Essays In Applied Econometrics

by

Abdoul Gadiry Sam

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ECONOMICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2005

The University of Arizona
Graduate College

As members of the Dissertation Committee, we certify that we have read the

dissertation prepared by Abdoul Gadiry Sam

entitled Essays in Applied Econometrics

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Alan P. Ker	date
Robert Innes	date
Ronald L. Oaxaca	date
George J. Jiang	date
Alfonso Flores-Lagunes	date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

We hereby certify that we have read this dissertation prepared under our direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Co-Director: Alan P. Ker date

Dissertation Co-Director: Robert Innes date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED:________________________

Acknowledgements

First and foremost I would like to thank my mother, my brother, and my sisters back in Senegal for their inexorable support and affection throughout my years of education. I cannot thank my brother Elhaj enough for making me the first person in the family to go to school.

My heartfelt gratitude goes to my advisors Alan Ker and Rob Innes whose assistance, guidance, and expertise has made it possible to finish this dissertation. Working with Alan and Rob has been a uniquely beneficial and enjoyable experience. Alan and I share a common passion for econometrics, he has given me the funding and facilities, allowing me to thrive in my work and education. Thank you guys.

I also want to extend my sincerest gratitude to the other committee members, Ron Oaxaca, George Jiang, and Alfonso Flores-Lagunes, for their time and very helpful inputs. I learnt a lot from the many discussions I have had with them in their respective offices or in class.

Dedication

To the memory of my father.

Table of Contents

TABLE OF CONTENTS—*Continued*

List of Figures

List of Tables

Abstract

The first essay of this dissertation studies the determinants and effects of firms' participation in a voluntary pollution reduction program (VPR) initiated by government regulators. This research presents empirical evidence in support of the "enforcement theory" for VPRs, which predicts that (1) participation is rewarded by relaxed regulatory scrutiny; (2) the anticipation of this reward spurs firms to participate in the program; and (3) the program rewards regulators with reduced pollution. The results also indicate that firms' VPR participation, and pollutant reductions themselves, were prompted by a firm's likelihood of becoming a boycott target and/or being subject to environmental interest group lobbying for tighter standards.

In the second essay, a nonparametric regression estimator which can accommodate two empirically relevant data environments is proposed. The first data environment assumes that at least one of the explanatory variables is discrete. In such an environment, a "cell" approach which estimates a separate regression for each discrete cell, has generally been employed. The second data environment assumes that one needs to estimate a set of regression functions that belong to different individuals. In both environments the proposed estimator attempts to reduce estimation error by incorporating extraneous data from the other individuals or "cells" when estimating the regression function for a given individual or "cell". The simulation results for the proposed estimator demonstrate a strong potential in empirical applications.

In the third essay, the nonparametric approach proposed in the second essay is used to estimate the parameters of the short-term interest rate diffusion. The nonparametric estimators of the drift of the short rate proposed by Stanton (1997) and Jiang (1998) can produce spurious nonlinearities due to the persistent dependence and limited sampling

period of interest rates. The simulations show that the proposed estimator significantly attenuates the spurious nonlinearities of Stanton's nonparametric estimator. An empirical study of the US term structure of interest rates is presented based on the proposed estimator and two other competing models. The results suggest that the estimation of the short rate diffusion parameters using additional data from yields of different maturities has significant economic implications on the valuation interest rate derivatives.

1. Dissertation Introduction

This dissertation is centered on nonparametric and applied econometrics with applications on environmental economics and financial economics.

The second chapter investigates the economic puzzle posed by profit-maximizing firms' voluntary overcompliance with pollution standards. Economists have advanced a number of theories explaining why private firms voluntarily agree to overcomply with environmental pollution standards given that such overcompliance comes at a cost. For example, Arora and Gangopadhyay (1995, 1996) argue that participation in these programs may be motivated by firms's desire to attract a clientele of "green consumers" who are willing to pay a premium for goods produced in an environmentally friendly way. Maxwell, Lyon and Hackett (2000) hypothesize that voluntary pollution reductions may also deter lobbying by environmental constituencies for a tightening of regulatory standards. The purpose of this chapter is to test the validity of these and other motives of self-regulation using the Environmental Protection Agency (EPA)'s 33/50 program as a research experiment. The chapter makes a significant contribution to the empirical literature of environmental self-regulation in more than one way, testing (1) the extent to which voluntary pollution reductions are driven by an enforcement bargain between the regulator (EPA) and the regulated firm, and (2) the effects of implicit boycott threats on a regulated firm's overcompliance with pollution standards. The results show that a firm's history of inspections and corrective actions constitute a significant determinant of participation in the 33/50 program and that the EPA reciprocated to voluntary participation by easing regulatory oversight on participants. The findings also indicate that firms' participation and performance in the 33/50 program were motivated by likelihood of becoming a boycott target.

The third chapter develops a new nonparametric estimator designed to attenuate the bias of ordinary nonparametric kernel estimators. Nonparametric methods have gained wide acceptance in applied econometrics because they circumvent the issue of functional form specification and are consistent under mild regularity conditions. Unfortunately, such flexibility comes at a cost of a finite sample bias, which may be large in applied work. Several estimators which reduce the bias have been proposed. For example Hjort and Glad (1995) propose a semiparametric estimator which combines a parametrically estimated pilot with a nonparametrically estimated correction factor in a multiplicative fashion. The parametric pilot can be thought of as a prior for the shape of regression function of interet whereas the correction factor adjusts the pilot if it does not satisfactorily capture the shape of the function being estimated. Consequently, the estimator behaves like the parametric start if the parametric assumption is correct, while resembling the nonparametric estimator otherwise. The estimator proposed in this chapter is similar in spirit to Hjort and Glad's; however, alternative data environments are considered, ones in which multivariate measurements from multiple experimental units (drawn from the same population) are available. In these data environments, the proposed estimator replaces the parametrically estimated pilot in the Hjort and Glad method with a nonparametrically estimated pooled function. This is motivated by the fact that if the different experimental units have identical regression curves then the optimal estimator would pool the data and estimate one regression curve. If those unknown functions are sufficiently similar, using the pooled estimator as a pilot in Hjort and Glad's framework would yield efficiency gains relative to the ordinary nonparametric estimators. Consequently, the proposed estimator behaves like the pooled estimator when the regression functions are identical while resembling the individual nonparametric estimator when the functions are dissimilar. The use of information from individuals possibly similar to that of interest in the form a nonparametric pooled start represents the key conceptual difference between the proposed estimator and the estimator of Hjort and Glad (1995) and the main contribution of this chapter. Consistency and asymptotic normality of the proposed estimator are established.

The fourth chapter is an application of the estimator proposed in the third chapter to the estimation of the parameters of the short term interest rate diffusion. The size of interest rate data (daily or higher-frequency data is readily available) and the low dimensionality of the vector of covariates, which is one for single factor models, makes it favorable to estimate the drift and diffusion functions of the short rate nonparametrically as done by Aït-Sahalia (1996a, 1996b), Stanton (1997), Jiang (1998) and others. Virtually all nonparametric estimators have found that the drift a the short rate in nonlinear, exhibiting a dramatic mean-reversion for high levels of the short rate. However, Chapman and Pearson (2000) argue that the nonlinear pattern of the nonparametrically estimated drift may be simply be an artifact of the bias of the nonparametric estimator caused by the strong persistence of the short term yields. Heuristically, the persistence of yield data implies that increasing the sample size does not necessarily translate into increased information. Since yields of different maturities are available along the yield curve and are known to have systematic co-movements, this chapter applies the nonparametric estimator proposed in chapter three to the estimation of the short-rate diffusion parameters. This chapter uniquely contributes to the term structure literature in that unlike any of the current parametric and nonparametric methods, the proposed estimator uses a panel of yields of different maturities instead of a single time-series or short rate observations to estimate the drift and the diffusion functions. The estimator is formally developed in the framework of correlated multivariate diffusion processes and its asymptotic properties derived.

The fifth chapter summarizes the findings in the three previous chapters and explores avenues of future research.

2. Voluntary Pollution Reductions and The Enforcement of Environmental law: An Empirical Study of the 33/50 Program

2.1. Introduction

Why do private firms voluntarily over-comply with environmental regulations? For example, over 1200 firms joined the U.S. Environmental Protection Agency's (EPA) 33/50 program. In this program, firms pledged to reduce emissions of 17 key toxic pollutants beyond targets required by law. Current voluntary EPA programs include "Energy Star," which seeks to decrease carbon dioxide emissions, and the "National Environmental Performance Track," designed to encourage environmentally proactive firms through rewards and public recognition.

Economists have offered a number of theories to explain why profit-driven firms might volunteer for costly pollution reduction efforts. Arora and Gangopadhyay (AG, 1995) argue that firms want to attract a clientele of "green consumers" who are willing to pay more for goods produced in an environmentally friendly way (see also Arora and Cason (AC), 1996). Voluntary pollution reductions may also deter lobbying by environmental groups for tighter legislative or regulatory standards (Maxwell, Lyon and Hackett (MLH), 2000); spur tighter environmental standards that "raise rivals' costs" (Salop and Scheffman, 1983; Innes and Bial, 2002); avoid future environmental liability; and/or deter boycotts by environmental interest groups (Baron, 2001; Innes, 2003).

However, there is another potential motive for voluntary over-compliance that has received relatively little attention in the literature: A firm's participation in a voluntary pollution reduction program (VPR) may lessen the scrutiny of environmental authorities, reducing the frequency of costly environmental inspections and enforcement actions. The

EPA officially claims that such rewards are not offered to program participants.[1] Nevertheless, such rewards, promised implicitly if not officially, may represent an optimal government policy to promote participation in a VPR. The societal benefit of the VPR is that it prompts participating firms to adopt management practices that reduce their costs of pollution abatement, leading ultimately to pollution reductions (Maxwell and Decker, 2002).[2] While intuitively compelling, the empirical strength of this "enforcement theory" for VPRs has yet to be studied.

The purpose of this chapter is to examine (1) the empirical validity of this enforcement based spur to participation in the EPA's 33/50 program, among many other potential participation motives, and (2) the related effects of program participation on both a regulated firm's pollution levels and the government's enforcement activity. In studying these issues, this work seeks to bridge two empirical literatures, one focusing on voluntary pollution reduction programs (e.g., AC; Khanna and Damon (KD), 1999; Videras and Alberini (VA), 2000; Anton, Deltas and Khanna, 2004) and the other investigating determinants and effects of government enforcement activities. The former literature suggests that participation in voluntary pollution reduction programs is motivated, in part, by green marketing (AC, KD, VA) and potential liability (KD, VA), with larger firms found to be more likely to participate (AC, VA). In contrast to the focus of this study, however, this literature does not study effects of voluntary over-compliance on government enforcement and does not consider potential effects of boycott threats and

[1] With regard to the 33/50 program, the EPA has stated (EPA, 1992, p. 11): "Participation in the program is enforcement neutral: a company will receive no special scrutiny if it elects not to participate and receive no relief from normal enforcement attention if it does elect to participate."

[2] Maxwell and Decker (2002) show that a reduced probability of enforcement may result from a firm's adoption of abatement-cost-reducing investments, thus spurring these investments apriori. If regulators can implicitly commit (a priori) to a flexible enforcement rate (as a function of the firm's investment) it can be shown that welfare is necessarily enhanced by the promise of reduced enforcement when firms invest more by, for example, participating in a voluntary pollution reduction program. Miceli and Segerson (1998) also stress benefits of voluntary pollution reduction programs in lessening tensions and facilitating negotiations between enforcement agencies and polluting firms.

incentives for regulatory preemption (MLH).[3]

A number of papers study determinants of the government's environmental enforcement activity, and its impact on pollution (e.g., Magat and Viscusi, 1990; Gray and Deily, 1996; Laplante and Rilstone, 1996; Nadeau, 1997). This work provides evidence that government enforcement efforts tend to prompt pollution reductions, a conclusion for which this work also finds support. However, most closely related to this study are papers that focus on the government's strategic use of enforcement tools to leverage desired conduct from regulated firms. Harrington (1988) argues that the apparent paradox of low and infrequent regulatory fines for environmental violations can be explained by the targeting of enforcement resources to "bad" firms that prompts desired conduct from "good" firms, despite low penalties for "good" firms' violations.[4] Helland (1998) studies an additional basis for targeting, the extent of a firm's self-reporting of violations. Decker (2003) studies an additional reward that may be offered to "good" firms: more rapid environmental permitting for new source construction. Both find evidence that these regulatory tools are exploited in enforcement practice. A key finding of this research is that that regulators use another instrument to target their enforcement activities: a firm's participation in voluntary pollution reduction programs.

The remainder of the chapter is organized as follows. Section II provides a summary of the 33/50 program. Section III discusses hypotheses on determinants of 33/50 participation, firm pollution decisions, and government inspections that are tested in this chapter. Section IV discusses the data and the econometric modeling. Section V presents

[3] VA consider the potential impact of prior regulatory fines on voluntary program participation, finding some evidence that such enforcement actions make participation more likely. This chapter studies the impact of regulatory inspections as well and, like KD, also models impacts on pollution.

[4] See also related work by Harford and Harrington (1991) and Heyes and Rickman (1999). In addition, consistent with this theory, Decker (2005) finds that government inspection activity responds to reductions in reported toxic pollutant releases as well as reductions in regulated pollutant releases and a good statutory compliance history.

the estimation results. Finally, Section VI concludes.

2.2. The 33/50 Program

Started in 1991, the 33/50 program was the EPA's first formal effort to achieve voluntary pollution reductions by regulated firms. The program sought to reduce releases of seventeen toxic chemicals by a third by 1992 and by 50 percent by 1995, measured from 1988 baseline levels. The seventeen 33/50 chemicals are listed in Appendix A. Roughly seventy percent of the 33/50 chemicals (by 1988 weight of releases) were air pollutants (AC). Two of the chemicals (carbon tetrachloride and 1,1,1-trichloroethane) depleted the stratospheric ozone layer and, hence, came under the Montreal Protocol's provisions for the phase-out of such substances; however, these two chemicals represented less than fifteen percent of total 33/50 releases (in 1988).

The EPA initiated the 33/50 program shortly after creating the Toxic Release Inventory (TRI), a database compiling information on toxic releases of all firms with ten or more employees producing one or more of 320 targeted pollutants. In early 1991, the EPA invited the 509 companies emitting the largest volume of 33/50 pollutants to participate in the program; these companies were responsible for over three-quarters of total 33/50 releases as of 1988. In July 1991, the 4534 other companies with reported 33/50 releases in 1988 were asked to participate as well. With additional enrollments through 1995, the EPA invited a total of 10,167 firms to join the 33/50 program, and 1294 firms accepted. The latter program participants accounted for 58.8 percent of 33/50 releases in 1990. In this research, the focus is exclusively on firms that were eligible for the 33/50 program in 1991, those invited in March and July of that year.

The 33/50 program was purely voluntary and its pollution reduction targets were not enforceable. Despite the absence of apparent regulatory teeth, the EPA (1999) cites some aggregate statistics as indicators of the program's success. Among reporting firms, total 33/50 releases declined by over 52 percent between 1990 and 1996, and net 33/50 releases, excluding the two ozone-depleting compounds, declined by over 45 percent. In contrast, non-33/50 TRI releases fell by 25.3 percent over this period. Moreover, rates of 33/50 release reductions were greater for program participants (down 59.3 percent between 1990 and 1996) than for non-participants (down 42.9 percent over the same interval). However, these numbers may mask other hidden determinants of firms' pollution. For example, participating firms may have been more apt to reduce pollution, regardless of participation in the 33/50 program. One of the goals in this chapter is to estimate the pollution abatement that is attributable to the 33/50 program, controlling for other relevant explanators and potential selection bias in program participation.

2.3. Hypotheses

Participation in the 33/50 program, although involving no enforceable commitment, required a firm to file a plan documenting how it proposed to reduce its emissions of target pollutants. Indeed, more than 82 percent of participants stipulated specific pollution reduction targets. In addition, the program was accompanied by some technical assistance to aid participants in realizing their target emission reductions (Khanna and Damon, 1999). Although the EPA, in its public statements, stressed the public recognition that participation could bring, there is little evidence that such recognition occurred in the broader public;[5] indeed, only with effort could a researcher obtain the names of program participants. However, the process of planning for emissions reductions, including possible managerial changes and environmental auditing procedures, could yield the very

[5] The EPA (1992) states that its "partnership programs offer recognition ... that can enhance corporate image with customers, regulators, neighbors, and the media."

reductions that were the program's objective. To spur these innovations, and the participation that promoted them, the EPA may have implicitly afforded participants less scrutiny in its enforcement of pollution control laws, leading to fewer costly inspections and enforcement actions for a participating firm (Maxwell and Decker, 2002). These potential enforcement benefits of participation are over and beyond any anticipated reductions in inspection rates due to reduced pollution.[6] The value of this regulatory reward to 33/50 participation is expected to have been higher for firms that otherwise anticipated greater regulatory scrutiny.

- Hypothesis 1. Firms with higher rates of government inspection and enforcement action in previous periods are more likely to have participated in the 33/50 program.

- Hypothesis 2. 33/50 participants should have experienced lower rates of government inspection and lower levels of pollution.

- Hypothesis 3. Government inspections should have prompted pollution reductions (Harrington, 1988). In addition to enforcement considerations, a number of theories suggest motives for participation in voluntary programs such as 33/50, and for desired pollutant reductions as well. Next these implications are summarized, followed by a discussion of each.

- Hypothesis 4. A firm was more likely to participate in the 33/50 program and to achieve pollution reductions if it:

  (a) had more contact with final consumers (green marketing);

  (b) was a more likely object of a consumer / environmental group boycott (boycott deterrence);

[6] This empirical work distinguishes between direct effects of 33/50 participation on government inspections, and indirect effects (when participation reduces emissions and thereby reduces government enforcement activity). Note also that firms may be averse to inspections because of their potential to ignite adverse public reaction in the media and financial markets (e.g., see Konar and Cohen, 2001).

(c) was more exposed to potential liability because it was larger (with deeper pockets) and/or operated in strict liability states (liability);

(d) was in a more concentrated industry and invested more in research and development (raising rivals' costs); and

(e) had a greater incentive and ability to preempt regulation because it was a larger firm and operated in states with larger environmentalist constituencies (preempting regulation).

A firm may be able to exploit "green consumerism" to establish a market niche for goods produced in an environmentally friendly way (Arora and Gangopadhyay (AG), 1995); if present at all, such an ability is tied to a firm's proximity to consumers (AC, KD, VA). Following KD, this link is measured using a dummy variable that takes a value of one if the firm sold a product directly to final consumers (FG for "final good");[7] due to potential economies of scale in "green marketing," an interaction variable between FG and a measure of firm size is also considered. AC indicate that green product differentiation incentives are likely to be stronger in less concentrated industries. This conjecture runs counter to Hypothesis 4(d) and is tested in this analysis using a standard measure of industry concentration (HERF for Herfindahl index).

Firms may also be the potential object of consumer boycotts organized by environmental interest groups (Baron, 2001; Innes, 2003; Henriques and Sadorsky, 1996). Voluntary pollution reductions and participation in the 33/50 program may be actions that a firm can take to deter such organized consumer action. The prospect of a boycott is greater and hence, more likely to motivate a firm's voluntary pollution reduction when the firm's products have good substitutes, are perishable, are sold publicly at a

[7] AC argue that industry-aggregated advertising expenditures may also measure closeness to consumers; however, because this measure may be an indicator of market power as well. Furthermore, because of the prevalence of missing values for advertising data, even industry-aggregated, makes it aan inneficient measure of closeness to consumers.

retail level, and are "visible" in the marketplace (Smith, 1990). For example, over the recent past, environmental and animal rights activists have successfully challenged large, "powerful" and visible firms such as McDonalds and Home Depot using boycott tactics (Innes, 2003).[8] To test for potential boycott threat effects in this chapter, a dummy variable is constructed which takes on a value of one if a firm is in an industry that was contemporaneously targeted for boycott.[9] This variable is denoted BC. Because boycott threats are likely to be more acute when firms operate in states with larger environmental constituencies, an interaction variable between BC and the average per capita Sierra Club membership in the home states of a firm's plants (SIERRA) is also used.[10]

Larger firms, with deeper pockets, may voluntarily reduce pollution in order to avoid potential liability for harm caused. Such incentives will be greater in states that levy strict liability for environmental harm, as opposed to negligence liability (Alberini and Austin, 1999). In an attempt to capture the liability motive for pollution reduction a dummy variable taking a value of one if a plant's home state has a strict liability statute (STRICT) is used; for a firm, the strict liability variable is constructed by averaging these zero-one values for the firm's plants. Because liability effects are likely to be more pronounced for larger firms, an interaction variable between STRICT and firm size (as

[8] In 1999, McDonalds agreed to significant reforms in its supplier protocols for handling chickens after boycott actions by the animal rights group PETA (People for the Ethical Treatment of Animals); Burger King and Wendy's quickly followed suit. Also in 1999, Home Depot agreed to phase out products using old growth timber and to give preference to timber certified by the Forest Stewardship Council; other major home improvement retailers, as well as home builders, have since made similar commitments.

[9] The 1992-1993 issue of the National Boycott News lists products subject to contemporaneous organized consumer boycott, including over 400 products made by over 100 firms. If a firm or plant in the sample is in an industry that produces a targeted product (based on the firm's or plant's primary SIC classification), the boycott variable is assigned a value of one for that firm or plant. It should be pointed out that actual boycotts are rare. In fact, theory predicts that boycotts will generally be deterred by cooperating firms (Baron, 2001). Hence, none of the firms in this sample were actually boycotted. Rather, the boycott variable attempts to measure the potential likelihood that a firm might face a boycott threat.

[10] In principle, boycott effects may also be more acute for larger polluters. To test that hypothesis, an interaction between BC and a firm's 33/50 releases was included in the regressions, but with no significant effects.

measured by its number of employees, LEMP) is also considered.

In a concentrated industry, a firm that has developed cost-effective pollution abatement methods may wish to over-comply with government environmental standards in order to prompt tighter standards that disadvantage its rivals (Salop and Scheffman, 1983; Innes and Bial, 2002). This "raising rivals' costs" motive for voluntary pollution reductions is likely to be greater for firms that invest more heavily in research and development, investments that make cost-saving innovations in environmental technologies more likely. These effects are captured with variables measuring industry concentration (HERF) and firm-level R&D expenditures (R&D).[11]

Finally, Maxwell, Lyon and Hackett (2000) argue that firms may voluntarily abate pollution in order to prevent the enactment of more costly environmental regulation. Environmental interest groups may, at a cost, lobby the government for tighter environmental regulation. By abating pollution voluntarily, firms can reduce these groups' incentive to lobby. Indeed, firms may be able to preempt lobbying by abating pollution to a lesser extent than would otherwise be compelled by a successful lobbying campaign. This motive for a firm's pollution reduction and participation in the 33/50 program is likely to be greater in states with larger environmental constituencies. In these states, the public sensitivity to a firm's pollution is likely to be greater, as are environmental groups' incentive and ability to successfully lobby the government for change. To test for these effects, the per-capita Sierra Club membership in a plant's home state (SIERRA) is used, averaged across plants to obtain a firm-level variable.

[11] For judging "raising rivals cost" effects, it is expected that a firm's total level of R&D is a relevant indicator of a firm's potential research advantage.

2.4. The Data and the Econometric Modeling

Three equations are estimated in order to explain (1) firms' participation in the 33/50 program (in 1991), (2) firms' annual emissions of 33/50 pollutants (by weight, 1989-1995), and (3) the government's (State and Federal) annual number of environmental inspections of firms' facilities (1989-1995).

Several data sources are used to estimate these equations. Financial and employment data was obtained from the Standard & Poor's Compustat database. The EPA's Office of Environmental Information Records provided data on 33/50 participation, Federal and State enforcement actions under the Clean Air Act (CAA) and the Resource Conservation and Recovery Act (RCRA) (1988-1990), and facility-level government inspections under the CAA (1988-1995).[12] The Toxic Release Inventory (TRI) provided facility-level data on 33/50 chemical releases, primary standard industrial codes (SIC), parent company names, and facility locations. Firm-level 33/50 pollutant releases and inspections were obtained by aggregating across each firm's facilities. The Sierra Club provided data on its state membership (from 1989-1995, measured per capita). The Maxwell, Lyon and Hackett (2000) dataset provided information on state characteristics (1988), such as per capita state spending on clean air laws, educational status (the number of bachelors degrees per capita), the number of lawyers per capita, and indicators for whether the state had a right-to-work law or strict environmental liability. The number of 1988 Superfund sites for which a firm was a potentially responsible party (PRP) was obtained from the EPA's Superfund Office. County attainment status (whether a facility's home county was designated by the EPA to be out of attainment with clean air laws) was obtained from the EPA's website (www.epa.gov/oar/oaqps/greenbk/anay.html). County

[12] Attention is restricted to CAA inspections because the 33/50 program was principally an air toxics program.

population density (1990) was obtained from the U.S. Census.

