs and output, the industry level production function shows how the capacities of micro units should be utilized to maximize the output of the industry. In this aggregate production function, each micro unit can have a different technology (vintage). Industry level production function utilizes the technologies of all micro units and finds an optimal relation between the aggregate industry input and output (Forsund and Eilev, 1983). The actual (observed) scenario for utilizing capacity and resources among micro units can deviate from the optimal solution defined by the industry level short-run production function. This provides a potential source of inefficiency for the micro units as well as the industry (Forsund, 2010). Note that the short run production function is static, but can provide snapshots of how an industry has evolved over time. Thus, in a dynamic perspective, the short run production function captures the history of the ex ante production function as well as the choices of capital and input coefficients over time (Forsund, 2010).

Forsund and Hjalmarsson (1974) discuss the concept of "optimal structural development path" for an industry where this development path can be constructed by understanding the objective function associated with industrial policy (e.g., minimizing the cost of production) and conditional upon the available information about future techniques, prices, and demand. The structure of an industry is captured by its capital equipment and its scale properties. The snap shot of the optimal development path at any given point of time defines the corresponding optimal or the efficient structure and can be compared with the actual structure at that point of time (Forsund and Hjalmarsson, 1974). These snapshots represent the short run production functions. This has a serious ramification for non-parametric methods, since one needs to further categorize the data set (i.e., smaller datasets) to make sure the a production unit is compared only with those production units that have the same capital vintage as the firm under analysis. This requires a large dataset, which is not usually available in most of the practical problems. Forsund and Hjalmarsson (1974) do not provide any discussion as to how one can implement such a path by timely investment in the new technologies and disposing of old ones.

Wibe (2008) also focuses on developing a dynamic production model when technology is embodied in capital equipments. The idea is that, in a dynamic context when technology evolves rapidly over time, it may be uneconomical for a firm to dispose its old capital any time that a new technology is introduced to the market. Thus, there is always a technical inefficiency due to difference between the observable technology of a firm (or observable average technology of an industry) and the best technology that is available in the market. Wibe (2008) discusses that this inefficiency is perfectly rational and measures the "rational inefficiency" in a dynamic context. Given a linear and homogeneous ex ante production function with capital (*K*) and labor (*L*) as the only production factors, Wibe (2008) defines the dynamic equilibrium as a situation in which the economic lifetime of all vintages are greater than zero and are the same. To obtain a dynamic equilibrium, it is assumed that wage, capital price, and output price are one at time zero and then increase with constant rates over time. It is also assumed that in equilibrium output increases at a rate *a*. Wibe (2008) then finds the economic lifetime *T* of capital, the optimum capital investment K_ν, and the optimum labor input L_ν at any investment time ν by maximizing the profit over the capital life time *T* given the ex ante production function as follows:

$$\underset{q_k(\nu),L_\nu}{Max}\ \Pi_\nu = \int_{\nu}^{\nu+T} e^{-r(t-\nu)}[y_\nu p(t) - L_\nu q(t)]dt - K_\nu q_k(\nu)$$
$$s.t.\quad y_\nu = e^{b\nu} f(K_\nu, L_\nu), \tag{19}$$

where *p(t), q(t),* and $q_k(\nu)$ are the time-varying output price, wage, and capital price, respectively, and *b* in the ex ante production function represents the rate of embodied technical change. It can be seen that this model is equivalent to Equation (18) developed by Forsund (2010) for deciding on optimal level of capital investment and variable inputs. Moreover, both models assume that firms are producing in full capacity over the economic life time of capital and that capital does not depreciate over time. Forsund (2010) assumes that economic life time (*T*) of capital is given, while Wibe (2008) estimates the optimal value for *T* and shows that *T* satisfies the familiar disposing condition meaning that a firm or a machine is scrapped when its variable cost (i.e., wage) grows faster than the output price. The measure of technical efficiency in equilibrium is then defined as the ratio of the observed total output Q^* (i.e., the sum of outputs over all vintages from 0 to *T*) to the output that could potentially be produced if the latest technology at time *T* would be used (i.e., $e^{bT} f(K^*, L^*)$), where K^* and L^* represent total capital and total labor, respectively, over all vintages from *0* to *T*. This measure captures the inefficiency resulted from economic rationality, when technology of a firm is less efficient than best practice. Wibe (2008) discusses that his measure of technical inefficiency complements Farrell's measure of technical efficiency and comparing these two measures can give valuable insights about the "real" technical inefficiency.

Färe and Grosskopf (1996) also extend the application of NDEA to evaluates technical efficiency of a firm while allowing the firms to have different vintages of capital. In their NDEA model, each vintage is modeled as a node that receives its own input and produces the output. Total output of the production unit is defined as the sum of outputs from all vintages. The NDEA model decides on allocation of the variable input among different vintages, thus provide the possibility for adopting new vintages and disposing the old ones at a specific time period t. Although their NDEA model accounts for heterogeneity of capital vintages as a potential source of efficiency differences among firms, it does not consider the inter-temporal nature of investment decision regarding adopting new vintages and disposing the old ones during a planning horizon and does not elaborate as how these decisions affect future production and investment decisions. Thus, their model does not capture any essential dynamics.

In sum, the models that were discussed in this section assume that adoption of new vintages of capital and disposing of old ones is the only source of technical change through time. As a result, the inefficiency of firms (or industries) can be attributed to the path they have undertaken for development of their capital (in terms of timing and amount of investment in new vintages of capital). These models assume that substitution possibilities between capital and variable inputs exist ex ante, but are restricted ex pots. Taking these assumptions into account, these models are basically inter-temporal optimization problems that decide on an ex ante decisions regarding the introduction of a new vintage (its magnitude and/or timing of investment) where the optimal decision requires information on future expected prices, demand, and other factors.

4.3.3.2 Investment in Capital

One of the important contributions in the literature for addressing the issue of capital investment and its inter-temporal effects on outputs is that of Sengupta (1995). His studies focus on generalizing the DEA framework by introducing the inter-temporal decision making context Sengupta (1994a; 1994b; 1996) adds a dynamic aspect to the DEA framework by explicitly introducing capital inputs in addition to variable inputs. The idea is that utilization of the capacity expanded by capital investment happens over future periods. Sengupta (1994b) uses the dual LP formulation of a CCR primal DEA model for this dynamic generalization and starts with primal and dual versions of Farrell's production frontier model as:

Primal:	Dual:	
$\underset{\beta}{Min}\ g = v'\beta$	$\underset{\lambda}{Max}\ J = e'\lambda$	**(20)**
$s.t.\ \beta' A \geq e',\ \beta \geq 0$	$s.t.\ A\lambda \leq v,\ \lambda \geq 0$	

The primal model uses the vector $v=(v_i)$ of aggregate inputs $v_i = \sum_{j=1}^{n} x_{ij}$ where there are n production units, matrix $A=(a_{ij})$ of input-output coefficients $a_{ij} = x_{ij} / y_j$, and e_n is the vector of ones. Moreover, β represents input shadow prices and λ represents vector of output weights. Thus, one possible interpretation of the dual model is that it tries to maximizes the value of total outputs of all production units given the observed input and output data (Sengupta, 1996). This formulation is similar to the industrial production function and industrial efficiency model developed by Johansen (1972). To generalize the dual model, Sengupta (1994b) assumes that aggregate input vector ν can partly be used for current inputs (current levels of both variable and

capital inputs) and partly for additional capital stocks that increase the output in the next period. Thus, the constraint and objective function of the dual model can be expanded to account for the additional output that has been generated due to the additional capital input as follows (Sengupta, 1994b):

$$\begin{aligned} &Max \;\; e'\lambda(t)+e'[\lambda(t+1)-\lambda(t)] \\ &s.t.\; A\lambda(t)+B(\lambda(t+1)-\lambda(t)) \le \nu(t), \;\; \lambda(t+1),\lambda(t) \ge 0 \end{aligned} \qquad \textbf{(21)}$$

where matrix $B=(b_{ij})$ represents incremental capital-output coefficients. It is assumed for simplicity that an increase in capital input in period *t* yields to an increase in output in the next period *t+1*. In this model the input vector $\nu(t)$ (including both variable and additional capital inputs) and matrices *A* and *B* are given for two periods. This model determines optimal implicit output weights $\lambda(t)$ and $\lambda(t+1)$. Note that in this model efficiency is defined in terms of both $\lambda(t)$ and $\lambda(t+1)$, meaning that the resource allocation should be output maximizing in both period *t* and *t+1* (Sengupta, 1994a). This prevents the problem of having a resource allocation that is optimal in period *t* but not in period *t+1*. Sengupta (1994b) also extends the two-period model (Equation (21)) over a planning horizon by fixing a time horizon T and allowing for the coefficient matrices *A* and *B* to change over time. Note that Sengupta's model only addresses the inter-temporal relation between two consecutive periods. A more general model should expand the inter-temporal relations over multiple periods.

Emrouznejad and Thanassoulis (2005) also develop a DEA model where there is an inter-temporal dependence between input and output levels due to effects of capital investment and capacity expansion on future outputs. In such setting, there is no correspondence between inputs and outputs observed during the same period, since outputs at period *t* may be caused by inputs in several previous periods. To address this issue in measuring performance of a production unit, Emrouznejad and Thanassoulis (2005) use input-output paths (i.e., the sequence $(x_j^t, y_j^t), t=1,...,T$) corresponding to production unit $j=1,...,N$ to construct the production possibility set in contrast to single period input-output levels. This sequence is defined as assessment path and reflects all inputs used by a production unit and all outputs produced irrespective of the time lag between inputs and outputs. Emrouznejad and Thanassoulis (2005) defines the assessment path of production unit *k* to be dynamically efficient if no other assessment path or combination of assessment paths produce more than the path of production

unit *k* of at least one output in at least one period t without using more of at least one input in at least one period or without producing less of some outputs in some periods.

Since the assessment window *T* usually captures part of the life of a production unit, one needs to account for the cases in which a capital investment made within an assessment window affect the outputs outside the window or where outputs inside the assessment window are results of capital investments prior to the window. To address this issue, Emrouznejad and Thanassoulis (2005) (i) treats initial capital stock (i.e., capital stock at period 1) as an input that can be converted to outputs within the assessment window; and (ii) treats the final stock of capital as another output at the end of assessment window. In contrast to Sengupta's model (Equation (21)), the model by Emrouznejad and Thanassoulis (2005) allows for inter-temporal effects of capital investment on outputs over multiple future periods (a more general case). Note that by using input-output paths (assessment paths) instead of single period input-output representation, the dimension of the production possibility set increases multiplicatively as number of time periods or input or outputs increases. This has serious consequences on discriminating power of the DEA formulations.

4.3.4 Adjustment Cost

4.3.4.1 Dynamic Programming (or Analytical Approach)

As it is discussed by Sengupta (1994a; 1996), development of an inter-temporal adjustment cost function is an alternative approach for involving an inter-temporal planning horizon into the dynamic framework for measuring efficiency. Sengupta (1994a) discusses two different ways for interpreting adjustment costs. One way is to attribute the adjustment costs to investment in capital or quasi-fixed inputs. The idea is that as investment in each quasi-fixed input increases, the level of foregone output increases. Thus capital inputs are available at increasing unit costs. Let *u* represents the change in capital input. Then the adjustment cost *C(u)* is defined to satisfy the convexity conditions presented in Equation (22) (Sengupta, 1994a). This approach has been followed in several studies (e.g., see Sengupta, 1994b; Silva and Stefanou, 2003; Silva and Stefanou, 2007) for modeling the adjustment cost in the form of foregone outputs due to capital expansions.

$$C(0) = 0,\ \ \partial C(u)/\partial u > 0,\ \ \partial^2 C(u)/\partial^2 u \geq 0 \qquad \textbf{(22)}$$

The second way is to model adjustment cost as a function of deviation of inputs or outputs from their target levels. This approach develops an adaptive process for adjustment of inputs or

outputs over time (Sengupta, 1996). In this approach, the adjustment cost usually is composed of two components. One component represents the cost of fluctuations in inputs or outputs through time; and the other component captures the cost of deviation from the desired path (disequilibrium cost). The adjustment cost in this approach is modeled using quadratic functions of the deviations of inputs or outputs from their desired levels. The choice of quadratic function for representing adjustment cost allows for smooth and stable adjustment of inputs or outputs over time (Sengupta, 1994a; Sengupta, 1996).

Sengupta (1992a; 1994a; 1994b; 1996; 1999) uses the adjustment cost approaches described above and develops a set of optimal control models for measuring efficiency of production units in an inter-temporal framework where both variable and capital inputs are used for producing the final outputs. More specifically, Sengupta (1994b) uses the first approach and generalizes the concept of Farrell efficiency in a dynamic framework where inter-temporal changes of capital inputs (through investment in capital and its depreciation over time) is defined as a constraint and the adjustment costs associated with capital investments are taken into account in the objective function. Assume there is a single output y_j, *m-1* variable inputs $x_{ij}, i = 1,\ldots,m-1$, and one capital variable x_{mj} for each production unit *j=1,...,K*. Let gross investment in capital for production unit *j* be captured by $\dot{x}_{mj} = v_j - \delta x_{mj}$, where v_j represents the investment in capital and δ is the constant rate of depreciation. By introducing the capital adjustment cost in the dual Farrell LP model, Sengupta (1994b) develops a dynamic model for measuring efficiency in production unit *k* as:

$$\begin{aligned} &\underset{\lambda_j(t), v_j(t)}{Max}\ z_k = \int_0^{\infty} \exp(-rt)\left[\sum_{j=1}^{K}\left\{\lambda_j(t) y_j(t) - C_j(\dot{x}_{mj})\right\}\right] dt \\ &s.t. \quad \sum_{j=1}^{K} x_{ij}(t)\lambda_j(t) \le x_{ik}(t),\ i = 1,\ldots,m, \\ &\lambda_j(t) \ge 0,\ j = 1,\ldots,K, \\ &\dot{x}_{mj}(t) = v_j(t) - \delta x_{mj}(t),\ j = 1,\ldots,K, \end{aligned} \qquad \textbf{(23)}$$

where *r* is the exogenous discount rate and $C_j(\dot{x}_{mj})$ shows the adjustment cost corresponding to the capital input and it satisfies the conditions presented by Equation (22). This is a standard problem in optimal control theory and aims to find the optimal trajectory for capital variable $x_m^*(t)$ that maximizes the present value of net output of production unit *k* minus adjustment cost. Note that this model is dynamic due to inter-temporal variation of capital input presented by the

last constraint of model (23) and the inter-temporal objective function (Sengupta, 1994b). The optimal investment path $x_m^*(t)$ obtained from model (23) defines the optimal trajectory $y^*(t)$. Thus, the deviations between $y^*(t)$ and observed *y(t)* can be used as a measure of efficiency.

Sengupta (1999) uses the second approach and incorporates dynamics of production technology into DEA models by introducing an inter-temporal adjustment cost functions associated with deviation of inputs from their target levels. In this model all inputs (both variable and capital inputs) can vary over time. Assuming $X_j(t)$ and $Y_j(t)$ represent the input and output vectors corresponding to production unit *j (j=1,…,K)* at time *t*, the inter-temporal optimization problem that minimizes the expected present value of a quadratic loss function for production unit *k* is developed as (Sengupta, 1999):

$$\begin{aligned} &\underset{x(t),\lambda(t)}{Min}\ L = E_t\left\{\sum_{t=1}^{\infty}\rho^t\left[q'(t)x(t)+\left(\frac{1}{2}\right)(d'(t)Wd(t))+\left(\frac{1}{2}\right)(z'(t)Hz(t))\right]\right\} \\ &s.t. \quad \sum_{j=1}^{K}X_j(t)\lambda_j(t)\le x(t), \\ &\qquad \sum_{j=1}^{K}Y_j(t)\lambda_j(t)\ge Y_k(t), \\ &\qquad \sum_{j=1}^{K}\lambda_j(t)=1,\ x(t)\ge 1,\ \lambda(t)\ge 0, \end{aligned} \qquad \textbf{(24)}$$

where *q(t)* is the vector of input prices varying over time, ρ is the discount factor, *W* and *H* are diagonal weight matrices. The term $d(t)=x(t)-x(t-1)$ captures fluctuations in inputs and the term $z(t)=x(t)-\hat{x}(t)$ captures deviations from the desired path $\hat{x}(t)$. Thus, the quadratic part of the objective function of Equation (24) captures the adjustment costs associated with time-varying input vector *x(t)* (Sengupta, 1999). Note that the objective function in this model involves an inter-temporal planning horizon, since (i) it is defined over the whole planning horizon; (ii) it maintains a relation between consecutive periods by introducing the term *d(t)*. By solving this inter-temporal optimization model, the optimal inter-temporal path of input variables $x^*(t)$ can be characterized. Difference between the observed path $X_k(t)$ and the optimal path $x^*(t)$ for production unit *k* at any time *t* represents the inter-temporal inefficiency (Sengupta, 1999). That future values for input prices *q(t)* and target level $\hat{x}(t)$ can be estimated by observing past trends using the method of least squares (Sengupta, 1994a).

