Essays on Nonparametric and Applied Econometrics

by

Ahmet Tolga Ergün

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

2 0 0 4

The University of Arizona ®
Graduate College

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Ahmet Tolga Ergun

entitled Essays on Nonparametric and Applied Econometrics

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

Alan P. Ker — date 6/24/04

Satheesh V. Aradhyula — date 6/24/04

Ronald L. Oaxaca — date 6/24/04

Alfonso Flores-Lagunes — date 6/24/04

Price V. Fishback — date 6/24/04

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.

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

Dissertation Director: Alan P. Ker — date 06/24/04

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: 

Acknowledgments

I would like to thank members of my committee; Alan P. Ker, Satheesh V. Aradhyula, Ronald L. Oaxaca, Alfonso Flores-Lagunes, and Price V. Fishback for their continuous support and encouragement during this research. Their comments and input have very significantly improved the quality of this dissertation.

Alan P. Ker and Satheesh V. Aradhyula deserve special thanks. I have known Dr. Aradhyula since day one at University of Arizona and he has always been a mentor and a friend. Finally, a very special thanks goes to my advisor Dr. Ker. Working with Alan and having him as a friend have been a true pleasure. Our conversations have taught me more than how to be a good researcher. I will always be grateful to him for taking this ride with me.

Dedication

Sevgili anne ve ablamın sonsuz desteği olmadan bu eğitimi tamamlamam mümkün olamazdı. Bu çalışmayı onlara adıyorum.

Table of Contents

TABLE OF CONTENTS—*Continued*

List of Figures

List of Tables

Abstract

This dissertation focuses on econometric methodology and its applications in insurance and the stock market.

The second chapter proposes a new semiparametric estimator for binary-choice single-index models. The estimator makes use of a "parametric start" idea from the statistics literature and applies it to econometric model estimation. Even though the chapter only focuses on binary-choice models, it is expected that the introduction of this idea to the econometrics literature is going to contribute to semiparametric estimation of econometric models in general, especially when one has (only) a rough initial guess about the shape of the unknown function. Consistency of the estimator is shown and the simulation results indicate that even though the parametric start is not correct in any of the simulation designs, the estimator's performance is very promising in the estimation of coefficients and probabilities.

The third chapter successfully applies this proposed estimator along with competing parametric and semiparametric estimators and is expected to expand our understanding of private insurance company involvement in the U.S. crop insurance program. This chapter stands almost alone in the literature as an overwhelming majority of other studies examine the involvement of producers in the program. Although preliminary, the results of this chapter show that the insurance company involvement in this program may be too costly to justify and that the program may not be as efficient in terms of premium rates and rating practices of the federal government.

The fourth chapter examines market volatility taking into account the New York Stock Exchange trading collar. The trading collar restricts certain forms of trade in component stocks of the S&P500 stock price index when there is "excess" volatility in the market. This important feature of the market has been ignored in the large volatility modeling literature and it is expected that this chapter contributes to this

literature by showing that after some data manipulation it is straightforward to incorporate this feature into standard econometric models. Another contribution of this chapter is the successful use of a polynomial specification to capture the well documented U-shaped pattern of intraday market volatility instead of a computationally more difficult two-step procedure.

Essays on Nonparametric and Applied Econometrics

Ahmet Tolga Ergün, Ph.D.
The University of Arizona, 2004

Director: Alan P. Ker

This dissertation focuses on econometric methodology and its applications in insurance and the stock market.

The second chapter proposes a new semiparametric estimator for binary-choice single-index models. The estimator makes use of a "parametric start" idea from the statistics literature and applies it to econometric model estimation. Even though the chapter only focuses on binary-choice models, it is expected that the introduction of this idea to the econometrics literature is going to contribute to semiparametric estimation of econometric models in general, especially when one has (only) a rough initial guess about the shape of the unknown function. Consistency of the estimator is shown and the simulation results indicate that even though the parametric start is not correct in any of the simulation designs, the estimator's performance is very promising in the estimation of coefficients and probabilities.

The third chapter successfully applies this proposed estimator along with competing parametric and semiparametric estimators and is expected to expand our understanding of private insurance company involvement in the U.S. crop insurance program. This chapter stands almost alone in the literature as an overwhelming majority of other studies examine the involvement of producers in the program. Although preliminary, the results of this chapter show that the insurance company involvement in this program may be too costly to justify and that the program may not be as efficient in terms of premium rates and rating practices of the federal government.

The fourth chapter examines market volatility taking into account the New York Stock Exchange trading collar. The trading collar restricts certain forms of trade in component stocks of the S&P500 stock price index when there is "excess" volatility

in the market. This important feature of the market has been ignored in the large volatility modeling literature and it is expected that this chapter contributes to this literature by showing that after some data manipulation it is straightforward to incorporate this feature into standard econometric models. Another contribution of this chapter is the successful use of a polynomial specification to capture the well documented U-shaped pattern of intraday market volatility instead of a computationally more difficult two-step procedure.

1. Dissertation Introduction

This dissertation focuses on econometrics with applications in insurance and the stock market. Although the semiparametric binary-choice model estimators that exist in the literature work well in most situations and provide flexibility that can not be obtained with parametric estimators, they do not incorporate reasonably reliable prior information about the shape of the unknown function. The motivation of the second chapter is the expectation that taking advantage of this fact can improve the performance of the semiparametric estimators. This is shown via simulations and an application with insurance data in the second and third chapters of this dissertation, respectively. Finally, the risk management aspect of this research is complimented in the fourth chapter with a stock market application. The feature of the stock market that is analyzed in this chapter has been ignored in the overwhelming majority of the finance and risk management literature. The motivation is to show that this feature can be easily incorporated into standard econometric models.