This study focuses on manufacturing firms that operated in SICs 20-39 and were invited to participate in the 33/50 program in 1991. Merging the Compustat and environmental datasets for these firms gave an unbalanced panel of 325 firms and 1257 facilities over the seven years, 1989-1995. The years 1989-1990 are included in order to capture pre-program trends.

Tables 2.1 and 2.2 present variable definitions and descriptive statistics for the sample.

2.4.1. The Participation Equation

This section presents the results of a probit model of firms' decisions to participate (or not) in the 33/50 program. Attention is restricted to firms that chose to join (or not join) the 33/50 program in 1991; hence, the probit estimation is performed using lagged cross-section data for the one-time 1991 firm-level decisions.[13] To test Hypothesis 1 (the enforcement motive for participation), regressors include (i) the number of government inspections of firm facilities in 1989-1990 (INSP89-90), (ii) an indicator that takes a value of one if a firm had an enforcement action in the period 1989-1990 (ENFORCE), and (iii) the number of Superfund sites for which a firm is a potentially responsible party (PRP). Potential enforcement-driven rewards to 33/50 participation and pollution reductions are expected to have been greater for firms with more Superfund involvement, as measured by the PRP variable.

Critics of the 33/50 program suggest that firms joined because their prior (1988-1990) emission reductions already placed them in near reach of the program's goals (KD). This effects are controlled for by including a variable measuring a firm's 33/50 pollutant reductions from 1988 to 1990 (DIFREL). In addition, industry effects are controlled for

[13]Firms invited to join the program in 1991 accounted for 88.6% of participants.

by including dummy variables for the seven industries most heavily represented in the sample (SICs 28, 33, 34, 35, 36, 37, and 38).

2.4.2. The Pollution Equation

Because the lagging and the use of a random effects model require that a firm has at least 3 years of observations, 6 firms are lost, leaving an unbalanced panel of 319 companies for a total of 1879 firm-year observations. A number of econometric issues arise in this panel.

First, there may be individual firm effects. Because of the relatively small sample of firms from the population of 33/50 polluters also becaus there is a good deal of cross-section data, the individual effects are modeled as random.

Second, as program participation occurred late in 1991, participation effects on pollutions levels are captured only from 1992 onwards. Although participation decisions were pre-determined in these years, there may nevertheless be sample selection bias. Specifically, if the error in the participation equation is correlated with the error in the pollution equation, then using actual participation decisions in the pollution equation leads to biased and inconsistent estimates. For example, due to attributes not observed in the data, 33/50 participants may have been more likely to reduce pollution even had they not joined the program (the endogenous treatment problem identified by Heckman (1978)). The data is allowed to reveal any such correlation by using actual participation decisions and constructing a selection correction (an augmented inverse Mills ratio) to remove any source of inconsistency.[14]

[14] The selection correction is achieved (following Vella, 1998) by constructing the fitted regressor, IMR_{it}, where $IMR_{it} = 0$ for $t \leq$1991 and, for $t \geq$1992,

$$IMR_{it} = P_i \frac{\phi(\hat{\gamma}' w_i)}{\Phi(\hat{\gamma}' w_i)} + (1 - Pi) \frac{-\phi(\hat{\gamma}' w_i)}{(1 - \Phi(\hat{\gamma}' w_i))}$$

where P_i is the participation dummy for firm i, is the estimated parameter vector for the probit estimation

Because participation effects may (or may not) wane over the course of the program, distinct effects are measured for each of the program years 1992-1995. This is done by constructing four participation variables that measure the incremental effect of participation on pollution in a given year; for example, the coefficient on the 1993 participation variable measures the pollution change from 1993 onwards that is attributable to a firm's participation in the 33/50 program.[15]

Third, per Hypothesis 3, firms may make pollution decisions in view of their recent history of government inspections. To test for these effects a firm's lagged inspections-per-facility (LINSPFAC) is used as an explanatory variable.[16]

Finally, because a predicted regressor (the augmented IMR) is included in the regression to obtain consistent parameter estimates, standard error estimates obtained by conventional methods are inconsistent (Murphy and Topel, 1985). To obtain consistent estimates of standard errors, the Murphy-Topel correction procedure is used.[17]

2.4.3. The Inspection Equation

For this equation, the dataset is an unbalanced panel of 1257 facilities over seven years, 1989-1995, yielding 5703 facility-year observations. The dependent variable, facility-level annual inspections, takes a count data form, with discrete and predominantly small

---

of the participation equation, wi is the firm i set of explanatory variables in the participation equation, and $\phi()$ and $\Phi()$ are the normal density and distribution functions respectively.

[15] The four regressors are constructed as follows: if P_t is the participation variable for year t (i.e taking a value of zero for all years other than t), then the regrrssors are $P_\tau^* = \sum_{t=\tau}^{1995} P_t, \tau = 1992, ..., 1995$. These variables are denoted PART92-PART95 (see table 2.1).

[16] Lagging, while logically sensible, avoids any potential problem of joint determination. Because scale effects on pollution are captured by including facility numbers as a regressor, the relevant measure of inspection ativity is a firm's annual inspections-per-facility.

[17] Additionally, bootstrapped standard errors were also computed based on 250 bootstrapped parameter estimates generated by the multi-spage estimation of the pollution equation. The standard error estimates from the bootstrap procedure are not reported but they are broadly consistent with the Murphy-Topel-adjusted results that are reported in table 2.4 (see note 24).

values.[18] To account for thes properties, the dependant variable is assumed to be distributed Poisson, and the individual (random) effect normally distributed; the model is estimated by maximum likelihood.[19] A notable advantage of the random effects specification, relative to the standard or the Fixed effects Poisson models, is that it accommodates over-dispersion.

Third, contemporaneous inspections are posited to depend upon firm performance (pollution and 33/50 program participation)with a lag. For example, program participation decisions were made by firms principally in the last two quarters of 1991, suggesting that any effects on annual government inspections would arise in 1992 and beyond. For these years, there is, in principle, the potential for sample selection bias with respect to 33/50 participation effects, as in the pollution equation. However, in the inspection equation, sample selection, if an issue at all, is expected to bias the estimates against the hypothesized effect (Hypothesis 2 that participation lowers inspection rates). The reason (per Hypothesis 1) is that participants are expected to be those who otherwise experience higher inspection rates. Nevertheless, sample selection bias was accounted for by implementing Terza's 1998 two-step estimator.[20] In doing so, no statistical evidence for sample selection bias is found. The estimation therefore proceeds under a maintained hypothesis of no selection correlation.

[18] In the annual plant-level inspection data, 9.7 percent of observations are zero's, 67.6 percent are ones, 14.5 percent are two's, and 8.2 percent are at least three, giving a clear case of count data.

[19] A Poisson with gamma-distributed random effects was also estimated the qualitative results were similar. Unfortunately, estimation of the model using the Hausman, Hall and Grilliches (1984) random effects negative binomial model did not converged. This is a common problem with the random effects negative binomial (see Cameron and Trivedi, 1998).

[20] To my knowledge, Terza's (1998) is the only known endogenous treatment correction for count data. As in this model, Terza's procedure assumes that the dependent variable is distributed Poisson, with a normal random effect. However, for the purpose of this study, a drawback of this estimator is that it assumes an observation-specific random effect, rather than the firm-specific effect that is posited in this chapter.

2.5. Results

2.5.1. The Participation Equation

Table 2.3 presents selected results from estimation of the participation equation. The dependant variable is the 33/50 program participation dummy. The dataset is a cross-section of 325 firms, with time-varying variables measured as of 1990. The hypothesis that all the slope coefficients are jointly insignificant is rejected. The likelihood ratio test of heteroscedasticity (Harvey 1976) due to firm differences in aggregate 33/50 releases fails to reject the null of homoscedasticity at the 5% level for Models I and II; for Model III, the null of homoskedasticity is rejected thus heteroskedasticiy-corrected estimates are presented instead. Squared variables are denoted by an addition of "^{2}" to the variable and interactions variables are denoted with hyphens.

The three presented models are distinguished by inclusion (or exclusion) of the interaction variables BC-SIERRA, HERF-RD, and FG-RELEASE; the alternate use of lagged 33-50 releases (RELEASE) and employment (LEMP) as measures of firm size; and the alternate use of total lagged inspections (INSP89-90) and its two components, inspections-per-facility (LINSPFAC) and a firm's number of facilities (FAC), to capture inspection effects. In the latter regard, it is expected that a firm's total scale of regulatory interaction is likely to matter in judging potential enforcement rewards to participation in the 33/50 program; hence the effects of facility numbers on participation (in the third model) are interpreted as enforcement-driven.

For regressors deemed to be particularly important determinants of 33/50 participation, their squares are included in the models in order to capture potential non-linearities. For example, one may expect regulatory scrutiny to be particularly acute when a firm is a PRP for a large number of Superfund cites; hence, it is expected that 33/50 participation would increase with the square of the PRP variable. For similar reasons, squared terms for the Sierra Club variable and the measures of firm scale (RELEASE, LEMP, and FAC) are also included.

In all three models, a test for heteroskedasticity was conducted.[21] In Models I and II, homoskedasticity could not be rejected and thus present probit results under this premise. In Model III, however, homoskedasticity was rejected and therefore the reported estimation results are heteroskedasticity-corrected.

Several implications of table 2.3 merit emphasis. First, larger firms with larger 33/50 releases are found to have been more likely to participate in the 33/50 program. These effects are consistent with a number of the theories / hypotheses discussed in Section III. Larger polluters are likely to have been more sensitive to any enforcement benefits of program participation; more able to preempt lobbying for tighter environmental regulations (MLH, 2000); more exposed to potential liability for environmental harm; and more exposed to potential harm from boycott threats.

Second, let us turn to explanatory variables which can distinguish between different hypothesized motives for program participation. Statistically significant (positive) parameter estimates on (1) the enforcement variables (PRP, ENFORCE, INSP89-90, LINSPFAC, FAC), (2) the measures of boycott sensitivity (BC), and (3) per-capita Sierra Club membership (SIERRA), suggest that the potential for implicit enforcement rewards, boycott deterrence, and regulatory preemption (MLH, 2000) were important motives for 33/50 program participation.

Firms with higher levels of R&D were more likely to participate as well. This effect is consistent with a "raising rivals cost" motive for program participation. However,

[21] In testing for heteroskedasticity, standard practice is followed (e.g., Greene, 2000, Chapter 19) by considering a variance that is a squared exponential function of exogenous data. In this study, the exogenous data that posited to possibly drive any heteroskedasticity is the level of prior 33/50 releases.

more research-intensive firms may also have participated because their costs of program obligations (in lowered pollution) were smaller; they could thereby obtain other program benefits (such as enforcement rewards and boycott deterrence) at lower cost. The results provide mixed evidence on the effects of liability law. While the indicator for strict environmental liability (STRICT) has an insignificant effect on participation, its interaction with firm size (STRICT-LEMP) is statistically significant. Hence, to some extent, larger firms with deeper pockets found a liability motive for 33/50 program participation.

However, this work did not find evidence that program participation was spurred by incentives for "green marketing" (with statistically insignificant effects of proximity to final consumers, FG, and its interaction with pollutant releases, FG-RELEASE). Thus, by accounting for three other motives for program participation which are enforcement, boycott deterrence, and regulatory preemption this study comes to a strikingly different conclusion about the impact of "green marketing" incentives than does prior work (AC, KD, VA).[22]

2.5.2. The Pollution Equation

Table 2.4 presents results from estimation of the pollution equation. The dependant variable is RELEASE. The Breush-Pagan LM test of OLS vs. Random Effects rejects the null of OLS. In all model variants, the coefficient on the augmented inverse Mills ratio is statistically significant, providing evidence for sample selection (from program participation decisions) in the predicted direction.

Several qualitative conclusions emerge from table 2.4. First and most important, the results indicate that firms' participation in the 33/50 program tends to lower pollution.

[22] The selection models were also estimated without any boycott, Sierra Club, or enforcement variables (other than PRPS). In these estimations, the interaction variables, FG-LEMP (in Models I and II) and FG-RELEASE (in Model III), had significant positive effects on participation, although FG did not.

These pollution reductions are statistically significant in the first year of program operation (1992), but persist throughout the sample period (to 1995). Second, as in prior work, this study finds that government inspections tend to lower firms' pollution levels. Although inspections have a direct effect on pollution that is statistically insignificant (table 2.4), note that they also indirectly spur pollution reductions by promoting participation in the 33/50 program (table 2.3). Third, firms may have been motivated to lower pollution in order to preempt regulation (with a statistically significant negative coefficient on SIERRA) and/or deter boycotts in states with large environmental constituencies (with a statistically significant negative coefficient on BC-SIERRA). Fourth, the threat of boycott tends to lower pollution, with a statistically insignificant direct effect in addition to an indirect effect - spurring pollution reductions by inducing 33/50 participation (table 2.3). Fifth, the results show some evidence that pollution reductions may have been motivated by research-intensive firms' incentive to "raise rivals costs" (with statistically significant negative coefficients on both firms' R&D expenditures and the measure of industry concentration, HERF). Finally, no statistically significant link is found between pollution and either a firm's proximity to final consumers or the presence of strict environmental liability (although the estimated effect of strict liability is negative, as predicted).[23] Hence, by accounting for potential effects of enforcement activity, boycott deterrence and regulatory preemption incentives, again no evidence that voluntary pollution reduction activity is motivated by incentives for "green marketing" is found.

[23] When using the bootstrap procedure to estimate standard errors (see note 15), the same conclusions are obtained, with two qualifications. Using the bootstrap procedure, parameter estimates for the LRD and BC-SIERRA regressors become statistically insignificant.

2.5.3. The Inspection Equation

Table 2.5 presents results from estimation of the inspections equation. They show that inspections tend to rise when a facility's prior period pollution is higher, with enforcement resources thus targeted to facilities for which inspectors can anticipate good prospects for pollutant reductions. Inspection rates tend to fall when there is more environmentalist pressure on firms, as measured by the Sierra Club variable; environmentalist pressure thus seems to substitute for government inspections in promoting environmental objectives.[24] In addition, inspection rates tend to be higher at the facilities of larger firms (with a statistically significant positive coefficient on the measure of firm size, EMP). However, most important from the estimations is the link between 33/50 program participation and government inspections. Program participation is estimated to have had only a marginal impact on inspection rates in 1992, perhaps because program-sponsored technical assistance took the form of some short-term government oversight. However, program participants experienced statistically and quantitatively significant reductions in their inspection rates from 1993 through 1995 (as indicated by the statistically significant negative coefficient on PART93). To help understand the quantitative significance of these effects, table 2.5 also presents the estimated marginal impacts of 33/50 participation on inspection rates in each of the program years, 1992-1995. It is estimated that a firm's 33/50 program participation translated into a cumulative reduction of .25 inspections over the 1992-1995 period or approximately 17 percent of the sample average inspection rate (1.5 per year). Note also that a firm's benefit of 33/50 participation, in a reduced inspection burden, tends to persist throughout the program years, 1992-1995,

[24] One might conjecture that environmentalism may spur pressure on government agencies for more inspections; the results suggest, in contrast, that government agencies recognize the salutary effects of environmentalism on firm performance and therefore reduce their inspection rates when firms are subject to more environmentalist pressure. Similar logic may explain why the "right to work" variable has a significant positive effect on inspection rates.

even though pollution reduction benefits of participation tend to wane (table 2.4).[25]

2.6. Conclusion

This chapter has studied why firms chose to participate in the EPA's voluntary 33/50 pollutant reduction program; effects that this program had on firms' pollution; and effects of program participation on subsequent government inspection activity. The study finds empirical support for the "enforcement theory" of voluntary pollution reductions (Maxwell and Decker, 2002). Specifically, program participation involves firm investments in environmental auditing and technology that lowers their pollution abatement costs and thereby prompts pollution reductions (the pollution equation effect of program participation). In view of this benefit, environmental authorities implicitly offer regulatory rewards to program participants (the inspection equation effect of program participation) that spurs participation by those firms who have the most to gain from such regulatory rewards (the participation equation effect of prior inspections and pollutant releases).

The results also support the hypotheses that firms participated in the 33/50 program in order to forestall potential boycotts by environmental groups (Baron, 2001; Innes, 2003) and/or to preempt lobbying by these groups for tighter environmental regulation and enforcement (MLH, 2000). Pollutant reductions, beyond those prompted by participation in the 33/50 program, were another means by which firms sought to preempt regulation and boycotts. However, in contrast to prior work that did not account for the potential enforcement, boycott deterrence or regulatory preemption incentives found to

[25] In the Poisson model with a gamma distributed firm effect, estimated impacts of 33/50 participation are similar to those presented in table 2.5, both in magnitude and statistical significance. However, impacts of some other variables are somewhat different. For example, the Sierra Club (SIERRA), boycott (BC), and right-to-work (RTW) variables do not have statistically significant effects in the Poisson-gamma model.

be important here, this work finds no support for the hypothesis that firms participated in the 33/50 program, and/or reduced their pollution levels, in order to obtain any "green marketing" advantages.

Overall, this work lends support to the view that voluntary pollutant reduction programs carefully combined with regulatory/ enforcement rewards for program participation can be useful and effective tools to reduce pollution and save government costs of overseeing firms' environmental performance. Voluntary programs may also offer firms the opportunity to signal their environmental commitment to potential political adversaries and thereby deter costly boycotts and political conflicts. As a result, even when consumer free-riding prevents firms from obtaining any "green premia" in the marketplace, a failure that would otherwise doom voluntary pollution reduction efforts, voluntary environmental programs can succeed.

TABLE 2.1. Variable Definitions

RELEASE	Total firm releases of 33/50 pollutants (millions of pounds)
LRELFAC	Lagged per-facility firm releases of 33/50 pollutants
PART92 - 95	Dummies that equal 1 if a firm is a 33/50 participant (note 16)
INSPECT	Number of a facility's CAA inspections (annual)
DIFREL	Change in total firm releases of 33/50 pollutants from 1988-1990
INSP8990	Number of CAA inspections of firm facilities, 1989-90
ENFORCE	Dummy that equals 1 if firm had an enforcement action in 1989-90
LINSPFAC	Firm lagged inspections per facility (annual)
PRP	Number of Superfund sites for which a firm is a PRP, 1990
SIC28 - SIC38	Dummies for a firm's primary two-digit SIC class
LRD	Lagged firm expenditures on R&D ($millions) (annual)
LEMP	Lagged number of firm employees (1000's) (annual)
FAC	Number of firm facilities (annual)
HERF	Herfindahl index for firm's two-digit SIC class
BC	Dummy that equals one if firm operates in an SIC that was subject to contemporaneous boycott, 1992
FG	Dummy that equals one if firm produces a final good (determined by a firm's primary four-digit SIC class)
SG	Firm percentage sales growth (annual)
SIERRA	Sierra Club members per capita in facility home state (annual), averaged across facilities for the firm
STRICT	Dummy that equals one if facility's home state has a strict liability statute, 1988, averaged for the firm
RTW	Dummy that equals one if facility's home state has a right-to-work statute, 1988, averaged for the firm
SPENDAQP	State expenditures on air quality programs in the facility's home state, 1988, averaged for the firm
LAWYERS	Number of lawyers per capita in facility home state, 1988, averaged for the firm
EDUC	Percentage of college degrees in facility home state population in 1990, averaged for the firm
NONATTAIN	Dummy that equals one if a facility's home county is out of attainment with clean air laws in any year, 1992-1995
CDENSITY	Population density of facility's home county, 1990

TABLE 2.2. Descriptive Statistics

	Participants		Non-participants	
Variable	Mean	std deviation	Mean	std. deviation
DIFREL	-0.1881	0.6243	-0.0576	0.1833
RELEASE	0.8284	1.534	0.1044	0.1722
LEMP	34.4284	71.4741	5.0099	7.1058
HERF	0.4481	0.1443	0.4939	0.1633
PRP	5.4061	9.7499	1.0875	2.2301
ENFORCE	0.4242	0.4957	0.1	0.3009
INSP8990	13.4545	19.9592	2.6	4.7731
SIERRA	2.2982	1.065	2.5442	1.7208
STRICT	0.7588	0.3117	0.7768	0.3836
BOYCOTT (BC)	0.3818	0.4873	0.25	0.4344
FINAL GOOD (FG)	0.6606	0.4749	0.625	0.4856
LRD	211.7544	549.1934	18.3815	46.8655
RTW	0.2984	0.3131	0.2589	0.3972
SPENDAQP	1.1798	0.5806	1.2274	0.6524
LAWYERS	2.8539	0.758	3.2358	1.0209
EDUC	19.9476	2.8079	20.3976	3.4283
SIC 28	0.2121	0.4101	0.125	0.3318
SIC 33	0.097	0.2968	0.0563	0.2311
SIC 34	0.0545	0.2278	0.1063	0.3091
SIC 35	0.1576	0.3655	0.1875	0.3915
SIC 36	0.1273	0.3343	0.1438	0.3519
SIC 37	0.1515	0.3596	0.0625	0.2428
SIC 38	0.0545	0.2278	0.1313	0.3387
No of observations	165		160	

TABLE 2.3. Probit Estimation of the Participation Equation

Hypothesis tested	Variable	Model I estimate	Model I t-ratio	Model II estimate	Model II t-ratio	Model III estimate	Model III t-ratio
Prior reductions	DIFREL	0.6511	1.30	0.6536	1.30	0.3359	3.26
	PRP	-0.1199	-1.46	-0.1158	-1.42	-0.0836	-2.86
	PRP^2	0.0123	1.78	0.012	1.75	0.0095	3.67
	ENFORCE	0.5037	1.96	0.4976	1.94	0.3201	3.45
Enforcement	INSP8990	0.0281	1.78	0.0273	1.73		
effects	LINSP-FAC					0.0442	1.98
	FAC					0.0387	4.75
	FAC^2					-0.001	-11.96
Raising rivals	HERF	1.6854	1.42	1.0804	0.92	1.073	3.56
costs	RD	0.0043	2.42			0.0024	1.68
	HERF-RD			0.0095	2.48		
Preemption of	SIERRA	0.594	1.96	0.5815	1.93	0.4303	2.97
regulations	$SIERRA^2$	-0.0792	-2.17	-0.0802	-2.21	-0.0578	-3
Liability effects	STRICT	-0.3064	-0.94	-0.2753	-0.84	-0.0343	-0.22
	STRICT-LEMP	0.0235	1.79	0.0221	1.71		
Boycott	BC	1.1176	1.67	0.9349	1.70	0.635	4.75
deterrence	BC-SIERRA	-0.082	-0.49				
Green marketing	FG	0.2981	0.81	0.3077	0.83	-0.048	-0.45
	FG-RELEASE					-0.009	-0.05
	RELEASE	1.0593	2.07	1.0879	2.13		
Firm-specific	$RELEASE^2$	-0.1044	-1.27	-0.1047	-1.27		
characteristics	LEMP					0.0123	2.26
	SPENDAQP	-0.0285	-0.13	-0.0288	-0.13	-0.1048	-0.9
	LAWYERS	-0.0971	-0.40	-0.0858	-0.36	-0.2068	-2.07
	EDUC	-0.0158	-0.24	-0.0183	-0.28	0.0118	0.39
	SIC 28	1.7053	2.40	1.7505	2.46	1.0189	4.65
	SIC 33	1.4025	1.82	1.4383	1.86	0.8082	3.95
Industry fixed	SIC 34	0.6709	1.16	0.6941	1.20	0.4876	3.13
effects	SIC 35	0.9262	1.69	0.9549	1.74	0.8343	4.04
	SIC 36	0.163	0.36	0.1683	0.37	0.4449	3.23
	SIC 37	0.0546	0.10	0.0691	0.13	-0.1108	-0.76
	SIC 38	0.9536	1.61	0.962	1.62	0.3442	1.51
	CONSTANT	-2.5068	-2.10	-2.4809	-2.07	-1.8096	-3.89
Chi-square (p-value)		176.74	(0.00)	181.78	(0.00)	182.6	(0.00)
LogL		-139.586		-139.581		-139.0058	
LogL (hetero)		-139.3694		-139.271		-134.5313	
No. observations		325		325		325	