Sengupta (1994a) also uses an adjustment cost approach similar to Equation (24) to develop an inter-temporal framework for measuring efficiency of risk averse production units where there are stochastic errors associated with input/output quantity.

Sengupta (1992b) extends the ideas of cointegration[11] and dynamic econometric modeling to the non-parametric methods to estimate the production and cost functions where the input-output data are non-stationary over time. The dynamic production frontier incorporates the adjustment cost as a function of the change in capital input where the adjustment cost function satisfies the convexity condition presented in Equation (22). The cost frontier incorporates the adjustment cost as a quadratic function of the deviation from the desired levels of outputs. The error terms in both frontiers are set to be one-sided and the error structures are directly incorporated in the constraints of the linear programming models. The estimations are done by minimizing the least sum of absolute value of error terms.

The other important contribution in the area of capital investment and adjustment cost as an inter-temporal decision making problem is attributed to the Dynamic NDEA models (Färe and Grosskopf, 1996). As it was discussed in Section 4.3.2, Dynamic NDEA models treat each period as an activity with its own technology or production process. Investment in capital at each period t can be modeled as a decision factor that is determined endogenously and affects the stock of capital at the end of that period. The stock of capital is then treated as an intermediate input at period $t+1$ and affects the production possibilities in future periods by expanding the production capacity. Given the available input levels at each period, an increase in investment in capital at each period t decreases the final output produced in the same period.

Assume there are K production units at each period t with exogenous inputs, $X_t=(x_{t1}, x_{t2}, \ldots, x_{tK})$, intermediate inputs $iY_{t-1}=(iy_{t-11}, iy_{t-12}, \ldots, iy_{t-1K})$, and final outputs $fY_t=(fy_{t1}, fy_{t2}, \ldots, fy_{tK})$. For simplicity assume all exogenous inputs are non-storable. Then, for a given production unit k, the dynamic efficiency can be computed by solving the following model (Färe and Grosskopf, 1997):

[11] A time series is defined to be integrated of order m if it is required to be differenced m times to be stationary. If several time series are each integrated but some linear combination of them leads to a time series with lower order of integration, then those time series are said to be cointegrated.

$$
\begin{aligned}
&\max_{\{\theta_t,\lambda_t,iy_t\}} \sum_{t=1}^{t=T} \theta_t \\
&s.t.\ \ \theta_t fy_{tk} + iy_t \le (fY_t + iY_t)\lambda_t \qquad t = 1,2,...,T, \\
&\qquad iY_{t-1}\lambda_t \le iy_{t-1} \qquad t = 1,2,...,T, \\
&\qquad X_t\lambda_t \le x_{tk} \qquad t = 1,2,...,T, \\
&\qquad \lambda_t \ge 0 \qquad t = 1,2,...,T.
\end{aligned}
\tag{25}
$$

In this model, the endogenous input iy_t which is the output from period t together with final output fy_t form the total output for period t. Moreover, the two inputs corresponding to period $t+1$ are exogenous inputs x_{t+1} and intermediate inputs iy_t. Moreover, the dynamic or time interdependence among consecutive periods arise from endogenous inputs $iy_1,\ldots,iy_T$ that are computed simultaneously by solving model (25) over the time horizon T. Färe and Grosskopf (1997) use model (25) to study the dynamic efficiency of APEC (Asian-Pacific Economic Community) countries where employment and nonresidential capital stock form the inputs and real GDP forms the total output. In their study, investment in capital stock in period t-1 is used as the intermediate inputs in period t. By solving model (25), one can also compare the optimal investment path $iy_1, \ldots, iy_T$ with actual investment path (Färe and Grosskopf, 1997). This comparison helps with analyzing the investment path as one of the potential sources of inefficiencies. Moreover, it is clear that investment in capital in any period t depends on capital stock in earlier periods and also affects the investment decision as well as final outputs of future periods.

Based on the discussion on NDEA, these models, to a great extent, can capture two important dynamic aspects of production: adjustment cost and inventory. Within a limited scope, NDEA models can also accommodate efficiency measurement in vintage specific production units. Moreover, NDEA models can be solved using linear programming, thus there is no need to use optimal control, dynamic programming, or other complex techniques (Färe and Grosskopf, 1997).

Nemoto and Goto (1999; 2003) also address the issue of inter-temporal decision-making and the costs of adjustment based on a similar concept as in dynamic NDEA models. In their model adjustment cost is imposed in terms of giving up some of the production of final goods at period t when it is necessary to increase the level of quasi-fixed capitals at the end of that period (or at the beginning of period $t+1$). Due to an increase in the level of quasi-fixed inputs at the beginning of period $t+1$, the level of final goods that can be produced in future periods increases. That is how a dynamic interdependence across time periods can be defined (Nemoto and Goto,

1999). To capture this resource allocation problem in DEA, they treat the quasi-fixed factors at the end of each period as outputs in that period that are placed into next period's production process as inputs. To represent the dynamic DEA model mathematically, they assume that there are N production units at period t with variable inputs, $X_t=(x_{t1}, x_{t2}, \ldots, x_{tN})$, quasi-fixed inputs at the start of period *t*, $K_{t-1}=(k_{t-11}, k_{t-12}, \ldots, k_{t-1N})$, quasi-fixed inputs at the end of period *t*, $K_t=(k_{t1}, k_{t2}, \ldots, k_{tN})$, and final products $Y_t=(y_{t1}, y_{t2}, \ldots, y_{tN})$. In addition, they assume $w_t=(w_{t1}, w_{t2}, \ldots, w_{tN})$ and $v_t=(v_{t1}, v_{t2}, \ldots, v_{tN})$ represent the prices for variable and quasi-fixed inputs respectively. Then the inter-temporal efficient cost frontier is developed as (Nemoto and Goto, 1999):

$$
\begin{aligned}
&\hat{C}(\bar{k}_0) = \min_{\{x_t, k_t, \lambda_t\}_{t=1}^{T}} \sum_{t=1}^{T} \gamma^t (w_t' x_t + v_t' k_{t-1}) \\
&s.t.\ X_t \lambda_t \le x_t, && t = 1,2,\ldots,T \\
&\quad K_{t-1}\lambda_t \le k_{t-1}, && t = 1,2,\ldots,T \\
&\quad K_t \lambda_t \ge k_t, && t = 1,2,\ldots,T \\
&\quad Y_t \lambda_t \ge y_t, && t = 1,2,\ldots,T \\
&\quad i'\lambda_t = 1, && t = 1,2,\ldots,T \\
&\quad k_0 = \bar{k}_0,\ x_t \ge 0,\ k_t \ge 0,\ \lambda_t \ge 0, && t = 1,2,\ldots,T
\end{aligned}
\qquad \textbf{(26)}
$$

where γ is the constant discount factor. Given the initial values $\bar{k}_0$ for quasi-fixed inputs, this model finds the optimal path for both variable and quasi-fixed inputs over the planning horizon T by constructing the inter-temporal efficient cost frontier. Nemoto and Goto (2003) build on their previous work (Nemoto and Goto, 1999) and show that comparison between the minimum cost obtained from model (26) and the discounted sum of actual cost from period 1 to period *T* can be used as a measure of overall efficiency of a production unit. Nemoto and Goto (2003) decompose overall efficiency into static measure of efficiency (by keeping the quasi-fixed factors at their observed level in model (26)) and dynamic efficiency. By isolating the static measure of efficiency from the overall efficiency, they obtain the dynamic measure of efficiency where the sources of underperformance are only attributed to the inefficient path of quasi-fixed factors. By developing the dual problem associated with model (26), Nemoto and Goto (2003) explicitly develop the path along which optimal values for variable and quasi-fixed factors evolve.

Note that dynamic DEA models proposed by Nemoto and Goto fail to address the case where the level of capital in some of the production units does change significantly. Note that capital in each period is used both as an input and as an output. A quick way to find efficient production units in each period is to look at various ratios of a single output to a single input and

pick the ones with highest ratio (Chen and Ali, 2002). If capital level in a specific period does change significantly in all production units, then the ratio of output (capital) over input (capital) in all production units will take its highest value and all production units can potentially be efficient. It is important to discuss how this issue can affect the firms that are defined as efficient production units and how much capital of a production unit will expect to change.

Comparing the dynamic NDEA model (Equations (25)) with the model developed Nemoto and Goto (Equation (26)), it is apparent that these models are identical to a large extent. Both of these models introduce some dynamic aspects of production into DEA modeling by allowing for intermediate inputs/outputs that link consecutive periods and solving for the time path of the endogenous or dynamic variables. For example, Färe and Grosskopf (1997), use network DEA to capture the impact of the investment in non-residential capital stock in the Asian-Pacific Economic Countries where this variable is determined endogenously. Thus, they can evaluate if dynamic mis-allocation of resources of the countries can be a potential source of inefficiency. Färe and Grosskopf (1997) mix the idea of the network model and the idea of using DEA models for developing a radial measure of efficiency, but fail to separate these two concepts clearly. On the other hand, Nemoto and Goto (1999; 2003) combine network DEA with cost minimization models and develop a concrete dynamic model by explicitly describing some capital measures as the linking variables.

Note that the dynamic models developed by Sengupta and the dynamic NDEA approach by Färe and Grosskopf (1996) and the models developed by Nemoto and Goto (1999; 2003) are similar in the sense that for all of these models there are some fixed inputs and outputs over multiple periods of time and these models try to estimate the intensity parameters $\lambda(t)$ for each period concurrently. In addition, in all these models the concept of dynamic production frontier is involved when outputs depend on both variable and capital inputs, where investment in capital inputs at any period t leads to expansion in potential outputs in consecutive periods. In Sengupta's models the inter-temporal behavior is captured by the interaction among consecutive periods incorporated either in the objective function (e.g., in the form of adjustment cost due to fluctuation of inputs over time) or in the constraints of the model (e.g., in the form of relations reflecting dynamics of capital expansion), but in dynamic NDEA or in Nemoto and Goto's models the inter-temporal relation has been captured by allowing for outputs from earlier periods to be used as inputs in future periods.

Moreover, Sengupta also considers an infinite horizon in framing these dynamic models and thus develops a long term perspective. In that sense, Sengupta's models have some advantages over other studies related to dynamic efficiency such as Nemoto and Goto's where a predefined time horizon T is considered in the analysis and it is assumed that production units reach a steady state at the end of the planning horizon. Therefore, the models developed by Sengupta can handle the systems that may be affected by constant shocks and never truly reach a steady state.

Another important contribution in the literature for analyzing the firm's dynamic behavior in the context of adjustment cost and inter-temporal cost minimization belongs to Silva, Stefanou, and Choi (see Silva and Stefanou, 2003; Choi et al., 2006; Silva and Stefanou, 2007). Silva and Stefanou (2003) develop a non-parametric dynamic dual cost framework for production analysis and use adjustment costs to distinguish between variable inputs and quasi-fixed factors. Adjustment costs are imposed in the form of reduction in physical output by diverting the resources from production to investment support activities. Thus, they can be incorporated in specifying the production technology over time (Silva and Stefanou, 2007). In other words, dynamics are involved in the production technology specification as an adjustment cost in the form of a change in production possibility set with respect to the change in quasi-fixed factors. Rapid adjustment in the quasi-fixed factors leads to greater loss at the time of investment. However, an increase in the stock of capital leads to an increase in the future output by increasing the future stock of capital (Silva and Stefanou, 2003).

The non-parametric dynamic dual cost approach to production analysis developed by Silva and Stefanou (2003) requires the data series to be consistent with inter-temporal cost minimizing behavior. Thus, several non-parametric tests are developed to (i) analyze the structure of a dynamic technology (tests for constant returns to scale and homotheticity), and (ii) to check if data is fully consistent with the inter-temporal cost minimizing behavior.

Silva and Stefanou (2007) build on the developed non-parametric framework for dynamic production analysis to construct dynamic measures of technical, allocative, and economic efficiency in the short-run and long run. The long-run efficiency measures evaluate the relative efficiency of both variable and quasi-fixed factors, while the short-run efficiency measures evaluate if variable inputs have been employed efficiently in the production process (Silva and Stefanou, 2007).

Choi et al. (2006) present a dynamic cost minimization model to capture the dynamics of efficiency improvement through input reallocation. This model relaxes the assumption that the inputs of (technically and allocatively) inefficient decision making units can be adjusted along an efficient input allocation path instantaneously at no cost. Changing the inputs mix requires that the firm reorganize the technique of the production, thus imposes a transition cost to the firm (e.g., in the form of a learning cost associated with reorganization of production techniques). This transition cost increases significantly as the absolute rate of change in inputs increases; thus firms tend to follow a gradual transition toward the fully efficient input allocation path (Choi et al., 2006). In such a setting, static frameworks are incomplete for modeling how firms gradually become efficient over time (Choi et al., 2006). In fact, the important factor for dynamic specification of the developed cost minimization model is the incorporation of transition costs, since these costs flow over time lead to a dynamic linkage of consecutive periods. Starting from the initial input bundle X_0 that is allocatively and technically inefficient (necessary assumption), this model finds the next period input bundle by improving technical and allocative efficiencies of the firm. The approach thus develops a transition path (efficiency improving trajectory) that leads to the target input bundle that is both technically and allocatively efficient. The cost associated with each input bundle (each period) in this transition path is composed of the (i) transition cost as a function of the magnitude of input bundle change; (ii) cost associated with the level of technical and allocative inefficiency of the new input bundle. The dynamic cost minimization model developed by Choi et al. (2006) minimizes the discounted stream of these costs and develops a long-run cost that evolves over time with an endogenously defined rate. Note that when using a static framework, all points along the transition path are technically and allocatively inefficient, but in a dynamic sense, these points are inter-temporally efficient due to satisfying the optimality conditions for an inter-temporal cost minimization model (Choi et al., 2006). Finally, the model developed by Choi et al. (2006) precludes the effect of technical change on input allocations and only allows for time-variant technical and allocative inefficiencies. One of the important limitations of this model is that one needs to assume the functional forms of the transition cost as well as the shadow cost imposed in each period.

It is interesting to compare and contrast the transition cost concept defined by Choi et al. (2006) and the adjustment cost concept used in the investment literature. Transition costs are estimated endogenously in the dynamic cost optimization model based on the input and output

variables corresponding to production units, while the adjustment costs are estimated exogenously. For example, the adjustment cost corresponding to increasing the number of workers includes the cost corresponding to hiring and training processes. In terms of capital, the adjustment cost of installing a new piece of capital can include the cost imposed by shutting down the production line to install a new machine and to adjust it with the rest of the production process as well as the cost for training the personnel how to work with the new machines. Obviously, the adjustment cost corresponding to some of the inputs is higher than others. In moving toward an allocatively efficient point, firms may choose to adjust the inputs that have smaller adjustment costs. As a result, the transition cost allocated to those inputs may be higher than the transition cost corresponding to inputs with higher adjustment costs.

Note that all models discussed in this section try to address the issue of adjustment cost and inter-temporal decision making in an ex post context. Given a historical dataset on a group of production units over time, these models try to estimate a dynamic production technology by capturing the adjustment cost explicitly. Using a different perspective, de Mateo et al. (2006) propose a range of dynamic DEA models that explicitly incorporate the constraints related to "cost of adjustment" and "investment budget" in a linear programming framework in an ex ante context. These models try to answer the main question of how an inefficient firm should attempt to reach its optimal target (i.e., to be fully efficient). The idea is that inefficient production units cannot change the level of their inputs instantaneously to reach to their efficient target on the frontier. For example, there is a cost associated with reducing the number of workers or increasing the number of its machines. Due to budget constraints firms are facing (investment budget constraint), changes in the level of inputs usually happen through time till a firm reaches its optimal target. Assume that firm k requires to produce the fixed and known output vector y_k. Starting from initial input vector x_{k0}, then firm k tries to find the input adjustment vectors $x_{kt}^{a},(t=1,...,t_a)$ that can transform the initial input x_{k0} to the target vector x_k^* during adjustment period t_a. Adjustment period t_a defines the number of periods that a firm allows itself to reach its target and is constrained to a prefixed maximum number of periods t_a^*. By solving an inter-temporal optimization problem, the models developed by de Mateo et al. (2006) specify an optimal point (set of targets) and an optimal path of adjustment (the sequence of input quantities $x_{kt}, t=1,...,t_a$) such that net profit value of the firm over the planning horizon is maximized. Let

w be the vector of input prices, w^a be the vector of cost of adjustments, p be the vector of output prices, and b_{kt} represents the available investment budget for production unit k at time t. Then the mathematical formulation that maximizes the net present value of profit of firm k over the evaluation period T is presented as (de Mateo et al., 2006):

$$
\begin{aligned}
&\underset{x^a_{kt},\lambda_t}{Max}\ \pi_k = \sum_{t=1}^{T}[s_t(p'y_k - w'x_{kt}) - s_{t-1}w^{a'}x^a_{kt}] \\
&s.t. \quad \left.\begin{array}{l} Y\ \lambda_t \geq y_k \\ X\ \lambda_t \leq x_{kt} \end{array}\right\} \qquad 1 \leq t \leq T \\
&\qquad x_{kt} = x_{k,t-1} + x^a_{kt} \\
&\qquad w^{a'}x^a_{kt} \leq b_{kt} \\
&\qquad s_t = (1 + r/100)^{-1} \\
&\qquad \lambda_t, x_{kt}, y_k \geq 0
\end{aligned}
\qquad \textbf{(27)}
$$

This inter-temporal optimization model, in one step, finds the optimal input target as well as the optimal path of adjustment that production units should take. One of the assumptions of model (27) is that adjustment costs are imposed at the start of each period, but gross income is received and input costs are imposed at the end of each period (de Mateo et al., 2006). They further extend the basic dynamic DEA model presented by model (27) to accommodate for asymmetric cost of adjustment (where the costs associated with increasing the quantities of inputs are different from the costs associated with decreasing the quantities of inputs), non-static output quantities (when the expected output quantities of a firm may change from period to period), non-discretionary inputs that may not be changed at management's discretion, capital investment constraints, and technical change.