The second chapter of this dissertation proposes a new semiparametric estimator for binary-choice single-index models which uses parametric information in the form of a known link (probability) function and nonparametrically corrects it. The estimator introduces (potentially) useful information in the form of a parametric link function which is its guide in the sense of Hjort and Glad [27]. The distinguishing characteristic of the proposed estimator is how the unknown link function is estimated using prior parametric information about its shape. It is shown that the estimator is consistent and the finite-sample properties of the proposed estimator are compared to those of parametric probit and semiparametric single-index model estimators of Klein and Spady [41] and Ichimura [35]. A distinguishing feature of the simulations is that besides using a fixed smoothing parameter, the method of Härdle et al. [25] is also used to choose the smoothing parameter for the proposed estimator. Results indicate that even though the parametric start is not correct in any of the simulation designs, the proposed estimator achieves significant bias reduction and efficiency gain. For purposes of estimating probabilities, the proposed estimator outperforms the other two single-index model estimators in nine out of ten simulation designs.

The third chapter of this dissertation empirically tests the revelation of private information by the insurance companies via their reinsurance decisions in the U.S. crop insurance program—a prominent facet of the U.S. farm policy worth more than $16 billion and gaining more financial and political importance. The participation of insurance companies as intermediaries in the U.S. crop insurance program, along with the producers and the federal government, can be justified on the basis of efficiency gains. These gains may arise from either decreased transaction costs through better established delivery channels and/or the revelation of private information that can be used to improve accuracy of the premium rates. Using both parametric and semiparametric estimators which are explained in detail in the second chapter, out-of-sample tests are conducted and it is found that the insurance companies do reveal private information. The results may prove useful for countries considering the use of intermediaries in their crop insurance programs and in analyzing other insurance programs where there is government involvement, e.g., insuring natural disaster risks.

In the fourth chapter of this dissertation, using five-minute data, market volatility in the Dow Jones Industrial Average is examined in the presence of trading collar. The trading collar, formally known as Rule 80A, restricts certain forms of trade in component stocks of the S&P500 stock price index when there is “excess” volatility in the market. Besides taking into account the trading collar, the model also captures intraday seasonality of market volatility via a polynomial specification. Use of this specification, which seems to be the first in the literature, is easier to estimate than the two-step Fourier transform procedure which is used in the literature. Results of this chapter indicate that market volatility is 3.4% higher in declining markets when the trading collar is in effect. Results also support a U-shaped intraday periodicity in market volatility as found in other studies.

Findings of this dissertation are concluded in the last chapter and some extensions and future research ideas are also provided. These extensions and ideas are expected to become parts of a research agenda on econometric methodology and their applications in finance, insurance, and risk management.

2. Semiparametric Estimation of the Link Function in Binary-Choice Single-Index Models

2.1. Introduction

Discrete choice models are commonly used in microeconometrics to analyze situations where a decision or choice has to be made. Another common use of these models is when dealing with selection bias; estimating a discrete choice model (to estimate the "selection" equation) is the first step in the popular 2-step Heckman procedure. These models usually take the following general form

$$y_i^* = v_i\beta + u_i \tag{2.1}$$

where y_i^* is a latent variable which is operationalized by defining $y_i = 1$ if $y_i^* \geqslant 0$ and $y_i = 0$ otherwise, β is a $(q+1)$x1 vector of unknowns, $v_i \equiv (1, x_i)$ where x_i is a 1xq vector of explanatory variables, and u_i is the error term. The literature is dominated by parametric estimation where the distribution function (cdf) of u_i, say $F(u)$, is assumed to be normal (probit model) or logistic (logit model). Because parametric assumptions that are not consistent with the data could invalidate the results,[1] some have considered nonparametric and semiparametric methods. In discrete choice models the estimated effects of the regressors are of interest as well as the estimated conditional mean. Thus instead of a fully nonparametric approach, semiparametric methods have been the focus to circumvent any distributional assumptions yet recover the desired estimates.[2]

There has been significant research on semiparametric estimation of single-index models that contain parametric discrete choice models as a special case (see Stoker

[1] For an exception see Ruud [57].

[2] Furthermore fully nonparametric methods suffer from the so-called "curse of dimensionality", i.e., as the number of regressors increases, estimation precision decreases rapidly. The single-index models, which are explained below, reduce this dimensionality problem to a scalar.

[63], Powell et al. [54] Ichimura [35], Klein and Spady [41], Horowitz and Härdle [34], Cosslett [10], Horowitz [32, chapters 2 and 3], Pagan and Ullah [52, chapter 7]). A single-index model has the following form

$$E(y|v) = F(v\beta) \tag{2.2}$$

where F is an unknown (not necessarily a distribution) function, called the *link* function. The term $v\beta$ is the *index*.[3] Note that if F is the normal or logistic distribution function, the function in (2.2) is the binary probit or logit model and if it is the identity function, equation (2.2) becomes the usual linear regression model. Among the advantages of single-index models is dimension reduction. The index $v\beta$ is a scalar and thus single-index models do not suffer from the curse of dimensionality; if β were known it would be possible to estimate F as the nonparametric mean regression of y_i on $z_i = v_i\beta$ which is a scalar. Therefore in single-index models it is possible to estimate F at the nonparametric rate as if there is a single regressor[4] and the coefficient vector β at the parametric rate $O(n^{-1/2})$ (see Horowitz [32, chapter 2] for further advantages of single-index models).

Along a completely different vein, a recent paper by Hjort and Glad [27] proposes a semiparametric method for density estimation which starts with a parametric estimator and multiplies this parametric start with a correction factor (unknown density divided by the parametric start) which is estimated nonparametrically. The idea is based on bias reduction. If the parametric start captures a sufficient amount of the curvature of the unknown density, the correction factor will be close to one and less rough. Thus the bias associated with the nonparametric estimation of this correction factor will be less than that associated with the underlying density. Neither this paper or a companion piece by Glad [17] consider single-index models.