TABLE 2.4. Random Effects Estimation of the Pollution Equation

Hypothesis tested	Variable	Model I estimate	Model I t-ratio	Model II estimate	Model II t-ratio	Model III estimate	Model III t-ratio
	PRP	-0.0195	-0.73	-0.0166	-0.60	-0.0271	-1.17
Enforcement	PRP^2	0.0014	1.83	0.0013	1.65	0.0017	2.43
effects	LINSPFAC	-0.0181	-1.36	-0.0187	-1.29	-0.021	-1.67
	FAC	0.0159	0.55	0.0183	0.55	0.0186	0.78
	FAC^2	0.0011	0.74	0.001	0.59	0.0011	0.94
Raising rivals	HERF	-0.611	-2.90	-0.4283	-1.93	-0.5427	-2.7326
costs	LRD	-0.0011	-8.03			-0.0012	-9.8267
	HERF-LRD			-0.002	-7.56		
Preemption of	SIERRA	-0.121	-2.03	-0.1202	-1.90	-0.1307	-2.36
	$SIERRA^2$	0.0135	1.71	0.0133	1.53	0.0142	1.93
regulations and	BC	-0.1395	-0.40	-0.1125	-0.33		
boycott deterrence	BC-SIERRA			0.0029	0.07		
Green marketing	FG	-0.0874	-0.31	-0.0552	-0.17	-0.0384	-0.13
	FG-LEMP					-0.0048	-1.10
	PART92	-0.2016	-4.68	-0.2265	-4.7511	-0.20	-5.22
Effects of the	PART93	-0.0249	-0.59	-0.0392	-0.92		
33/50 program	PART94	-0.0318	-0.76	-0.0393	-0.94		
	PART95	-0.059	-1.27	-0.0819	-1.73		
Self-selection bias	IMR	0.0888	2.48	0.1127	2.97	0.0649	1.94
	LEMP	0.0271	10.19	0.0252	7.73	0.0309	8.49
Liability effects	$LEMP^2$	-3.30E-05	-8.10	-3.20E-05	-7.36	-3.20E-05	-7.94
	STRICT	-0.0276	0.40	-0.021	-0.30	-0.0301	-0.45
Firm-specific	SG	0.0001	1.05	0.0001	1.07	0.0001	1.04
effects and state	RTW	-0.0889	-1.16	-0.0962	-1.13	-0.0885	-1.22
characteristics	EDUC	-0.0003	-0.01	-0.0008	-0.05	0.0013	0.09
	LAWYERS	0.0389	0.70	0.043	0.66	0.0287	0.53
	SPENDAQP	-0.0287	-0.60	-0.0342	-0.60	-0.0233	-0.51
	SIC 28	-0.1779	-0.51	-0.1675	-0.48	-0.123	-0.40
	SIC 33	-0.1639	-0.34	-0.0833	-0.16	-0.0971	-0.23
Industry fixed	SIC 34	-0.2321	-0.63	-0.1975	-0.54	-0.1628	-0.50
effects	SIC 35	-0.2655	-0.85	-0.2458	-0.79	-0.1902	-0.79
	SIC 36	-0.0955	-0.32	-0.1109	-0.37	-0.1565	-0.62
	SIC 37	0.1935	0.56	0.1939	0.55	0.1339	0.43
	SIC 38	-0.035	-0.08	0.0117	0.027	0.0309	0.08
	TIME	-0.006	-0.85	-0.0029	-0.41	-0.0131	-2.03
	CONSTANT	1.1712	1.50	0.7644	0.95	1.6864	2.36
	No. obs		1879		1879		1879
	R^2		0.34		0.34		0.34
OLS vs RE test statistic			26640.88		2450.93	2634.86	
Chi-square (1)			3.84		3.84		3.84

TABLE 2.5. Random Effects Poisson Estimates of the Inspections Equation

	Variable	Model I estimate	Model I t-ratio	Model II estimate	Model II t-ratio
	PART92	-0.048	-0.524		
Effects of the 33/50 program	PART93	-0.2895	-2. 807	-0.1793	-2.23
	PART94	0.0934	0.889		
	PART95	-0.2619	-2.47		
	SIERRA	-0.1628	-3.917	-0.1801	-4.526
	BC	0.3086	1.515	0.5282	2.827
	NONATTAIN	-0.1844	-1.583	-0.0892	-0.82
	CDENSITY	0.0019	0.698	-0.0008	-0.293
	LRELFAC	0.00012	2.086	0.0002	3.296
	LEMP	0.0033	6.615	0.0025	4.575
Firm-specific and County characteristics	SPENDAQP	0.5973	5.972	0.629	6.807
	RTW	0.2785	2.268	0.1975	1.639
	EDUC	0.0042	0.136	0.0013	0.047
	STRICT	-0.1409	-1.109	-0.1106	-0.942
	LAWYERS	-0.0967	-0.856	-0.096	-0.941
	SIC 28	-0.6679	-3.352	-0.4923	-2.887
	SIC 33	-0.8669	-4.059	-0.6948	-3.718
Industry fixed effects	SIC 34	-1.0863	-5.1	-0.7484	-4.033
	SIC 35	-0.7271	-3.44	-0.5534	-3.084
	SIC 36	-1.3818	-6.212	-1.4104	-6.475
	SIC 37	-0.8024	-4.784	-0.9492	-5.739
	SIC 38	-1.0381	-3.147	-0.7565	-2.574
	CONSTANT	-9.4063	-3.146	-6.5026	-2.823
	TIME	0.0865	2.678	0.0537	2.184
	Number of Observations		5703		5703
	Log-likelihood		-4026.99		-4031.17

Marginal Effects of Participation for Each Year

Year	Marginal effect	t-ratio
1992	-0.0348	-0.852
1993	-0.1698	-3.436
1994	-0.1255	-2.269
1995	-0.2469	-3.436

3. Nonparametric Regression under Alternative Data Environments

3.1. Introduction

Let (X_{ij}, Y_{ij}), $i = 1, ..., n_j, j = 1, ..., J$ be a sample of R^{p+1} valued random vectors where Y_{ij} represents a response variable and X_{ij} is a p-dimensional vector of explanatory variables. In many empirical situations it is necessary to estimate a set of regression curves, say one for each experimental unit of interest, which can be arranged as

$$Y_{ij} = m_j(X_{ij}) + \epsilon_{ij} \tag{3.1}$$

where j denotes the j^{th} experimental unit, and ϵ_{ij} is a zero-mean and finite-variance error process. This research is concerned with the estimation of the conditional mean $E(Y|X = x)$. Kernel regression estimators based solely on individual samples such as the Nadaraya-Watson and the locally linear kernel estimators have become widespread because they circumvent the risk of functional misspecification inherent to their parametric counterparts and provide consistent estimates under mild regularity conditions.

The standard Nadaraya-Watson estimator of the conditional mean $m_j(x)$ is given by

$$\tilde{m}_j(x) = \frac{\sum_{i=1}^{n_j} Y_{ij} K_{h_j}(X_{ij} - x)}{\sum_{i=1}^{n_j} K_{h_j}(X_{ij} - x)} \tag{3.2}$$

where h_j is the smoothing parameter and $K_{h_j}(u) = \frac{1}{h_j} K(\frac{u}{h_j})$ with $K(u)$ being the kernel function. Denoting $\mu_2 = \int z^2 K(z) dz$ and $\mathrm{R}(K) = \int K^2(z) dz$, the standard properties of the Nadaraya-Watson estimator are

$$E[\tilde{m}_j(x) - m_j(x)] = \frac{1}{2}\mu_2 h_j^2 \{m_j''(x) + 2m_j'(x)\frac{f_j'(x)}{f_j(x)}\} + o(h_j^2), \tag{3.3}$$

$$\mathrm{Var}[\tilde{m}_j(x)] = \frac{\sigma^2 R(K)}{(nh_j) f_j(x)} + O(h_j/n_j) \tag{3.4}$$

where $f_j(x)$ is the marginal density function of X_{ij} evaluated at support point x. Since the bias is $O(h_j^2)$ and $h_j = h_j(n_j)$ goes to 0 as n_j goes to ∞, it follows that the Nadaraya-Watson estimator is consistent. However, a drawback is its finite sample bias which can be quite large. Several papers have proposed estimators which reduce the bias (Hardle and Browman, 1988; Hjort and Glad, 1995; Glad, 1998, among others). One such bias-correction estimator, Hjort and Glad (1995), is of particular interest because of the ease with which it can be implemented. Hjort and Glad (1995) propose a semiparametric estimator which combines a parametrically estimated pilot with a nonparametrically estimated correction factor. The parametric pilot can be thought of as a prior for the shape of $m_j(x)$ whereas the correction factor adjusts the pilot if it does not satisfactorily capture the shape of $m_j(x)$. Consequently, the estimator behaves like the parametric start if the parametric assumption is correct, while resembling the nonparametric estimator otherwise.

The estimator proposed in this chapter is in the same realm as Hjort and Glad's (1995); however, alternative data environments are considered where the researcher has data from possibly similar regression functions. If those unknown functions are identical, the optimal estimator would pool the data and estimate one regression curve. If, however, those unknown functions are sufficiently similar, using the pooled estimator as a pilot in Hjort and Glad's framework would yield efficiency gains relative to the Nadaraya-Watson estimator. The use of extraneous data in the form a nonparametric pooled start represents the key conceptual difference between the proposed estimator and the estimator of Hjort and Glad (1995).

Two empirically relevant data environments are considered. The first data environment assumes that at least one of the explanatory variables is discrete. While this situation is easily accommodated in a parametric framework, the continuity assumptions required for nonparametric regression are violated. As a result, a separate nonparametric regression estimation is required for each discrete value. For example, if one of the explanatory variables is discrete and may take values $\{0, 1, 2, 3\}$, the sample must be partitioned according to the four discrete values into four cells where a separate regression function is undertaken for each. Recently, Racine and Li (2004) developed a nonparametric estimator that smoothes across the discrete values, thereby reducing variance at a cost of increased bias. Conversely, the proposed estimator attempts to reduce bias by utilizing the entire data set. The second data environment assumes that one needs to estimate a set of regression curves rather than a single regression curve. Empirically, this situation arises often and led Altman and Casella (1995) to develop a Stein-type Bayesian nonparametric estimator that uses empirical Bayes techniques pointwise across the function space to reduce estimation error. This latter data environment can be viewed as a generalization of the former with each of the discrete cells representing an experimental unit.

The remainder of this chapter is organized as follows. The second section introduces the proposed estimator and investigates its asymptotic properties. The third section presents the simulation results. The final section summarizes the findings.

3.2. A Nonparametric Estimator with a Pooled Start

Underlying the proposed estimator is that there exists a prior belief that the conditional means are similar in shape. If the curves were identical, that is, if $m_1(x) = m_2(x) = ...m_J(x) = m(x)$, one would simply pool the data and estimate a common curve. Conversely, if the conditional means were dissimilar, the pooled estimator is inappropriate. A primary strength of the proposed estimator is that the form or extent of similarity among the curves is not required; in most empirical applications the form or extent of similarity is unknown. The Hjort and Glad estimator is adapted to the context of model (3.1) by combining pooled and individual nonparametric estimators. As a result, the proposed estimator, which is denoted the nonparametric estimator with a pooled start (NEPS), resembles the pooled estimate if the curves are identical or similar and the individual (Nadaraya-Watson) estimate if the curves are dissimilar. The NEPS estimator of conditional mean $m_j(x)$ is

$$\hat{m}_j(x) = \hat{m}_p(x)\hat{r}_j(x) = \frac{\sum_{i=1}^{n_j} Y_{ij}[\frac{\hat{m}_p(x)}{\hat{m}_p(X_{ij})}]K_{h_j}(X_{ij}-x)}{\sum_{i=1}^{n_j} K_{h_j}(X_{ij}-x)}. \tag{3.5}$$

The estimator is implemented in two steps. The first step pools the data from all experimental units to estimate a single curve denoted $\hat{m}_p(x)$. This step introduces extraneous information from the pooled dataset that is potentially relevant to the estimation of the conditional mean of interest. The second step consists of multiplying the pooled estimate by a nonparametrically estimated correction factor $\hat{r}_j(x)$ to account for individual effects. The NEPS estimator is designed to outperform the standard Nadaraya-Watson estimator when the hypothesis of similarity is tenable, but also produce reliable estimates when the curves are dissimilar.

Note that both the pooled and the individual functions can be estimated by a higher order kernel-weighted polynomial, for example the locally linear kernel (degree 1). The locally linear estimator is preferable in most applications as it does not have the boundary bias problem of the Nadaraya-Watson estimator (degree 0). In the simulations results to follow, the locally linear kernel is used to estimate both the pooled pilot and the individual estimators. However, the asymptotic results are presented only for the Nadaraya-Watson estimator (see for example Ruppert and Wand (1995) for asymptotic properties of the higher order kernel-weighted polynomial).

3.2.1. Asymptotic Properties of the NEPS Estimator

In deriving the asymptotic properties of the NEPS estimator, the following assumptions are required:

A1. The X_{ij}s are i.i.d. and independent of the error process ϵ_{ij}, which is also i.i.d.

A2. The density function $f_j(x)$ and the conditional mean $m_j(x) \in \mathcal{C}^2(\Theta)$ with finite second derivatives and $f_j(x) \neq 0$ in Θ, the neighborhood of point x.

A3. The kernel function $K(z)$ is bounded, real-valued, with the following characteristics: (i) $\int K(z)dz = 1$, (ii) $K(z)$ is symmetric about 0, (iii) $\int z^2 K(z)dz < \infty$, (iv) $|z|K(|z|) \to 0$ as $|z| \to \infty$, (v) $\int K^2(z)dz \leq \infty$.

A4. $h_j \to 0$ and $n_j h_j \to \infty \ \forall \ j = 1, ..., J$.

A5. It is assumed that $h_p \to 0$ and $n_j h_p \to \infty \ \forall \ j = 1, ..., J$ where h_p is the smoothing parameter for the pooled estimator.

A6. $E\left|\frac{m_p(x)}{m_p(X_i)}\right|^{2+\delta}$, $E|\epsilon_{ij}|^{2+\delta}$, and $\int |K(\omega)|^{2+\delta}$ are finite for some $\delta > 0$.

Proposition 3.1

1. Under assumptions A1-A4, the mean and variance of the proposed estimator are:

$$E[\hat{m}_j(x) - m_j(x)] = \frac{1}{2}\mu_2 h_j^2\{r_j''(x) + 2r_j'(x)\frac{f_j'(x)}{f_j(x)}\}m_p(x) + o(h_j^2) \tag{3.6}$$

$$\mathrm{V}ar[\hat{m}_j(x)] = \frac{\sigma^2 R(K)}{(nh_j)f_j(x)} + O(h_j/n_j + (Nh_p)^{-1}). \tag{3.7}$$

where $N = \sum_{i=1}^{J} n_j$.

2. Under assumptions A1-A6, $\hat{m}_j(x)$ has a limiting normal distribution

$$\sqrt{n_j h_j}(\hat{m}_j(x) - m_j(x)) \to N(B(h_j), \Sigma_j) \tag{3.8}$$

where $B(h_j) = \frac{1}{2}\mu_2 h_j^2[m_p(x)r_j''(x) + 2m_p(x)r_j'\frac{f_j'(x)}{f_j(x)}]$ and $\Sigma_j = \frac{\sigma^2}{f_j(x)}R(K)$.

Proof: See appendix.

Equations 3.4 and 3.7 show that the variances of the Nadaraya-Watson estimator and the NEPS estimator differ by $O(\frac{1}{n_1 h_p + n_2 h_p + \ldots\ldots + n_J h_p})$, which is negligible by A5. The bias of the NEPS estimator is not a function of the slope and curvature of the true regression function as it is for the Nadaraya-Watson estimator (see equation 3.3). Rather, the bias is a function of the slope and second derivative of the correction factor $r_j(x)$. If the nonparametric pilot m_p coincides with or is proportional to the true function m_j, then $r_j(x)$ will be a straight line and $r_j' = r_j'' = 0$. This implies that the leading terms of the bias will vanish. Similarly if $m_p(x)$ and $m(x)$ are sufficiently similar, the correction factor will be less variable than the individual conditional mean, hence leading to bias reduction. Interestingly, the pooled start does not have to be a good approximation of $m_j(x)$ for the NEPS estimator to remain competitive to the Nadaraya-Watson estimator in moderate samples.

3.2.2. Computational Issues

The ratio $\frac{\hat{m}_p(x)}{\hat{m}_p(X_{ij})}$ can be highly influential in regions when X_{ij} is far from x. Also, it is possible that $\hat{m}_p(X_{ij})$ and $\hat{m}_p(x)$ have different signs. Following Hjort and Glad (1995), the ratio $\frac{\hat{m}_p(x)}{\hat{m}_p(X_{ij})}$ should be substituted by

$$\left|\frac{\hat{m}_p(x)}{\hat{m}_p(X_{ij})}\right|_{\frac{1}{10}}^{10},$$

that is, truncating values below $\frac{1}{10}$ and above 10 to make the estimator robust to these local effects. Additionally, when the number of curves J is large, selecting the "optimal" extraneous data to be included in the nonparametric pilot is not trivial. This problem is analogous to the choice of instruments in instrumental variable estimation when the number of instruments is large or the choice of the functional form of the parametric pilot in Hjort and Glad (1995). A cross-validation procedure to select the extraneous data to be included in the pooled start can be used. The cross-validation procedure consists of alternating the pooled start from the set formed by the J curves, for a total of 2^J possible pooled guides, and then choosing the one whose loss function is the lowest.

3.2.3. Smoothing Parameter Considerations

Implementation of the proposed estimator requires the selection of two smoothing parameters. There are many ways to choose the smoothing parameters. Both h_p and h can be chosen by minimizing the asymptotic mean integrated squared error (AMISE) of the proposed estimator. The smoothing parameters can also be selected by cross-validation, either sequentially or simulteneously. Sequential selection of the smoothing parameters consists of first choosing h_p to minimize the cross-validation function of $\hat{m}_p(x)$ and then picking h to minimize the cross-validation function of $\hat{m}_j(x)$. Alternately, h_p and h can be chosen simultaneously by minimizing the cross-validation function of $\hat{m}_j(x)$:

$$\sum_{i=1}^{n_j}(\hat{m}_j^{-i}(x) - m_j(x))^2$$

where $\hat{m}_j^{-i}(x)$ is the usual "leave-one-out" estimator of $\hat{m}_j(x)$. It is noted that the latter approach is preferrable and, although more computationally intensive, is easily applicable in empirical work.

3.3. Finite Sample Simulations

This section conducts Monte Carlo simulations to investigate the empirical applicability of the NEPS estimator compared to the Nadaraya-Watson and other related estimators. Prior to the simulation results a terse review of two related estimators is provided.

3.3.1. The Racine and Li Estimator

The objective of the Racine and Li estimator is to nonparametrically estimate regression functions with discrete independent variables without having to partition the data. Suppose the researcher has data on one experimental unit: Y_i a scalar response variable, X_i^c a vector of continuous variables, and X_i^d an r-dimensional vector of discrete regressors. The Racine and Li estimator smoothes the continuous variables by a c-variate kernel while the discrete variables are smoothed as follows

$$S(X_{it}^d, x_t^d, \lambda) = \begin{cases} 1 & \text{if } X_{it}^d = x_t^d \\ \lambda & \text{otherwise }, 0 \leq \lambda \leq 1 \end{cases} \tag{3.9}$$

where X_{it}^d is the t^{th} component of the vector X_i^d. The Racine and Li estimator is

$$\tilde{m}^{RL}(x^c, x^d) = \frac{\sum_{i=1}^{N} Y_i W_{h,\lambda}(X_i^c, x^c, X_i^d, x^d)}{\sum_{i=1}^{n} W_{h,\lambda}(X_i^c, x^c, X_i^d, x^d)} \tag{3.10}$$

where $W_{h,\lambda}(X_i^c, x^c, X_i^d, x^d) = K_h(X_i^c - x^c) \prod_{t=1}^{r} S(X_{it}^d, x_t^d, \lambda)$.

In a context of multiple curve estimation as laid out in equation (3.1), the "discrete" smoother $S(., \lambda)$ controls the inclusion of extraneous information by assigning a weight of 1 to observations belonging to the experimental unit of interest and a weight of λ to observations from the remaining experimental units. The boundedness of λ within the unit interval allows the Racine and Li estimator to nest both the pooled (λ=1) and Nadaraya-Watson (λ=0) estimators.

3.3.2. The Altman and Casella Estimator

The Altman and Casella model assumes a fixed and balanced design for the predictor variable so that (3.1) can be rewritten as $Y_{ij} = m_j(X_i) + \epsilon_{ij}$ with $X_i = i/n_j$. It is also assumed that each curve can be written as $m_j(X_i) = m(X_i) + \eta_j(X_i)$; that is, the curve for experimental unit j at design point X_i is the population mean curve plus a term which captures the deviation from the population mean curve. Underlying this last assumption is the fact that the curves are all sampled from the same population. Denote $\tilde{m}_j$ the nonparametric estimate of m_j. Given that $\tilde{m}_j$ is biased, it can be expressed as $\tilde{m}_j = \phi_j + \upsilon_j$ where υ_j is an error term such that $E[\upsilon_j] = 0$ and $Var[\upsilon_j] = \alpha^2/n$. Altman and Casella form a hierarchical model ($\tilde{m}_j|\phi_j$ is normally distributed, and ϕ_j and m_j are jointly normally distributed) and derive the posterior mean of m_j as

$$\tilde{m}_j(x) = \bar{m}(x) + \alpha(x)[\tilde{m}_j(x) - \phi(x)]. \tag{3.11}$$

In practice, the hyperparameters are replaced by sample estimates, which leads to the Altman and Casella estimator for experimental unit j

$$\tilde{m}_j^{AC}(x) = \bar{y}_x + \tilde{\alpha}(x)[\tilde{m}_j(x) - \overline{\tilde{m}}(x)] \tag{3.12}$$

where $\bar{y}_x = \frac{1}{J}\sum_{j=1}^{J} y_{xj}$ is the cross-individual sample mean of the data at design point x, $\tilde{\alpha}(x) = \frac{\tilde{\sigma}_{y(x)m(x)}}{\hat{\sigma}^2_{\tilde{m(x)}}}$ is the ratio of the covariance between the data and the nonparametric estimates and the variance of the nonparametric estimates, and $\overline{\tilde{m}}(x) = \frac{1}{J}\sum_{j=1}^{J}\tilde{m}_j(x)$. The reader is directed to Altman and Casella (1995) for a complete derivation of their model. Note that this estimator uses the data from the other experimental units in the population in the regression of the curve of interest through $\overline{\tilde{m}}(x)$ and $\bar{y}_x$. If the individual curves are similar, then $[\hat{m}_i(t) - \overline{\tilde{m}}(t)]$ goes to zero and the final estimates behave like $\bar{y}_x$ which is unbiased for the population mean curve. Altman and Casella note that their estimator performs better when the number of experimental units is sufficiently large so that $\bar{y}_x$ provides a good approximation to the population mean.

3.3.3. Simulation Design

In the first experiment a random design regression is considered whith the explanatory variable being uniformly distributed on the [0,1] interval. The second experiment forces the explanatory variable to be equispaced on the [0,1] interval as required by the Altman and Casella estimator. For each experiment two scenarios are investigated. In the first scenario, which is referred to as the "case of identical curves," four identical curves were generated: $m_1(x) = m_2(x) = m_3(x) = m_4(x) = sin(5\pi x)$. Individual-specific errors differentiate the data across experimental units. This is the ideal case for the NEPS estimator. In the second scenario, which is referred to as the "case of dissimilar curves," four very dissimilar curves were generated (see figure 3.1). The four curves are

$$m_1(x) = sin(15\pi x); \tag{3.13}$$

$$m_2(x) = sin(5\pi x); \tag{3.14}$$

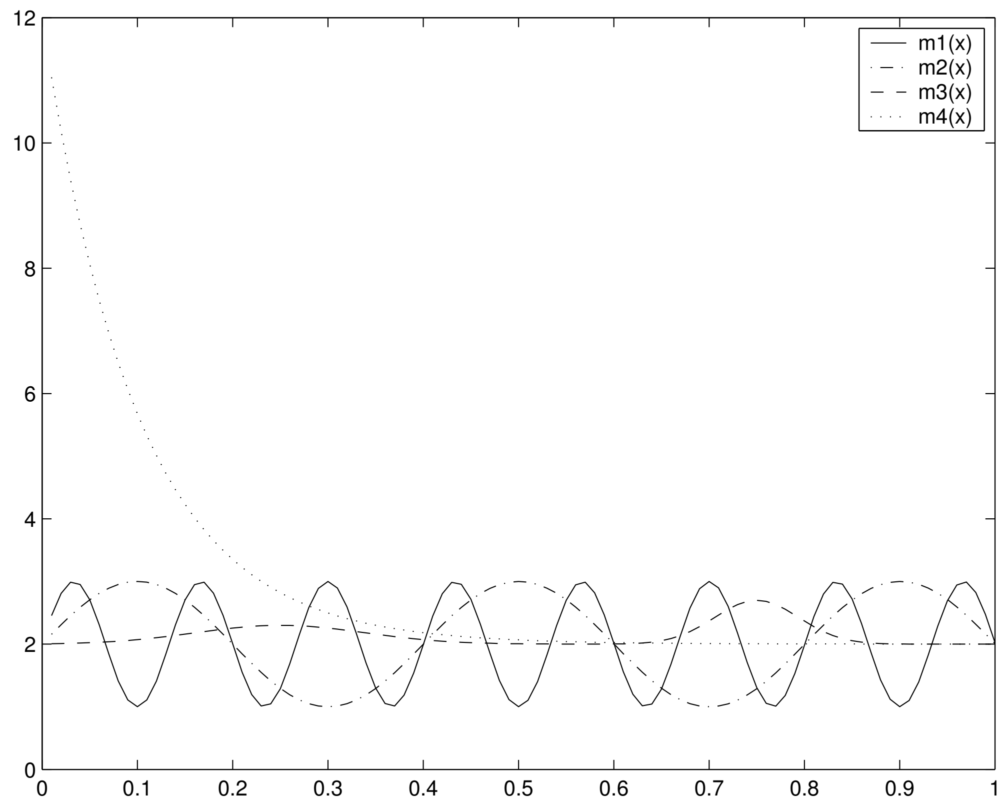
$$m_3(x) = .3e^{(-64(x-.25)^2)} + .7e^{(-256(x-.75)^2)}; \text{ and} \tag{3.15}$$

$$m_4(x) = 10e^{-10x}. \tag{3.16}$$

Unlike in density estimation where the Marron and Wand densities (1992) are commonly used to study the finite sample performance of density estimators, there are no standard test functions in the regression case. However, the curves used here have also been employed in similar simulations (Hurvitch and Simonoff, 1998; Ruppert, Sheather, and Wand, 1995; Herrmann,1997). The choice of these two extreme scenarios is motivated by the fact that in empirical settings it is impossible to know if the conditional means are similar, identical, or dissimilar.