As discussed by de Mateo et al. (2006), several remarks regarding these dynamic DEA models are in order. First, the DEA models that are presented are among the first models to address the inter-temporal decision making in an ex ante context, meaning these dynamic DEA models are designed as ex ante management tools that not only tell the decision makers where they should be, but also specify the steps they need to take in the future to get to the optimal point. This is in contrast to most of the previous studies in which a historical panel data on a group of firms is used to estimate a dynamic production function and to study the past variations in productivity of a particular industry in an ex post context. Second, the adjustments costs for all inputs (w^a) are given in the dynamic DEA models developed by de Mateo et al. (2006), while in other studies the adjustment costs are estimated in the model estimation process (the shadow costs of adjustment are implicitly derived using historical data). Third, the optimal targets in

dynamic DEA models developed by de Mateo et al. (2006) do not necessarily lie on the frontier considering the constraints related to the adjustment costs and budget limitations. Thus potential savings that are suggested by these models are less than those that are suggested by static DEA models in which adjustment costs are assumed to be zero and optimal targets lie on the frontier. Fourth, most of the discussion and models provided by de Mateo et al. (2006) assume that the industry under analysis is not facing any technological change over time, thus the frontier and the shape of the production possibility sets are not changing over time. This implies that the type of efficiency change for a firm through time until it reaches its target is related to managerial/engineering change rather than technical change. Thus, the ex ante concept discussed in this paper is different from the ex ante concept that has been discussed by some of the previous studies (Forsund and Hjalmarsson, 1974; Forsund, 2010) where the production unit under analysis is undergoing real technological change and consequently change in the shape of its production possibility set.

4.3.4.2 Dynamic Modeling Using System Dynamics Simulation

There are also two other studies in the literature (Vaneman and Triantis, 2003; Vaneman and Triantis, 2007) with a focus on the future (ex ante context) and how a firm can identify an optimal adjustment path (transition path) to reach its optimal target. Vaneman and Triantis introduce a methodological approach that uses system dynamics techniques (Sterman, 2000) to explore productive efficiency in a dynamic environment. They start by expanding the static production axioms (e.g., input and output disposability, scarcity, convexity, etc.) into dynamic production axioms in terms of set of rules that can explain transformation of inputs into outputs in a dynamic system. These axioms are developed for three distinct classes of dynamical systems, namely: (i) dynamic and historical (systems that correlate the initial and final conditions of a system without considering any information about the structure of the system); (ii) dynamical and causal (systems that allow for the inputs to be added at any intermediate time, but there is no feedback mechanism from within the system); and (iii) dynamical, causal, and closed (systems that let the results from past actions to influence the future decisions based on a feedback mechanism).

Vaneman and Triantis (2007) build on their dynamic axiomatic framework by Vaneman and Triantis (2003) and introduce the "Dynamic Productive Efficiency" concept. Dynamic productive efficiency is a measure of a system's ability for transforming inputs into outputs at a

specific time *t* during a transient period in the life-cycle of a system. The transient period refers to the time between introducing a new disturbance into a system (e.g., introduction of a new technology or start of a performance improvement program) and the time that system seeks a new steady-state (Vaneman and Triantis, 2007). Considering that systems usually experience some temporary productivity loss during their transient periods, the Dynamic Productive Efficiency Model (DPEM) (Vaneman and Triantis, 2007) provides an approach for following the most efficient path (the optimal path) for achieving the new steady state (the optimal state). Following the optimal path leads to the minimum productivity loss during the transient period. The optimal path can be used as a benchmark for comparing with the system's actual progress during a transient period. The DPEM assumes that one can define and measure process parameters (e.g., system reliability), use of inputs as well as outputs produced during the transient period. Thus, performance of a system at any time t should be compared with the optimal path (the dynamic production frontier) at the same time period. Thus, unlike the traditional performance measurement models, DPEM does not use the ex post data to make a historical comparison among various production units (Vaneman and Triantis, 2007).

To construct the DPEM, Vaneman and Triantis (2007) start with hill-climbing system dynamics structure presented by Sterman (2000) and expand it by introducing the notion of production function to capture the optimal relationship between inputs and outputs of the system. The key reason for using the production function representation of a production environment is to make sure that DPEM computes the optimal performance of a system during a transient period. This is also a limitation of this study, because of the assumptions that are imposed by the production function functional forms. However, the assumption of incorporating a production function within this framework was relaxed by Pasupathy (2006). Estimating a well-behaved production function for most of the practical problems in the world is not an easy task. This deficiency can be addressed by using the causal relationships within system dynamics model to explicitly represent how inputs are converted into outputs. Looking inside the black box of the production process provides a unique insight to good and poor operating practices and potential sources of inefficiency (Färe and Grosskopf, 1996; Vaneman and Triantis, 2003). The DPEM can have various representations (input decreasing or output increasing). It can also accommodate production processes with multiple inputs and outputs.

The DPEM finds the optimal path based on the hill-climbing heuristic. Thus, it is not guaranteed to always provide an optimal solution (optimal combinations of inputs or outputs). The DPEM assumes that structural relations between inputs and outputs can either be represented by well-behaved production functions or by the causal relationships and input/output structure of the system. Thus, all variations can be attributed to technical inefficiencies (e.g. engineering and managerial problems) (Vaneman and Triantis, 2007). Another important characteristic of the DPEM is the introduction of adjustment time. Adjustment time that is defined exogenously is the time required to place the input or output variables into their steady state or optimal values. Thus, the period by period projections of inputs/outputs on the optimal path are defined based on the adjustment time. Note that DPEM also allows for the incorporation of adjustment cost. As illustrated in an application of the DPEM (Vaneman and Triantis, 2010), the adjustment of inputs and outputs can be defined based on costs.

As it was stated earlier, the dynamic DEA models (de Mateo et al., 2006) along with DPEM (Vaneman and Triantis, 2007) lie in the category of models that are looking to the future and determine how a firm can follow an optimal adjustment path to reach their targets. This is in contrast to the other groups of models in the efficiency measurement that use historical data to study the variations in the productivity of firms observed in the past.

It is also important to compare and contrast the DPEM model (Vaneman and Triantis, 2007) with the model developed by Choi et al. (2006). In the DPEM, the initial states of a production unit as well as a set of parameters (e.g. transition time) that defines how a production unit is allowed to change are defined exogenously. This model then develops a transition path to the desired steady state (determined endogenously) by defining the input or output levels at any step in the transition process such that efficiency loss during the transition path is minimized. This is assured by minimizing the possible inputs that are used at any step during the transition path (given a set of outputs) or by maximizing the outputs that are produced (given a fixed set of inputs). The adjustment time concept defined in the DPEM model does not allow for instantaneous change in inputs or outputs levels, thus leads to the development of the transient period that firms must pass in order to reach their steady state. This concept is similar to the discussion provided by Choi et al. (2006) that firms cannot reach their technically and allocatively efficient target simultaneously at no cost. Then, instead of considering an adjustment time, adjustment costs associated with the changes in input bundles are estimated that leads to

the development of the transition path. The model by Choi et al. (2006) endogenously defines the target point as the allocatively efficient point on the frontier that is constructed using inputs and outputs of production units in the last period. Then, the transition costs associated with the change in input bundles in each period are calculated such that the total cost of the transition to the allocatively efficient target point is minimized.

Finally, there are also some similarities between the concept of adjustment cost that Sengupta defines in Equation (24) in terms of deviations from the desired path $(z(t) = x(t) - \hat{x}(t))$ and the concept of the transition path to the desired steady state (determined endogenously) that has been discussed in DPEM by Vaneman and Triantis (2007). In both these models the adjustment path for inputs (outputs) are obtained based on the deviations from desired target levels of inputs (outputs).

4.3.5 Learning Models (Disembodied Technical Change)

This section focuses on the effects of learning models on productivity growth over time. Learning has been discussed in many industries (Benkard, 2000), such as aircraft production, shipbuilding, etc. Based on the nature of the industry, learning can take different forms. For example, in aircraft production, which is labor intensive, leaning is attributed to the improvement in proficiency by the workforce due to practice and repetition (Benkard, 2000). This type of learning is also referred to as learning-by-doing and has been shown to be an important factor justifying productivity improvement in certain industries.

Wright (1936) was the first person to develop an empirical "learning curve" or "progress function" concept for the production of military aircraft. He showed that as the cumulative output increases, the average direct man-hours per unit time decreases, but he did not provide any relationship between the developed empirical progress function and economic theory. Alchian (1963) and Rapping (1965) were among the first who linked progress functions with economic theory. More specifically, Rapping (1965) studied the effects of learning on shipbuilding output improvement per man-hour in fifteen emergency yards over a five-year period. The idea is that the stock of available knowledge (captured by accumulated or achieved output) affects the rate of output for any given levels of inputs (e.g., capital and labor). To investigate the effects of learning on the shipbuilding output, Rapping (1965) postulated the following production function:

$Y_{it} = A\ L_{it}^{\beta_1} K_{it}^{\beta_2} C_{it}^{\beta_3} V_{it}$, where C_{it} takes one of the following forms:

$$C_T = \sum_{t=0}^{T} Y_t \quad or \quad C_T = \sum_{t=0}^{T-1} Y_t + Y_{T/2} \quad or \quad C_T = \sum_{t=0}^{T-1} Y_t \tag{28}$$

where *Y* represent the annual rate of physical output, *L* represent the annual rate of physical labor input, *K* represent the annual rate of capital input, *C* represents the accumulated output, and *V* is a random noise term. Equation (28) represents a dynamic production function where the dependent variable is a function of explanatory variables as well as the lagged values of the dependent variable. Moreover, by considering the stock of knowledge (accumulated output) as one of the explanatory variable, Rapping (1965) basically adjusts the efficiency frontier to account for the presence of disembodied technical changes.

There are several studies in the literature that have augmented planning problems by incorporating learning functions (e.g., see Rosen, 1972; Womer, 1979; Gulledge and Womer, 1986). Rosen (1972) provides one of the first studies for integrating progress functions with cost theory by developing a general dynamic production function that considers learning as one of its inputs. Assuming *q(t)* represents the output rate, *x(t)* represents the composite resource use rate, and *Z(t)* represents the cumulative knowledge at time *t*, then Rosen (1972) maximizes the present value of the firm's profit over the planning horizon given the firm's production function as follows:

$$\begin{aligned} &\underset{x(t)}{Max} \sum_t (p\,q(t) - w\,x(t)) / (1+r)^t \\ &s.t.\ \ q(t) = F(x(t), Z(t)), \\ &\qquad Z(t) = Z_0 + \beta \sum_{j=0}^{t-1} q_j(t) \end{aligned} \tag{29}$$

where *p* and *w* are output and input prices, respectively, and *r* is the discount rate, and Z_0 is the initial stock of learning. By maximizing the objective function, an optimal time path for the output is defined and it can be compared with the observed time path for the output of a production unit. Womer (1979) also uses a similar model in a make-to-order production framework in order to find the optimal resource required over time given that the production unit needs to produce *V* units as output in *T* time periods. A measure of dynamic efficiency can then be defined by comparing the optimal time path of the required input with the observed (actual) time path of that input.

All previous approaches assume that learning is a consequence of production experience. Consequently, they augment a production function by considering learning as one of the inputs,

where learning is defined as a function of accumulated output. Gulledge and Womer (1986) focus on a new approach and assume that firms can explicitly assign some of their resources to learning. Thus, learning is treated as a separate output that increases the stock of knowledge at each period. The achieved stock of knowledge at the end of each period is used as an input to the production in the next or future periods. Then the inter-temporal optimization problem is defined to find the optimum resource allocation between the final output and learning (increase in the stock of knowledge) in each period, given that the production unit needs to produce V units as output in T time periods. Note that in the model by Gulledge and Womer (1986), learning in not defined as automatic result of production, but a result of the appropriate allocation of resources. Moreover, the appropriate allocation of resources at any time t affects the learning at time t and, consequently, the accumulated knowledge at the end of period t. The accumulated knowledge then affects the output rate in future periods. Thus, the inter-temporal relation among consecutive periods is maintained by using the stock of knowledge (obtained at the end of each period) as an input for producing output in future periods. This approach implements an idea similar to NDEA where the stock of knowledge is treated as an intermediate input.

Benkard (2000) with a different point of view discusses that accumulated experience may depreciate over time due presence of organizational forgetting (depreciation of production experience over time due to layoffs, turnovers, etc.) and incomplete spillovers of production experience (i.e., incomplete transfer of skills and knowledge required to produce a new model of an aircraft) for the aircraft production industry that is characterized by the presence of organizational learning. Benkard (2000) basically advances the idea that learning is not a function of accumulated output, but it is a function of the accumulated experience that is constantly depreciating over time due to both organizational forgetting and incomplete spillovers and that both of them depend on the current production output.

In sum, the models discussed in this section capture the effects of learning on productivity growth over time through two main approaches. One approach is to incorporate the stock of knowledge (commonly represented as a function of accumulated output) as one of the input variables in the firm's production function. This adjusted production function then can be used for production planning purposes (e.g., to find the optimal path for inputs given the effects of learning and also given that the production unit needs to meet a specific level of demand over time). The other approach is to define learning or the stock of knowledge as one of the outputs

produced at the end of each period. The stock of knowledge can be then used as an intermediate input to production in future periods. In this case, the inter-temporal optimization problem is defined as finding the optimum resource allocation between the final outputs and learning (increase in the stock of knowledge) in each period, given that the production unit needs to meet a specific level of demand over time. Appendix B summarizes the methods discussed in Section 4.3, and briefly describes each method, and its strengths/limitations.

4.4 Conclusions and Future Research Directions

Dynamic efficiency measurement frameworks account for an inter-temporal dependence between input consumption and output realization for production units over consecutive periods. Reviewing the literature revealed that this inter-temporal dependence for a production unit can be attributed to one or a combination of the following dynamic aspects associated with production processes, namely, (i) material and information delays; (ii) inventory (inventories of exogenous inputs, inventories of intermediate and final products, etc.); (iii) capital or generally quasi-fixed factors; (iv) adjustment costs; and (v) incremental improvement and learning models (disembodied technological change).

Each of the five stated dynamic issues associated with production has not received the same level of attention in the performance measurement literature. Most of the studies have focused on the issues of adjustment cost and capital as the sources of inter-temporal dependence. Moreover, among the studies that have focused on capital investment, a limited number of them have treated capital as a vintage specific factor. Considering that the heterogeneity of capital equipment (due to different vintages) can be an important source of efficiency differences among production units, more research on vintage models and their application in real world problems is needed.

It is also surprising that inventory as one of the important time-varying aspects of production has not received enough attention in the literature of dynamic performance. In order to capture more realistic production processes, studies should focus on various dynamic aspects of the production process instead of only one of them. Thus, developing a general framework that can integrate all five dynamic aspects of the production process in evaluating dynamic performance of a production unit will be a significant contribution to the dynamic efficiency literature. In that case, dynamic performance of production units that only contain one or a combination of the dynamic aspects of production can be easily evaluated as special cases of the

developed general framework. When one considers the discussion provided in Section 4.3 regarding the applicability of NDEA for capturing several dynamic aspects of production (e.g., adjustment cost, inventory, as well as vintage specific capitals), NDEA models are suggested as a potential technique that can be used for developing the proposed general dynamic performance measurement framework.

Finally, most of the non-parametric models presented in the literature for measuring dynamic efficiency are based on the unrealistic assumption that there is a perfect anticipation/knowledge of future variables and prices. Generalizing stochastic DEA models by incorporating dynamic aspects of the production process is a potential approach for addressing this issue so that the resulting dynamic measures of efficiency account for possible forecasting errors.