[3]This study will be concerned with a linear index as in (2.2) instead of a general form $h(v;\beta)$ where h is a scalar valued function. Ichimura [35] has a general analysis of single-index models and Ichimura and Lee [36] extend that general framework to multiple-index models.

[4]The fact that β is unknown and has to replaced with an estimator does not change this result as long as the estimator of β is $\sqrt{n}$-consistent (see Horowitz [32, pp.21-22]).

Unlike the semiparametric papers in the literature that estimate the link function nonparametrically, this study proposes to estimate the link function semiparametrically by employing the Hjort and Glad [27] bias reduction idea. Potentially relevant information is introduced in the form of a parametric function which is the parametric guide for the link function. The distinguishing characteristic of the proposed estimator is how the unknown link function F is estimated using prior parametric information about its shape.

There are a number of papers in the literature which compare some of the semiparametric estimators for single-index models but these studies are applications and compare different methods with the particular data set in hand. As these semiparametric methods are adopted, finite sample comparison of these methods by means of simulations can provide valuable information. A second contribution of this study is that an extensive simulation analysis is conducted in which the finite sample performance of the more popular of these semiparametric estimators, namely the estimators of Klein and Spady [41] and Ichimura [35] are compared with the proposed estimator. Note that finite sample simulation is especially important because bias reduction is not always realized in samples of reasonable size (see Jones and Signorini [37]).

The choice of smoothing parameter is also addressed for the new estimator from a practical point of view following the idea of Härdle et al. [25]. In single-index models, the asymptotic distribution of the centered and normalized estimator does not depend on the smoothing parameter so, asymptotically, any sequence of smoothing parameters is going to give the same estimator as long as it satisfies certain conditions. But in finite samples the performance of the estimators can be very sensitive to the choice of this smoothing parameter. For the proposed estimator, besides using a fixed smoothing parameter, the objective function is optimized with respect to the smoothing parameter as well as the unknown coefficients following Härdle et al. [25].

The chapter is organized as follows: In the next section, some of the semiparametric and nonparametric estimators for binary data are reviewed. The emphasis in this section is on the semiparametric estimators based on the index restriction. Then the new estimator is presented in section 2.3. In section 2.4 the applied literature comparing existing parametric and semiparametric methods for the estimation of discrete choice models is briefly reviewed. The mixed results of these studies motivate the need for a general simulation comparison of these estimators. Section 2.5 contains the simulation results where the finite sample performance of the single-index model estimators are compared. Finally, concluding thoughts are discussed in section 2.6.

2.2. Semi and Nonparametric Estimators of Binary Data

In this section the semi and nonparametric models for binary data are reviewed. This review is by no means exhaustive, instead, reflects their importance and relevance for the proposed estimator in this study. For a more comprehensive review, see, for instance, Horowitz [30], Pagan and Ullah [52], and Powell [53].

2.2.1. Semiparametric Estimators

Common features of these models are that it is assumed that the distribution of u depends on x only through the index[5] (index restriction) and that F is completely unknown (no centering assumptions will be made thus the intercept term can not be identified).[6]

[5]Single-index models, unlike models which assume independence of u and x, allow for limited forms of heteroscedasticity (general but known form and unknown form if it depends only on the index). This limitation can be serious since, for instance, the assumption that $Pr(y = 1|v)$ depends only on the index does not allow a certain form of heteroscedasticity, called random coefficients model, which may be important in applications.

[6]The maximum score estimator of Manski [45] and its smoothed version by Horowitz [29] make zero conditional median assumption (median$(u|x) = 0$) which identifies the intercept term (zero conditional mean assumption is not sufficient for identification in a binary response model, see Manski [46, p.731] and Horowitz [32, section 3.2]). These models allow for different forms of heteroscedasticity including random coefficients models although at the cost of a rate of convergence slower than $\sqrt{n}$. In fact under this conditional median independence assumption, $\sqrt{n}$ consistency is not possible, see Pagan and Ullah [52, p.278] and Horowitz [30].

Quasi-Maximum Likelihood Estimator The semiparametric single-index model of Klein and Spady [41] can be considered as quasi log-likelihood estimation. Note that for binary data $Pr(y = 1|v) = F(v\beta)$ and if F (the link function) were known the maximum likelihood estimator would maximize the log-likelihood

$$\sum_{i=1}^{n} (y_i \log[F(v_i\beta)] + (1 - y_i) \log[1 - F(v_i\beta)]). \tag{2.3}$$

The idea is to consider the link function in (2.3) unknown and replace it with a nonparametric estimator. Since β in single-index models is not fully identified, a location and scale normalization is required (see Horowitz [32, section 2.4]). Location normalization is achieved by requiring v to contain no intercept term and scale normalization is achieved by setting the β coefficient of a (continuous) regressor equal to one.[7] The β vector after scale and location normalizations is denoted by b, i.e., $b \equiv (1, \beta_2, \ldots, \beta_q)^T$ assuming the first regressor has a continuous distribution. The following Nadaraya-Watson nonparametric estimator for the link function is used

$$\hat{F}(x_i b) = \sum_{j \neq i} y_j K\left(\frac{x_i b - x_j b}{h}\right) \Big/ \sum_{j \neq i} K\left(\frac{x_i b - x_j b}{h}\right) \tag{2.4}$$

where K is the kernel function (usually a symmetric density function) and $h = h(n)$ is the smoothing parameter such that $h \to 0$ as $n \to \infty$.[8] By replacing the unknown link function F in (2.3) by (2.4), the quasi log-likelihood function is obtained and by maximizing the quasi log-likelihood with respect to $\tilde{b}$ where $\tilde{b}$ is b without its first component, the semiparametric estimator $\hat{\tilde{b}}$ is obtained. Klein and Spady [41] show that $\hat{\tilde{b}}$ satisfies $\sqrt{n}(\hat{\tilde{b}} - \tilde{b}_0) \xrightarrow{d} N(0, \Omega_{QL})$ where Ω_{QL} can be consistently estimated by the Hessian and the outer product of the gradient matrices and it attains the

[7] An alternative scale normalization would be $||\beta|| = 1$ where $||\cdot||$ is the Euclidean norm.