Following Hurvitch and Simonoff, for both scenarios, the error term for individual j is assumed to be normally distributed with mean zero and variance equal to $0.25 \times$ (range of $(m_j(x))$. Throughout the simulations, a Gaussian kernel is used. For computational expediency, the smoothing parameters for the pooled and the proposed estimator are chosen by sequentially minimizing the integrated squared errors of $m_p(x)$ and $m_j(x)$, with the comfort that the proposed estimator would have performed at least as well had we chosen the smoothing parameters simultaneously.

FIGURE 3.1. Graph of the Four Conditional Means



3.3.4. Results

Tables 3.1 and 3.2 report the average mean integrated squared error (MISE) of the four curve estimates respectively for the random and fixed designs for samples sizes 50, 100, and 500 with 500 simulations. The average mean integrated squared bias (MIB^2) is also reported as the NEPS estimator is designed to reduce bias.

Table 3.1 presents the results of the first simulation experiment where a random design regression is used. LLK denotes the locally linear kernel estimator, R&Li is the Racine and Li estimator, and NEPS is the proposed estimator. For the "case of similar curves" the NEPS estimator significantly outperformed the LLK estimator in all sample sizes. The superior performance of the NEPS estimator is attributable to a lower bias as seen in table 3.1, confirming the derived theory. The R&Li estimator also outperformed the LLK estimator but not to the extent of the NEPS. Interestingly, the NEPS estimator also has a lower MISE than the LLK estimator for the sample sizes of 50 and 100 in the "case of dissimilar curves." An intuitive explanation of this somewhat surprisingly good performance is that the LLK estimator is a special case of the NEPS estimator with $m_p(x)$ being equal to a constant $\forall\ x$. However, a "flat start" is quite conservative for most curves, including those curves considered in these simulations. Therefore $\hat{m}_p(x)$ need not be a great approximation of the conditional mean of interest for the NEPS estimator to perform well. This result was also found by Hjort and Glad (1995) and Glad (1998) and represents a strength of their idea. Formally, if $m_p(x)$ is such that $|r_j''(x)m_p(x) + 2r_j'(x)m_p(x)\frac{f_j'(x)}{f_j(x)}| < |m_j''(x) + 2m_j'(x)\frac{f_j'(x)}{f_j(x)}|$, the NEPS estimator will have a smaller asymptotic mean squared error than the Nadaraya-Watson estimator as the variances are essentially the same for large J. The R&Li estimator remained competitive because of its ability to revert to the LLK estimator by having $\hat{\lambda}_j \to 0$ when the curves are dissimilar.

Table 3.2 reports the results of the second experiment where a fixed design is used. A&C denotes Altman and Casella's nonparametric empirical Bayes estimator. The NEPS estimator outperformed the LLK estimator and the A&C estimator when the conditional means are identical. As in the random design case, the NEPS remained competitive to the LLK estimator for the samples sizes of 50 and 100 even when the similarity assumption is inappropriate. The performance of the A&C estimator is somewhat disappointing, which could be explained by the small number of experimental units ($J = 4$) considered in the simulations. Altman and Casella (1995) noted that J needs to be large for their estimator to perform well relative to the LLK estimator. In general empirical Bayes estimators require a large number of experimental units and few design points per unit to perform well.

3.4. Conclusion

This chapter has proposed a computationally simple nonparametric regression method which admits two empirically relevant data environments. The method was designed to achieve bias reduction by incorporating extraneous information from curves which are thought to be similar to the curve of interest. Consistent with the derived theory, the simulation results indicate that the NEPS estimator has a strong practical potential in small to moderate samples. It outperformed the LLK estimator when the curves were identical and did not lose much efficiency when the curves were very dissimilar. The proposed estimator also performed admirably against the related estimators of Racine and Li and Altman and Casella.

TABLE 3.1. Average Estimation Error of the Four Furves: Case of a Random Design

Case of similar curves

n	LLK		R&Li		NEPS	
	MISE	MIB^2	MISE	MIB^2	MISE	MIB^2
50	10.644	5.6111	3.8884	2.6365	2.6987	0.3431
100	5.7383	3.5223	2.3816	1.7222	1.4638	0.1710
500	1.9022	1.3000	0.5883	.3064	0.3915	0.0553

Case of dissimilar curves

n	LLK		R&Li		NEPS	
	MISE	MIB^2	MISE	MIB^2	MISE	MIB^2
50	18.5100	14.3130	20.3510	16.2460	18.1560	12.7090
100	14.8060	12.3190	15.5110	12.9280	14.5970	11.0740
500	12.2757	11.3895	15.1869	14.8708	13.4409	12.2039

Table 3.2. Average Estimation Error of the Four Curves: Case of a Fixed Design

Case of similar curves

n	LLK		A&C		NEPS	
	MISE	MIB^2	MISE	MIB^2	MISE	MIB^2
50	5.6786	1.5812	7.8547	0.0152	2.7299	0.3296
100	3.1335	0.6486	7.4178	0.0149	1.5254	0.1980
500	0.8969	0.2003	6.6718	0.2652	0.6154	0.0273

Case of dissimilar curves

n	LLK		A&C		NEPS	
	MISE	MIB^2	MISE	MIB^2	MISE	MIB^2
50	12.4690	8.9958	18.3277	9.1848	11.9579	7.3078
100	4.4908	2.0037	11.1970	3.1989	5.1113	1.7449
500	1.1113	0.2649	11.1993	4.6252	2.7831	0.3238

4. Nonparametric Estimation of the Short Rate diffusion from a Panel of Yields

4.1. Introduction

Substantial efforts have been devoted to modeling the short-term interest rate, which is generally believed to be the most important state variable driving the dynamics of interest rate term structure. Continuous-time univariate models of short rate, r_t, are typically specified as the following time-homogenous Itô process:

$$dr_t = \mu(r_t)dt + \sigma_t(r_t)dw_t$$

where w_t is the standard Brownian motion with $t \in [0, T]$. Most existing interest rate models are nested in the parametric specification of Aït-Sahalia (1996b):

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1})dt + \sigma r_t^{\gamma} dw_t;$$

where both the drift function and diffusion function are specified to capture potential nonlinearities. Various restrictions on the parameters of the above model lead to the Vasicek (1977) model ($\alpha_2 = \alpha_3 = 0, \gamma = 0$), the Brennan and Schwartz (1980), the Courtadon (1982) model ($\alpha_2 = \alpha_3 = 0, \gamma = 1$), the Cox Ingersoll Ross (CIR) (1985) model ($\alpha_2 = \alpha_3 = 0, \gamma = 1/2$), the Cox, Ingersoll and Ross (1980) model ($\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 0$, $\gamma = 3/2$), the CEV models of Chan, Karolyi, Longstaff and Sanders (1992) ($\alpha_2 = \alpha_3 = 0$), etc.

While the main appeal of the parametric models is the tractability and potential closed form bond prices and prices of interest rate derivative securities, the main concern about these models is the risk of model misspecification. Empirical studies have provided evidence rejecting most popular parametric diffusion models of the short rate, see e.g.

Chan, Karolyi, Longstaff and Sanders (1992), Aït-Sahalia (1996b), Andersen and Lund (1997), and Hong and Li (2005) among others. Further analysis also confirms that mispecified models can have significant economic implications on the pricing of interest rate derivative securities, see e.g. Backus Foresi and Zhin (1995) and Canabarro (1995). For the above reasons, nonparametric modeling of the short rate dynamics has received considerable attention in the finance literature in recent years. In a pioneering work, Aït-Sahalia (1996a) proposes a nonparametric estimator of the diffusion function based on a parametrically specified drift function. Jiang and Knight (1997) propose a nonparametric kernel estimator of the diffusion function, and a nonparametric estimator of the drift function using the diffusion function estimator as well as the nonparametric marginal density estimator. Nicolau (2003) refines the estimators in Jiang and Knight (1997) to reduce the finite sample bias of the diffusion function estimator. Stanton (1997) proposes nonparametric estimators of the drift and diffusion functions based on various orders of approximation of the Itô process. Bandi and Phillips (2003) generalize Stanton's nonparametric approach to the situation of recurrent diffusion processes, circumventing the necessity of a stationary density.

The empirical evidence from the model specification tests in Aït-Sahalia's (1996b) and nonparametric drift function estimates in Stanton (1997) and Jiang (1998) suggest that the drift function of the short rate is nonlinear. In particular, the nonparametric test in Aït-Sahalia (1996b) provides evidence that the linear drift function of popular parametric models is the principal source of model misspecification. The nonparametric drift function estimates in both Stanton (1997) and Jiang (1998) share the feature that the short-term interest rate exhibits very little mean reversion, that is behaves like a random walk for interest rate below 15% but has a dramatically high mean reversion at higher levels of the short rate. Conley, Hansen, Luttmer, and Scheinkman (1997) report similar results where the estimated drift function is nonzero only for rates below 3% or above 11%. The findings of a nonlinear drift have, however, been challenged by Pritsker (1998) and Chapman and Pearson (2000) among others. Pritsker (1998) shows that the nonparametric test in Aït-Sahalia's (1996b) has a poor finite sample performance because of persistent dependence in daily interest rates resulting in a slow convergence of the nonparametric density function estimator. The Monte Carlo simulations in Chapman and Pearson (2000) show that Stanton's drift function estimator can produce spurious nonlinearities even when the underlying drift function is truly linear. What is troublesome is that the spurious nonlinearlity has a similar pattern as in the empirical application in Stanton (1997). Chapman and Pearson argue that a combination of the truncation of the "observed" short rates and a finite sample creates artificial patterns of nonlinearity near the boundaries of the support. Abhyankar and Basu (2001), and Li, Pearson and Poteshman (2004) provide formal results supporting Chapman and Pearson's explanation. They show that if the truncation of the observed short rate process is accounted for, the resulting drift is nonlinear even if the drift of the unrestricted process is linear. Using a Bayesian approach, Jones (2003) shows that the nonlinearity of the estimated drift is contingent on the assumed prior distribution and the stationarity or lack thereof of the interest rate process. The assumption of stationarity and the use of "flat" priors contribute strongly to the finding of nonlinear mean reversion.

The ongoing debate about the linearity or nonlinearity of the drift function underscores the difficulty of identifying and estimating the drift function. This chapter applies the nonparametric method developed in the previous chapter to the estimation of the short rate diffusion process with a focus on the drift function estimation. The key conceptual difference between the proposed estimator and existing parametric and nonparametric estimators of the short rate diffusion is that the new estimator uses extraneous interest rate observations beyond the short rate. That is, it uses information from a panel of yields of different maturities instead of a single time series of short rates. Intuitively, if the drift functions of interest rates with different maturities are identical or sufficiently close, then the efficient estimator of the drift function would pool the data of all the interest rates. If the drift functions are dissimilar, using the pooled estimator as a pilot may still yield efficiency gains relative to the ordinary estimator which relies only on the short rate observations.

It is known in the literature that while increasing sampling frequency is helpful for the identification and estimation of the diffusion function, it is the increasing of sampling period that is crucial for the identification and estimation of the drift function. Evidence in Pritsker (1998), Chapman and Pearson (2000) and Jones (2003) suggests that as a result of the strong persistence of interest rates, identifying the drift function requires a long sampling period. Thus, the pooled data offers more incremental information about the drift function than the diffusion function. Evidence from the simulations herein suggests that the proposed estimator significantly attenuates the spurious nonlinearities of the Stanton drift function estimator. It is noted that the proposed estimator is particularly appealing for the estimation of short rate diffusion process as the bond yields of different maturities are available along the yield curve. Moreover, while the yields of different maturities may be determined by different economic factors, they tend to be highly correlated with systematic co-movements. This is because the short rate is the most important factor driving the dynamics of the term structure.

To further investigate the short rate diffusion process, the proposed method is applied to US data using a time series of 50 years of daily 3-month T-bill yields. This is compared to the typical 20 to 30 years of data used in most studies. Additionally, five time series of yields with maturities ranging from 6 months to 10 years are also employed to implement the proposed estimator. Each of these additional series has 42 years of daily observations. The empirical results suggest that the short rate drift function is nonlinear at high levels of interest rate. However, the level of mean reversion is significantly weaker than that documented in Stanton (1997) and Jiang (1998).

The remainder of the chapter is structured as follows. The next section outlines the proposed estimator of the drift function and derive its asymptotic properties. Section II performs Monte Carlo simulations to assess the finite sample properties of the proposed estimator, in comparison with the Stanton (1997) estimator. In section III, an empirical application of the proposed estimator is undertaken using US interest rate data. Economic implications of alternative short rate drift functions are also examined using simulated bond prices and prices of interest rate derivative securities. Section IV concludes.

4.2. Nonparametric Estimation of the Short Rate Diffusion from a Panel of Yields

Consider the following one-factor diffusion model for the short rate $\{r_t^{(1)}, t \geq 0\}$:

$$dr_t^{(1)} = \mu_1(r_t^{(1)})dt + \sigma_1(r_t^{(1)})dw_t^{(1)}, \qquad \text{(The model)} \tag{4.1}$$

where $w_t^{(1)}$ is a standard Brownian motion, $\mu_1(.)$ and $\sigma_1^2(.)$ are respectively the drift and diffusion functions. In addition to the above short rate, let us assume that there are additional $J-1$ interest rates $\{r_t^{(j)}, t \geq 0\}, j = 2, ..., J$ which follow the diffusion processes:

$$dr_t^{(j)} = \mu_j(r_t^{(j)})dt + \sigma_j(r_t^{(j)})dw_t^{(j)}, \qquad \text{(The auxiliary models)}, \tag{4.2}$$

where $\mu_j(.)$ and $\sigma_j^2(.)$ are respectively the drift and diffusion functions of $r_t^{(j)}$, and the standard Brownian motions $w_t^{(j)}, j = 1, \cdots, J$, are potentially correlated. The short rate model in (4.1) is termed as "The model" since it is the model of interest and the one needed to be estimated, and the models in (4.2) as "The auxiliary models" since these models, as illustrated later in this chapter, are used only to improve the estimation of "The model". If the short term rate ($r_t^{(1)}$) is taken to be, say, the yields of the three-month Treasury bill, then the auxiliary rates could be the yields with longer than three-month maturities, including the yields on Treasury bills, Treasury notes, and Treasury bonds. It is noted that the only restriction on the auxiliary model is that the state variable $r_t^{(j)}$ also follows a diffusion process. Thus interest rate observations of other countries with similar economic fundamentals and monetary policies can also be used. Without loss of generality, all the realized rates are assumed to be equispaced over the time period $[0, T]$ with $\delta = T/n$ being the sampling interval.

Nonparametric modeling of the short rate diffusion has generated a great deal of interest in recent years mainly because it eliminates the potential risk of misspecifying the functional forms of the drfit and diffusion functions under mild regularity conditions. Various nonparametric estimators of the drift and diffusion functions have been proposed in the literature (see Aït-Sahalia, 1996; Jiang and Knight, 1997; Stanton, 1997; Bandi and Phillips, 2003; Nicolau, 2003 etc). A fully nonparametric estimation of the drift and diffusion functions based on a discretization of (4.1) was first proposed by Stanton (1997). In particular, with the discretization interval $\delta > 0$, a first-order approximation of the discretized process yields

$$\mu_1(r) = \frac{1}{\delta} E\left[\left(r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)}\right) | r_{t\delta}^{(1)} = r\right] + O(\delta), \tag{4.3}$$

$$\sigma_1^2(r) = \frac{1}{\delta} E\left[\left(r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)}\right)^2 | r_{t\delta}^{(1)} = r\right] + O(\delta) \tag{4.4}$$

which can be consistently estimated by

$$\tilde{\mu_1}(r) = \frac{1}{\delta} \frac{\sum_{t=0}^{n-1} \left(r^{(1)}_{(t+1)\delta} - r^{(1)}_{t\delta} \right) K_h \left(r^{(1)}_{t\delta} - r \right)}{\sum_{t=0}^{n-1} K_h \left(r^{(1)}_{t\delta} - r \right)}, \tag{4.5}$$

$$\tilde{\sigma_1^2}(r) = \frac{1}{\delta} \frac{\sum_{t=0}^{n-1} \left(r^{(1)}_{(t+1)\delta} - r^{(1)}_{t\delta} \right)^2 K_h \left(r^{(1)}_{t\delta} - r \right)}{\sum_{t=0}^{n-1} K_h \left(r^{(1)}_{t\delta} - r \right)} \tag{4.6}$$

where $K_h(\cdot)$ is kernel density function that satisfies common regularity conditions.[1] The statistical properties, in particular the bias and variance, of the drift function estimator are given in the following proposition.

Proposition 4.1

Given the common regularity conditions for the short rate process (see appendix) and for the function $K_h(\cdot)$ (see chapter three), suppose that $h \to 0, \delta \to 0, nh \to \infty$, as $n \to \infty$ and $Th \to \infty$, the bias and variance of the Stanton estimator are:

$$E[\tilde{\mu_1}(r) - \mu_1(r)] = \frac{h^2}{2} m(K)[\mu_1''(r) + 2\mu_1'(r)\frac{p_1'(r)}{p_1(r)}] + q(r)\delta + o(h^2)$$

$$Var[\tilde{\mu_1}(r) - \mu_1(r)] = \frac{\sigma_1^2(r)R(K)}{Th\hat{p}_1(r)} + o(Th)^{-1}$$

where $q(r) = \frac{1}{2}\{\mu_1'(r)\mu_1(r) + \frac{1}{2}\sigma^2(r)\mu_1''(r)\}, m(K) = \int z^2 K(z)dz,$
and $\mathrm{R}(K) = \int K^2(z)dz.$

Proposition 4.1 shows that the bias of the Stanton estimator can be large in regions where the slope and curvature of the underlying drift function are influential, in other words when the true drift function is nonlinear. Thus using the Stanton model to estimate the drift of the short rate diffusion process has the potential of generating erroneous results. Simulation results in Chapman and Pearson (2000) show that Stanton's drift estimator can produce spurious nonlinearities for high levels of the short rate. Chapman

[1] Estimators based on higher order approximations are also constructed in Stanton (1997). Fan and Zhang (2003) indicate that higher order approximations will reduce discretization bias, but at a cost of an increased variance.

and Pearson (2000) argue that a combination of the truncation of the "observed" short rates and a small sample create artificial patterns of nonlinearity near the boundaries of the support.

To further understand the difficulty of estimating the drift function, let us examine the relevant properties regarding the drift and diffusion functions. It is well known (see e.g. Merton (1980)) that based on data sampled over a short time interval, even though the diffusion term can be estimated very precisely when the sampling interval is small, the estimate of the drift coefficient tends to have low precision (see Aït-Sahalia (1996) for an example of the Geometric Brownian motion process). The intuition behind the relative difficulty of identifying and estimating the drift term versus the diffusion term is as follows. It is noted that in (1) the drift term is of order dt and the diffusion term is of order $\sqrt{dt}$, as $(dw_t)^2 = dt + O((dt)^2)$, i.e., the diffusion term has lower order than the drift term for infinitesimal changes in time. Therefore, the local-time dynamics of the sampling path reflects more of the properties of the diffusion term than those of the drift term, which suggests the possibility of identifying the diffusion term from high-frequency discrete sampling observations. For the same reason, the drift term cannot be estimated precisely based on the local-time dynamics of such sampling paths without further constraints. Not surprisingly, approximations of the drift function from high frequency data, such as the ones suggested by Stanton (1997), can be very non-robust and the estimates can be very sensitive to the sampling path.

The simulation results in this chapter lend support to the findings in Chapman and Pearson (2000) and provide further evidence that the identification and estimation of the drift function ultimately depends on that of the unconditional distribution of the interest rate process. The link between the identification and estimation of the drift function and the unconditional density of the short rate process is formally established in the following relation. Consider the case where the diffusion process is either strictly stationary or has a limiting probability density function. From the Kolmogorov forward equation, the drift term of the diffusion process in (1) is related to the diffusion function and the marginal density of the process as follows:

$$\mu_1(r) = \frac{1}{2p_1(r)} \frac{\partial}{\partial r}[\sigma_1^2(r)p_1(r)] \tag{4.7}$$

where $p_1(\cdot)$ is the marginal density function of the short rate process. It is obvious that with a well identified diffusion function, the identification and drift function estimation is equivalent to that of the marginal density of the process. A well-known property of the short rate is its high persistence over time.[2] The high persistence makes interest rate stay around certain levels over extended time period. Such property leads to the fact that interest rate observations over even reasonably long time period can only offer us restricted or truncated information of the short rate distribution. This observation is consistent with the argument in Chapman and Pearson (2000), Abhyankar and Basu (2001), and Li, Pearson and Poteshman (2004). For instance, from the plot of the daily 3-month T-bill yields, the yields are never below 2% for the entire period of 1960s to 1990s. In addition, the high interest rate observations only occur during the late 1970s and early 1980s with relatively fewer observations of large daily interest rate changes. Existing studies, such as Stanton (1997), have typically used interest rate observations during this sampling period. Extending the sampling period, however, beyond this period back to the 1950s or to early 2000s offers significant number of interest rates observations below 2% with the lowest observation at 0.55%.

The hope thus rests in extending the sampling period instead of sampling frequency

[2] A statistical issue involved in the nonparametric estimation of the drift function from highly persistent data is the optimal choice of bandwidth which can be substantially different from that under the iid conditions. The simulations as well as those in Chapman and Pearson (2000) both confirm that the optimal choice of bandwidth helps to reduce the spurious bias. However, the improvement is limited.

in order to provides reliable estimates of the drift term. The time period of historical observations of interest rate is nevertheless inevitably limited, making the new approach of using an expanded information set to improve the estimation of the drift function particularly appealing.

The proposed method is similar in spirit to the method of indirect inference in that the parameter estimates of the short rate are obtained using information about the remaining $J-1$ diffusion processes whose parameters are possibly similar to those of the short rate diffusion. For this reason the additional diffusion processes are referred to as "auxiliary models". While both the drift and diffusion functions are estimated in the empirical application, the theoretical results and simulations are focussed on the estimation of the drift. This is motivated by the difficulty of identifying the drift nonparametrically as underlined earlier.

Let $\mu_p(r)$ be a function of r, the proposed estimator of the drift function builds on the identity

$$\mu_1(r) = \mu_p(r)\frac{\mu_1(r)}{\mu_p(r)} \tag{4.8}$$

$$= \mu_p(r)c(r) \tag{4.9}$$

where $c(r)$ is called the "correction factor" and is defined on a set where $\mu_p(r) \neq 0$ (see Hjort and Glad, 1995; Glad, 1998a, 1998b). If the functional form of $\mu_p(r)$ was known then the function $c(r)$ can be estimated nonparmetrically by

$$\tilde{c}(r) = \frac{1}{\delta}\frac{\sum_{t=0}^{n-1}[\frac{(r^{(1)}_{(t+1)\delta}-r^{(1)}_{t\delta})}{\mu_p(r^{(1)}_{t\delta})}]K_h(r^{(1)}_{t\delta}-r)}{\sum_{t=0}^{n-1}K_h(r^{(1)}_{t\delta}-r)} \tag{4.10}$$

motivated by the fact that $E[\frac{(r^{(1)}_{(t+1)\delta}-r^{(1)}_{t\delta})}{\delta\mu_p(r^{(1)}_{t\delta})}|r^{(1)}_{t\delta}=r]=c(r)$.