References

Alchian A. 1963. Reliability of Progress Curves in Airframe Production. Econometrica **31** (4): 679-694.

Baltagi BH, Griffin JM. 1988. A general index of technical change. The Journal of Political Economy **96** (1): 20-41.

Battese GE, Coelli T. 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. The Journal of Productivity Analysis **3**: 153-169.

Battese GE, Coelli TJ. 1995. A model for technical efficiency effects in a stochastic frontier production function for panel data. Empirical Economics **20**: 315-332.

Benkard CL. 2000. Learning and Forgetting: The Dynamics of Aircraft Production American Economic Review **90** (4): 1034-1054.

Charnes A, Clark T, Cooper WW, Golany B. 1985. A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in U.S. Air Forces. In R. Thompson and R.M. Thrall (Eds.). Annals of Operational Research **2**: 95-112.

Chen C-M. 2009. A network-DEA model with new efficiency measures to incorporate the dynamic effect in production networks. European Journal of Operational Research **194** (3): 687-699.

Chen C-M, Dalen Jv. 2010. Measuring dynamic efficiency: Theories and an integrated methodology. International Journal of Production Economics **203**: 749–760.

Chen Y, Ali AI. 2002. Output–input ratio analysis and DEA frontier. European Journal of Operational Research **142**: 476–479.

Choi O, Stefanou SE, Stokes JR. 2006. The dynamics of efficiency improving input allocation. Journal of Productivity Analysis **25** (159-171):

Cornwell C, Schmidt P, Sickles RC. 1990. Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels. Journal of Econometrics **46** (1-2): 185-200.

de Mateo F, Coelli T, O'Donnell C. 2006. Optimal paths and costs of adjustment in dynamic DEA models: with application to Chilean department stores. Annals of Operations Research **145**: 211-227.

Diewert WE. 1992a. The Measurement of Productivity. Bulletin of Economic Research **44** (3): 163-198.

Emrouznejad A, Thanassoulis E. 2005. A mathematical model for dynamic efficiency using data envelopment analysis. Applied Mathematics and Computation **160**: 363-378.

Färe R. 1986. A dynamic non-parametric measure of output efficiency Operations Research Letters **5** (2): 83-85.

Färe R. 1992. Productivity Changes in Swedish Pharmacies 1980-1989: A Non-parametric Malmquist Approach. Journal of Productivity Analysis **3** (3): 85-101.

Färe R, Grosskopf S. 1996. Intertemporal Production Frontiers: With Dynamic DEA. in: R Färe, Grosskopf S (Ed.), Kluwer Academic Publishers: Boston.

Färe R, Grosskopf S. 1997. Efficiency and productivity in rich and poor countries. Dynamics, Economic Growth, and International Trade. BS Jensen, Wong KY. Ann Arbor, MI, The University of Michigan Press.

Färe R, Grosskopf S. 2000. Network DEA. Socio-Economic Planning Science **34**: 35-49.

Fare R, Grosskopf S, Lovell CAK. 1994b. Production Frontiers. in: (Ed.), Cambridge University Press: Cambridge.

Färe R, Grosskopf S, Norris M, Zhang Z. 1994a. Productivity growth, technical progress, and efficiency change in industrialized countries. The American Economic Review **48** (1): 66-83.

Färe R, Grosskopf S, Roos P. 1996. On two definitions of productivity. Economics Letter **53**: 269-274.

Farrell MJ. 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A (General) **120** (III): 253-281.

Forsund FR. 2010. Dynamic Efficiency Measures. Department of Economics, University of Oslo Working Paper.

Forsund FR, Eilev S. 1983. Technical progress and structural change in the Norwegian primary aluminium industry. Scandinavian Journal of Economics **85** (2): 113-126.

Forsund FR, Hjalmarsson L. 1974. On the measurement of productive efficiency. The Swedish Journal of Economics **76** (2): 141-154.

Frenger P. 1992. Comment on M.D. Intriligator, productivity and embodiment of technical progress. Scandinavian Journal of Economics **94 Supplemet**: 89-93.

Gulledge TR, Womer NK. 1986. The economics of made-to-order production. in: (Ed.), New York: Springer-Verlag:

Hackman ST. 1990. An axiomatic framework of dynamic production. The Journal of Productivity Analysis **1**: 309-324.

Johansen L. 1959. Substitution versus fixed production coefficients in the theory of economic growth: A synthesis. Econometrica **27** (2): 157-176.

Johansen L. 1972. Production Functions. in: (Ed.), North Hollad Publication Co.: Amsterdam.

Kumbhakar SC. 1990. Production frontiers, panel data and time-varying technical inefficiency. Journal of Econometrics **46**: 201-212.

Nemoto J, Goto M. 1999. Nemoto-Dynamic data envelopment analysis: Modeling intertemporal behavior of a firm in the presence of productive inefficiencies. Economics Letter **64**: 51-56.

Nemoto J, Goto M. 2003. Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities. Journal of Productivity Analysis **19**: 191-210.

O'Donnell C. 2008. An aggregate quantity-price framework for measuring and decomposing productivity and profitability change. Centre for Efficiency and Productivity Analysis Working Papers WP07/2008, University of Queensland.

O'Donnell C. 2010a. Measuring and decomposing agricultural productivity and profitability change. Australian Journal of Agricultural and Resource Economics: in press.

O'Donnell C. 2010b. Nonparametric Estimates Of The Components Of Productivity And Profitability Change In U.S. Agriculture. Centre for Efficiency and Productivity Analysis Working Papers: WP02/2010, University of Queensland.

Pasupathy K. 2006. Ph.D. Industrial and Systems Engineering. Falls Chruch, VA.

Pitt M, Lee LF. 1981. The measurement and sources of technical inefficiency in Indonesian weaving industry. Journal of Development Economics **9**: 43-64.

Rapping L. 1965. Learning and World War II Production Functions. Review of Economics and Statistics 81-86.

Rosen S. 1972. Learning by experience as joint production. Quarterly Journal of Economics **86**: 366-382.

Schmidt P, Sickles R. 1984. Production frontiers and panel data. Journal of Business and Economic Statistics **2**: 367-374.

Sengupta JK. 1992a. Adjustment costs in production frontier analysis. Economic Notes **21**: 316-329.

Sengupta JK. 1992b. Non-parametric approach to dynamci efficnecy: A non-parametric aplication of cointegration to production frontiers. Applied Economics **24**: 153-159.

Sengupta JK. 1994a. Measuring dynamic efficiency under risk aversion European Journal of Operational Research **74**: 61-69.

Sengupta JK. 1994b. Evaluating dynamic efficiency by optimal control. International Journal of Systems Science **25** (8): 1337-1353.

Sengupta JK. 1996. Dynamic aspects of data envelopment analysis. Economic Notes **25** (1): 143-164.

Sengupta JK. 1999. A dynamic efficiency model using data envelopment analysis. International Journal of Production Economics **62**: 209-218.

Sickles R, Good D, Johnson R. 1986. Allocative distortions and the regulatory transition of the airline industry. Journal of Econometrics **33**: 143-163.

Silva E, Stefanou SE. 2003. Nonparametric dynamic production analysis and the theory of cost. Journal of Productivity Analysis **19**: 5–32.

Silva E, Stefanou SE. 2007. Dynamic efficiency measurement: Theory and application. American Journal of Agrecultural Economics **89** (2): 398-419.

Solow RM. 1957. Technical change and the aggregate production function. Review of Economics and Statistics **39**: 312-320.

Sterman JD. 2000. Business Dynamics: System Thinking and Modeling for a Complex World. in: (Ed.), McGraw-Hill: Boston, MA.

Vaneman WK, Triantis K. 2003. The dynamic production axioms and system dynamics behaviors: The foundation for future integration. Journal of Productivity Analysis **19**: 93-113.

Vaneman WK, Triantis K. 2007. Evaluating the productive efficiency of dynamical systems. IEEE Transactions on Engineering Management **54** (3): 600-612.

Vaneman WK, Triantis K. 2010. Application of DPEM.

Wibe S. 2008. Efficiency: A dynamic approach. International Journal of Production Economics **115**: 86–91.

Womer NK. 1979. Learning curves, production rate, and program cost. Management Science **25**: 312-319.

Wright TP. 1936. Factors Affecting the Costs of Airplanes. Journal of Aeronautical Science **3**: 122-128.

Appendix A: Comparison of Methods that Consider Time but not Inter-temporal Relations when Quantifying Change in Performance of a Firm

Category	Sub-category	References	Description	Strengths	Limitations
Measuring productivity change	Index umber approach	Diewert 1992a	Discusses the approaches that have been suggested in the literature for measuring productivity change of production units with multiple inputs and outputs where productivity is defined as the ratio of an aggregate output to an aggregate input. The Laspeyres, Paache, Fisher ideal, and Tornqvist aggregate input/output quantity and price indexes are used to construct different measures of productivity change. The axiomatic and economic approaches are discussed as ways of choosing the most appropriate functional forms for quantity and price indexes.	• No assumption regarding the functional form of the technology that transforms inputs into outputs is required. • The approach can estimate productivity change even if data on inputs and outputs of only one production unit is available, in contrast to DEA approach (as an example) that requires input/output data on a reasonable number of production units to be able to come up with any measure of performance.	• The approach requires the assumption of competitive profit maximizing behavior for production units. • It does not account for any inter-temporal dependence between inputs and outputs. Thus, it follows a comparative static framework for evaluating productivity change. • It requires data on prices for all inputs and outputs which may not easily be available. • Choosing different forms for input and output aggregator functions lead to different measures for productivity change.
	Malmquist productivity index	Caves et al. 1982	Develop the Malmquist productivity index for measuring productivity change as the ratio of distance functions of a production unit at period t+1 and t, respectively, where the distance functions can be developed with respect to the frontier for period t or t+1. To estimate the distance function, a specific functional form is assumed.	• Distance functions as the aggregator functions provide a very general representation of the production technology.	• There is no technical or allocative inefficiency in each period. • The approach requires the assumption of cost minimizing or profit maximizing behavior for production units depending on the orientation of the Malmquist index. • It requires a specific assumption about the functional form of technology. • It is not easy to develop the Malmquist index when production units produce multiple outputs. • It does not account for any inter-temporal dependence between inputs and outputs of a production unit. So, it uses a comparative static framework for evaluating productivity change.

Category	Sub-category	References	Description	Strengths	Limitations
		Färe et al. 1989; Färe 1992; Färe et al. 1994a; Fare et al. 1994b	To prevent choosing an arbitrary reference technology (frontier), the Malmquist productivity index is defined as the geometric mean of the two alternative Malmquist productivity indices defined by Caves et al., 1982. The non-parametric technical efficiency measures are used to estimate the distance function component of the Malmquist index. This productivity index is then decomposed into two components capturing "change in technical efficiency" and "shifts in frontier" over time. The efficiency change component is further decomposed into "change in scale efficiency" and "change in pure technical efficiency".	• No specific assumption is required about the form of the production function. • This index can also be used for measuring productivity change when production units produce multiple outputs. • Technical/allocative inefficiencies are allowed.	• The approach does not account for any inter-temporal dependence between inputs and outputs of a production unit. Thus, it follows a comparative static framework for evaluating productivity change. • The constant returns to scale assumption is imposed.
	Total factor productivity index	O'Donnell 2008; O'Donnell 2010a; O'Donnell 2010b	Develops a conceptual static framework for decomposing the complete productivity indexes into technical change, change in pure technical efficiency, change in scale efficiency, and change in mix efficiency. Mix efficiency measures the change in productivity if restrictions on input and output mixes are removed. The last two terms capture the changes in economies of scale and scope of the production units. The DEA methodology is used to compute and decompose two distance-based complete productivity indexes, namely Hickes-Moorsteen and Lowe productivity indices.	• This index can also be used for measuring productivity change when production units produce multiple outputs.	• It does not account for any inter-temporal dependence between inputs and outputs of a production unit. Thus, it follows a comparative static framework for evaluating productivity change. • It is limited to complete productivity indexes (TFP indexes that can be expressed as the ratio of an output quantity index to an input quantity index are referred to as multiplicatively complete.)
Measuring technical		Diewert 1980	Categorizes four approaches for measuring technical change: (i) econometric estimation of production function and cost	• These methods follow a sequential approach by constructing the production	• These approaches are based on the assumption that that inputs and outputs corresponding to period t are on the

Category	Sub-category	References	Description	Strengths	Limitations
change			function; (ii) divisia approach; (iii) exact index number approach; and (iv) nonparametric approach. - The **econometric approach** assumes a convenient functional form for production or cost function and uses the regression equations to estimate the unknown parameters of production or cost functions. Derivatives of production or cost functions with respect to time estimate the technical change over time. - The **divisia approach** uses time series data of inputs, output, and prices and assumes that production or cost functions exist and they are differentiable at any moment of time. Then the continuous derivative of the production function or cost function with respect to time is used as a measure of shifts in productivity over time. The continuous time differences of the divisia approach can be approximated by discrete differences. - The **exact index number approach** assumes that the firm's production or cost function has a specific functional form and then derives an index number formula that is consistent with the assumed functional form. - The **non-parametric approach** uses a time-series data on inputs and outputs and assumes that (i) technology sets are convex and accommodate free disposal assumption; and (ii) technology does not show any technical regress. Then non-parametric technique is used for estimating the maximum possible output for a firm at any point of time.	reference set by adding new observations into the reference set period by period. Thus, as new observations are obtained, firms never forget how they have performed before. • Advantage of the exact index number approaches is (i) that it can be implemented even if there is a large set of variables; and (ii) it leads to an exact formula that can be used for discrete data. • Advantage of the parametric and non-parametric approaches: that they provide an approximation of the true production frontier or production possibility set. • Advantage of the non-parametric approach: no assumption about the functional forms of production function is needed.	firm's period t production/cost function. Thus they assume there in no technical/allocative inefficiency. • No inter-temporal relation is assumed between consecutive periods in any one of these approaches. • These models enforce technical progress and do not allow for technical regress. • The limitation of the divisia approach is that it cannot provide a definite formulation for measuring changes in the technology, since there are various approaches for estimating the continuous time difference by discrete time difference. • The limitation of exact index number approach is that one needs to assume a specific form for the cost function (e.g., translog cost function). It is also assumed that firms follow a profit maximizing behavior. • The limitation of econometric approach is that specific functional forms should be assumed for production or cost functions. • Econometric estimation of the cost function requires the assumption of cost minimizing behavior.