[8] Since there is no location restriction on u in single-index models and thus the intercept term is not identified, (2.4) is actually a nonparametric estimator of the distribution of $u + \beta_0$ where β_0 is the intercept. Also, Klein and Spady [41] have additive terms in the numerator and denominator of (2.4) to control the rate at which numerator and denominator tend to zero. Ichimura [35] utilizes indicator variables to trim those observations which correspond to small density values. A similar trimming function and an indicator variable enter multiplicatively to objective functions in (2.3) and (2.5) respectively. In this presentation those terms are ignored for simplicity.

semiparametric efficiency bound of Cosslett [10] if the errors are independent of the regressors.[9]

Note that in (2.4) the denominator can get arbitrarily close to zero so care must be taken. Klein and Spady [41] use complicated trimming procedures without restricting x to be in a specific set as in Ichimura [35] (see below).

(Weighted) Semiparametric Least Squares Estimator The single-index model of Ichimura [35] is, in contrast to Klein and Spady [41], based on minimizing a nonlinear least squares (NLS) loss function. The NLS estimator of b minimizes

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{F}(x_i b)]^2 \tag{2.5}$$

where $\hat{F}$ is the nonparametric estimator for the unknown link function in (2.4). Ichimura [35] denotes this model semiparametric least squares (SLS) and shows that $\hat{\tilde{b}}$ is consistent and $\sqrt{n}(\hat{\tilde{b}} - \tilde{b}_0) \xrightarrow{d} N(0, \Omega_{SLS})$ and gives a consistent estimator of $\Omega_{SLS} = \Gamma^{-1}\Sigma\Gamma^{-1}$. Γ and Σ can be consistently estimated by

$$\hat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i\tilde{x}_i^T \hat{F}'(x_i\hat{b})^2,$$
$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i\tilde{x}_i^T \hat{F}'(x_i\hat{b})^2[y_i - \hat{F}(x_i\hat{b})]^2$$

where $\tilde{x}_i \equiv (x_{2i}, \ldots, x_{qi})$, $\hat{b} \equiv (1, \hat{\tilde{b}}^T)^T$, and $\hat{F}'$ is the derivative of $\hat{F}$.

Ichimura [35] also considers weighted SLS (WSLS) in which he weights the objective function (2.5) and the summands in $\hat{F}$ by a weight function $W(x_i)$. As in parametric NLS, efficiency considerations play a role: the choice of the weight function does not affect the consistency and rate of convergence of the estimator of b

[9] A recent paper by Chen [9] builds on Klein and Spady [41] and shows that the intercept can be consistently estimated and there are possible efficiency gains in the estimation of slope coefficients although at the cost of stronger assumptions: a location restriction in the form of conditional symmetry, i.e., the density of u conditional on the regressors is symmetric around zero and an index restriction stronger than the one in Klein and Spady [41], namely, the conditional density depends on x only through the squared index.

but does affect its efficiency. Optimally weighted (the weight function is a consistent estimator of $Var(y|x)^{-1}$) WSLS achieves the semiparametric efficiency bound.[10] Horowitz [32, p.31] explains how a consistent estimator of $Var(y|x)$ can be obtained. Note that the first order conditions from (2.3) are the same conditions that one can obtain from (2.5) with the *estimated*[11] weight function $W(x) = \{\hat{F}(x\hat{b})[1-\hat{F}(x\hat{b})]\}^{-1}$.

As in the Klein and Spady [41] estimator, care must be taken to prevent the denominator of (2.4) from getting arbitrarily close to zero. Ichimura [35] restricts the summands in (2.4) and (2.5) to those observations for which the density of the index is not too small (see Ichimura [35] for details).

2.2.2. Nonparametric Estimators

A completely nonparametric analysis of binary-choice models, one that not only assumes that the distribution of the error term is unknown but also the conditional mean is just some unknown function of the covariates, i.e., $E(y|x) = m(x)$ where m is an unknown smooth function, is not very appealing to economists since the β coefficients (which can not be recovered in a completely nonparametric analysis) may have behavioral significance and may contain important information. Furthermore, as mentioned above, a completely nonparametric analysis would suffer from the curse of dimensionality and impractical sample sizes may be needed to obtain reliable conditional mean estimates. Hence, only one nonparametric estimator is going to be reviewed (which is applicable to data generating processes more general than just binary-choice models) which is also based on using prior parametric information in attempts to reduce bias similar to the estimator proposed in this chapter. Note that this estimator has not been used in an application.

[10] For this efficiency result, the weight function should depend on x only through the index as in binary-choice models where $Var(y|x) = Var(y|xb)$.

[11] Ichimura [35] treats $W(\cdot)$ as a known function.