However, in empirical applications, the functional form of $\mu_p(r)$ is unknown. Hence a feasible estimator of the drift is obtained by replacing $\mu_p(r)$ by its estimate. Under the above conceptual framework discribed above, $\mu_p(r)$ can be estimated nonparametrically by pooling the yields of all the J securities, i.e

$$\hat{\mu}_p(r) = \frac{1}{\delta}\frac{\sum_{j=1}^{J}\sum_{t=0}^{n-1}\left(r_{(t+1)\delta}^{(j)} - r_{t\delta}^{(j)}\right)K_{h_p}\left(r_{t\delta}^{(j)} - r\right)}{\sum_{j=1}^{J}\sum_{t=0}^{n-1}K_{h_p}\left(r_{t\delta}^{(j)} - r\right)} \tag{4.11}$$

With an estimator of $\mu_p(r)$ at hand, the new estimator of the drift of the short rate $\mu_1(r)$ is given by

$$\begin{aligned}\hat{\mu_1}(r) &= \hat{\mu}_p(r)\hat{c}(r)\\ &= \hat{\mu}_p(r)\left[\frac{1}{\delta}\frac{\sum_{t=0}^{n-1}[\frac{(r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)})}{\hat{\mu}_p(r_{t\delta}^{(1)})}]K_h(r_{t\delta}^{(1)} - r)}{\sum_{t=0}^{n-1}K_h(r_{t\delta}^{(1)} - r)}\right]\end{aligned} \tag{4.12}$$

The proposed estimator is designed so as to reduce the finite sample bias of the Stanton estimator. Intuitively, if the drift functions are identical, that is if $\mu_1 = \mu_2 = ...\mu_J = \mu_p$ then $\hat{c}(r)$ is an estimate of unity and the efficiency gains in this case are substantial because the proposed estimator behaves like the pooled pilot. In general when $\mu_p(r)$ is not distant from $\mu_j(r)$ then $\hat{\mu}_p(r)$ is more efficient than the Stanton estimator because the correction factor is less rough than $\mu_1(r)$ resulting in bias reduction. The correction factor $\hat{c}(r)$ adjusts the pooled pilot if the drift functions are dissimilar so that the proposed estimator resembles the Stanton estimator.

Proposition 4.2

Under the assumptions of proposition 4.1, the bias and variance of the proposed estimator are:

$$\begin{aligned}E[\hat{\mu_1}(r) - \mu_1(r)] &= \frac{h^2}{2}m(K)[c''(r) + 2c'(r)\frac{p_1'(r)}{p_1(r)}]\mu_p(r) + l(r)\mu_p(r)\delta + o(h^2)\\ Var[\hat{\mu_1}(r) - \mu_1(r)] &= \frac{\sigma_1^2(r)R(K)}{(Th)\hat{p}_1(r)} + O((JTh_p)^{-1} + h^2h_p^2) + o(Th)^{-1}\end{aligned}$$

where h_p is the smoothing parameter of the pooled data, and $l(r) = \frac{1}{2}\{c'(r)c(r) + \frac{1}{2}\sigma^2(r)c''(r)\}$.

The variance of the new estimator is approximately the same as that of the Stanton estimator when J is large, but the biases may substantially differ. The bias of the new estimator depends on the slope and curvature of the correction factor $c(r)$ while the bias of the Stanton estimator is a function of the slope and curvature of the drift function $\mu_1(r)$ in (4.1). Consequently, when the "prior function" is identical to the drift of the short rate, that is $\mu_p(r) = \mu_1(r)$, the correction factor $c(r) = \frac{\mu_1(r)}{\mu_p(r)}$ is a line about 1 hence $c'(r) = c''(r) = 0$. As a result the finite sample bias of the new estimator is reduced to a negligible order. Intuitively, if $\mu_p(r)$ is not too different from $\mu_1(r)$, the correction factor will ocsillate around unity thus having less curvature than $\mu_1(r)$ which leads to bias reduction.

Proposition 4.3

In addition to the assumptions of proposition 4.1, suppose that $nh^5 \to 0$, and $\sqrt{nh}\delta \to 0$, the new estimator has a limiting normal distribution:

$$\sqrt{Th}(\hat{\mu}_1(r) - \mu_1(r)) + o_p(1) \to N(0, \Theta(r)) \text{ where } \Theta(x) = \frac{\sigma_1^2(r)R(K)}{\hat{p_1}(r)}$$

Proposition 4.3 shows the asymptotic equivalency of $\hat{\mu}_1(r)$ and $\tilde{\mu}_1(r)$ under slightly stronger conditions than those in propositions 1 and 2. Both estimators converge to the same normally distributed random variable. Intuitively there exists a large enough sampling period T_0 beyond which any marginal information is inconsequential in improving the precision of the Stanton estimator. Thus any bias-correction by the new estimator must be in finite samples as discussed above.

The approach proposed in this chapter is particularly relevant for the estimation of the short rate diffusion as the yields of different maturities are available along the yield curve. While the yields of different maturities may be determined by different economic factors or market forces, empirical evidence suggests that the yield curve tends to behave systematically. In particular, the expected changes of yields of different maturities are highly correlated. This is supported by the evidence of the principal component analysis in Litterman and Scheinkman (1991). Litterman and Scheinkman report that there are three major factors driving the dynamics of the US yield curve, namely the level of short rate, the slope of the yield curve, and the curvature of the yield curve, a result for which this study also finds support. In particular, the short rate as the level factor itself accounts for nearly 80% variations of the whole yield curve.

4.3. Simulations

In this section Monte Carlo simulations are performed to assess the finite sample performance of the proposed drift function estimator. The simulations are designed with two goals. The first goal is to revisit the findings of Chapman and Pearson (2000). To evaluate the finite sample properties of the Stanton estimator, Chapman and Pearson simulate interest rate data from a CIR process and find that the drift function estimate is linear only in the center of the support, deviating severely from the true (linear) drift on both boundaries, in particular for interest rates above 14%. The second goal of the simulation is to investigate whether the use of additional interest rate data from similar diffusion processes can attenuate the bias of the short rate drift function estimate. For computational expediency, only one auxiliary model is considered in the simulation experiment. The CIR and Vasicek diffusion processes are used to model the dynamics of the short and auxiliary rates.

The CIR processes for the short rate and the auxiliary rate are specified respectively as

$$
\begin{aligned}
dr_t^{(1)} &= 0.8537(0.08571 - r_t^{(1)})dt + 0.1566 dw_t^{(1)} \\
dr_t^{(2)} &= \kappa * 0.8537(0.08571 - r_t^{(2)})dt + 0.1566 * \sqrt{\gamma} dw_t^{(2)} \\
dw_t^{(1)} dw_t^{(2)} &= \rho dt, \qquad (4.13)
\end{aligned}
$$

where the parameter values are set to equal those in Chapman and Pearson. The Vasicek processes for the short rate and and the auxiliary rate are respectively defined as:

$$
\begin{aligned}
dr_t^{(1)} &= 0.261(0.0717 - r_t^{(1)})dt + 0.02237\sqrt{r^{(1)}} dw_t^{(1)} \\
dr_t^{(2)} &= \kappa * 0.261(0.0717 - r_t^{(2)})dt + 0.02237\sqrt{\gamma * r^{(2)}} dw_t^{(2)} \qquad (4.14) \\
dw_t^{(1)} dw_t^{(2)} &= \rho dt, \qquad (4.15)
\end{aligned}
$$

where the parameter values are set to equal those in Aït-Sahalia (1999b) estimated for the 7-day Eurodollar rate. The same parameters are also used by Nicolau (2003) in his simulations.[3]

The parameters κ, γ, ρ in the auxiliary model are specified at different values in the simulations in order the investigate the effect of different factors. The following five cases are considered in the simulations.

- Case I (benchmark): $\kappa = \gamma = 1$ and $\rho = 0$. This is the ideal situation for the use of the proposed method since both the drift and diffusion functions of the auxiliary model are identical to those of the short rate model. The information from the auxiliary model is no different than that from the short rate model.

- Case II (mean reversion effect): $\gamma = 1$, $\rho = 0$, $\kappa = 0.75, 1.25$, and 1.5. In this case, the short rate model and the auxiliary model converge to the same long-run mean, but at different speeds. This case focuses on the effect of the mean-reversion level

[3] Chapman and Pearson (2000) only used a square-root diffusion process in their simulations.

of the auxiliary process, or the difference in the drift of the auxiliary process, on the performance of the proposed estimator.

- Case III (diffusion effect): $\kappa = 1$, $\rho = 0$, $\gamma = 0.75, 1.25$, and 1.5. In this case, the short rate process and the auxiliary have identical drift functions, but different diffusion functions. The auxiliary process is more (less) volatile when $\gamma > (<) 1$.

- Case IV (correlation effect): $\kappa = \gamma = 1$ and $\rho = 0.2, 0.4$, and 0.8. This case allows us to examine the impact of level of correlation between the short rate model and auxiliary model on the performance of the new drift function estimator. As mentioned earlier, the yields of different maturities along the yield curve tend to be correlated with systematic co-movements.

- Case V (mixed processes): Instead of restricting both the short rate model and auxiliary model to follow the same type of processes as in cases I through IV, in this simulation it is assumed that the short rate $r_t^{(1)}$ follows a Vasicek process while the auxiliary interest rate $r_t^{(2)}$ follows a CIR process. This case is particularly interesting because in empirical applications the short rate model and the auxiliary model may have different functional forms for the diffusion component. For example, empirical evidence suggests that yields with longer maturities are noisier than short term yields

Two thousand paths were generated for the short rate model and the auxiliary models. Each path simulates 31 years of daily data. Observations in the first year are discarded to eliminate the start-up effects, which results in daily observations of 30 years. When $\rho \neq 0$, the short rate model and the auxiliary model are simulated jointly. More specifically, for the Vasicek processes, starting values are drawn from the marginal normal distributions and subsequent values are obtained from the transitional bivariate normal density. For the CIR processes, the starting values are drawn from the marginal Gamma distributions and the subsequent values are simulated using the Milshtein discretization scheme. With the simulated sampling paths, the Stanton estimator as well as the proposed estimator for the drift function of the short rate process are implemented. Throughout the simulations, a Gaussian kernel is used and the bandwidth is the one that minimizes the integrated squared error (ISE).

The drift function estimates of the Vasicek and CIR short rate processes for the benchmark case are plotted in Figures 4.1 and 4.2 (Panels A and B), together with the true drift function. In Figure 4.1, the naive bandwidth is used for both the Stanton estimator and the new proposed estimator. The results are overall consistent with the findings in Chapman and Pearson (2000). In particular, the Stanton estimates display a significant nonlinearity at the end-points or boundaries of the support. Chapman and Pearson (2000) argue that such nonlinearity in the estimated drift function results from two sources, the truncation of the observed interest rates within the interval $\left[r_{min}^{(1)}, r_{max}^{(1)}\right]$, and the well-known boundary bias of the nonparametric estimator. While the new method has a much reduced nonlinearity towards the boundaries, it remains a poor estimate of the linear drift function.

Figure 4.2 plots the drift function estimates of both the Stanton estimator and the new proposed estimator for the Vasicek and CIR models using the ISE-minimizing bandwidth. The results of both estimators are substantially improved compared to those in Figure 4.1 where the naive bandwidth is used. The "spurious" nonlinearities in the Stanton estimates are dramatically reduced. Mild nonlinearities are observed at the end points of the support because of the well-known boundary bias problem. The difference in the results also highlights the importance of the smoothing parameter. When the smoothing parameter is appropriately chosen the Stanton estimator performs well in the interior of the support. As expected, the new estimator significantly outperforms the Stanton estimator with visibly less bias towards the boundaries. Since in this simulation, the auxiliary model is identical to the short rate model, the effect of using the auxiliary model is equivalent to doubling the sample size or the sampling period. The improvement of the new estimator is similar to that in Chapman and Pearson where 15,000 or 60 years of daily observations instead of 7,500 or 30 years of daily observations are used.

Tables 4.1 and 4.2 report the mean integrated absolute bias (MIAB) and the root mean integrated squared error (RMISE) of alternative estimates for the drift function of the Vasicek and CIR processes. Panels A and B (the benchmark case) of tables 4.1 and 4.2 confirms that the proposed estimator has substantially lower bias relative to the Stanton estimator for both the Vasicek and CIR models. With the optimal bandwidth, the MIAB of the new estimator is just over 20% of the bias of the Stanton estimator with the RMISE also substantially reduced.

The bias and RMISE of the new and Stanton estimators for the remaining cases are also reported in tables 4.1 and 4.2. No plots are produced for these estimates for brevity. Panel C (case II) reports the performance of the new estimator when the auxiliary model has the same type of process but a different mean reversion parameter for the Vasicek and CIR models, respectively. The results indicate that allowing the auxiliary process to have a somewhat faster or lower mean-reversion (case II) does not significantly affect the bias of the new estimator relative the benchmark case for the Vasicek model. For the CIR model, when $k = 1.5$, the bias has more than doubled relative to the benchmark case but remains considerably under the bias of the Stanton estimator. The admirable performance of the proposed estimator in this case can be explained by the fact that the drift functions $\mu_1(r)$ and $\mu_1(r)$ are not far apart hence the correction factor oscillates around unity, thus is easier to estimate than the underlying true drift of the short rate.

The results reported in Panle D (case III) show that the performance of the new estimator relative to the benchmark is improved when the auxiliary rate is more volatile than the short rate. This is because an increased volatility also means less persistence of the auxiliary rates, thus more information.

The results in Panel E (case IV) of tables 4.1 and 4.2 indicate that a low ($\rho = .2$) to moderate ($\rho = .4$) correlation between the Wiener processes $\omega_t^{(1)}$ and $\omega_t^{(2)}$ does not drastically affect the performance of the new method relative to the benchmark case. However, for $\rho = 0.8$, the RMISE if the new estimator deteriorates for both the Vasicek and the CIR models. Intuitively, as $\rho \to 1$ the observations become more and more similar, reducing the information added by the auxiliary process.

How does the new estimator perform when the auxiliary model is different from the model to be estimated? To answer this question, a case where the auxiliary model is the CIR process while the model to be estimated is the Vasicek process is studied. The Vasicek model is simulated using the parameter values in Nicolau and the auxiliary CIR process is simulated using the parameters in Chapman and Pearson. The mean integrated absolute bias (MIAB) and root-mean integrated squared error (RMISE) of the new estimator with 2,000 replications are reported in Panel F of Table 4.1, which are equal to 0.0989 and 1.087, respectively. The MIAB is about twice that of the new estimator for case I where both the auxiliary and the short rate processes followed a Vasicek diffusion but is less than half of the bias of the Stanton estimator. The root-mean integrated squared error (RMISE) of the new estimator is also considerably lower than that of the Stanton estimator. The good performance of new estimator is explained by the functional similarity between the underlying drifts even though the processes are not identical.

The performance of the proposed estimator in these finite sample experiments is quite satisfactory. The incorporation of additional observations from auxiliary interest rate processes has the effect of expanding the local information even in the edges of the support thus leading to bias reduction. The simulation experiments also show that the new estimator performs most admirably against the Stanton estimator even if the drift functions of the short and the auxiliary rates are not identical. This represents a strength of the proposed estimator since in empirical estimation, it impossible to know if the processes are dissimilar, similar and if so to what extent. Overall, the simulations indicate a strong potential for the new estimator in empirical applications.

4.4. Empirical Application

4.4.1. The Data and Estimation Results of the Short Rate Diffusion Parameters

The data in this empirical analysis consists of daily US 3-month T-bill rates from January 1954 to November 2004 with 12,704 daily observations over more than 50 years. This data has the longest sampling period among all existing studies. The additional yields along the yield curve are from February 1962 to November 2004 with 10,676 observations for the 6-month T-bill, 1-year T-bill, and 3-, 5-, and 10-year T-notes. The 3-month T-bill yields are used to proxy for the short rate. In other words, there are 12,704 observations on the short rate process which is the one to estimate and 10,676 observations on each of the five additional "auxiliary" diffusion processes. Figure 4.3 plots the time series of the daily 3-month Treasury yields and the daily changes in panels A and B, respectively. The time series plot reflects the wide range of observations of the 3-month T-bill yields over the sampling period, and the first difference reflects some large changes of the 3-month T-bill yields from day to day. The visibly large daily changes of 3-month T-bill yields are associated with the unusually high levels of yields during the later 70's and early 80's, indicating different behavior of interest rate at different levels. Descriptive statistics of the data are reported in table 4.3. The average yields of different maturities suggest that the yield curve is overall upward sloping, and the standard deviations of the daily yield changes suggest that the yield curve is more volatile over the short end than the long end. Both skewness and kurtosis statistics indicate that interest rates are non-normally distributed. The minimum and maximum observations confirm again the wide range of interest rates over the sampling period. For instance, the 3-month T-bill yield has a minimum value of 0.55% and a maximum value of 17.14%. Although the autocorrelations in the interest rate level decays very slowly, those of the day-to-day changes are generally small and are not consistently positive or negative. The augmented Dickey-Fuller nonstationarity test is used to check whether the interest rate follows a unit root process. The augmented Dickey-Fuller test statistic for the 3-month T-bill rates has a value of -2.86 which is compared to the 10% critical value of -2.57. That is, the null hypothesis of nonstationarity is rejected at the 10% significance level for the 3-month T-bill time series. Since the test is known to have low power, even a slight rejection means that stationarity of the series is very likely. Similar results are obtained for the time series of yields with longer maturities.

Table 4.4 also reports the correlation matrix and principal components of daily interest rate changes. The correlation matrix suggests that the changes of interest rates along the yield curve are highly correlated. For instance, the daily changes of 3-month and 6-month T-bill yields have a correlation of nearly 86%. The principal component analysis reports similar results as in existing studies. Namely, there are mainly three factors driving the dynamics of yield curve, the level, the slope and the curvature. In particular, the short rate explains more than 80% of the total variation of the yield curve dynamics, suggesting the importance of modeling the short rate process.

With the panel of yields, the proposed estimator is implemented along with the CIR and the Vasicek models. Figure 4.4 plots the nonparametric drift function estimates based on the new estimator using 3-month and 6-month yields (Panel A) as well as the yields of all maturities (Panel B), together with the Stanton nonparamnetric estimates and the CIR linear drift function using only the 3-month T-bill yield. The 95% pointwise confidence bands of the new estimator are also plotted. Panel A of Figure 4.4 suggests that both the Stanton nonparametric drift function estimates and the new drift function estimates are highly nonlinear and significantly different from the CIR linear drift function, particularly at high level of interest rate. However, even though the Stanton estimates are visibly different from the new estimates with the new drift function estimates exhibiting less nonlinearity or mean reversion, the Stanton estimates stay within the 95% confidence band of the new estimates. This is expected since the 6-month t-bill yields are highly correlated with the 3-month t-bill yields and add only limited information to the drift function estimation of the 3-month t-bill process. On the other hand, in Panel B where all the yields of various maturities up to 10 years are used in the drift function estimation, there is further difference between the Stanton estimates and the new estimates of the drift function. Noticeably, the new drift function estimates become more flat relative to those plotted in Panel A. The use of information from the yields of different maturities produce more efficient estimates of the drift function. More interestingly, the Stanton estimates are now outside the 95% confidence band of the new estimates for high level of interest rate.

The diffusion function estimate using the yields of all maturities is plotted in Panel A of Figure 4.5, together with the Stanton estimate and CIR diffusion function using only the 3-month T-bill yield. Note that the diffusion function estimates have a much narrower 95% confidence band relative to the drift function estimates, further confirming that the inference of the diffusion function is much reliable. Again, both the Stanton and new diffusion function estimates are highly nonlinear, with the volatility of short rate process increasing sharply first as interest rate level increases and then drops off slightly at the very high level of interest rate. Relative to the Stanton estimates, however, the new diffusion function estimates suggest less dramatic increase of volatility as interest rate increases. In particular, the Stanton estimates are substantially above new diffusion function estimates for high level of interest rate. This is consistent with the drift function estimates. It is also noted that the CIR diffusion function is extremely low compared to both the Stanton and new estimates.

4.4.2. Implications of the Drift Function on Bond Pricing and the Valuation of Interest Rate Derivatives

This section examines the implications of drift function estimates on the prices of both zero-coupon bonds and interest rate derivatives. Unlike assets such as equity or foreign currency where the drift function of the underlying asset return process does not appear in the risk-neutral process, the drift function of the interest rate enters directly into its risk-neutral counterpart. The risk-neutral process corresponding to the short rate diffusion defined in (4.1) is given by

$$d\tilde{r}_t = (\mu(\tilde{r}_t) - \lambda(\tilde{r}_t))dt + \sigma(r_t)d\tilde{w}_t \tag{4.16}$$

where $\lambda(\tilde{r}_t) = \lambda_0(\tilde{r}_t)\sigma(r_t)$ is the market price of interest rate, and $\tilde{w}_t$ is a standard Brownian motion under the equivalent martingale measure Q.

The market price of interest rate can be nonparametrically estimated following the procedure in Jiang (1998). Since the market price of interest rate risk is fully determined by the short rate, it can be straightforwardly estimated from any two non-dividend paying assets. Suppose $Y(r_t, \tau)$ represents the yield with r_t at t and maturity $\tau = T - t$, and

$$dY(r_t, \tau) = \alpha(r_t, \tau)dt + \kappa(r_t, \tau)d\tilde{w}_t$$

Following Itô's lemma and using (6), an estimator of time-stationary $\lambda_0(r_t)$ can be derived as

$$\hat{\lambda}_0(r_t) = \frac{Y_d(r_t, \tau_1, \tau_2) + \frac{1}{2}(\tau_1^2\kappa^2(r_t, \tau_1) - \tau_2^2\kappa^2(r_t, \tau_2)) + \tau_2\alpha(r_t, \tau_2) - \tau_1\alpha(r_t, \tau_1)}{\tau_2\kappa(r_t, \tau_2) - \tau_1\kappa(r_t, \tau_1)} \tag{4.17}$$

where $\tau_i = T_i - t, i = 1, 2, Y_d(r_t, \tau_1, \tau_2) = Y(r_t, \tau_1) - Y(r_t, \tau_2)$ is the yield spread between maturities τ_1 and τ_2. The 3-month and the 10-year yields are used to estimate the market price of interest rate risk, which is plotted in Panel B of Figure 4.5.

The prices of interest rate derivative securities can be computed based on the simulation of the risk-neutral process in (4.16). Simulations of the sample path are often performed using either the Euler scheme or the Milshtein scheme in the literature. In this chapter, the Milshtein scheme is employed:

$$\begin{aligned} r^n_{(m+1)T/n} = r^n_{mT/n} \quad & +(\mu(r^n_{mT/n} - \lambda(r^n_{mT/n}))T/n + \sigma(r^n_{mT/n})(W_{(m+1)T/n} - W_{mT/n}) \\ & +\tfrac{1}{2}\sigma^2(r^n_{mT/n})((W_{(m+1)T/n} - W_{mT/n})^2 - T/n) \end{aligned} \tag{4.18}$$

with $r_0^n = r_0$. It is noted that the Milshtein scheme has better convergence rates than the Euler scheme for the convergence in $L^p(\Omega)$ and the almost sure convergence (see Talay, 1996). In financial applications of the Monte Carlo simulation methods, a number of variance reduction methods have been proposed, e.g. the control variate approach, the antithetic variate method, the moment matching method, the importance sampling method, the conditional Monte Carlo methods, and quasi-random Monte Carlo methods (see, e.g. Boyle, Broadie and Glasserman, 1996). In these simulations, the antithetic variate method is employed in reducing the sample variance. The conditional expectation of the final payoff under the risk-neutral dynamics gives the prices of the interest rate contingent claims. Since the behavior of the drift is only debated at the relatively high levels of short rate, it has potentially the highest impact on the valuation of interest rate caps. For this reason the proposed estimator is used to price various interest rate caps. First the impact of the proposed drift estimator on the valuation of bonds is analyzed.

Valuation of Bond Prices The price of a zero-coupon bond with face value $P(r_t, T, T) = 1$ is given by

$$P(r_t, t, T) = E_t^Q[exp\{-\int_t^T \tilde{r}_u du\}]$$

To investigate the impact drift function estimate on the bond price, zero-coupon bond prices are generated using three alternative estimations of the drift and diffusion parameters, the CIR estimates, the Stanton estimates, and the estimates of the proposed method. The prices are computed based on 2000 simulated risk-neutral interest rate paths using the Milshtein scheme and the antithetic variate method to reduce the variability of the results. Two thousand simulated bond prices with maturities of 3, 6-month, 1, 3, 5, 10, and 30 years are computed using the proposed estimator, the Stanton estimator and CIR estimator. Converting the prices to yields, the yield curves with three starting values of interest rate are plotted in Figure 4.6. The 95% confidence bands of the proposed estimator are also calculated and plotted with the yield curves. As expected, the yields of the Stanton estimator are significantly higher than the yields implied by the proposed estimator for higher interest rates, which is consistent with the fact that the proposed estimator has a lower mean-reversion at higher levels of interest rate.