Category	Sub-category	References	Description	Strengths	Limitations
		Baltagi and Griffin 1988	Develop an econometric method for estimating a general index of technical change when the underlying technology is general. This approach uses a panel data set with observations of firms in the same industry and utilizes time dummy variables to estimate econometrically a pure index of technical change.	• Relaxes the assumption that technical change has to happen at a constant rate. So allows for an econometrically rich pattern of technical change. • No restrictive assumption is needed regarding the underlying technology. • This approach enables technical effects to be decomposed into pure, non-neutral and scale-augmenting technical change.	• This approach increases the number of parameters that should be estimated. Thus, it requires the number of firms (observations) to be larger relative to the size of the time horizon under analysis to be able to use the large sample properties for estimation purpose.
Measuring technical efficiency change		Pitt and Lee 1981	Estimate a stochastic frontier production function using random effects panel data models. The estimated production frontier includes a production function of the regular regression type and an error term composed of a usual statistical noise and a nonpositive term representing the technical inefficiency. The technical inefficiency is defined to be firm-specific and time-varying. The method of maximum likelihood is then used to estimate technical inefficiency component.	• It allows for the time varying firm specific technical inefficiencies.	• A specific functional form for the production frontier is required. • Requires distributional assumptions about the technical inefficiency term. • No inter-temporal relation is assumed between inputs and output over consecutive periods/ • It does not allow for allocative inefficiencies. • It accommodates only single output firms. • It requires the technical inefficiency component to be uncorrelated with all regressors and the error component. It is difficult to justify these assumptions in most real world situations.
		Cornwell et al. 1990	Estimates a stochastic frontier production function while allowing for both cross-sectional and temporal variation in efficiency levels. Technical inefficiency term is defined as a quadratic function of time whose parameters depend on firms. The efficient instrumental variables	• Can capture productivity growth over time with different rates for different firms, meaning there are cross-sectional variations in the rate of productivity growth.	• A specific functional form for the production frontier is required. • The approach does not allow for allocative inefficiencies. • No inter-temporal relation is assumed between inputs and output over consecutive periods/

Category	Sub-category	References	Description	Strengths	Limitations
			estimation of a panel data model is used. This method allows for coefficients in addition to intercepts to vary over firms (observations).	• Does not require strong distributional assumptions about technical inefficiency or random error noise.	• The firm-specific technical inefficiencies either stay constant or are increased/decreased with a constant rate.
		Sickles et al. 1986	Evaluates efficiency growth over time by modeling a profit function that captures allocative efficiency using a flexible function of time. The efficiency component is assumed to be firm-invariant. To define a time-varying allocative inefficiency component, price distortions are defined as a function of time.	• Does not require strong distributional assumptions about technical inefficiency or random error noise.	• A specific functional form for the profit frontier is required. • It does not allow for allocative inefficiencies. • No inter-temporal relation is assumed between inputs and output over consecutive periods. • Does not capture technical inefficiency.
		Kumbhakar 1990	Uses a cost minimization framework for modeling firm-specific and time-varying technical and allocative inefficiencies. The inefficiency component is modeled as the product of a deterministic function of time and a non-negative time invariant firm effect. The functional form for the component that captures time is chosen to allow for various types of behaviors for the inefficiency term. The maximum likelihood method is used for estimation.	• Models firm-specific and time-varying technical and allocative inefficiencies.	• The need for distributional assumptions on technical and allocative inefficiencies is the main drawback of this study. • No inter-temporal relation is assumed between inputs and output over consecutive periods.
		Battese and Coelli 1992; Battese and Coelli 1995	Extend the stochastic frontier production function estimation to (i) explicitly explain the technical inefficiencies in terms of appropriate explanatory variables; and (ii) capture technical change in the stochastic frontier by explicitly incorporating a regressor that represents the year of the observations involved. The method of maximum likelihood is used to simultaneously estimate the parameters of the stochastic frontier model as well as the technical inefficiency effect model.	• Estimates both technical change and firm-specific time-varying technical inefficiency.	• A specific functional form for the production frontier is required. • The approach requires that the technical inefficiency component to be uncorrelated with the error component. • It requires distributional assumptions of technical inefficiency. • No inter-temporal relation is assumed between inputs and output over consecutive periods.
		Charnes et	Develops the window analysis framework	• When there are limited	• Uses a static framework for capturing

Category	Sub-category	References	Description	Strengths	Limitations
		al. 1985	to, given a panel dataset, evaluate performance of each production unit over time by defining it as a separate production unit in each time period. By defining window size W, the production units in the first W time periods are first evaluated using a relevant DEA model. Next, a new period is added to the window and the earliest period is dropped and the DEA model is run again over the new set of production units. This is repeated till the end of the time horizon.	number of production units in the panel dataset, window analysis improves the number of degree of freedom for efficiency rating in each DEA. • Provides the possibility for studying the trends of efficiency scores. • Analyzing each production unit's efficiency score in each time period obtained through separate DEA runs helps to examine stability of efficiency scores across and within windows.	the technical efficiency variations over time. Thus, it does not account for any inter-temporal relation between inputs and output over consecutive periods. • It drops any observation that is older than the size of the window and forgets about how firms have performed in the past. • Although it provides a larger comparison set for performing a nonparametric estimation of the technical efficiency, but there is no theoretical underpinning of why firms would forget after a specific time window.

Appendix B: Comparison of Dynamic Frameworks that Explicitly Consider Inter-temporal Dependence when Quantifying Change in Performance of a Firm

Category	Sub-category	References	Description	Strengths	Limitations
Delays (Lagged output)		Pitt and Lee 1981; Schmidt and Sickles 1984; Sickles et al. 1986; Cornwell et al. 1990; Kumbhakar 1990; Battese and Coelli 1992; Battese and Coelli 1995	One potential approach to account for the production delays when measuring technical efficiency is to add lagged variables to any of the parametric methods that use panel data to measure time-variant or time-invariant technical efficiency.	• By considering lagged dependent variable as an explanatory variable, the approach allows for inter-temporal dependencies between inputs and outputs.	• Similar limitations that hold for parametric models of production or cost function (presented in Table 2) hold for these models as well.
		Chen and Dalen 2010	Develop a methodology for incorporating lagged productive effects of input consumption into the DEA models where inputs at period t relate to the outputs of the current period t and future m periods (i.e., periods t, t+1,…,t+m). The productive effects of each specific input over current and future periods are represented by a collection of lag parameters. Then the dynamic output associated with inputs used by production unit k at period t is defined as a function of outputs over periods t,t+1,…,t+m where the relative intensity of each of the future outputs is defined by lagged parameters. Using the inputs and newly defined dynamic outputs corresponding to production units at each time period t, the dynamic efficiency of each production unit at each period is estimated.	• This framework uses the empirical panel-data and applies the panel vector autoregressive (PVAR) method to estimate the lagged parameters to estimate the dynamic efficiency scores.	• Lagged parameters are defined to be constant over all time periods and for all production units. However, if one allows for variable returns to scale, it is more realistic to let the lagged parameters to change with respect to the scale of the production or let them to vary over time. • The functional form chosen for the function that calculates the dynamic output can potentially affect the dynamic efficiency scores associated with the inputs.

Inventory	Hackman 1990	Develops an axiomatic framework for modeling a dynamic production process as a collection of interrelated production activities operating together to produce final outputs. The developed set of axioms show how to handle inventory (inventory of exogenous inputs, intermediate and final outputs) in a production process. The axioms are developed to generate the production static axioms such as no free lunch, closeness and convexity of production possibility set, etc.	• This framework allows for technical change since production function corresponding to each activity is represented by a dynamic production function that transforms time-varying inputs into time-varying outputs in a given period. This relation can change as the time period changes.	• This framework does not consider capital as one of the time-varying aspects of the production process. It is assumed that capital inputs change in a longer term. • Each activity is assumed to produce only one output. Thus, more nodes are necessary to model multiple outputs.
	Färe and Grosskopf 1996;	Present a set of Network DEA (NDEA) models (Consistent with the axioms of Hackman's framework) to explicitly capture the network of activities and intermediate products inside the transformation process. By allowing activities or production nodes to represent production in different time periods, NDEA can be used for modeling dynamic production processes. Dynamic NDEA models can be expanded by allowing for exogenous inputs to be stored (modeling inventory of inputs) for use in future periods.	• Allows for technological change throughout the time horizon under analysis by estimating the production possibility in each period using period specific inputs/outputs and intensity factors. • There are two sources of interdependence among consecutive periods: intermediate inputs and storable inputs.	• Limits the scope of their analysis to a single production process connected over time periods using the intermediate inputs. • Allows for inputs to be stored only for one period. It can also be expanded to allow the stored input to be used in multiple future periods, but computationally it will be complicated.
	Chen 2009	Uses the idea of NDEA to model production units that use several production processes to produce final outputs. The production processes at any period t are connected through intermediate inputs. The intermediate inputs produced by sub-production processes are allowed to be stored and used in future periods. This model also allows for perishability of storable intermediate inputs. After developing dynamic efficiency of each sub-process, the dynamic efficiency of a production nit is defined as multiplication of dynamic efficiency of sub-production processes.	• By explicit modeling of the internal structure of a production unit, this model allows for detecting the potential sources of inefficiency. • This model allows for perishability of storable inputs over time.	• This model does not discuss how the parameters that capture perishability of storable inputs can be estimated. • This model does not consider storability of exogenous inputs.

		Färe 1986	Uses a nonparametric approach to develop a measure of output efficiency for the whole planning horizon T (by comparing the observed total output over the time horizon T with the largest sum of potential output obtained from the model) where two types of inputs are available: one that is given at any time t and one whose total volume is given for the whole time horizon T. This model determines the allocation of the second type of inputs over finitely many time periods.	• By allowing for the second type of inputs allocated among periods, a dynamic aspect is introduced to the general non-parametric models.	• The input and output matrices that are used for estimating the production possibility set are time independent. This assumption can be relaxed by allowing for technical change (through using time-varying input/output matrices).
Capital	The vintage model	Johansen 1972; Forsund and Hjalmarsson 1974; Forsund and Eilev 1983; Forsund 2010	Focus on studying dynamic development of a firm or industry due to technological change where technology is embodied in the capital equipment and capital equipment is essential for production. Progress in the technology happens due to investment in new capital and scrapping the old capital. The vintage approach is used in these settings to explain efficiency differences among firms due to possessing different vintages of capital equipment. An optimization model is developed for deciding on the level of capital investment and the initial level of variable inputs by distinguishing between substitution possibilities between capital and variable inputs before the investment takes place (micro unit ex ante production function) and substitution possibilities after the time of the investment (micro unit ex post production function). This model maximizes the present value of net profit of investment in capital over the economic lifetime of capital.	• By treating capital as a vintage specific factor, there is no need to use replacement cost or book value of the capital as a measure of capital. • It allows for constructing homogeneous measures of capital when, for example, one accounts for the age of machines in a production unit.	• This approach has a serious ramification for nonparametric methods, since one needs to further categorize the data set (i.e., smaller datasets) to make sure the a production unit is compared only with those production units that have the same capital vintage as the firm under analysis. • No discussion is provided as to how one can implement such a structural development path for an industry by timely investment in the new technologies and disposing of old ones. • It is assumed that firms are producing in full capacity over the economic life time of capital and that capital does not depreciate over time.

Wibe 2008	Evaluates inefficiency of firms when technology evolves rapidly over time and it is uneconomical for a firm to dispose its old capital any time that a new technology is introduced to the market. To decide on optimal level of capital investment and variable inputs as well as optimal value for capital economic life time *T*, discounted profit over the economic lifetime of capital is maximized given the ex ante production function. The measure of technical efficiency is then defined as the ratio of the observed total output (i.e., the sum of outputs over all vintages from 0 to T) to the output that could potentially be produced if the latest technology at time T would be used (obtained from the ex ante production function using total capital and total labor over all vintages from 0 to T). This measure captures the inefficiency resulted from economic rationality, when technology of a firm is less efficient than best practice (rational inefficiency).	• Provides a measure of rational inefficiency that complements Farrell's measure of inefficiency. These two measures can be compared to see what percentage of inefficiency of a firm is due to its economically rational behavior and what percentage of it is due to real technical inefficiency.	• It is assumed that capital and labor are fully used (i.e., there is no slack associated with capacity) and that capital does not depreciate over time. • It is assumed that firms/industries produce maximum possible output (based on their frontier). Although, technology of the firm can be less efficient than the best practice.
Färe and Grosskopf 1996	Extend the application of NDEA models by developing a multi-vintage NDEA model that can decide on introducing new vintage technologies and the discontinuation of old ones. In this model each vintage is represented by a node in the network model. Each node uses both variable inputs and durable inputs. Plus, the total output is defined as the sum of outputs from all vintages. NDEA model decides on allocating the variable inputs among different vintages, thus provide the possibility for adopting new vintages and disposing the old vintages.	• It evaluates technical efficiency of a firm while allowing the firms to have different vintages of capital. Thus, accounts for heterogeneity of capital vintages as a potential source of efficiency differences among firms. • It allows for durable inputs to age through time and lose their efficiency. This affects the nondurable inputs that are used and the outputs that are produced at any time t greater than the investment time.	• It is assumed that all vintages in a firm produce the same outputs (i.e., there is only one type of capital equipment). This model can be expanded to allow different pieces of capital (each of which with several vintages) producing different outputs. • This model does not consider the inter-temporal nature of investment decision during a planning horizon regarding adopting new vintages and disposing the old ones.

Investment in capital	Sengupta 1994a; Sengupta 1994b; Sengupta 1995; Sengupta 1996	Develop an optimal control models by adding a dynamic aspect to the DEA framework to explicitly introduce capital inputs in addition to variable inputs. The idea is that the utilization of the capacity expanded by capital investment happens over future periods. Starting from dual LP formulation of a CCR DEA model, current levels of variable and capital inputs as well as additional investment in capital stocks are introduced into the model. It is assumed for simplicity that investment in capital input in period t yields to an increase in output in the next period t+1. Objective function of the dual model is then expanded to incorporate the additional output that has been generated to the additional capital input. This model determines optimal implicit output weights over consecutive periods.	• This model requires the resource allocation procedure to be output maximizing in periods t and t+1. This prevents the problem of having a resource allocation that is optimal in period t but not in period t+1 (the situation that happens in static frameworks).	• This model only accounts for inter-temporal relation between capital investment and capacity expansion over two consecutive periods. A more general model should expand the inter-temporal relations over multiple periods.
	Emrouznejad and Thanassoulis 2005	Develop a DEA model to capture the effect of capital investment on future outputs. To account for the correspondence between inputs and outputs observed in different period, an assessment path is defined to reflect all inputs used by a production unit and all outputs produced irrespective of the time lag between inputs and outputs. Also, to account for the cases in which a firm starts with higher level of capital at the start of analysis horizon (or assessment window) or ends up with higher level of capital stock at the end of the analysis horizon (T), the level of capital stock at the start period is defined as an input and its level at the end of period T is defined as an output. Each period also has variable inputs and investment in capital (as two input) and produce final outputs.	• This model evaluates each production unit over the whole assessment window. Thus, accounts for inter-temporal dependence between inputs and outputs due to capital investment. • This model allows for inter-temporal effects of capital investment on for outputs over multiple future periods (a more general case). • This model accounts for aging of capital by allowing for increase in variable inputs that are used and decrease in outputs that are produced at any time t greater than the investment time.	• Size of the assessment window should be chosen such that it gives a reasonable approximation of the correspondence on the inputs and outputs of production units. Such information may not easily be obtained in practice. • By using input-output paths instead of single period input-output data, the dimension of the production possibility set increases multiplicatively as number of time periods or input and outputs increases. This has serious consequences on the discriminating power of the DEA models.

Analytical Approach	Sengupta 1994b	Develops an optimal control model by generalizing the concept of Farrell efficiency in a dynamic framework where inter-temporal changes of capital inputs and their corresponding adjustment costs are taken into account. The adjustment costs associated with investment in capital is imposed in the form of foregone output (the more investment in capital goods, the more the lost output). By maximizing the present value of net output minus adjustment cost given the dynamics of capital stock (due to both investment in capital and its depreciation over time), a dynamic model for defining the optimal trajectory for capital investment is developed.	• This model accounts for depreciation of capital through an equation that captures the dynamics of the capital stock (i.e., the effects of investment in capital and its depreciation at any time t). • An infinite time horizon is considered in framing the dynamic models. Thus this model develops a long term perspective in contrast to models that consider a predefined time horizon T. • By allowing for inputs, outputs, and intensity variables to change over time, this model can capture technological changes over time.	• The explicit formulation of the adjustment cost function has not been defined. Instead of introducing adjustment cost in the objective function (to reduce the value of net output), one can modify the production possibility set to allow for a decrease and then an increase in maximum possible output at the time of investment and future periods, respectively.
	Sengupta 1999 Sengupta 1996	Develops an optimal control model to capture the dynamics of production technology when there is an inter-temporal adjustment cost functions associated with deviation of inputs (both variable and capital inputs) or outputs from their target levels. The adjustment cost is composed of two components: one represents the cost of fluctuations in inputs or outputs through time; and the other captures the cost of deviation from the desired target (disequilibrium cost). Given the desired target level of inputs or outputs, the inter-temporal optimization problem that minimizes the expected present value of a quadratic loss function for a production unit is developed. By solving this inter-temporal optimization model, the optimal paths of inputs or outputs are characterized.	• An infinite time horizon is considered in framing the dynamic models. Thus, this model develops a long term perspective and can handle the systems that may be affected by constant shocks and never truly reach a steady state.	• The target level of inputs or outputs is defined exogenously through historical data. • The optimal paths of inputs (outputs) among other factors depend on the input (output) prices over future periods and the input (output) target levels. Any error associated with these factors can affect the validity of the optimal solution. • Developments of the weight matrices in the objective function that have been assigned to the two components of adjustment cost have not been explained.

Adjustment cost	Sengupta 1994a Sengupta 1992a	Uses an inter-temporal framework for measuring efficiency of risk averse firms when there is stochastic errors associated with input and output data and their prices. The risk attitudes of firms and the effects of output fluctuations are incorporated into the dual version of the DEA model in the form of adjustment costs to develop an optimal control model and define an optimal inter-temporal adjustment path for output toward a desired target.	• Accounts for the stochastic noise associated with inputs output data and their corresponding prices. Thus, allows risk averse firms to follow a cautious policy to get to a more efficient production frontier.	• Developments of the weight matrices in the objective function that have been assigned to the two components of adjustment cost have not been explained.
	Sengupta 1992b	Extends the ideas of cointegration and dynamic econometric modeling to the non-parametric methods to estimate the production and cost functions where the input-output data are non-stationary over time. The dynamic production frontier incorporates the adjustment cost as a function of the change in capital input. The cost frontier incorporates the adjustment cost as a quadratic function of the deviation from the desired levels of outputs. The error terms in both frontiers are set to be one-sided and the error structures are directly incorporated in the constraints of the linear programming models. The estimations are based on the LVA procedure that minimizes the least sum of absolute value of error terms.	• Addresses the econometric issues related to estimation of dynamic production/cost functions when input and output data are non-stationary over time.	• The target level of inputs or outputs is defined exogenously through historical data.