Local Nonlinear Least Squares Gozalo and Linton [20] propose a nonparametric procedure which is analogous to local likelihood estimation and can be centered at any parametric regression function. Their main objective is to estimate the unknown regression function at an interior support point. Their main idea, as is the proposed estimator's in this study, is bias reduction. Their setup is to minimize a nonlinear least squares loss function where they introduce prior parametric information. The performance of the estimator depends on this parametric model: the bias of the estimator is proportional to the distance between the second derivative of this parametric model and the second derivative of the true regression function (whereas the variance is the same as the variance of Nadaraya-Watson and local linear estimators). Also, like the local linear estimator, their estimator is design adaptive, i.e., its bias does not depend on the design density $f(x)$ (the bias of Nadaraya-Watson estimator does). For each point v, they find $\hat{\beta}(v)$ and hence $\hat{F}(v) = F(v\hat{\beta}(v))$ by minimizing

$$\sum_{i=1}^{n}\{y_i - F(v_i\beta)\}^2 K_H(v_i - v) \tag{2.6}$$

with respect to β. In (2.6), $K_H(\cdot) = det(H)^{-1}K(H^{-1}\cdot)$ where H is a nonsingular smoothing parameter matrix and $K(\cdot)$ is a kernel function. Their prior parametric information is in F, that is, F is a parametric function. Note that in the single-index models considered above F is unknown and β fixed. Here F is a known parametric function and β varies with v. As Gozalo and Linton [20, p.81] explain, when they fit a local probit, "...this [model] can be interpreted as a random coefficient probit, except that the variation in parameters is driven by the conditioning information rather than by some arbitrary distribution unrelated to the covariates." Loosely speaking, single-index models and parametric probit and random coefficient probit models are special cases of their model. If the estimated β's are constant over x, the result would be the probit model whereas if the single-index or the random coefficient probit model is the underlying truth, the ratios of β's corresponding to slope coefficients should be constant over x. Thus one can use estimated β coefficients from this model along with calculated confidence intervals to test against these alternatives.

Note that implementation of this estimator requires n optimizations where n is the sample size and thus computationally can be very expensive to calculate. Also, this estimator—being a generalization of the linearity part in local linear estimator to any parametric regression function—has a slower rate of convergence $O((nh^q)^{1/2})$, where q is the number of regressors, than single-index models and suffers from curse of dimensionality due to its complete nonparametric nature.

2.3. Parametrically-Guided Single-Index Model

The proposed semiparametric estimator is first motivated by the estimator of Hjort and Glad [27] and Glad [17]. Suppose one wishes to estimate the conditional mean function $E(y|x) = m(x)$. They first start with a parametric estimator $m(x, \hat{\beta})$ which could be, for instance, a simple linear regression or a more complex maximum likelihood estimation, and then multiply it with a correction factor $r(x) = m(x)/m(x, \hat{\beta})$ which is estimated nonparametrically. The idea is based on bias reduction: if the parametric start is close to the truth, the correction factor will be close to a constant and thus smoother and (bias wise) easier to estimate than m itself. Hence the bias associated with nonparametric estimation of this correction factor would be less than the bias from direct nonparametric estimation of the unknown regression function. Their estimator is

$$\hat{m}(x) = m(x, \hat{\beta})\hat{r}(x).$$

When the correction factor is estimated by the Nadaraya-Watson estimator[12] their parametrically guided estimator is

$$\hat{m}(x) = \sum_{i=1}^{n} y_i \frac{m(x, \hat{\beta})}{m(x_i, \hat{\beta})} K_h(x_i - x) \Big/ \sum_{i=1}^{n} K_h(x_i - x).$$

[12] Glad [17] generalizes this to local pth order polynomial estimator which reduces to the Nadaraya-Watson for $p = 0$.

Glad [17] shows that this estimator has the same large sample variance as the standard nonparametric estimators (Nadaraya-Watson and local linear) while bias reduction is possible if the parametric start belongs to a neighborhood around the true regression curve.

The proposed semiparametric estimator for the single-index model is based on the above idea, that is introducing (potentially) relevant information in the form of a parametric function in attempts to reduce bias. The estimator starts with a parametric model for the link function, say $G(xb)$, where $G(\cdot)$ is a known function (for instance normal cdf) and multiplies it with the correction function $r(xb) = F(xb)/G(xb)$ which is estimated nonparametrically. Note that the information that this estimator starts with is only related to the shape of the link function and not to the coefficient estimates. In this sense the estimator is using a fixed start vis-à-vis Hjort and Glad [27] and Glad [17]. When the Nadaraya-Watson estimator is used to estimate the correction factor the proposed model is

$$\hat{F}(x_i b) = \sum_{j \neq i} \left\{ y_j \frac{G(x_i b)}{G(x_j b)} \right\} K\left(\frac{x_i b - x_j b}{h}\right) \Big/ \sum_{j \neq i} K\left(\frac{x_i b - x_j b}{h}\right). \tag{2.7}$$

This model will be referred to as parametrically-guided single-index model (PGSIM). The unknown F in (2.3) is replaced with (2.7) and the resulting quasi log-likelihood function is maximized with respect to $\tilde{b}$.

Note that the semiparametric estimator of the link function in (2.7) does not nest the nonparametric estimator in (2.4) so the Klein and Spady [41] model is not nested in the model proposed model.[13] However, a bias and variance comparison of (2.4) and

[13]For this to happen, the parametric start $G(\cdot)$ should be a constant function. In density estimation, the Hjort and Glad [27] estimator nests the usual kernel density estimator if the parametric start is the uniform density over the space. But here a distribution function which is constant and which satisfies the continuity (of the index) assumption can not be found. One obvious example of such a constant distribution function, which does not satisfy the continuity assumption, is the unit point mass at a when $z = a$ a.s.:

$$G(z) = \begin{cases} 0 & \text{if } z < a \\ 1 & \text{if } z \geqslant a. \end{cases}$$

Of course here not only is the continuity assumption not satisfied but also β is not identified with this $G(\cdot)$.