Valuation of Interest Rate Cap Prices The value of a interest rate cap is determined by its cash flow over its contract period. The cash flows of an interest rate cap with a notional principal equal to $100 at time t are:

$$100 \times \max[(Y(t - \Delta t; t) - Y_S)\Delta t; 0] \tag{4.19}$$

where $t = t_1; t_2; \cdots; t_n$ are the payment dates (reset of Δt occurs in advance before the payment date), t_n is the last payment date and is often referred as the cap tenor, n is the number of payment dates, $Y(t-\Delta t; t)$ is the annualized cap interest rate over the period $(t-\Delta t; t]$, Y_S is the annualized cap strike rate. The cap prices are reported in table 4.5 for different strikes prices (in basis points), tenors (in years), and annualized spot rates. The results are based on 2000 simulated interest rate paths under the risk neutral measure. Again the antithetic variate method is used to reduce the variance of the simulated prices. The prices implied by the proposed estimator are found to be significantly different from the Stanton prices and the CIR prices at the 95% level. Consistently with the plots of the drifts, the prices implied by the new estimator are higher that the Stanton prices for higher interest rates as expected.

4.5. Conclusion

This chapter has proposed a nonparametric estimator for the short rate diffusion process. The proposed estimators attempts to reduce the bias of Stanton's nonparametric estimator by exploiting information from a panel of yields of different maturities including the observations proxying for the short rate. The simulation results indicate that the proposed estimator significantly attenuates the spurious nonlinearities of the Stanton estimator. Using the longest time-series of short rate observations used in empirical studies (50 years of the US three-month treasury bill) and six additional time-series on US treasury bills and bonds, the estimator proposed in chapter III is applied to the estimation of the parameters of the US short-term interest rate. The empirical findings corroborate the results of Stanton (2000) and Jiang (2000) about the nonlinearity of the short rate drift function for high levels of interest rate, however, the mean-reversion is significantly weaker in the results of this study.

TABLE 4.1. The Simulation Results of Vasicek Model

MIAB	RMISE	MIAB	RMISE	MIAB	RMISE
Panel A: Benchmark Stanton (1997) estimator					
h_{iid}		h_{opt}			
1.8317	7.8938	0.2400	2.3810		
Panel B: Benchmark new panel estimator (Case I: $\kappa = \gamma = 1$, $\rho = 0$)					
h_{iid}		h_{opt}			
0.6001	5.4872	0.0530	0.8769		
Panel C: Effect of mean reversion (Case II: $\gamma = 1$, $\rho = 0$)					
$\kappa = 0.75$		$\kappa = 1.25$		$\kappa = 1.5$	
0.0521	0.8650	0.0526	0.9997	0.0571	1.0533
Panel D: Effect of diffusion (Case III: $\kappa = 1$, $\rho = 0$)					
$\gamma = .75$		$\gamma = 1.25$		$\gamma = 1.5$	
0.0873	1.3223	0.0.0459	0.8759	0.0477	0.7237
Panel E: Effect of correlation (Case IV: $\kappa = \gamma = 1$)					
$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.8$	
0.05550	1.0431	0.0656	1.1520	0.0878	1.4972
Panel F: Effect of mixed models (Case V: $\kappa = \gamma = 1$, $\rho = 0$)					
Auxiliary model: CIR					
0.0.0989	1.087				

Note: MIAB and RMISE denote mean integrated absolute bias and root mean integrated squared error, respectively.

Table 4.2. The Simulation Results of CIR Model

MIAB	RMISE	MIAB	RMISE	MIAB	RMISE
Panel A: Benchmark Stanton (1997) estimator					
h_{iid}		h_{opt}			
1.2597	10.4216	0.3581	3.7215		
Panel B: Benchmark new panel estimator (Case I: $\kappa = \gamma = 1$, $\rho = 0$)					
h_{iid}		h_{opt}			
0.4025	4.4719	0.0798	1.9198		
Panel C: Effect of mean reversion (Case II: $\gamma = 1$, $\rho = 0$)					
$\kappa = 0.75$		$\kappa = 1.25$		$\kappa = 1.5$	
0.1038	1.7411	0.1275	1.9506	0.1998	2.2551
Panel D: Effect of diffusion (Case III: $\kappa = 1$, $\rho = 0$)					
$\gamma = .75$		$\gamma = 1.25$		$\gamma = 1.5$	
0.1433	2.1651	0.0459	1.6588	0.0402	1.8696
Panel E: Effect of correlation (Case IV: $\kappa = \gamma = 1$)					
$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.8$	
0.0828	2.1345	0.0952	2.3223	0.1441	2.9607

Note: MIAB and RMISE denote mean integrated absolute bias and root mean integrated squared error, respectively.

TABLE 4.3. Summary Statistics of Interest Rates

τ	Mean*	StDev	Skew	Kurt	Min	Max	ρ_1	ρ_2	ρ_3	ρ_4
Panel A: Summary Statistics of Daily Interest Rates										
r3M	5.226	2.848	1.060	1.684	0.550	17.140	0.999	0.998	0.996	0.995
r6M	5.910	2.730	0.927	1.415	0.800	15.930	0.999	0.998	0.997	0.996
r1Y	6.365	2.932	0.945	1.301	0.880	17.310	0.999	0.998	0.997	0.996
r3Y	6.795	2.741	0.907	0.989	1.340	16.590	0.999	0.998	0.997	0.996
r5Y	7.009	2.630	0.959	0.878	2.080	16.270	0.999	0.998	0.997	0.997
r10Y	7.229	2.515	0.969	0.656	3.130	15.840	0.999	0.999	0.998	0.997
Panel B: Summary Statistics of Daily Interest Rate Changes										
Δr3M	-0.609	0.102	0.202	26.11	-1.270	1.340	0.139	0.018	-0.024	0.039
Δr6M	-0.618	0.090	0.296	23.74	-1.100	1.170	0.116	0.039	-0.001	0.031
Δr1Y	-0.758	0.092	-0.194	21.39	-1.080	1.100	0.115	0.045	-0.001	0.033
Δr3Y	-0.665	0.081	-0.180	13.26	-0.790	0.920	0.117	0.029	0.003	0.004
Δr5Y	-0.440	0.076	-0.305	11.42	-0.770	0.720	0.112	0.026	-0.001	-0.006
Δr10Y	0.103	0.068	-0.273	10.08	-0.750	0.650	0.088	0.029	-0.005	-0.018

Note: * The mean for the daily change of interest rate has a magnitude of 10^{-4}.

TABLE 4.4. Correlation Matrix and Principal Components of Daily Interest Rate Changes

Panel A: Correlation Matrix of Daily Interest Rate Changes

	Δr3M	Δr6M	Δr1Y	Δr3Y	Δr5Y	Δr10Y
Δr3M	1.0000					
Δr6M	0.8564	1.0000				
Δr1Y	0.7462	0.8720	1.0000			
Δr3Y	0.6076	0.7451	0.8546	1.0000		
Δr5Y	0.5684	0.7063	0.8121	0.9401	1.0000	
Δr10Y	0.5039	0.6343	0.7434	0.8754	0.9251	1.00000

Panel B: Principal Components of Daily Interest Rate Changes

Factor							%
1	8.8759	8.6152	9.0454	7.6667	7.0018	5.8645	0.809
2	-5.1705	-2.3409	0.2146	2.8958	3.2404	3.2790	0.121
3	2.1203	-1.4016	-2.5534	0.2010	0.9380	1.4054	0.034
4	0.7978	-2.2113	1.4981	0.4569	-0.2175	-0.6072	0.017
5	-0.0270	0.2172	-0.7848	1.7024	0.2252	-1.5623	0.013
6	-0.0048	0.0429	-0.0851	0.7668	-1.3537	0.6892	0.006

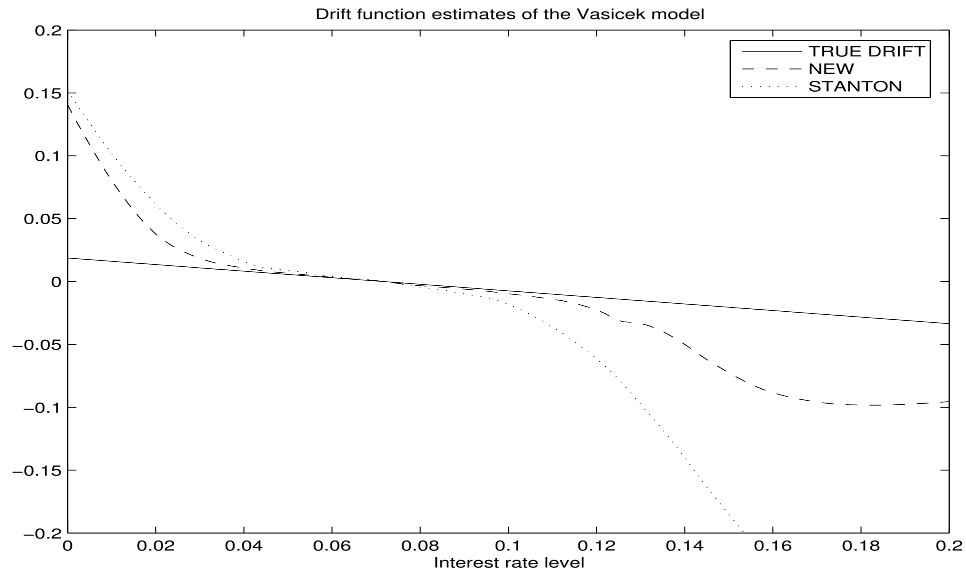
TABLE 4.5. Parametric and Nonparametric Valuation of Interst rate Caps

Spot Rate	Cap Tenor	Strike Price				
		-50	-25	0	25	50
0.05	3	3.0199	2.4089	1.8709	1.4165	1.0466
		(0.2464)	(0.3383)	(0.4424)	(0.5407)	(0.5632)
		3.2331	2.6177	2.0711	1.6060	1.2258
		3.1532	2.6323	2.1743	1.7814	1.4520
	5	6.4768	5.4566	4.5331	3.7213	3.0345
		(0.4099)	(0.5556)	(0.7283)	(0.8718)	(0.9703)
		7.3886	6.3579	5.4090	4.5544	3.8082
		6.6175	5.7646	4.9992	4.3284	3.7409
0.10	3	2.6807	2.1420	1.6708	1.2729	0.9543
		(0.4617)	(0.5656)	(0.6414)	(0.6857)	(0.6718)
		2.5740	2.0554	1.6015	1.2294	0.9374
		3.4744	3.0183	2.6048	2.2379	1.9177
	5	5.3489	4.9561	3.6810	3.0033	2.4348
		(0.9528)	(2.2740)	(1.2479)	(1.3360)	(1.3735)
		5.5414	4.2644	3.9160	3.2447	2.6825
		7.1835	6.8973	5.7762	5.1683	4.6162
0.15	3	2.7397	2.3277	1.9669	1.6567	1.3906
		(1.0298)	(1.0851)	(1.1081)	(1.0933)	(1.0464)
		2.2901	1.9074	1.5744	1.2932	1.0589
		3.6975	3.2755	2.8883	2.5391	2.2284
	5	5.5923	4.9561	4.3834	3.8767	3.4329
		(2.2146)	(2.2740)	(2.2988)	(2.2866)	(2.2392)
		4.8808	4.2644	3.7174	3.2435	2.8393
		7.5667	6.8973	6.2738	5.7041	5.1817

Note: Cash flows are paid over a one-year period according to the formula: $100 \times \max[(Y(t-1;t)-Y_S)\Delta t; 0]$ with $Y(t-1,t)$ being the annually compouded 12-month interest rate, and Y_S the strike price of the cap, which is the difference between the strike rate and the prevailing spot rate. The fours elements from top to bottom are: the nonparametric prices using the proposed estimator, its standard errors (in parentheses), the nonparametric prices using the Stanton estimator, and finally the CIR prices.

FIGURE 4.1. The Drift Function Estimates of the Vasicek and CIR Models with the Naive Bandwidth

Panel A: The drift function estimates of the Vasicek model



Panel B: The drift function estimates of the CIR model

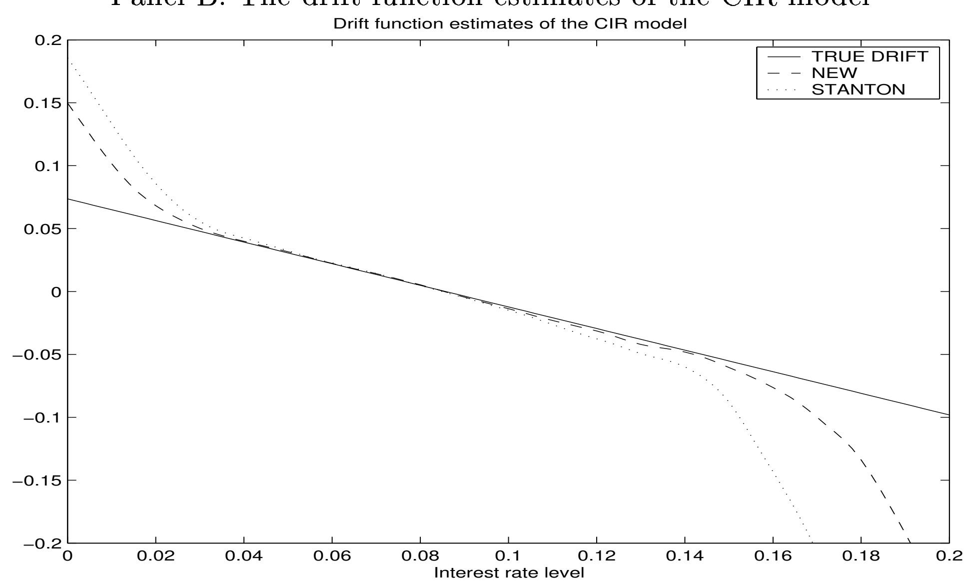
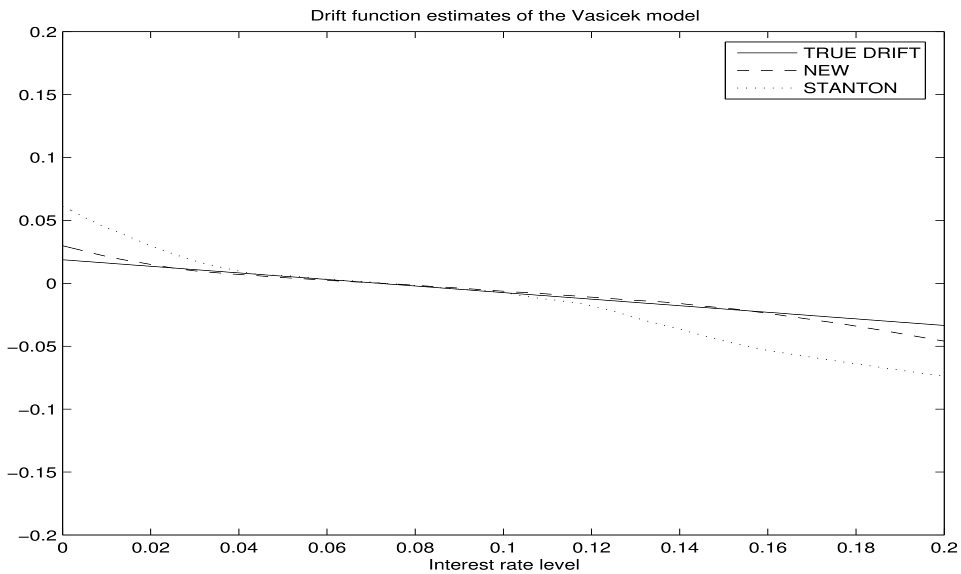


FIGURE 4.2. The Drift Function Estimates of the Vasicek and CIR Models with the Optimal Bandwidth

Panel A: The drift function estimates of the Vasicek model



Panel B: The drift function estimates of the CIR model

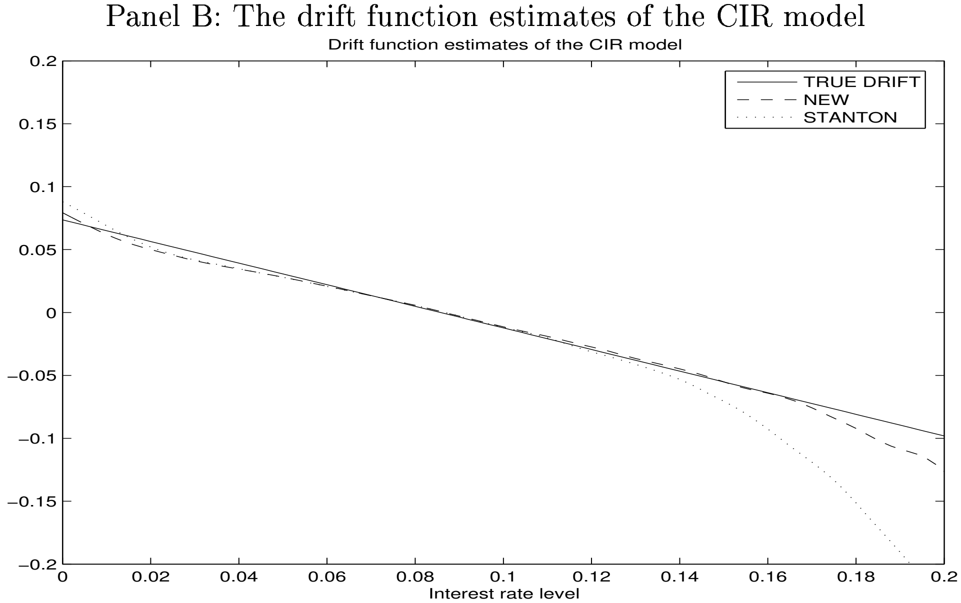


FIGURE 4.3. Daily and Daily Changes of US 3-month T-bill Yields

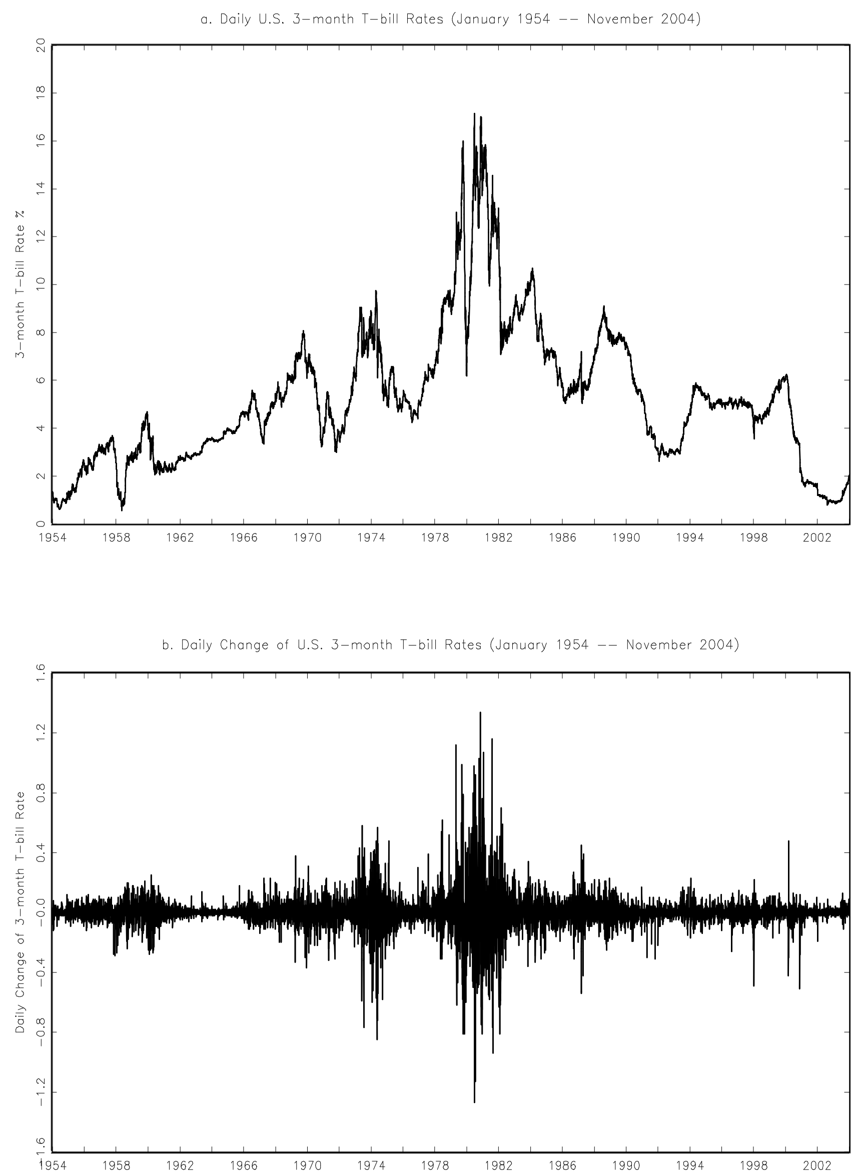
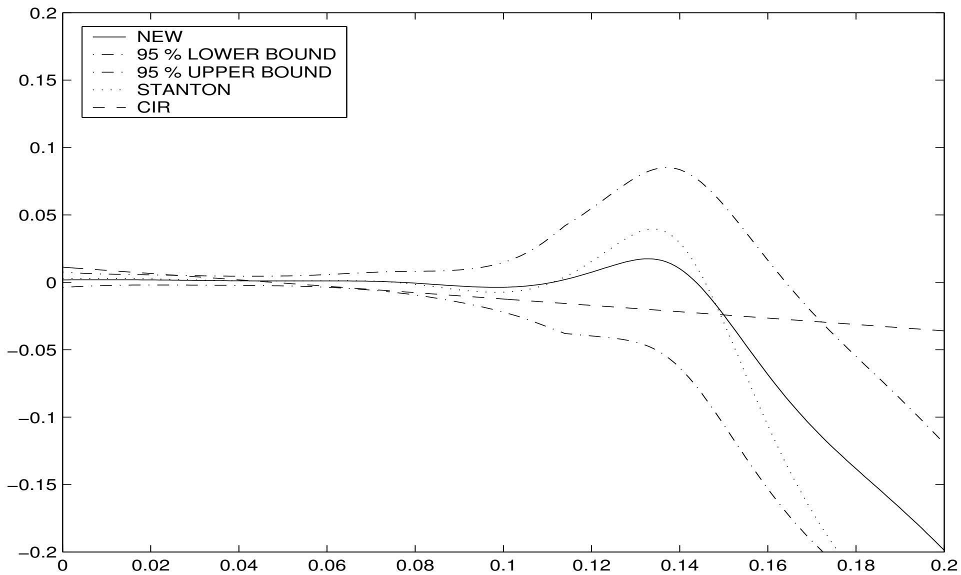


Figure 4.4. The Nonparametric Drift Function Estimates



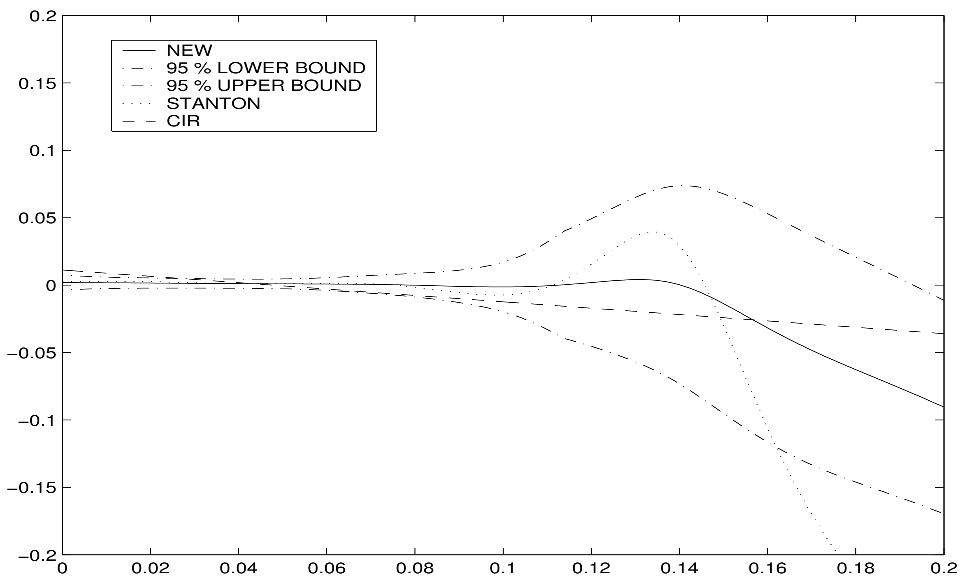


FIGURE 4.5. The nonparametric Diffusion Function and Market Price of Risk Estimates