Färe and Grosskopf 1996	Dynamic NDEA is extended to model investment in capital. At each period t investment in capital is modeled as a decision factor that is determined endogenously and affects the stock of capital at the end of that period. The stock of capital is then treated as an intermediate input at period t+1 and affects the production possibilities in future periods by expanding the production capacity. Moreover, given the input levels at each period, increase in investment in capital at each period t decreases the final output produced in the same period. In this model the dynamic or time interdependence among consecutive periods arise from endogenous intermediate inputs (capital levels) that are computed simultaneously for all time periods.	• Introduce some dynamic aspects of production into DEA modeling by allowing for intermediate inputs/outputs that link consecutive periods and solve for the time path of the endogenous or dynamic variables. • Allow for analyzing the investment path of a production unit as one of the potential sources of inefficiencies. • Can be solved using linear programming, thus there is no need to use optimal control, dynamic programming, etc. • By allowing inputs, outputs, and intensity parameters to change in each period, this model allows for technical change.	• Dynamic NDEA model does not necessarily cover the capital deterioration concept popular in capital investment literature where capital deteriorates over time. • Mixes the idea of the network model and the idea of using DEA models for developing a radial measure of efficiency, but fails to separate these two concepts clearly. • A predefined time horizon T is assumed in the analysis. It is also assumed that production units reach a steady state at the end of the planning horizon.
Nemoto and Goto 1996; Nemoto and Goto 2003	Use network DEA to develop a dynamic framework for modeling the investment behavior when a firm cannot adjust the levels of its quasi-fixed inputs instantaneously with no adjustment cost. Adjustment cost is imposed in terms of giving up some of the production of final goods at period t when it is necessary to increase the level of quasi-fixed capitals at the end of that period (or at the beginning of period t+1). Due to an increase in the level of quasi-fixed inputs at the beginning of period t+1, the level of final goods in future periods increases. To capture this resource allocation problem in DEA, the quasi-fixed factors at the end of each period are treated as outputs in that period that are placed into next period's production process as inputs. The developed model finds the optimal path for both variable and quasi-fixed inputs over the planning horizon T by constructing the inter-temporal efficient cost frontier.	• Introduce some dynamic aspects of production into DEA modeling by allowing for intermediate inputs/outputs that link consecutive periods and solving for the time path of the endogenous or dynamic variables. • Combine network DEA with cost minimization models and develop a concrete dynamic model by explicitly describing some capital measures as the linking variables. • By allowing inputs, outputs, and intensity parameters to change in each period, this model allows for technical change.	• Fail to address the circumstances where the level of capital in some of the production units is not going to change significantly, considering that capital in each period is used both as an input and as an output. It is important to discuss how this issue can affect the firms that are defined as efficient production units and how much capital of a production unit is expected to change. • A predefined time horizon T is assumed in the analysis. It is also assumed that production units reach a steady state at the end of the planning horizon.

Silva and Stefanou 2007	Uses the non-parametric dynamic dual cost approach to construct dynamic measures of technical, allocative, and economic efficiency in the short-run and long run. The long-run efficiency measures evaluate the relative efficiency of both variable and quasi-fixed factors, while the short-run efficiency measures evaluate if variable inputs have been employed efficiently in the production process.	• This model develops lower and upper bounds for efficiency measure of each production unit at each point of time.	• Data should be fully consistent with the inter-temporal cost minimization behavior. • Implementation of the model is very complicated.
Choi et al. 2006	Present a dynamic cost minimization model to capture the dynamics of efficiency improvement through input reallocation. This model relaxes the assumption that (technically and allocatively) inefficient decision making units can be adjusted along an efficient input allocation path instantly without imposing any transition cost. Starting from an initial input bundle that is allocatively and technically inefficient (necessary assumption), this model finds the next period input bundle by improving technical and allocative efficiencies of the firm. The cost associated with each input bundle (each period) in this transition path is composed of the (i) transition cost as a function of the magnitude of input bundle change; (ii) cost associated with the level of technical and allocative inefficiency of the new input bundle. The dynamic cost minimization model then minimizes the discounted stream of these costs at each period and treats the long-run cost as a stock that evolves over time with an endogenously defined rate	• The important factor for dynamic specification of the developed cost minimization model is the incorporation of transition costs, since these costs flow over time and lead to dynamic linkage of consecutive periods.	• This model precludes the effect of technical change on input allocation and only allows for time-variant technical and allocative inefficiencies. • One needs to assume the functional forms of the transition cost as well as the shadow cost imposed in each period.

	de Mateo et al. 2006	Propose a range of dynamic DEA models that explicitly incorporate the constraints related to the "cost of adjustment" and "investment budget" in a linear programming framework in an ex ante context. The idea is that inefficient production units cannot change the level of their inputs instantly to reach to their efficient target on the frontier due to the cost associated with changing inputs (e.g. labors or machines). By solving an inter-temporal optimization problem, these models specify an optimal point (target) and an optimal path of adjustment (the sequence of input quantities) such that net profit value of the firm over the planning horizon is maximized.	• These DEA models are among the first models designed as ex ante management tools that not only to inform the decision makers where they should be, but to also specify the steps they need to take in the future to get to the optimal point. • The optimal target does not necessarily lie on the frontier given the constraints related to adjustment costs and budget limitations. Thus potential savings that are suggested by these models are more realistic and less than those suggested by static DEA models in which adjustment costs are assumed to be zero and optimal targets lie on the frontier.	• The adjustments costs for all inputs are given in these dynamic DEA models, while in other studies the adjustment costs are estimated in the model estimation process (the shadow costs of adjustment are implicitly derived using historical data). • It is assumed that industry under analysis is not facing any technological change over time, thus the frontier and the shape of the production possibility sets are not changing over time.
System Dynamics Simulation	Vaneman and Triantis 2003; Vaneman and Triantis 2007	Introduce the "Dynamic Productive Efficiency" concept as a measure of a system's ability for transforming inputs into outputs at a specific time t during a transient period in the life-cycle of a system. The transient period refers to the time between introducing a new disturbance into a system and the time that system seeks a new steady-state. Dynamic Productive Efficiency Model (DPEM) uses system dynamics simulation and hill-climbing heuristic to find the most efficient path for achieving the new steady state that is defined endogenously. Following the optimal path leads to the minimum productivity loss during the transient period. The optimal path can be used as a benchmark for comparison with the system's actual progress during a transient period.	• DPEM does not use the ex post data to make a historical comparison among various production units. Thus, it can be used as an ex ante management tool to tell the decision makers where they should be and how they can get to the optimal point. • The causal relationships within system dynamics model can be used to explicitly represent how inputs are converted into outputs where estimating a well-behaved production function is not an easy task. • Looking inside the black box of the production process provides a unique insight to good and poor operating practices and potential sources of inefficiency.	• The DPEM finds the optimal path based on the hill-climbing heuristic. Thus, it is not guaranteed to always provide an optimal solution (optimal combinations of inputs or outputs). • The DPEM assumes that structural relations between inputs and outputs can either be represented by well-behaved production functions or by the causal relationships and input/output structure of the system. Thus, all variations are attributed to technical inefficiencies. • Adjustment time/cost are defined exogenously.

Learning models (disembodied technical change)	Wright 1936	Develops an empirical “learning curve” or “progress function” concept for the production of military aircraft by showing that as the average direct man-hours F per unit time decreases. The slope in this relation captures learning or progress and the cumulative output N captures past experience. The idea is that learning-by-doing (due to increases in cumulative production) can decrease the unit cost of military aircrafts.	• This estimation led to the widely accepted "80-percent learning curve" in the airctaft industry, meaning that a production process with a learning curve that has a slope of 80% needs 80% of its resources for every doubling of the cumulative past production.	• Did not provide any relationship between the developed empirical progress function and economic theory
	Alchian 1963	Links progress functions with economic theory to captures the relationship between direct man hours per pound of airframe used to produce a specific type of airframe and the number of airframes of that specific type produced in a production facility. The model in general implies that as more airframes are produced, the required number of direct man-hours decreases.	• An attempt to contrast progress function with economic theory by analyzing the nature of the progress functions (in the context of the relationship between labor requirements and production volume).	• Not explicit measure of dynamic efficiency is discussed.
	Rapping 1965	Studies the effects of learning on shipbuilding output improvement per man-hour during WWII. The idea is that accumulated or achieved output affects the rate of output for any given level of inputs (e.g., capital and labor), since management and labor learn or progress as production experience accumulates on a particular vessel type (a particular output mix). Accumulated output (representing the stock of available knowledge) is then used as a regressor in addition to capital and labor to estimate the output improvement per man-hour using regression models.	• An attempt to contrast progress function with economic theory by analyzing the nature of the progress functions (in the context of the relationship between labor requirements and production volume). • Used panel data and showed that learning can be different for different groups of products. • Adjusts the efficiency frontier to account for the presence of disembodied technical changes.	• A predefined functional form for the production function is needed to define how learning can potentially affect the outputs.

Rosen 1972	Provides one of the first studies for estimating a dynamic production function where output rate at time t is modeled as a function of composite resource use rate and cumulative knowledge at time t. Plus, learning is also modeled as a linear function of current output. Then the present value of the firm's profit over the planning horizon is maximized given the following dynamic production function. By maximizing the objective function, an optimal time path for the output at each period t is defined.	• One of the first studies providing a general dynamic production function that considers learning as one of its inputs.	• A predefined functional form for the production function is needed to define how learning can potentially affect the outputs.
Womer 1979	Assuming that a make-to-order production framework needs to produce *V* units as output in *T* time periods, the total discounted cost of required resources is minimized given a dynamic production function where output rate at time t is a function of resource use rate and cumulative output at time t (representing the stock of knowledge) and assuming that resource prices are constant. This is an optimal control problem and can be solved to define the optimal resource required at any point in time.	• Integrates progress functions with cost functions in a make-to-order production framework where firms need to produce *V* units as output in T time periods.	• This model uses the unrealistic assumption that input prices do not change over time.

	Gulledge and Womer 1986	Assuming that firms can explicitly assign some of their resources to learning, learning is treated as a separate output that increases the stock of knowledge at each period. The achieved stock of knowledge at the end of each period is used as an input to the production in the next or future periods. They consider the situation in which the production rate (or time path of output) as well as the rate of change of knowledge (or time path of learning) can be modeled as functions of composite resources and cumulative knowledge. Assuming that *V* units of output should be produced in time horizon *T* and the relative prices of inputs do not change, the objective is to minimize the production cost given the production functions (capturing both production rate and the rate of change of knowledge).	• Learning in this model is not an automatic result of production, but a result of the appropriate allocation of resources. • Learning in each period increases the stock of knowledge at the end of that period. Stock of knowledge is used as an intermediate input between consecutive periods and introduces an inter-temporal relation among consecutive periods (an idea similar to NDEA).	• The representation and measurement of the stock of knowledge is not discussed. The same issue holds for the amount of learning that is taking place in each period.
	Benkard, 2000	Implements that idea that layoffs and turnover of workforce lead to the loss of the accumulated experience. Plus, the incomplete experience affects firms when they decide to produce a new model of a product. This results in a setback in learning and a higher production cost for the overall aircraft production. Thus, experience at any point of time is modeled by incorporating both organizational forgetting and incomplete spillover across models. Then assuming that production frontier takes a Cobb-Douglas form, the current production rate is modeled as a function of variable input labor, fixed capital, and the firm's experience.	• One of the first studies to incorporate organizational forgetting and incomplete spillovers of production experience in estimating a firm's experience for the aircraft industry.	• Not explicit measure of dynamic efficiency is discussed.

Chapter 5 Conclusions

5.1 Summary of the Research and Major Findings

With the construction of highway systems mainly completed, the preservation and maintenance of the existing road infrastructure is becoming much more critical (Ozbek, 2007). In particular, road authorities are being challenged to improve the performance of their highway maintenance policies and practices to preserve a safe, reliable, and efficient road infrastructure that can support society's needs (TRB, 2006). Successful improvement of existing maintenance policies as well as successful implementation of new ones requires state DOTs to measure the performance of their highway maintenance practices. Given this scenario, implementing performance monitoring tools to evaluate the efficiency and effectiveness of highway maintenance operations as well as road authorities (e.g., State Departments of Transportation, districts, or counties) who are responsible for the maintenance of the road in their administrative area has significant benefits. In a performance measurement system, the effectiveness dimension focuses on the achieved level of service and the efficiency dimension focuses on the amount of resources consumed to achieve a given level of service. Within this context, this research has focused on the development of multiple frameworks that can evaluate the cost *efficiency* of road maintenance operations in relation to their achieved level-of-service (*effectiveness*).

In the first research thrust, recent developments on non-parametric frontier estimation (a combination of the meta-frontier approach as well as the two-stage bootstrapping technique) were utilized to develop a five-stage analytical approach for evaluating and comparing the performance of "performance-based" highway maintenance contracting versus "traditional" contracting with a focus on the fundamental relation between the maintenance level of service (effectiveness) and maintenance expenditures (efficiency). The meta-frontier approach accounts for the heterogeneity that exists among different types of highway maintenance contracts due to different limitations and regulations. The two-stage bootstrapping technique accounts for the large set of uncontrollable (environmental and operational) factors that affect the highway deterioration and maintenance processes. It also enables statistical analysis of the estimated efficiency scores, such as the correction for the bias and construction of the confidence interval for the efficiency scores.

The preliminary findings, based on the historical data of 180 miles of Virginia's Interstate highways maintained by Virginia Department of Transportation (VDOT) using traditional maintenance practices and 250 miles of Virginia's Interstate highways maintained using a performance-based maintenance approach, suggest that road authorities (i.e., counties) that have used traditional contracting have been more efficient than road authorities that have used performance-based contracting (PBC). These findings may be explained by the fact that the pilot project performed via performance-based contracting has been VDOT's first experience with performance-based contracting, while traditional contracting has been used many times. Thus, it is very possible that VDOT's shift to the needed PBC culture had not been fully developed during the execution of this first performance-based maintenance pilot project. Another plausible explanation is that the road-builder contracting industry has not been used to think in terms of "life-cycle" costing models (a performance-based contracting characteristic), which are radically different from "first cost" models (a traditional contracting characteristic). This research recommends that road authorities use hybrid contracting approaches that include best practices from both traditional and performance-based highway maintenance contracting.

In the second research thrust, the focus was placed on using an "engineering approach" (i.e., focusing on the transformation process inside the black box for a better understanding of the sources of inefficiencies) to find the superior blueprint that can be used as benchmark for evaluating the performance of highway maintenance contracting. To this end, a dynamic micro-level simulation model of highway deterioration and renewal processes was developed. This model was calibrated using empirical data to estimate the unknown parameters of the model. The calibrated model was coupled with an optimization module to find the best policy for allocating the maintenance budget to different maintenance operations.

The policies that were found through optimization pointed to alternative priorities where preventive maintenance is preferred over corrective maintenance. However, the overlap in the priority of preventive and corrective maintenance operations increases as the total maintenance budget decreases. This shows the need for sharing the budget between preventive maintenance and corrective maintenance rather than satisfying preventive maintenance first and then allocating leftover resources to corrective maintenance. The extent of this contribution is however limited by the small improvements that were found to be feasible.

The outcome from this research track provides a blueprint for designing optimal highway maintenance practices. There are two important considerations regarding the developed optimum maintenance policy: (i) the optimum policy depends on the objective of the road authorities, for example, minimizing the area under distress in the highway network in their administrative area; and (ii) the optimum policy covers a time horizon under analysis, thus takes into account the fact that the utilization of the maintenance budget and the treatment that is performed in a road section in a specific year directly affects the road condition and required maintenance operations in consecutive years. Thus, the optimum maintenance policy leads to an optimal development path for the road condition over time. This optimal development path can be used as a benchmark to develop a measure of dynamic efficiency where the inter-temporal dependence between inputs and outputs are explicitly taken into account.

To obtain a better understanding of the concept of dynamic efficiency as well as the related models/approaches, in the third research thrust, an overview of the most relevant studies in the dynamic efficiency literature was developed. This literature review proposed a classification taxonomy for dynamic performance measurement frameworks according to five issues (i.e., production delays, inventories, capital, adjustment costs, and learning effects).

Building on the insights obtained from this classification taxonomy as well as from further exploration of the dynamics of road conditions revealed that there is a time lag between spending maintenance expenditures (i.e., inputs) and its effects on road condition (i.e., output) as well as on the required maintenance operations/budget in consecutive periods. That is, the maintenance expenditure spent in each year affects the road condition in the same year and also affects the transformation process in the consecutive years. Such lag spans over several time periods, thus makes it difficult to find an explicit relationship between maintenance spending and changes in road condition over time. Based on these characteristics, the source of inter-temporal dependence between inputs and outputs in highway maintenance operations can be attributed to the inherent "delay" that exists in the highway deterioration/renewal transformation processes. In such a setting, there is a need for a dynamic efficiency measurement framework that accounts for the inter-temporal dependence between inputs and outputs while evaluating performance of highway maintenance operations. The next section provides a description of some possible future research areas that build on the finding of this research.