(2.7) is useful to see the bias reduction. Equation (2.4) is the usual Nadaraya-Watson estimator of the link function whereas (2.7) corresponds to fixed start of Hjort and Glad [27, section 2] and Glad [17, section 2] as mentioned above. While both have the same variance

$$Var(\hat{F}(z)) = (nh)^{-1}f(z)^{-1}\sigma^2(z)R(K) + o_p((nh)^{-1})$$

where $f(z)$ is the density of z, $\sigma^2(z) = Var(y|z)$, and $R(K) = \int K^2(t)dt$, the bias of (2.4) is

$$Bias(\hat{F}_{KS}(z)) = \frac{h^2\mu_2(K)}{2f(z)}(F''(z)f(z) + 2f'(z)F'(z)) + o_p(h^2)$$

where $\mu_2(K) = \int t^2K(t)dt$, whereas the bias of (2.7) is

$$Bias(\hat{F}_{PGSIM}(z)) = \frac{h^2\mu_2(K)}{2f(z)}(r''(z)G(z)f(z) + 2f'(z)r'(z)G(z)) + o_p(h^2).$$

Thus, for the same h and K, bias reduction is possible if the parametric start G can be chosen such that

$$|r''(z)G(z)f(z) + 2f'(z)r'(z)G(z)| < |F''(z)f(z) + 2f'(z)F'(z)|. \tag{2.8}$$

If the parametric start is proportional to F, the correction factor r is going to be a constant and thus $r' = r'' = 0$. If it is sufficiently close to F, roughness of r will be less than roughness of F and r will have a smaller second derivative. Thus (2.8) defines a "neighborhood" of F where bias reduction is possible by choosing a parametric start from this neighborhood.[14]

The above argument indicates that even though the estimator can not (asymptotically) outperform Klein and Spady [41] with respect to the coefficients as they attain the semiparametric efficieny bound, in finite samples there is potential to increase

[14] Obviously (2.8) is a neighborhood for the link function and not for the coefficient estimates. For the latter, it is not straightforward to obtain a neighborhood like (2.8) as there is no closed form solution but it is reasonable to expect bias reduction for coefficient estimates as well since they would be obtained from first order conditions which are functions of less biased estimates of the link when (2.8) is satisfied ($\hat{b}(\hat{F}_{KS})$ vs. $\hat{b}(\hat{F}_{PGSIM})$).

efficieny by using a parametric guide. On the other hand, the bias of $\hat{F}_{PGSIM}$ reduces to a smaller order than Klein and Spady [41] and Ichimura [35] probability estimates when the parametric start is proportional to F.

2.3.1. The Estimator and Consistency

Following the intuition behind the estimator that is presented above, here, a formal description of the estimator and its consistency result is provided.

As mentioned above, in the semiparametric estimators of Ichimura [35] and Klein and Spady [41] care must be taken as the nonparametric density estimator in the denominator can get arbitrarily small. Also, in Klein and Spady [41], any estimator of F should be kept in the $(0,1)$ open interval. They use complicated trimming procedures which they denote likelihood trimming (to downweight observations for which the corresponding densities are small) and probability trimming (to control the rate at which numerator and denominator of $\hat{F}$ tend to zero). Ichimura [35], on the other hand, restricts x to a set by indicator variables on which the above mentioned problems are avoided. This latter approach is easier to deal with in asymptotics.[15] Hence this approach is employed. Here, these restrictions should include one more thing, namely, that the parametric start $G(\cdot)$ should be nonzero throughout the support of the index. In what follows, the notation for the trimming terms will follow Ichimura [35, p.78] and Horowitz [32, pp.23-24] closely.

The proposed semiparametric estimator for binary-choice models maximizes

$$\hat{Q}_n(b) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[x_i \in A_x]}(y_i \log[\hat{F}(x_i b)] + (1-y_i)\log[1-\hat{F}(x_i b)]) \qquad (2.9)$$

where $\mathbf{1}_{[\cdot]}$ is an indicator variable, $A_x \subset \mathbb{R}^q$ is such that $A_x = \{x : p(xb) \geq \eta \quad \forall\, b \in$

[15] In actual estimations, trimming has very little effect on the performance of the estimators. Klein and Spady article reports simulation results from untrimmed estimator: "...the estimate obtained without any trimming performed quite similar to that under the trimming that we employed. Accordingly, we report results for the semiparametric estimator obtained without probability or likelihood trimming (Klein and Spady [41, p.406])."

$B\}$, $p(\cdot)$ is the density of the index, η is a positive constant, and

$$\hat{F}(x_i b) = \frac{\frac{1}{(n-1)h}\sum_{j\neq i} y_j \mathbf{1}_{[x_j\in A_{nx}]}\left\{\frac{G(x_i b)}{G(x_j b)}\right\} K\left(\frac{x_i b - x_j b}{h}\right)}{\frac{1}{(n-1)h}\sum_{j\neq i} \mathbf{1}_{[x_j\in A_{nx}]} K\left(\frac{x_i b - x_j b}{h}\right)}$$

$$\text{if} \quad \sum_{j\neq i} \mathbf{1}_{[x_j\in A_{nx}]} K\left(\frac{x_i b - x_j b}{h}\right) \neq 0, \qquad \text{and} \tag{2.10}$$

$$= \begin{cases} 0.9 & \text{if} \quad y_i = 0 \\ 0.1 & \text{otherwise} \end{cases}$$

where $K : \mathbb{R} \to \mathbb{R}$ is a density function, $h > 0$ and $h \to 0$ as $n \to \infty$, and A_{nx} is a set such that, as Ichimura [35, p.79] explains "...includes $[A_x]$ in such a way that all boundary points in $[A_x]$ are interior to $[A_{nx}]$, in a neighborhood of x, with probability approaching 1, there are data in all directions to take a local average". Clearly the purpose is to reduce bias that may otherwise result close to the boundary points. His suggestion is to use $A_{nx} = \{x : ||x - x'|| \leqslant 2h \text{ for some } x' \in A_x\}$. Note that as $n \to \infty$, A_{nx} seems to get smaller but as the sample size increases there will be more and more sample points close to boundary as well and sufficient to take a local average.

The identification of single-index models, in the most general context, has been analyzed by Ichimura [35]. Manski [45] looks at identification of binary-choice models with linear index under different assumptions including index restriction. Klein and Spady [41] give conditions for identification of single-index models in binary response models where the index is a general but known function. The reader is referred to the original papers for details. Here the following assumptions are made:

Assumption I1 The model in (2.1) satisfies the index restriction.