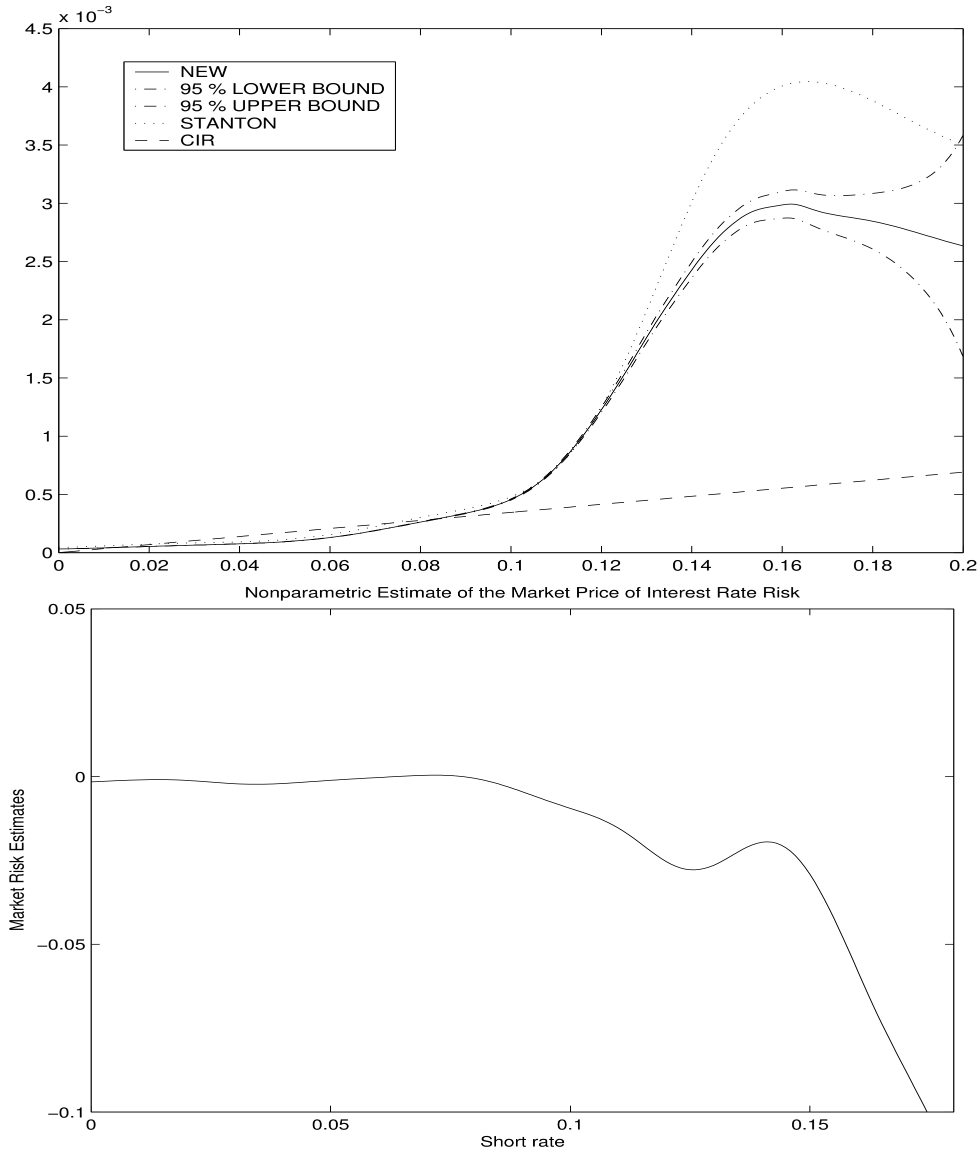
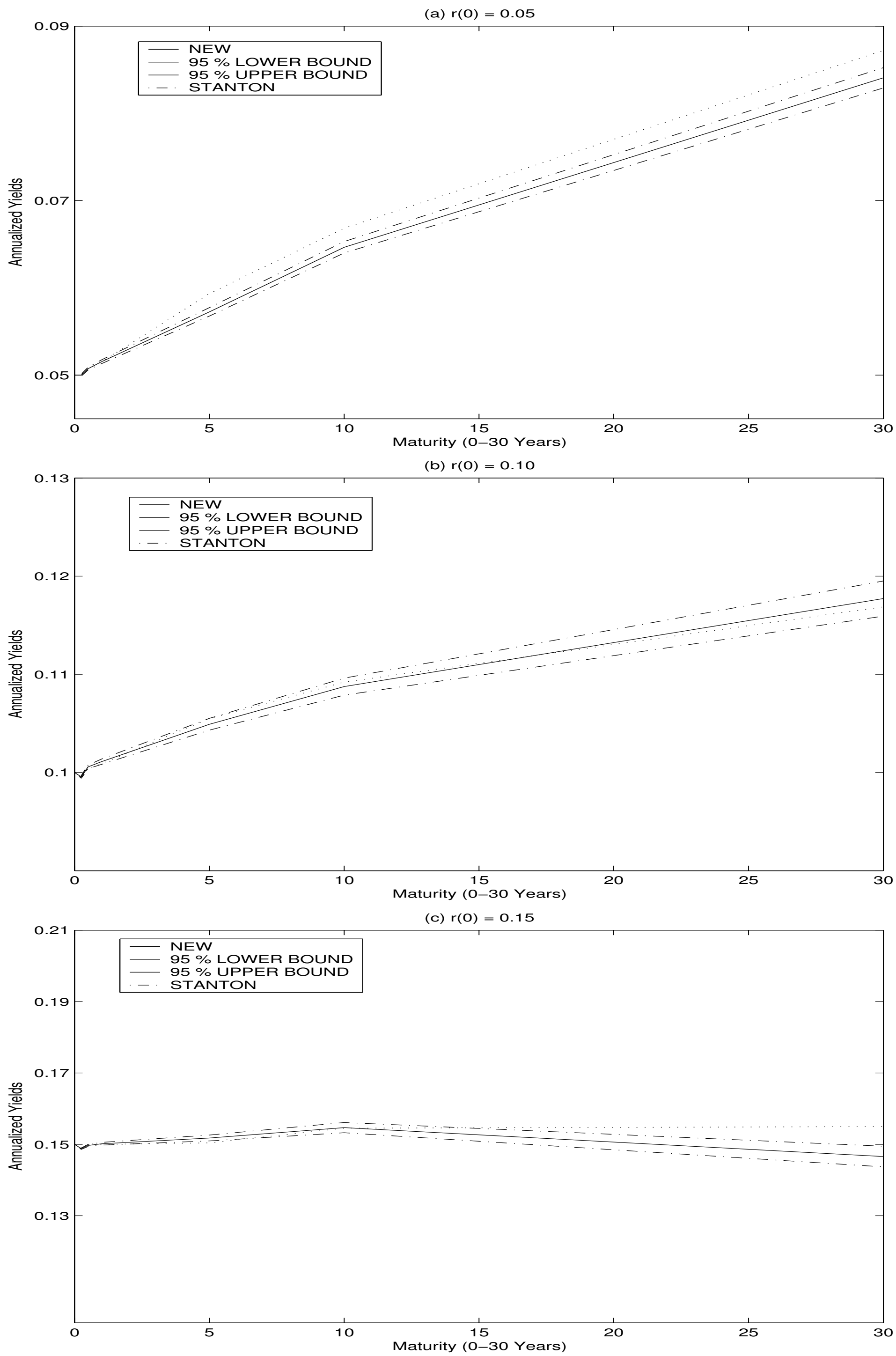


FIGURE 4.6. The simulated Yield Curves from Alternative Short Rate Processes



5. Dissertation Conclusions

The passage of the Emergency Planing and Community Right-To-Know Act in 1986 and the subsequent creation of the Toxic Release Inventory have dramatically lowered informational asymmetries between consumers and the financial markets on one hand and polluting firms on the other. Consequently, corporate environmental responsibility is no longer limited to compliance with mandatory pollution regulations. As evidence has suggested in recent empirical literature and in this study, it also includes proactive policies of self-regulation. The research in the second chapter can be extended in more than one direction. Since the EPA-sponsored VPRs are not enforceable, most of the empirical literature has focused on studying the effectiveness of such programs. However, the mechanism by which a voluntary program achieves pollution reductions is yet to be investigated empirically. It could be that participation in a VPR, in and of itself, creates greater environmental consciousness on the part of corporate management, thus leading to pollution reductions. Or it could be that participation in these programs leads to the adoption by firms of environmental management policies such as the creation of an environmental audit department or the adoption of total quality environmental management, which are responsible for long-term environmental improvements. Another potential research avenue is the empirical study of the bi-directional links between environmental innovation and VPRs. It would be very interesting to know if participation in these overcompliance programs triggers an increase in environmental technological innovation in terms of an increase in environmental patents or if only firms with the technological edge at the onset are more likely to participate.

As seen in the simulations herein, the new estimator developed in the third chapter of this manuscript works admirably well against the ordinary nonparametric estimator as well as a number of related estimators when the regressions functions are identical. Equally importantly, the proposed estimator did not lose much efficiency against the individual nonparametric estimator even when the curves are quite different. This and the fact that knowledge of the extent of similarity is not required to be known or modeled represent the main strengths of the estimator. The empirical application of the new estimator in chapter four illustrates the ease with which it can be implemented. The new estimator can be applied to a variety of empirical problems. For example, when rating crop insurance policies, the proposed method can be used to estimate the temporal process of county yields more precisely by incorporating information from neighboring counties within, say the same crop-reporting district. Other applications of the proposed estimator include the estimation of the growth curves in general when more than one experimental unit is available.

A. List of The Chemicals Targeted By The 33/50 Program.

Benzene	Lead and Compounds	Tetrachloroethylene
Cadmium and Compounds	Mercury and Compounds	Toluene
Carbon Tetrachloride	Methyl Ethyl Ketone	Trichloroethane
Chloroform	Methyl Isobutyl Ketone	Trichloroethylene
Chromium and Compounds	Methylene Chloride	Xylenes
Cyanides	Nickel and Compounds	

Source: 33/50 Program The final record, EPA march 1999

B. Proof of Proposition 3.1.

In what follows the subscript j is dropped for simplicity.

1. Under assumptions A1-A4, the mean and variance of the proposed estimator are

$$E[\hat{m}(x) - m(x)] = \frac{1}{2}\mu_2 h^2 \left[r''(x) + 2r'(x)\frac{f'(x)}{f(x)} \right] m_p(x) + o(h^2) \qquad \text{(B.1)}$$

$$\mathrm{V}ar[\hat{m}(x)] = \frac{\sigma^2 R(K)}{(nh)f(x)} + O(h/n + (Nh_p)^{-1}). \qquad \text{(B.2)}$$

Proof: $\hat{m}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)\left(\frac{Y_i}{\hat{f}(x)}\right)\left(\frac{\hat{m}_p(x)}{\hat{m}_p(X_i)}\right)$. A Taylor series expansion of $\frac{\hat{m}_p(x)}{\hat{m}_p(X_i)}$ around $\frac{m_p(x)}{m_p(X_i)}$ yields

$$\begin{aligned} \hat{m}(x) &\simeq \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)\frac{Y_i}{\hat{f}(x)}\left(\frac{m_p(x)}{m_p(X_i)} + \frac{\hat{m}_p(x) - m_p(x)}{m_p(X_i)}\right. \\ &\quad \left. - \left(\frac{m_p(x)}{m_p(X_i)}\frac{\hat{m}_p(X_i) - m_p(X_i)}{m_p(X_i)}\right)\right). \end{aligned}$$

The expressions

$$\frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)\frac{\epsilon_i}{\hat{f}(x)}\frac{\hat{m}_p(x) - m_p(x)}{m_p(X_i)}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)\frac{\epsilon_i}{\hat{f}(x)}\frac{m_p(x)}{m_p(X_i)}\frac{\hat{m}_p(X_i) - m_p(X_i)}{m_p(X_i)}$$

are of order $o_p(h_p^2)$; hence,

$$\begin{aligned} \hat{m}(x) &\simeq \frac{1}{n\hat{f}(x)}\sum_{i=1}^{n} K_h(X_i - x)Y_i\left[\frac{m_p(x)}{m_p(X_i)}\right] \\ &\quad + \frac{1}{n\hat{f}(x)}\sum_{i=1}^{n} K_h(X_i - x)r(X_i)\left[\hat{m}_p(x) - m_p(x) - \hat{m}_p(X_i) - m_p(X_i)\frac{m_p(x)}{m_p(X_i)}\right] \\ &\quad + o_p(h_p^2) \\ &= \frac{A_n}{\hat{f}(x)} + \frac{B_n}{\hat{f}(x)} + o_p(h_p^2) \end{aligned}$$

where $\frac{A_n}{\hat{f}(x)}$ is the proposed estimator with a non-random start. First let's find the bias of the proposed estimator.

$$E(\hat{m}(x) - m(x)) \quad = \quad \frac{1}{f(x)} E(A_n - m_p(x) r(x) \hat{f}(x)) + \frac{1}{f(x)} E(B_n) + o_p(1)$$

Let $\epsilon_i^* = \frac{\epsilon_i}{m_p(X_i)}$, then

$\frac{m_p(x)}{n} \sum_{i=1}^n K_h(X_i - x)(r(X_i) + \epsilon_i^* - r(x))$

$$\begin{aligned}
E(A_n - m_p(x) r(x) \hat{f}(x)) &= m_p(x) E\left(n^{-1} \sum_{i=1}^n K_h(X_i - x)(r(X_i) + \epsilon_i^* - r(x))\right) \\
&= m_p(x) E\left(n^{-1} \sum_{i=1}^n K_h(X_i - x)(r(X_i) - r(x))\right) \\
&\quad \text{by the law of iterated expectations} \\
&= m_p(x) \int K_h(X_1 - x)\left(r(X_1) - r(x)\right) f(X_1) dX_1 \\
&= m_p(x) \int K(\omega)\left(r(x + h\omega) - r(x)\right) f(x + h\omega) d\omega \\
&\quad \text{after a change of variable} \\
&= \frac{h^2}{2}\left(m_p(x) f(x) r''(x) + 2 m_p(x) f'(x) r'(x)\right) \mu_2(K) + o(h^2).
\end{aligned}$$

Denote B_n^1 and B_n^2 respectively the first and second terms of B_n.

$$\begin{aligned}
E(B_n^1) &= E \frac{1}{n} \sum_{i=1}^n K_h(X_i - x) r(X_i) E_{X_i}\left(\hat{m}_p(x) - m_p(x)\right) \\
&= \frac{1}{2} \mu_2 h_p^2 \left(m_p''(x) + 2 m_p'(x) \frac{g'(x)}{g(x)} + O(h_p^2)\right) E\left(\frac{1}{n} \sum_{i=1}^n K_h(X_i - x) r(X_i)\right) \\
&= Bias(\hat{m}_p(x)\left(r(x) f(x) + O(h^2)\right).
\end{aligned}$$

where $g(x)$ is the density of the pooled data.

Similarly, it can be seen that $E(B_n^2) = Bias(\hat{m}_p(x)\left(r(x) f(x) + O(h^2)\right)$.

Since $B_n = B_n^1 - B_n^2$, it follows that $E(B_n) = 0$ ignoring terms of $O(h^4)$ and smaller order (s.o) terms. Hence $E(\hat{m}(x) - m(x)) = \frac{1}{f(x)} E(A_n - m_p(x) r(x) \hat{f}(x)) + o_p(1)$.

Now turning to the variance. It is easily verified that

$$Var[A_n] = \sigma^2 (nh)^{-1} R(K) f(x) + O(h/n).$$

Also, $Var(B_n)) = Var(B_n^1) + Var(B_n^2) - 2Cov(B_n^1, B_n^2))$.

$$\begin{aligned} Var[E_{X_i}(B_n^1)] &= Var[n^{-1}\sum_{i=1}^{n} r(X_i)K_h(X_i - x)E_{X_i}(\hat{m}_p(x) - m_p(x))] \\ &= O(h_p^4)Var[\frac{1}{n}\sum_{i=1}^{n} r(X_i)K_h(X_i - x)] \\ &= o(h_p^4) \end{aligned}$$

$$\begin{aligned} E[Var_{X_i}(B_n^1)] &= E(Var_{X_i}[\frac{1}{n}\sum_{i=1}^{n} r(X_i)K_h(X_i - x)(\hat{m}_p - m_p)]) \\ &= O[(Nh_p)^{-1}]\frac{1}{n^2}E\sum_{i=1}^{n} r^2(X_i)K_h^2(X_i - x) + \\ &\quad O[(Nh_p)^{-1}]E\frac{1}{n^2}\sum_{i\neq j} r(X_i)r(X_j)K_h(X_i - x)K_h(X_j - x) - E(B_n^1)^2 \\ &= O[(Nh_p)^{-1}]\frac{1}{n^2}E\sum_{i=1}^{n} r^2(X_i)K_h^2(X_i - x) + \\ &\quad O[(Nh_p)^{-1}]\frac{n(n-1)}{n^2}[Er(X_1)K_h(X_1 - x)]^2 + O(h_p^4) \\ &= o((Nh_p)^{-1}) + O((Nh_p)^{-1} + h_p^4) \end{aligned}$$

Since $Var[B_n^1] = Var[E_{X_i}(B_n^1)] + E[Var_{X_i}(B_n^1)]$ and using the fact that $h_p^4 \propto (Nh_p)^{-1}$ for optimal rate of convergence of the asymptotic MISE of $\hat{m}_p(x)$, it follows that $Var[B_n^1] = O(Nh_p)^{-1}) + o(Nh_p)^{-1}$.

Similar calculations give $Var[B_n^2] = O(Nh_p)^{-1}) + o(Nh_p)^{-1}$. Let us find the covariance between B_n^1 and B_n^2. First, note that

$E_{X_i,X_j}[(\hat{m}_p(x) - m_p(x)(\hat{m}_p(X_j) - m_p(X_j)] = Cov_{X_i,X_j}(\hat{m}_p(x), \hat{m}_p(X_i))+$

$E_{X_i,X_j}[(\hat{m}_p(x) - m_p(x)]E_{X_i,X_j}[(\hat{m}_p(X_j) - m_p(X_j)]$.

Furthermore, the covariance between $\hat{m}_p(x)$ and $\hat{m}_p(X_i)$ is given by

$$Cov(\hat{m}_p(x), \hat{m}_p(X_i)) = \frac{\sigma^2}{Nh_p g(x)} K \circ K(\frac{x - X_i}{h}), \quad \text{where } K \circ K \text{ is the convolution Kernel.}$$

It follows that

$$Cov(B_n^1, B_n^2) = O((Nh_p)^{-1}) + o((Nh_p)^{-1}).$$

Using the same procedure as above, it is found that

$$Cov(A_n, B_n) = O(Nh_p)^{-1} \quad \text{after ignoring s.o terms.}$$

It follows that $Var[\hat{m}(x)] = \sigma^2(nhf(x))^{-1}R(K) + O(h/n) + O(Nh_p)^{-1}$. This completes the first part of the proof.

2. Under the assumptions A1-A6, $\hat{m}(x)$ has a limiting normal distribution:

$$\sqrt{nh}(\hat{m}(x) - m(x) - B(h)) \rightarrow N(0, \Sigma) \tag{B.3}$$

where $B(h) = \frac{1}{2}\mu_2 h^2[m_p(x)r''(x) + 2m_p(x)r'\frac{f'(x)}{f(x)}]$ and $\Sigma = \frac{\sigma^2}{f(x)}R(K)$.

Write $(\hat{m}(x) - m(x))\hat{f}(x) = C_n + D_n + o_p(h_p^2)$ where

$$\begin{aligned} C_n &= \frac{m_p(x)}{n}\sum_{i=1}^{n} K_h(X_i - x)(r(X_i) - r(x)) \\ &+\frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)r(X_i)(\hat{m}_p(x) - m_p(x)) \\ &-\frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)\frac{m_p(x)}{m_p(X_i)}r(X_i)(\hat{m}_p(X_i) - m_p(X_i)) \\ &\text{and} \\ D_n &= \frac{m_p(x)}{n}\sum_{i=1}^{n}\frac{\epsilon_i}{m_p(X_i)}K_h(X_i - x). \end{aligned}$$

From the first part of the proof of the theorem, it can be seen that

$$E(C_n) = \frac{h^2}{2}\left(m_p(x)f(x)r''(x) + 2m_p(x)f'(x)r'(x)\right)\mu_2(K) + o(h^2)$$

and that

$$Var(C_n) = o(h^4) + O(\frac{1}{n_1h_p + n_2h_p + ... + n_Qh_p}).$$

By assumption A5, $n_jh_p \rightarrow \infty \ \forall \ j = 1, ..., J$; hence the last term of the variance of C_n can be ignored for a reasonably large J. Combining the expectation and variance of C_n, it follows that

$$\begin{aligned} C_n &= E(C_n) + o_p(h^2) \\ &= \frac{h^2}{2}\left(m_p(x)f(x)r''(x) + 2m_p(x)f'(x)r'(x)\right)\mu_2(K) + o_p(h^2) \\ &= f(x)B(h) + o_p(h^2); \end{aligned}$$

Similarly, $E(D_n) = 0$ and $Var(D_n) = (nh)^{-1}\left(\sigma^2 R(K)f(x) + o(1)\right)$. D_n is a triangular array of i.i.d. random variables; thus, under assumption A6,

$$(nh)^{-\delta/2} E\left|\frac{m_p(x)}{m_p(X_i)}\right|^{2+\delta} E\left|\epsilon_i\right|^{2+\delta} E\left|K_h(X_i - x)\right|^{2+\delta} h^{-1} \to 0 \quad \text{as} \quad n \to \infty.$$

Hence we can apply Liapounov's central limit theorem to obtain $\sqrt{nh}(D_n) \to N(0, f^2(x)\Sigma)$. Since $plim\hat{f}(x) = f(x)$, it also follows that

$$\sqrt{nh}(\hat{m}(x) - m(x) - B(h)) = \sqrt{nh}\frac{D_n}{f(x)} + o_p(1) \to N(0, \Sigma). \tag{B.4}$$

This completes the second part of the proof.

C. Proof of Propositions 4.2 and 4.3.

The methodology of proof adopted here closely follows closely that of Nicolau (2003). The following regularity conditions are required:

(1) each of the J interest rate processes satisfies assumption A1 in Nicolau (2003), i.e each process has a unique strong solution,

(2) each of the J interest processes satisfies assumptions A1 and A2 in Nicolau (2003), i.e each process has an invariant density $p_j(r), j = 1...J$,

(3) under assumption A4 in Nicolau (2003), the short rate diffusion process is ρ-mixing, meaning that for square integrable functions $v_n(r)$ and $w_n(r)$, $\rho(v_n(r), w_n(r)) \to 0$ as $n \to \infty$,

(4) the Kernel function $K(u)$ satisfies assumption A4 in chapter two.

Again for notational convenience the superscript (1) indicating the short rate and the subscript (1) indicating its drift and density are omitted. The proof of propositon 4.1 is similar to the proof of proposition 4.2 thus not provided.