5.2 Areas of Further Research

5.2.1 Developing a Framework for Evaluating Highway Maintenance Utilizing the Concept of Dynamic Efficiency

The idea of developing a framework for measuring the dynamic efficiency of highway maintenance operations aims at capturing the time interdependence between inputs and outputs while evaluating performance of highway maintenance operations. Based on the previous discussions about the dynamics of the road condition, the condition/state of a highway network (or the level of the variable that represents the road condition) at any period t depends on the state of the highway network at the end of period t-1, the exogenous inputs (maintenance budget and deterioration factors) in period t, and production/policy factors in period t. Equation (30) shows this relation. In this equation, *F* represents the function that governs the dynamics of the road condition.

State of the road (t) = F(Exogenous Inputs (t), Policy Factors (t), State of the road (t-1)) **(30)**

When a variable (e.g., condition of the road) changes dynamically over time, the main interest lies in evaluating the "development over time" of the variable. A key requirement for this evaluation is to have a benchmark to evaluate the path of development of that variable (Vaneman and Triantis, 2007; Forsund, 2010). As a result of this evaluation/comparison, one can develop a dynamic measure of efficiency of the road authorities in spending the limited available maintenance budget. The measure of efficiency in such a dynamic setting needs to be a function of the level of the efficiency in previous periods. To this end, the two important steps that need to be considered are (i) the "development of a benchmark"; and (ii) the development of the measure of dynamic efficiency.

Characterizing the benchmark requires some judgments about particular objectives of the production process (Forsund and Hjalmarsson, 1974). In the highway maintenance context, the main objective of the maintenance operations is to improve the condition of the roads (as the main output of the maintenance operations) to its best possible condition (or to a condition that meets the required level of service) given a maintenance budget. Building on the System Dynamics research thrust in this research (second essay), the benchmark is constructed using the micro-level system dynamics model of road deterioration and maintenance. Given specific budget constraints as well as operational conditions (climate condition, traffic load, etc.), the

optimization module of the micro-level system dynamics model will lead to the optimum maintenance budget allocation policy. This will lead to the optimal path for the road condition development in the time horizon under analysis. This is a key step, since the fundamental dynamics are involved when one solves for the time path of the variable under analysis (i.e., road condition). Moreover, through optimization process one can explicitly link maintenance activities to performance goals over time.

As stated before, the main objective of the road authorities is to maintain the highway network in their administrative area in the best possible condition given the limited available maintenance budget. This objective can be translated to minimizing the area of the road sections that are under distress given the available maintenance budget and uncontrollable environmental condition (traffic load and climate condition). In this optimization problem, the decision variables are related to the maintenance policies that govern how the limited available budget should be allocated between the road sections that are in need of maintenance. Let $A_i(t)$ represents the area under distress in road section $i = 1,...,N$ at time *t*, $I(t) = (X(t), U(t))$ represents the vector of exogenous inputs at time *t* which is composed of the available maintenance budget $X(t)$ at the network level (i.e., total budget available for the road authority) and the uncontrollable environmental condition $U(t)$ that a road authority is facing, and P^t represents the road authority's dynamic maintenance policy. Assume that condition or state of road section *i* can be defined by the level of distress on that road section. Then, the optimization model can be defined as the minimization of the total area under distress at the highway network given the dynamics of the condition of the road sections defined based on Equation (30) as follows:

$$\begin{aligned} &\underset{P^T}{Min} \int_{t=0}^{T} \sum_{i=1}^{N} A_i(t)\,dt \\ &s.t.\ A_i(t) = F(I(t), P^T, A_i(t-1)) \quad \forall i \end{aligned} \tag{31}$$

The function *F* in this framework represents the dynamics of the road deterioration/renewal processes captured by the micro-level system dynamics model. By solving this optimization model using the system dynamics framework, the optimal maintenance policy P^{T*} and, consequently, the optimal trajectories $A_i^*(t), i = 1,...,N$ for a dynamically efficient road authority

are obtained. A weighted average of the road sections' optimal trajectories (i.e., $A_i^*(t)$) gives the optimal development path (optimal trajectory) for the whole highway network under analysis, where the lengths of road sections can be used as the weights factors. Note that optimal policy P^{T*} is dynamic since it changes as the time horizon *T* changes. In other words, P^{T*} maps an array of exogenous inputs *I(t)* to an array of optimal road conditions $A_i^*(t)$. As time horizon *T* changes, the corresponding arrays change, and an alternative estimation for the optimal policy P^{T*} is obtained. The estimated policy P^{T*} is dynamically optimum in that sense and it maintains the inter-temporal relation between inputs and outputs over the assessment window by performing the optimization over *T*. These policies may not be optimal if one looks at snapshots at each time period, but they are dynamically optimal.

The final step of the dynamic performance measurement framework involves the development of the measure of dynamic efficiency given the constructed benchmark (optimal path for the development of the road condition). The measure of efficiency has to have several characteristics, in particular: (i) it is a function of the gap between the benchmark and the actual road condition that has been observed by the road authority under analysis; and (ii) it is a function of time since it is a dynamic measure of efficiency. Let *A(t)* be the actual condition of the highway network observed over time. If a road authority does not follow the optimal maintenance policy P^{T*}, the observed road condition *A(t)* deviates from the optimal trajectory $A^*(t)$. The gap between optimal and observed road condition over time (i.e., $\int_{t=0}^{T} |A(t) - A^*(t)|$) can be used as the measure of dynamic efficiency of the road authority when a finite planning horizon *T* is used. An alternative measure of dynamic efficiency of a road authority can be developed as $\int_{t=0}^{T} A^*(t) \Big/ \int_{t=0}^{T} A(t)$. Obviously, it is expected that the total amount of distress on the highway network over time under optimal maintenance policy to be less than or equal to the total observed distress on the highway network. This second measure of dynamic efficiency changes between zero and one where one represents that the road authority under analysis has followed the optimal maintenance policy and has maintained the highway network in the best possible condition given the available maintenance budget.

Note that the optimization model presented in Equation (31) captures the inter-temporal dependence between inputs and outputs in consecutive periods through the first order dynamic

constraint representing the evolution of the state of the system and also through the objective function. Moreover, the developed optimization model is an optimal control model that deals with the problem of finding a control policy for a given system such that a certain optimality criterion is achieved. A typical control problem includes an objective function (e.g., a cost function) that is a function of state and control variables. Optimal control variables then are defined as set of differential equations describing the paths of the control variables that optimizes the objective function. In the case of Equation (31), $A_i(t)$ represents the state variables and P^T represents the control variable(s). As it can be seen, the objective function is also a function of the state variables. Assuming an explicit formulation for the function *F*, the optimal control P^{T*} can then be developed analytically.

In sum, the framework for evaluating dynamic efficiency of highway maintenance operations combines dynamic simulation and optimization to evaluate and improve the performance of road maintenance policies. To achieve this objective the following steps need to be performed:

1) A simulation model of the road deterioration process is built using the causal relationships and feedback mechanisms available in the system dynamics modeling paradigm. This allows us to explicitly link maintenance activities, to uncontrollable drivers and performance goals over time.

2) This model is calibrated with data from real road operations and maintenance to estimate model parameters.

3) The calibrated model is then used for policy analysis through an optimization approach where the pavement deterioration is coupled with the highway maintenance and renewal decision-making process that decides the allocation of a limited budget to maintenance operations.

4) An optimization module is developed and applied to find the best policy (benchmark) for a budget allocation for a given set of environmental and operational conditions.

5) The optimal path for development of the road condition over time obtained through optimization process is used as a benchmark to develop the measure of dynamic efficiency performance of the current maintenance practices.

5.2.2 Modifying the Meta-frontier Framework for Evaluating the Efficiency of Heterogeneous Production Units

The non-parametric meta-frontier has been extensively used in the literature for evaluating the efficiency of heterogeneous production units. The idea is that a separate frontier can be created for each group of homogeneous production units. The group frontiers are then compared through the development of meta-technology ratios (O'Donnell et al., 2008). Meta-technology ratios evaluate how a group frontier is located with respect to the meta-frontier where the production units are not limited to the production technology of their own group. There are two important shortcomings with this approach. The first shortcoming is that the development of the meta-frontier is based on the assumption that the convex combination of two production units that belong to two different groups (production technologies) is technologically possible and that the virtual production unit belongs to the production possibility set that is developed by the meta-frontier. The second shortcoming is that the meta-technology ratio only evaluates technical efficiency of production units (i.e., the amount of increase in the output given the same amount of input that has already been used) while comparing different groups and no attention is paid to the scale inefficiencies that can be imposed by different group frontiers. For a better understanding, Figure 5-1 provides a schematic view of the output-oriented technical as well as scale efficiency of production unit *A* captured by its aggregate input and output.

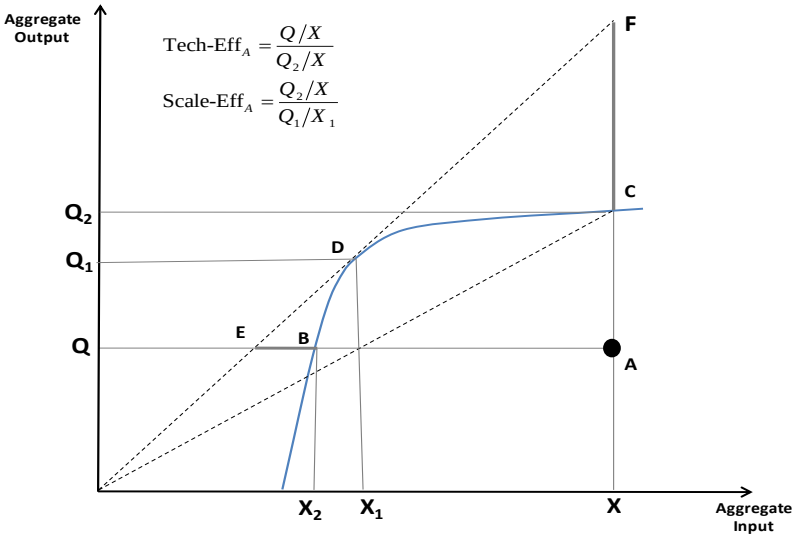


Figure 5-1: Output-Oriented Technical and Scale Efficiencies (O'Donnell, 2008)

In this figure, the curved line that passes through point *C* defines a variable-returns-to-scale production frontier. Note that by increasing the output from *Q* to Q_2 (i.e., the technically efficient

point) the total factor productivity (TFP) of the production unit *A* increases, where total factor productivity represents the ratio of an aggregate output to an aggregate input. If the mix of inputs and outputs are held fixed, the TFP of production unit *A* would be maximized by moving to point *D*, where point *D* represents the optimal scale of production given the production technology defined by the frontier. Point *D* gives the max TFP that can be achieved and is defined as the tangent of the variable-return-to-scale frontier and the constant-returns-to-scale frontier (which is represented by a ray that passes through origin). The meta-frontier framework can be modified to take into account both technical and scale inefficiency of production units in each group to develop a more comprehensive measure for evaluating performance of different groups (production technologies).

5.2.3 Expanding the Micro-level Simulation Model of Highway Deterioration/Renewal for Macro-level Analysis

Improvement programs and maintenance policies for road infrastructure not only affect the existing condition of the highway system, but also affect the environmental, societal, and economic factors. Thus, the micro-level simulation model presented in Chapter 3 can be augmented to establish a comprehensive set of utility functions that incorporates the effects of the maintenance policies and practices on improvement of deteriorated roads as well as on the direct user benefits and non-user benefits. The direct user benefit module could potentially capture the effect of the road deterioration on vehicle operating costs, travel time costs, safety, and accident costs of motorists. The non-user benefits capture the economic effects of the road condition on the community and include increase in economic activities and business opportunities in that area, the effects on land value as well as the population living in that area, etc. Building on the micro-level simulation model of highway deterioration/renewal process and connecting it to macro-level highway maintenance policies (e.g., non-user benefits) could potentially lead to alternative resource allocation policies and can be defined as an important area for further research.

Bibliography

ADOT. 2010. Traffic Data by Arizona Department of Transportation, http://tpd.az.gov/data/aadt.php.

Alchian A. 1963. Reliability of Progress Curves in Airframe Production. Econometrica **31** (4): 679-694.

Anastasopoulos PC, McCullouch BG, Gkritza K, Mannering FL, Kumares SC. 2009. A Cost Saving Analysis for Performance-based Contracts For Highway Maintenance Operations. ASCE Journal of Infrastructure Systems, http://dx.doi.org/10.1061/(ASCE)IS.1943-555X.0000012:

ASCE. 2009a. American Society of Civil Engineers: Reportcard for America's Infrastructure, http://www.infrastructurereportcard.org/fact-sheet/roads. Retrieved October 2009,

ASCE. 2009b. American Society of Civil Engineers: Facts About Roads, www.asce/reportcard. Retrieved October 2009,

Baltagi BH, Griffin JM. 1988. A general index of technical change. The Journal of Political Economy **96** (1): 20-41.

Banker RD, Morey RC. 1986a. Efficiency Analysis for Exogenously Fixed Inputs and Outputs. Operations Research **34** (4): 513-521.

Battese GE, Coelli T. 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. The Journal of Productivity Analysis **3**: 153-169.

Battese GE, Coelli TJ. 1995. A model for technical efficiency effects in a stochastic frontier production function for panel data. Empirical Economics **20**: 315-332.

Battese GE, Rao D. 2002. Technology Gap, Efficiency, and a Stochastic Metafrontier Function. International Journal of Business and Economics **1** (2): 87-93.

Battese GE, Rao D, O'Donnell CJ. 2004. A Metafrontier Production Function for Estimation of Technical Efficiency and Technology gaps for Firms Operating Under Different Technologies. Journal of Productivity Analysis **21** (1): 91-103.

Benkard CL. 2000. Learning and Forgetting: The Dynamics of Aircraft Production American Economic Review **90** (4): 1034-1054.

Bjornsson HC, de la Garza JM, Nasir MJ. 2000. A decision support system for road maintenance budget allocation. In Proceedings of the 8th International Conference on Computing in Civil and Building Engineering. Palo Alto, CA.

Borger B, Kerstens K, Staat M. 2008. Transit Cost and Cost Efficiency: Bootstrapping Non-parametric Frontiers. Research in Transportation Economics **23**: 53-64.

Butt AA, Shahin MY, Carpenter SH, Carnahan JV. 1994. Application of Markov process to pavement management systems at network level. In Proceedings of the 3rd International Conference on Managing Pavements. San Antonio, Texas.

Chang Albitres C, Krugler P, Smith R. 2005. A knowledge approach oriented to improved strategic decisions in pavement management practices. First Annual Inter-university Symposium of Infrastructure Management. Waterloo, Canada.

Charnes A, Clark T, Cooper WW, Golany B. 1985. A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in U.S. Air Forces. In R. Thompson and R.M. Thrall (Eds.). Annals of Operational Research **2**: 95-112.

Charnes A, Cooper WW, Rhodes E. 1978. Measuring the efficiency of decision making units. European Journal of Operational Research **2** (4): 429-444.

Chasey AD. 1995. A framework for determining the impact of deferred maintenance and/or obsolescence of a highway system. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. Blacksburg.

Chasey AD, de la Garza JM, Drew DR. 1997. Comprehensive level of service: Needed approach for civil infrastructure systems. Journal of Infrastructure Systems **3** (4): 143-153.

Chasey AD, de la Garza JM, Drew DR. 2002. Using simulation to understand the impact of deferred maintenance. Computer-Aided Civil and Infrastructure Engineering **17**: 269–279.

Chen C-M. 2009. A network-DEA model with new efficiency measures to incorporate the dynamic effect in production networks. European Journal of Operational Research **194** (3): 687-699.

Chen C-M, Dalen Jv. 2010. Measuring dynamic efficiency: Theories and an integrated methodology. International Journal of Production Economics **203**: 749–760.

Chen Y, Ali AI. 2002. Output–input ratio analysis and DEA frontier. European Journal of Operational Research **142**: 476–479.

Choi O, Stefanou SE, Stokes JR. 2006. The dynamics of efficiency improving input allocation. Journal of Productivity Analysis **25** (159-171):

Chowdhury T. 2007. Supporting document for pavement models and decision matrices development process used in the needs-based budget. Richmond, VA, Virginia Department of Transportation Asset Management Division.

Cook WD, Kazakov A, Roll Y. 1990. A DEA Model for Measuring the Relative Efficiency of Highway Maintenance Patrols. INFOR **28** (2): 131-124.