Assumption I2 F is continuously differentiable and not a constant function of the index over its support.

Assumption I3 At least one regressor, with nonzero coefficient, has a continuous distribution. Its distribution conditional on the remaining regressors is absolutely continuous.

Assumption I4 Varying the values of discrete regressors must not divide the support of the index into disjoint subsets.

Assumption I5 $Pr(y = 1|xb_0) = Pr(y = 1|xb_*) \Rightarrow b_0 = b_*$.

For I4, see Horowitz [32, pp.16-17] for an example how its violation turns the slope coefficient on the discrete regressor into an intercept term which is not identified. Assumption I5 is a necessary restriction for identification of binary-choice maximum-likelihood models in particular. Klein and Spady [41, pp.395-397] provide sufficient conditions under this assumption.[16]

The consistency proof requires showing uniform convergence, over x and b, of $\hat{F}$ to F. Once this is accomplished, the estimator that maximizes the quasi-likelihood in (2.9) asymptotically behaves like the estimator that maximizes the likelihood function for a known F since $\sup_{b \in B} |\hat{Q}_n(b) - Q_n(b)| = o_p(1)$ where

$$Q_n(b) = n^{-1} \sum_{i=1}^{n} \mathbf{1}_{[x_i \in A_x]}(y_i \log[F(x_i b)] + (1 - y_i) \log[1 - F(x_i b)]).$$

The estimator which maximizes this likelihood can be analyzed by standard results for parametric estimators. First the following assumptions are made:

Assumption 1 Observed sample (x_i, y_i), $i = 1, \ldots, n$ is *i.i.d.*

Assumption 2 $B \subset \mathbb{R}^q$ is compact and the true parameter vector b_0 is in the interior of B.

Assumption 3 A_x is compact.

Assumption 4 $K(s)$ is a density. Furthermore $\int sK(s)ds = 0$, $K(s) = 0$ for $s < -1$ and $s > 1$, and its second derivative satisfies a Lipschitz condition.

[16] There are two cases to consider: When the link function F is monotonic in the index and when it is not. If the underlying distribution is heteroscedastic, for instance, F need not be monotonic in the index.

Assumption 5 Parametric start G is uniformly bounded over x, b and $G(xb) \neq 0 \; \forall x, b \in A_x \times B$.

Assumption 6 $\int |\phi(t)|dt < \infty$ where $\phi(t)$ is the characteristic function of K.

Assumption 7 There exist $\underline{F}$ and $\overline{F}$ that do not depend on x such that $0 < \underline{F} \leqslant F(xb) \leqslant \overline{F} \; < \; 1 \quad \forall b \in B$.

Lemma 1 shows the uniform convergence of $\hat{F}$ and is proved in the appendix.

Lemma 1. *Under assumptions 1-7, if $h \to 0$ and $h\sqrt{n} \to \infty$ as $n \to \infty$, then for any $\epsilon > 0$*

$$Pr\left(\sup_{(x,b)\in A_x \times B} |\hat{F}(xb) - F(xb)| > \epsilon\right) \to 0 \quad as \quad n \to \infty.$$

Proof. See the appendix. □

As Bierens [5, p.115] notes, the best uniform convergence rate is obtained when $\min(h\sqrt{n}, h^{-2})$ (see the appendix) is maximum so $h \propto n^{-1/6}$ and thus $\min(h\sqrt{n}, h^{-2}) \propto n^{2/6}$.[17] This is not the fastest uniform convergence rate achievable for nonparametric regression (see Schuster and Yakowitz [60]), however, this conservative approach has been chosen for its simplicity. Also, lemma 1 does not provide uniform convergence of derivatives of $\hat{F}$ which are required in asymptotic normality arguments.

With lemma 1, it can be shown that $\hat{Q}_n(b)$ in (2.9) converges in probability, uniformly in b, to a likelihood function for a known F. This limiting likelihood, in turn, converges in probability, uniformly in b, to its expectation which is maximized by b_*. This b_* satisfies (see Klein and Spady [41, p.400]) $Pr(y = 1|x) = Pr(y = 1|xb_0) = Pr(y = 1|xb_*)$ where the first equality is the index restriction and the second equality is a necessary condition for b_* to be maximum. Hence, from I5, $b_* = b_0$ is unique maximum. Theorem 1 below gives the result and is proved in the appendix.

[17] Klein and Spady [41] obtain a similar uniform convergence rate.

Theorem 1. *Under assumptions 1-7,*

$$\hat{b} \equiv \arg\sup_{b} \hat{Q}_n(b) \xrightarrow{p} b_0.$$

Proof. See the appendix. □

2.4. Review of Parametric vs. Semiparametric Comparison

There are a number of empirical papers that compare different parametric and semiparametric estimators for binary data. These studies show that semiparametric methods which reduce the risk of misspecification by avoiding strong distributional assumptions may give quite different results than parametric methods. But depending on the application and data set used, these studies have mixed results and thus can not be generalized. In this section these results are reviewed.

Horowitz [31] uses parametric probit and random-coefficients probit with the single-index model of Klein and Spady [41] and the smoothed maximum score estimator of Horowitz [29]. He conducts specification tests and rejects the fixed coefficient probit and the semiparametric single-index models. Gerfin [16] compares the parametric probit model with the semiparametric models of Gabler et al. [15], Klein and Spady [41], and Horowitz [29]. He finds that the coefficient estimates do not differ substantially across models. His specification tests and within sample prediction exercises are not conclusive. Fernandez and Rodriguez-Poo [13] estimate the parametric probit and logit models with the semiparametric single-index model of Klein and Spady [41]. They observe important differences in coefficient estimates from the parametric models and the semiparametric model. Newey et al [51] are mainly interested in estimating a selection model. They estimate the selection equation by the semiparametric estimators of Ichimura [35] and Klein and Spady [41] and find the coefficient estimates to be quite close to the probit estimates. Goodwin and Holt [19] are interested in labor supply and find that the probit and the Ichimura [35] models give quite similar results though marginal effects are generally more elastic in the single-index model. These mixed results suggest the need for a comprehensive simulation study whose results can be generalized.