C.1. Proof of Proposition 4.2

A Taylor series expansion of $\frac{\hat{\mu}_p(r)}{\hat{\mu}_p(r_{t\delta})}$ around $\frac{\mu_p(r)}{\mu_p(r_{t\delta})}$ yields

$$
\begin{aligned}
\hat{\mu}(r) \quad &\simeq \quad \frac{1}{n\hat{p}(r)} \sum_{t=0}^{n-1} K_h(r_{t\delta} - r) \frac{(r_{(t+1)\delta} - r_{t\delta})}{\delta} \left[\frac{\mu_p(r)}{\mu_p(r_{t\delta})} \right] \\
&\quad + \frac{1}{n\hat{p}(r)} \sum_{t=0}^{n-1} K_h(r_{t\delta} - r) \frac{(r_{(t+1)\delta} - r_{t\delta})}{\delta} \left[\frac{\hat{\mu}_p(r) - \mu_p(r)}{\mu_p(r_{t\delta})} - \frac{\hat{\mu}_p(r_{t\delta}) - \mu_p(r_{t\delta})}{\mu_p(r_{t\delta})} \frac{\mu_p(r)}{\mu_p(r_{t\delta})} \right] \\
&= \quad \frac{1}{\hat{p}(r)} [A_n + B_n]
\end{aligned}
$$

$$
E[\hat{\mu}(r) - \mu(r)] \quad = \quad \frac{1}{p(r)} E(A_n - \mu_p(r) c(r) \hat{p}(r)) + \frac{1}{p(r)} E(B_n) + o_p(1)
$$

$$
\begin{aligned}
E(A_n - \mu_p(r)c(r)\hat{p}(r)) &= \frac{\mu_p(r)}{n} E\left(\sum_{t=0}^{n-1} K_h(r_t - r)[\frac{(r_{(t+1)\delta} - r_{t\delta})}{\delta \mu_p(r_t)} - c(r)]\right) \\
&= \frac{\mu_p(r)}{n} E\left(\sum_{t=0}^{n-1} K_h(r_t - r)[c(r_t) + l(r_t)\delta - c(r)]\right) \\
&= \mu_p(r) \int K(z)[c(r+hz) + l(x+hz)\delta - c(r)]f(r+hz)dz \\
&= \mu_p(r)\{l(r)\delta + \frac{h^2}{2}\mu_2(k)[c''(r)p(r) + 2c'(r)p'(r)]\} + o(h^2 + \delta)
\end{aligned}
$$

It is can be seen that $E(B_n) = 0$ after ignoring terms of the orders h_p^4, $h_p^2\delta$ and smaller. Hence

$$
E[\hat{\mu}(r) - \mu(r)] = \mu_p(r)\{l(r)\delta + \frac{h^2}{2}\mu_2(k)[c''(r) + 2c'(r)\frac{p_1'(r)}{p_1(r)}]\} + o(h^2 + \delta). \tag{C.1}
$$

where $l(r)$ is as defined in proposition 4.2. Now turning to the variance, we have:

$$
\begin{aligned}
Var[\hat{\mu}(r)] &= [Var(\frac{A_n}{\hat{p}(r)}) + Var(\frac{B_n}{\hat{p}(r)}) + 2Q] \quad \text{where } Q = Cov(\tfrac{A_n}{\hat{p}(r)}, \tfrac{B_n}{\hat{p}(r)}) \\
Var\left[\frac{A_n}{\hat{p}(r)}\right] &\simeq Var\left(\frac{A_n - \hat{p}(r)\mu(r)}{p(r)}\right) \\
&= \frac{\mu_p^2(r)}{p^2(r)} Var\{\frac{1}{n}\sum_{t=0}^{n-1} K_h(r_t - r)[\frac{(r_{t+\delta} - r_t)}{\delta\mu_p(r_t)} - c(r)]\} \\
&= \frac{1}{p(r)^2}\frac{1}{Th} Var\{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g(r_{t\delta}, r_{(t+1)\delta})\}
\end{aligned}
$$

where

$$
g(r_{t\delta}, r_{(t+1)\delta}) = \mu_p(r)\sqrt{\frac{\delta}{h}} K(\frac{r_{t\delta} - r}{h})[\frac{(r_{(t+1)\delta} - r_{t\delta})}{\delta\mu_p(r_{t\delta})} - c(r)].
$$

Let

$$
\dot{g}(r_{t\delta}, r_{(t+1)\delta}) = g(r_{t\delta}, r_{(t+1)\delta}) - E(g(r_{t\delta}, r_{(t+1)\delta})).
$$

$E\dot{g}(r_{t\delta}, r_{(t+1)\delta}) = 0$ and $E(\dot{g}^2(r_{t\delta}, r_{(t+1)\delta})) = E(g^2(r_{t\delta}, r_{(t+1)\delta})) - [E(g(r_{t\delta}, r_{(t+1)\delta}))]^2$.

$$
\begin{aligned}
E(g(r_{t\delta}, r_{(t+1)\delta})) &= \mu_p(r)\sqrt{\frac{\delta}{h}} E\{\frac{1}{n}\sum_{t=0}^{n-1} K_h(r_{t\delta} - r)[\frac{(r_{(t+1)\delta} - r_{t\delta})}{\delta\mu_p(r_{t\delta})} - c(r)]\} \\
&= \mu_p(r)\sqrt{h\delta}\, O(h^2 + \delta) \to 0.
\end{aligned}
$$

$$
\begin{aligned}
E(g^2(r_{t\delta}, r_{(t+1)\delta})) &= \mu_p^2(r)\frac{\delta}{h}EK^2(\frac{r_{t\delta}-r}{h})E[\{\frac{r_{(t+1)\delta}-r_{t\delta}}{\delta\mu_p(r_{t\delta})} - c(r)\}^2|r_{t\delta}] \\
&= \mu_p^2(r)\int K^2(z)[\sigma^2(r+hz)+O(h^2\delta)]\frac{G(r+hz)}{\mu_p^2(r+hz)}dz \\
&= \sigma^2(r)R(K)f(r)+O(h^2+\delta) < \infty.
\end{aligned}
$$

The next step is to find the autocovariance function of $g(r_{t\delta}, r_{(t+1)\delta})$. For $t=0$,

$$
\begin{aligned}
E(g(r_0;r_\delta)g(r_\delta;r_{2\delta})) &= \mu_p^2(r)\frac{\delta}{h}E[K(\frac{r_0-r}{h})K(\frac{r_\delta-r}{h})\{\frac{r_\delta-r_0}{\delta\mu_p(r_0)} - c(r)\} * \\
&\quad \{\frac{r_{2\delta}-r_\delta}{\delta\mu_p(r_\delta)} - c(r)\}]
\end{aligned}
$$

By the law of iterated expectations,

$$
\begin{aligned}
E(g(r_0;r_\delta)g(r_\delta;r_{2\delta})) &= \mu_p^2(r)\frac{\delta}{h}E[K^2(\frac{r_0-r}{h})\{(c(r_0-c(r))^2+O(\delta\}] \\
&= \mu_p^2(r)\delta\int K^2(z)[(c(r)+O(h)-c(r))^2+O(\delta)][p(r)+O(h)]dz \\
&= \mu_p^2(r)O(h^2+\delta)\delta \to 0.
\end{aligned}
$$

Under assumption 4 of Nicolau, the process is ρ-mixing, hence for $t \geq 1$

$$
|E(g(r_0;r_\delta)g(r_{t\delta};r_{(t+1)\delta}))| \leq |E(g(r_0;r_\delta)g(r_\delta;r_{2\delta}))| \to 0. \tag{C.2}
$$

Applying lemma 9 of Nicolau (2003), it follows that

$$
\begin{aligned}
Var[\frac{A_n}{\hat{p}(r)}] &\simeq \frac{1}{(Th)}\frac{1}{p(r)^2}(E[\dot{g}^2(r_{t\delta}, r_{(t+1)\delta})]+o(1)) \\
&= \frac{\sigma^2(r)R(K)+o(1)}{(Th)p(r)}.
\end{aligned}
$$

Now turning to the covariance function Q.

$$
\begin{aligned}
Q &\simeq Cov\left(\frac{A_n}{p(r)}, \frac{B_n}{p(r)}\right) \\
Cov(A_n, B_n) &= \tfrac{1}{n^2}\sum_{t=0}^{n-1}E(g_1(r_{t\delta})g_2(r_{t\delta})) + \frac{2}{n^2}\sum_{t<t'}E(g_1(r_{t\delta})g_2(r_{t'\delta})) - E[A_n]E[B_n]
\end{aligned}
$$

where $g_1(r_{t\delta}) = K_h(r_{t\delta}-r)\frac{(r_{(t+1)\delta}-r_{t\delta})}{\delta}\frac{\mu_p(r)}{\mu_p(r_{t\delta})}$, and

$g_2(r_{t\delta}) = K_h(r_{t\delta}-r)\frac{(r_{(t+1)\delta}-r_{t\delta})}{\delta}[\frac{\hat{\mu}_p(r)-\mu_p(r)}{\mu_p(r_{t\delta})} - \frac{\hat{\mu}_p(r_{t\delta})-\mu_p(r_{t\delta})}{\mu_p(r_{t\delta})}\frac{\mu_p(r)}{\mu_p(r_{t\delta})}]$.

First let us focus on the second expression of the covariance. By the law of iterated expectations, it follows that for $t = 0$ and $t' = 1$, $E(g_1(r_0)g_2(r_\delta)) = O(h^2h_p^2 + h^2\delta)$, driven by the fact that $E(\hat{\mu}_p(r) - \mu_p(r)) = O(h_p^2 + \delta)$.

Again by assumption 4 of Nicolau, the process is ρ-mixing, thus, for $t' > 1$

$$|E(g_1(r_0)g_2(r_{(t')\delta}))| \leq |E(g_1(r_0)g_2(r_\delta))| = O(h^2h_p^2 + h^2\delta) \to 0. \tag{C.3}$$

It follows that $\frac{2}{n^2}\sum_{t<t'} E(g_1(r_{t\delta})g_2(r_{t'\delta}))$ is negligible. Furthermore, the first term in the covariance expression, $\frac{1}{n^2}E(g_1(r_{t\delta})g_2(r_{t\delta}))$, is of the order of $o((Th)^{-1})$ while the last term $E(A_n)E(B_n) = O(1)O(h_p^4 + h_p^2\delta)$

Similar calculations show that

$$Var(B_n) = 2\left(\frac{\sigma_p^2(r)}{(Nh_p)g(r)}\right)\gamma(r) \tag{C.4}$$

where $g(r)$ is the density of the pooled data, $\gamma(r) < \infty$, and $N = JT$.

C.2. Proof of Proposition 4.3

Let $\hat{B}_p(r) = \hat{\mu}_p(r) - \mu_p(r)$. It can be easily shown that the pooled drift function estimator is weakly consistent, that is $\hat{B}_p(r) = \epsilon_N \to 0$ in propability as $N \to \infty$.

Write

$$\begin{aligned}
\hat{\mu}(r) - \mu(r) &= \frac{(\hat{\mu}(r) - \mu(r))\,\hat{p}(r)}{\hat{p}(r)} \\
&= \frac{\left(\hat{\mu}(r) - \mu(r) - c(r)\hat{B}_p(r) + c(r)\hat{B}_p(r)\right)\hat{p}(r)}{\hat{p}(r)} \\
&= \frac{\mu_p(r)}{nh\hat{p}(r)}\sum_{t=0}^{n-1} K(\frac{r_{t\delta} - r}{h})\{\frac{r_{(t+1)\delta} - r_{t\delta}}{\delta\mu_p(r_{t\delta})} - c(r)\} \\
&\quad + \frac{1}{nh\hat{p}(r)}\sum_{t=0}^{n-1} K(\frac{r_{t\delta} - r}{h})\{\frac{r_{(t+1)\delta} - r_{t\delta}}{\delta\mu_p(r_{t\delta})} - c(r)\}(\epsilon_N) \\
&\quad - \frac{1}{nh\hat{p}(r)}\sum_{t=0}^{n-1} K(\frac{r_{t\delta} - r}{h})\{\frac{r_{(t+1)\delta} - r_{t\delta}}{\delta\mu_p(r_{t\delta})}\frac{\mu_p(r)}{\mu_p(r_{t\delta})}) - c(r)\}(\epsilon_N)
\end{aligned}$$

Hence

$$\sqrt{\frac{Th}{R(K)}\hat{p}(r)}\left(\frac{\hat{\mu}(r)-\mu(r)}{\sigma(r)}\right) = \frac{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g(r_{t\delta};r_{(t+1)\delta})}{\sigma(r)\sqrt{R(K)\hat{p}(r)}} + \sqrt{Th}[C_n] + \sqrt{Th}[D_n]$$

where $g(r_{t\delta};r_{(t+1)\delta})$ is as defined earlier, and

$$\begin{aligned} C_n &= \frac{1}{\sigma(r)\sqrt{R(K)\hat{p}(r)}}\frac{1}{nh}\sum_{t=0}^{n-1} K(\frac{r_{t\delta}-r}{h})\{\frac{r_{(t+1)\delta}-r_t}{\delta\mu_p(r_{t\delta})} - c(r)\}\epsilon_N \\ D_n &= \frac{1}{\sigma(r)\sqrt{R(K)\hat{p}(r)}}\frac{1}{nh}\sum_{t=0}^{n-1} K(\frac{r_{t\delta}-r}{h})\{\frac{r_{(t+1)\delta}-r_{t\delta}}{\delta\mu_p(r_{t\delta})}\frac{\mu_p(r)}{\mu_p(r_{t\delta})}) - c(r)\}\epsilon_N \end{aligned}$$

Using Slutsky's theorem, it can be proved that $\sqrt{Th}[C_n] = \sqrt{Th}[D_n] = o_p(1)$, thus asymptotically negligible.

Since $\frac{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g(r_{t\delta};r_{(t+1)\delta})}{\sigma(r)\sqrt{R(K)\hat{p}(r)}}$ and $\frac{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g(r_{t\delta};r_{(t+1)\delta})}{\sigma(r)\sqrt{R(K)p(r)}}$ have the same asymptotic distribution, it is enough to show that $\frac{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g(r_{t\delta};r_{(t+1)\delta})}{\sigma(r)\sqrt{R(K)p(r)}} \to N(0,1)$

$$\begin{aligned} E(g(r_{t\delta};r_{(t+1)\delta})) &= \mu_p(r)\sqrt{h\delta}\{O(h^2+\delta)\}(\text{ see proof of proposition 4.2}), \text{ hence} \\ E(\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g(r_{t\delta};r_{(t+1)\delta})) &= \mu_p(r)(\sqrt{nh^5}+\sqrt{nh}\delta)O(\sqrt{\delta}) \to 0 \end{aligned} \tag{C.5}$$

under the assumptions of proposition 4.3, and

$$E(g^2(r_{t\delta};r_{(t+1)\delta})) \quad \to \sigma^2(r)R(K)p(r) \text{ (see proof of proposition 4.2)} \tag{C.6}$$

Futhermore

$$E(g(r_0;r_\delta)g(r_{t\delta};r_{(t+1)\delta})) \to 0 \ \forall t \geq 1 \text{ (see proof of proposition 4.2)} \tag{C.7}$$

It follows from equations (C.5),(C.6),(C.7), and Lemma 9 of Nicolau (2003) that:

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{n} g(r_{t\delta};r_{(t+1)\delta}) \to N(0,\sigma^2(r)R(K)p(r)) \tag{C.8}$$

References

[1] Abhyanker, A., and D. Basu (2001): "Does Conditioning Information Matter in Estimating Continuous-time Interest Rate Diffusions," *Journal of Financial and Quantitative Analysis*, 36, 335-344.

[2] Aït-Sahalia, Y. (1996a): "Nonparametric pricing of interest rate derivative securities," *Econometrica*, 64, 527-560.

[3] Aït-Sahalia, Y. (1996b): "Testing Continuous-time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385-426.

[4] Alberini, A. and D. Austin (1999): "Strict Liability as a Deterrent in Toxic Waste Management: Empirical Evidence from Accident and Spill Data," *Journal of Environmental Economics and Management*, 38, 20-48.

[5] Altman, N.S., and G. Casella (1995): "Nonparametric Empirical Bayes Growth Curve Analysis," *Journal of the American Statistical Association*, 90, 508-514.

[6] Andersen, T. G. and J. Lund (1997): "Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate," *Journal of Econometrics*, 77, 343-377.

[7] Anton, W., G. Deltas, and M. Khanna (2004): "Incentives for Environmental Self-Regulation and Implications for Environmental Performance," *Journal of Environmental Economics and Management*, 48, 632-654.

[8] Arora, S. and T. Cason (1996): "Why Do Firms Volunteer To Exceed Environmental Regulations? Understanding Participation in EPA's 33/50 Program," *Land Economics*, 72, 413-432.

[9] Arora, S. and S. Gangopadhyay (1995): "Towards a Theoretical Model of Voluntary Overcompliance," *Journal of Economic Behavior and Organization*, 28, 289-309.

[10] Backus, D.K., S. Foresi, and S.E. Zin (1995): "Arbitrage Opportunities in Arbitrage-free Models of Bond Pricing," *Working paper*, New York University.

[11] Bandi, F., P. Phillips (2003): "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, 71, 241-283.

[12] Baron, D. (2001): "Private Politics, Corporate Social Responsibility, and Integrated Strategy," *Journal of Economics and Management Strategy*, 10, 7-45.

[13] Black, F., E. Dermon, and W. Toy (1990): "A one factor model of interest rates and its application to Treasury bond options," *Financial analysts' Journal*, 46, 33-39.

[14] Black, F. and P. Karasinski (1991): "Bond and Option Pricing When Short Rates are Lognormal," *Financial Analysts' Journal*, July/August, 52-59.

[15] Boyle, P., M. Broadie, and P. Glasserman (1996): " Monte Carlo Methods for Security Pricing," *Journal of Economic Dynamics and Control*, 21, 1267-1321.

[16] Brandt, M. W., and A. Yaron (2003): "Time-consistent no-arbitrage models of the term structure," *Working paper,* NBER.

[17] Brennan, M. J., and E. S. Schwartz (1979): "A continuous time approach to the pricing of bonds," *Journal of Banking and Finance*, 3, 133-155.

[18] Cameron, C. and P. Trivedi (1998): *Regression Analysis of Count Data*, Cambridge University Press.

[19] Canabarro, E. (1995):"Where do one-factor interest rate models fail?," *Journal of Fixed Income*, 5, 31-52.

[20] Chan, K. C., A. G. Karolyi, Francis A. Longstaff, and Anthony B. Sanders (1992): "An empirical comparison of alternative models of the short-term interest rate," *Journal of Finance*, 47, 1209-1227.

[21] Chapman, David A., and Neil D. Pearson (2000): "Is the short rate drift actually nonlinear?," *Journal of Finance*, LV, 355-388.

[22] Conley, T.G., L.P. Hansen, E.G. Luttmer, and J.A. Scheinkman (1997): "Short-term Interest Rates as Subordinated Diffusions," *Review of Financial Studies*, 10, 525-577.

[23] Courtadon, G. (1982): "The Pricing of Options on Default-free Bonds," *Journal of Financial and Quantitative Analysis*, 17, 75-100.

[24] Cox, John, J. E. Ingersoll, and S. A. Ross (1980): "An Analysis of Variable Rate Loan Contracts," *Journal of Finance*, 35, 389-403.

[25] Cox, J., J. E. Ingersoll, and S. A. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-408.

[26] Decker, C. (2003): "Corporate Environmentalism and Environmental Statutory Permitting," *Journal of Law and Economics*, 46, 103-29.

[27] Decker, C. (2004): "Do Regulators Respond to Voluntary Pollution Control Efforts? A Count Data Analysis," *Working Paper*, University of Nebraska.

[28] Fan, J., and C. Zhang (2003): " A re-examination of Stanton's Diffusion Estimations with Applications to Financial Model Validation ," *Journal of the American Statistical Association*, 98, 118-134.

[29] Glad, I. (1998a): "Parametrically-Guided Nonparametric Regression," *Scandinavian Journal of Statistics*, 25, 649-668.

[30] Glad, I. (1998b): "A Note on Unconditional Properties of a Parametrically Guided Nadaraya-Watson Estimator," *Statistics and Probability Letters*, 37, 101-108.

[31] Gray, W., and M. Deily (1996): "Compliance and Enforcement: Air Pollution Regulation in the U.S. Steel Industry," *Journal of Environmental Economics and Management*, 31, 96-111.

[32] Greene, W. (2000): *Econometric Analysis, Fourth Edition*, Prentice-Hall.

[33] Hamilton, J. (1995): "Pollution as News: Media and Stock market Reactions to the Toxic Release Inventory Data," *Journal of Environmental Economics and Management*, 28, 98-113.

[34] Hardle, W., and S. Browman (1988): "Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bounds," *Journal of the American Statistical Association*, 83, 102-110.

[35] Harford, J. and W. Harrington (1991): "A Reconsideration of Enforcement Leverage When Penalties Are Restricted," *Journal of Public Economics*, 45, 391-95.

[36] Harrington, W. (1988): "Enforcement Leverage When Penalties Are Restricted," *Journal of Public Economics*, 37, 29-53.

[37] Harvey, A. (1976): "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica*, 44, 461-5.

[38] Hausman, J., B. Hall, and Z. Grilliches (1984): "Economic Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52, 909-38.

[39] Heckman, J. (1978): "Dummy Endogenous Variables In a Simultaneous Equation System," *Econometrica*, 46, 931-959.

[40] Heckman, J. and J. Hotz (1989): "Choosing Among Alternative Non-experimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84, 862-874.

[41] Helland, E. (1998): "The Enforcement of Pollution Control Laws: Inspections, Violations, and Self-Reporting," *Review of Economics and Statistics*, 80, 141-53.

[42] Henriques, I. and P. Sadorsky (1996): "The Determinants of an Environmentally Responsive Firm: An Empirical Approach," *Journal of Environmental Economics and Management*, 30, 381-385.

[43] Herrmann, E. (1997): "Local Bandwidth Choice in Kernel Regression Estimation," *Journal of Computational and Graphical Statistics*, 6, 35-54.

[44] Heyes, A. and N. Rickman (1999): "Regulatory Dealing - Revisiting the Harrington Paradox," *Journal of Public Economics*, 72, 361-78.

[45] Hjort, N.L., and I. Glad (1995): "Nonparametric Density with a Parametric Start," *The Annals of Statistics*, 23, 882-904.

[46] Hurvitch, C., and J.S. Simonoff (1998): "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion," *Journal of Royal Statistical Society Series B*, 60, 271-293.

[47] Innes, R (2003), "Corporate Social Responsibility Under the Threat of Consumer Boycott," *Working paper*, University of Arizona.

[48] Innes, R., and J. Bial (2002): "Inducing Innovation in the Environmental Technology of Oligopolistic Firms," *Journal of Industrial Economics*, 50, 265-287.

[49] Jamshidian, F. (1990): "The Preference-Free Determination of Bond and Option Prices from the Spot Interest Rate," *Advances in Futures and Options Research*, 4, 51-67.

[50] Jamshidian, F. (1991): "Bond and Option Valuation in the Gaussian Interest Model," *Research in Finance*, 9,131-170.

[51] Jiang, G., J. L. Knight (1997): "A Nonparametric Approach to the Estimation of Diffusion Processes with an Application to a Short-Term Interest Rate Model," *Econometric Theory*, 13, 615-645.

[52] Jiang, G. (1998); "Nonparametric Modeling of U.S Interest Rate Term Structure Dynamics and Implications on the Price of Derivative Securities," *Journal of Financial and Quantitative Analysis*, 33, 465-497.

[53] Jones, C.S. (2003): " Nonlinear Mean Reversion in the Short-term Interest Rate," *Review of Financial Studies*, 16, 793-843.

[54] Karlin, S., and H. M. Taylor (1981): *A Second Course in Stochastic Processes*, Academic Pess, New York.

[55] Khanna, M. and L. Damon (1999): "EPA's Voluntary 33/50 Program: Impact on Toxic Releases and Economic Performance of Firms," *Journal of Environmental Economics and Management*, 37, 1-25.

[56] LaPlante, B. and Rilstone, P. (1996): "Environmental Inspections and Emissions of the Pulp and Paper Industry in Quebec," *Journal of Environmental Economics and Management*, 31, 19-36.

[57] Li, M., N. D. Pearson, and A. M. Poteshman (2004): "Conditional Estimation of Diffusion Processes," *working paper*, University of Illinois at Urbana-Champaign.

[58] Magat, W. and K. Viscusi (1990): "Effectiveness of the EPA's Regulatory Enforcement: The Case of Industrial Effluent Standards," *Journal of Law and Economics*, 33, 331-60.

[59] Marron, J.S., and M.P. Wand (1992): "Exact Mean Integrated Squared Error," *The Annals of Statistics*, 20, 712-736.

[60] Maxwell, J., and C. Decker (2003): "Voluntary Environmental Investment and Regulatory Responsiveness," *Working paper*, University of Nebraska.

[61] Maxwell, J., T. Lyon, and S. Hackett (2000): "Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism," *Journal of Law and Economics*, 43, 583-617.

[62] Merton , R.C. (1980): " On Estimating the Expected Return on the Market: An Explanatory Investigation," *Journal of Financial Economics*, 8, 323-361.

[63] Murphy, K., and R. Topel (1985): "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics*, 3, 370-379.

[64] Nadeau, L. (1997): "EPA Effectiveness at Reducing the Duration of Plant-Level Noncompliance," *Journal of Environmental Economics and Management*, 34, 54-78.

[65] Nicolau, J. (2003): " Bias Reduction in Nonparametric Diffusion Coefficient Estimation," *Econometric Theory*, 19, 754-777.

[66] Øksendal, B. (1985): *Stochastic Differential Equations: An Introduction with Applications, third edition*, Springer-Verlag, New York.

[67] Pagan, A., and A. Ullah (1999): *Nonparametric Econometrics*, Cambridge University Press.

[68] Pritsker, M. (1998), " Nonparametric Density Estimation and Tests of Continuous-time Interest Rate Models," *Review of Financial Studies*, 11, 449-487.

[69] Racine, J.S., and Qi Li (2004): "Nonparametric Regression with Both Categorical and Continuous Data," *Journal of Econometrics*, 119, 99-130.

[70] Ruppert, D., S.J. Sheather, and M.P. Wand (1995): "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, 90, 1257-1270.

[71] Salop, S., and D. Scheffman (1983): "Raising Rivals' Costs," *American Economic Review*, 73, 267-271.

[72] Segerson, K., and T. Miceli (1998): "Voluntary Environmental Agreements: Good or Bad News for Environmental Protection?," *Journal of Environmental Economics and Management*, 36, 109-130.

[73] Smith, N. (1990): *Morality and the Market: Consumer Pressure for Corporate Responsibility*, London: Rutledge.

[74] Stanton, R. (1997): "A nonparametric Model of Term Structure Dynamics and the Market Price of Interest Eate Risk," *Journal of Fiance*, 52, 1973-2002.

[75] Talay, D. (1996): "Probabilistic Numerical Methods for PDEs," In *Nonlinear Partial Differential Equations* Talay, D. and L. Tubaro eds. Berlin, Germany, Springer.

[76] Terza, J. (1998): "Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Switching Effects," *Journal of Econometrics*, 84, 129-154.

[77] U.S. Environmental Protection Agency (1999): *33/50 Program: The Final Record*.

[78] Vasicek, O. A. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177-188.

[79] Vella, F. (1998): "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources*, 33, 127-169.

[80] Videras, J. and A. Alberini (2000): "The Appeal of Voluntary Environmental Programs: Which Firms Participate and Why," *Contemporary Economic Policy*, 18, 449-461.