Cook WD, Kazakov A, Roll Y. 1994. "Chapter 10: On the Measurement and Monitoring of Relative Efficiency of Highway Maintenance Patrols". Data Envelopment Analysis: Theory, Methodology and Applications. A. Charnes, W. Cooper, A. Y. Lewin and L. M. Seiford, . Boston, Kluwer Academic Publishers. 195-210.

Cornwell C, Schmidt P, Sickles RC. 1990. Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels. Journal of Econometrics **46** (1-2): 185-200.

Daraio C, Simar L. 2005. Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach. Journal of Productivity Analysis **24** (1): 93–121.

Daraio C, Simar L. 2007. Conditional Nonparametric Frontier Models for Convex and Nonconvex Technologies: A Unifying Approach. Journal of Productivity Analysis **28**: 13–32.

de la Garza JM, Drew DR, Chasey AD. 1998. Simulating highway infrastructure management policies. Journal of Management in Engineering **14** (5): 64-72.

de la Garza JM, Fallah-Fini S, Triantis K. 2009. Efficiency Measurement of Highway Maintenance Strategies Using Data Envelopment Analysis. In Proceedings of the Proceedings of 2009 NSF Engineering Research and Innovation Conference. Hawaii, USA.

de la Garza JM, Krueger DA. 2007. Simulation of highway renewal asset management strategies. In Proceedings of the International Workshop on Computing in Civil Engineering, American Society of Civil Engineers. Carnegie Mellon University, Pittsburgh.

de Mateo F, Coelli T, O'Donnell C. 2006. Optimal paths and costs of adjustment in dynamic DEA models: with application to Chilean department stores. Annals of Operations Research **145**: 211-227.

Dekker R. 1996. Applications of maintenance optimization models: a review and analysis. Reliability Engineering and System Safety **51**: 229-240.

Diewert WE. 1992a. The Measurement of Productivity. Bulletin of Economic Research **44** (3): 163-198.

Emrouznejad A, Thanassoulis E. 2005. A mathematical model for dynamic efficiency using data envelopment analysis. Applied Mathematics and Computation **160**: 363-378.

Fallah-Fini S, Triantis K. 2009. Evaluating the productive efficiency of highway maintenance operations: environmental and dynamic considerations. The XI European Workshop on Efficiency and Productivity Analysis. Pisa, Italy.

Fallah-Fini S, Triantis K, de la Garza JM. 2009. Performance measurement of highway maintenance operation using data envelopment analysis: Environmental considerations. In Proceedings of the IIE Annual Conference. Miami, FL.

Färe R. 1986. A dynamic non-parametric measure of output efficiency Operations Research Letters **5** (2): 83-85.

Färe R. 1992. Productivity Changes in Swedish Pharmacies 1980-1989: A Non-parametric Malmquist Approach. Journal of Productivity Analysis **3** (3): 85-101.

Färe R, Grosskopf S. 1996. Intertemporal Production Frontiers: With Dynamic DEA. in: R Färe, Grosskopf S (Ed.), Kluwer Academic Publishers: Boston.

Färe R, Grosskopf S. 1997. Efficiency and productivity in rich and poor countries. Dynamics, Economic Growth, and International Trade. BS Jensen, Wong KY. Ann Arbor, MI, The University of Michigan Press.

Färe R, Grosskopf S. 2000. Network DEA. Socio-Economic Planning Science **34**: 35-49.

Fare R, Grosskopf S, Lovell CAK. 1994b. Production Frontiers. in: (Ed.), Cambridge University Press: Cambridge.

Färe R, Grosskopf S, Norris M, Zhang Z. 1994a. Productivity growth, technical progress, and efficiency change in industrialized countries. The American Economic Review **48** (1): 66-83.

Färe R, Grosskopf S, Roos P. 1996. On two definitions of productivity. Economics Letter **53**: 269-274.

Farrell MJ. 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A (General) **120** (III): 253-281.

Feighan KJ, Shahin MY, Sinha KC. 1987. A dynamic programming approach to optimization for pavement management systems. In Proceedings of the Second North American Conference on Managing Pavement. Toronto, Ontario.

FHWA. 2003. Special Experimental Project Task Force 14. Retrieved October 2009,

Forrester JW. 1971. World Dynamics. in: (Ed.), Wright-Allen Press: Cambridge, Mass.

Forsund FR. 2010. Dynamic Efficiency Measures. Department of Economics, University of Oslo: Working Paper.

Forsund FR, Eilev S. 1983. Technical progress and structural change in the Norwegian primary aluminium industry. Scandinavian Journal of Economics **85** (2): 113-126.

Forsund FR, Hjalmarsson L. 1974. On the measurement of productive efficiency. The Swedish Journal of Economics **76** (2): 141-154.

Frenger P. 1992. Comment on M.D. Intriligator, productivity and embodiment of technical progress. Scandinavian Journal of Economics **94 Supplemet**: 89-93.

Fwa TF, Chan WT, Hoque KZ. 2000. Multiobjective optimization for pavement maintenance programming. Journal of Transportation Engineering **126** (5): 367–374.

Fwa TF, Sinha KC, Riverson JDN. 1988. Highway routine maintenance programming at network level. Journal of Transportation Engineering **114** (5): 539–554.

Gao L, Tighe SL, Zhang Z. 2007. Using markov process and method of moments for optimizing management strategies of pavement infrastructure. 86th Annual Meeting of the Transportation Research Board. Washington, D.C.

Gao L, Zhang Z. 2008. Robust optimization for managing pavement maintenance and rehabilitation. Transportation Research Record: Journal of the Transportation Research Board **2084**: 55–61.

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

Greene WH. 2002. Econometric Analysis. in: (Ed.), Fifth edition, Prentice Hall: Upper Saddle River, NJ.

Gulledge TR, Womer NK. 1986. The economics of made-to-order production. in: (Ed.), New York: Springer-Verlag:

Hackman ST. 1990. An axiomatic framework of dynamic production. The Journal of Productivity Analysis **1**: 309-324.

Huang YH. 2004. Pavement Analysis and Design. in: (Ed.), Pearson/Prentice Hall: Upper Saddle River, NJ.

Ibbs W, Liu M. 2005. System dynamic modeling of delay and disruption claims. AACE Construction Engineering **47** (6): 12-15.

Ismail N, Ismail A, Atiq R. 2009. An overview of expert systems in pavement management. European Journal of Scientific Research **30** (1): 99-111.

JLARC. 2002. Adequacy and Management of VDOT's Highway Maintenance Program. Richmond.

Johansen L. 1959. Substitution versus fixed production coefficients in the theory of economic growth: A synthesis. Econometrica **27** (2): 157-176.

Johansen L. 1972. Production Functions. in: (Ed.), North Hollad Publication Co.: Amsterdam.

Kazakov A, Cook WD, Roll Y. 1989. Measurement of Highway Maintenance Patrol Efficiency: Model and Factors. Transportation Research Record **1216**:

Kim K. 1998. A transportation planning model for state highway management: A decision support system methodology to achieve sustainable development. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. Blacksburg.

Kumbhakar SC. 1990. Production frontiers, panel data and time-varying technical inefficiency. Journal of Econometrics **46**: 201-212.

Lee S, Pena-Mora F. 2005. System dynamics approach for error and change management in concurrent design and construction. In Proceedings of the Winter Simulation Conference. Orlando, FL.

Lee S, Pena-Mora F, Park M. 2005. Quality and change management model for large scale concurrent design and construction projects. Journal of Construction Engineering and Management **131** (8): 890–902.

Lytton RL. 1985. From ranking to true optimization. In Proceedings of the North American Pavement Management Conference. Toronto, Ontario.

Mahoney JP, Uhlemeyer J, Morin P, Luhr D, Willoughby D, Mouench ST, banker T. 2010. Pavement preservation funding and performance in Washington State. Transportation Research Board Annual Meeting. Washington, D.C.

McCullouch BG, Anastasopoulos PC. 2009. Performance-based Contracting-Yes or No: An in Depth Analysis. 12th AASHTO-TRB Maintenance Management Conference.

NCDC. 2010. Natinal Climate Data Center Archive. Retrieved May 2009,

NCHRP. 2009. NCHRP Synthesis 389 : Performance-based Contracting for Maintenance, Transportation Research Board of National Academies.

Nemoto J, Goto M. 1999. Nemoto-Dynamic data envelopment analysis: Modeling intertemporal behavior of a firm in the presence of productive inefficiencies. Economics Letter **64**: 51-56.

Nemoto J, Goto M. 2003. Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities. Journal of Productivity Analysis **19**: 191-210.

NRCAN. 2010. Climate change impacts and adaptation: A Canadian perspective (impacts on transportation infrastructure). Retrieved March 2010,

O'Donnell C. 2008. An aggregate quantity-price framework for measuring and decomposing productivity and profitability change. Centre for Efficiency and Productivity Analysis Working Papers WP07/2008, University of Queensland.

O'Donnell C. 2010a. Measuring and decomposing agricultural productivity and profitability change. Australian Journal of Agricultural and Resource Economics: in press.

O'Donnell C. 2010b. Nonparametric Estimates Of The Components Of Productivity And Profitability Change In U.S. Agriculture. Centre for Efficiency and Productivity Analysis Working Papers: WP02/2010, University of Queensland.

O'Donnell CJ, Prasada Rao DS, Battese GE. 2008. Metafrontier Frameworks for the Study of Firm-level Efficiencies and Technology Ratios. Empirical Economics **34**: 231-255.

Ogunlana S, Li H, Sukhera F. 2003. System dynamics approach to exploring performance enhancement in a construction organization. Journal of Construction Engineering and Management **129** (5): 528–536.

Ogunlana SO, Lim J, Saeed K. 1998. DESMAN: A dynamic model for managing civil engineering design projects. Computers & Structures **67** (5): 401-419.

Oliva R. 2003. Model calibration as a testing strategy for system dynamics models. European Journal of Operational Research **151**: 552-568.

Ozbek EM. 2007. Development of a comprehensive framework for the efficiency measurement of road maintenance strategies using data envelopment analysis. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. Blacksburg.

Ozbek M, de la Garza JM, Triantis K. 2010a. Data and Modeling Issues Faced during the Efficiency Measurement of Road Maintenance using Data Envelopment Analysis. ASCE, Journal of Infrastructure Systems **16** (1): 21-30.

Ozbek ME, de la Garza JM, and Triantis K. 2010b. Efficiency Measurement of Bridge Maintenance using Data Envelopment Analysis. ASCE, Journal of Infrastructure Systems **16** (1): 31-39.

Ozbek ME, de la Garza JM, Triantis K. 2009. Data Envelopment Analysis as a Decision Making Tool for the Transportation Professionals. Journal of Transportation Engineering **135** (11): 822-831.

Pasupathy K. 2006. Ph.D. Industrial and Systems Engineering. Falls Chruch, VA.

Pitt M, Lee LF. 1981. The measurement and sources of technical inefficiency in Indonesian weaving industry. Journal of Development Economics **9**: 43-64.

Priest AL, Timm DH. 2006. Methodology and calibration of fatigue transfer functions for mechanistic-empirical flexible pavement design. National Center for Asphalt Technology. Auburn University, AL.

Rapping L. 1965. Learning and World War II Production Functions. Review of Economics and Statistics 81-86.

Rosen S. 1972. Learning by experience as joint production. Quarterly Journal of Economics **86**: 366-382.

Rouse P, Chiu T. 2008. Towards Optimal Life Cycle Management in a Road Maintenance Setting Using DEA. European Journal of Operational Research doi: 10.1016/j.ejor.2008.02.041:

Rouse P, Putterill M, Ryan D. 1997. Towards a General Managerial Framework for Performance Measurement: A Comprehensive Highway Maintenance Application. Journal of Productivity Analysis **8**: 127–149.

Ruggiero J. 1998. Non-discretionary Inputs in Data Envelopment Analysis. European Journal of Operational Research **111** (461-469):

Salter WEG. 1960. Productivity and Technical Change. in: (Ed.), Cambridge University press: London.

Schmidt P, Sickles R. 1984. Production frontiers and panel data. Journal of Business and Economic Statistics **2**: 367-374.

Seaver WL, Triantis K. 1992. A Fuzzy Clustering Approach for Measuring Technical Efficiency in Manufacturing. Journal of Productivity Analysis **3** (337-363):

Sengupta JK. 1992a. Adjustment costs in production frontier analysis. Economic Notes **21**: 316-329.

Sengupta JK. 1992b. Non-parametric approach to dynamci efficnecy: A non-parametric aplication of cointegration to production frontiers. Applied Economics **24**: 153-159.

Sengupta JK. 1994a. Measuring dynamic efficiency under risk aversion European Journal of Operational Research **74**: 61-69.

Sengupta JK. 1994b. Evaluating dynamic efficiency by optimal control. International Journal of Systems Science **25** (8): 1337-1353.

Sengupta JK. 1996. Dynamic aspects of data envelopment analysis. Economic Notes **25** (1): 143-164.

Sengupta JK. 1999. A dynamic efficiency model using data envelopment analysis. International Journal of Production Economics **62**: 209-218.

Sickles R, Good D, Johnson R. 1986. Allocative distortions and the regulatory transition of the airline industry. Journal of Econometrics **33**: 143-163.

Silva E, Stefanou SE. 2003. Nonparametric dynamic production analysis and the theory of cost. Journal of Productivity Analysis **19**: 5–32.

Silva E, Stefanou SE. 2007. Dynamic efficiency measurement: Theory and application. American Journal of Agrecultural Economics **89** (2): 398-419.

Simar L. 1992. Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-parametric, and Semi-parametric Methods with Bootstrapping. Journal of Productivity Analysis **3**: 167-203.

Simar L, Wilson P. 2007. Estimation and Inference in Two-stage Semi-parametric Models of Production Processes. Journal of Econometrics **136** (1): 31-64.

Simar L, Wilson P. 2008. Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives. The Measurement of Productive Efficiency and Productivity Chaneg. H Fried, Lovell CAK, Schmidt S. New York, Oxford University Press. 421-521.

Smadi OG. 1994. Network pavement management system using dynamic programming: Application to Iowa interstate network. In Proceedings of the 3rd International Conference on Managing Pavements. San Antonio, Texas.

Smith RE, Nazarian S. 1992. Defining pavement maintenance and distress precursors for pavement maintenance measurement, maintenance of pavements, lane markings, and road sides. Transportation Research Record: Journal of the Transportation Research Board **1334**: 16-18.

Solow RM. 1957. Technical change and the aggregate production function. Review of Economics and Statistics **39**: 312-320.

Sterman J. 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World. in: (Ed.), Irwin/Mc-Graw Hill: Homewood, IL.

Sterman JD. 2000. Business Dynamics: System Thinking and Modeling for a Complex World. in: (Ed.), McGraw-Hill: Boston, MA.

Thanassoulis E. 1993. A Comparison of Regression Analysis and Data Envelopment Analysis as Alternative Methods for Performance Assessments. Journal of Operational Research Society **44** (I1): I 29-1144.

Thompson BP, Bank LC. 2010. Use of system dynamics as a decision-making tool in building design and operation. Building and Environment **45** (4): 1006-1015.

TRB. 2006. Maintenance and Operations of Transportation Facilities 2005 Strategic Vision. E-C092.

Triantis K. 2004. Engineering Applications of Data Envelopment Analysis. Handbook on Data Envelopment Analysis, W. W. Cooper, L. M. Seiford, J. Zhu. Boston, Kluwer Academic Publishers.

Triantis K, Seaver WL, Sarayia D. 2010. Using Multivariate Methods to Incorporate Environmental Variables for Local and Global Efficiency Performance Analysis. forthcoming, Informational Systems and Operational Research:

Vaneman WK. 2002. Evaluating system performance in a complex and dynamic environment. Ph.D. Industrial and Systems Engineering. Falls Church, VA.

Vaneman WK, Triantis K. 2003. The dynamic production axioms and system dynamics behaviors: The foundation for future integration. Journal of Productivity Analysis **19**: 93-113.

Vaneman WK, Triantis K. 2007. Evaluating the productive efficiency of dynamical systems. IEEE Transactions on Engineering Management **54** (3): 600-612.

Vaneman WK, Triantis K. 2010. Application of DPEM.

VDOT. 2002. 2002 Windshield Survey Program User Manual. Virginia, Pavement Management Program & Virginia Department of Transportation Asset Management.

VDOT. 2010. Virginia traffic data publications, http://www.virginiadot.org/info/ct-TrafficCounts.asp. Retrieved March 2010,

Wang F, Zhang Z, Machemehl RB. 2003. Decision-making problem for managing pavement maintenance and rehabilitation projects. Transportation Research Record: Journal of the Transportation Research Board **1853**: 21–28.

Wibe S. 2008. Efficiency: A dynamic approach. International Journal of Production Economics **115**: 86–91.

Womer NK. 1979. Learning curves, production rate, and program cost. Management Science **25**: 312-319.

Wright TP. 1936. Factors Affecting the Costs of Airplanes. Journal of Aeronautical Science **3**: 122-128.