2.5. Simulations

In this section, finite sample performance of the estimators which are based on index restriction are compared. The procedures that are employed to pick the smoothing parameters are also discussed in detail.

2.5.1. Design

The model generating the data is

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

and y_i takes a value of 1 if the latent $y_i^* \geqslant 0$ and a value of 0 otherwise. The values of the true parameters are $\beta_0 = 0$, $\beta_1 = 1$, and $\beta_2 = 1$. The regressors x_1 and x_2 are independently and identically distributed. The data generating process (DGP) for x_1 is chi-square distribution with 3 degrees of freedom truncated at 6. The DGP for x_2 is standard normal truncated at ± 2. Both x_1 and x_2 are first trimmed by 2%, i.e., lower and upper tails of their empirical distributions are trimmed by one percent and then standardized to have zero mean and unit variance. For the DGP of the error term, four homoscedastic normal mixtures and a heteroscedastic distribution are considered. Normal mixtures are (1) standard normal, (2) 0.75·N(0,1) + 0.25·N(0,25), (3) 0.75·N(-0.5,1) + 0.25·N(1.5,25), and (4) 0.5·N(3,1) + 0.5·N(-3,1). The second distribution is leptokurtic, the third distribution is skewed and leptokurtic, and the fourth distribution is bimodal. The second (third) (fourth) distribution has standard error 2.65 (2.78) (3.16), skewness 0 (1.29) (0), and kurtosis 6.61 (6.29) (1.38). The heteroscedastic distribution is normal with zero mean and variance $0.25(1+(v_i\beta)^2)^2$ where $v_i\beta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$. [18]

Figure 2.1 gives a graph of the link function for different errors in the simulation design. Note that when the error distribution is heteroscedastic, the link function is not monotonic in the index.

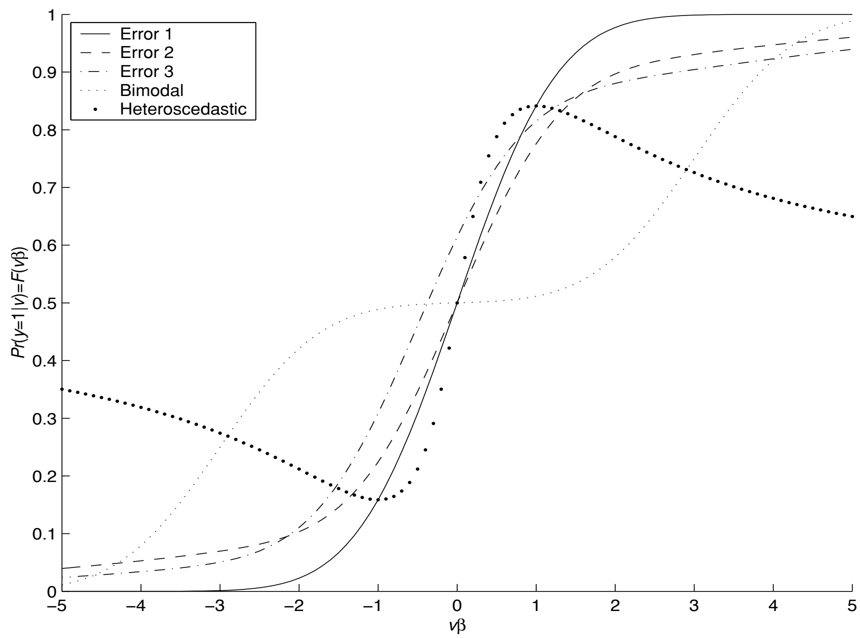

FIGURE 2.1. Link Function for Design Errors

[18]Even though it is not in the simulation design, at this point, it is instructive to digress to discuss maximum likelihood estimation of misspecified binary choice models. Ruud [57] showed that when the explanatory variables are multivariate normal, maximum likelihood estimates of slope coefficients can still be estimated consistently up to scale even when the distributional assumption is not correct. More generally the result holds when

$$E(\tilde{x}|xb = t) = c_0 + c_1 t \tag{2.11}$$

where c_0 and c_1 are constants, which is satisfied when the explanatory variables are multivariate normal. Note that this consistency result does not hold for the probability estimates.

Another interesting implication of (2.11) is that when it holds, the semiparametric efficiency bound is the same as the parametric (Cramér-Rao) efficiency bound (see Cosslett [10]).

2.5.2. Smoothing Parameter

In general, it is usually standard to use a version of cross-validation (CV) or plug-in methods (see, for instance, Wand and Jones [65, chapter 3]). In semiparametric single-index models, however, selection of the smoothing parameter has not been well studied.[19] One exception is the paper by Härdle et al. [25] where they show that equation (2.5) with $W(x_i) = 1$, i.e., SLS, can be expanded as $A(b) + B(h)$ and can be minimized simultaneously with respect to both b and h. This is like separately minimizing $A(b)$ with respect to b and $B(h)$ with respect to h. The end result is a $\sqrt{n}$-consistent estimator of b and an asymptotically optimal estimator of h in the sense that $\hat{h}/h_0 \to 1$ as $n \to \infty$ where h_0 is the optimal bandwidth for estimating F when b is known and is proportional to $n^{-1/5}$ as usual in nonparametrics (see Härdle et al. [25] for technical details).

For the Klein and Spady [41] estimator, the smoothing parameter h has to satisfy $nh^8 \to 0$ and $nh^6 \to \infty$ and for the Ichimura [35] estimator, the smoothing parameter h has to satisfy $nh^8 \to 0$ and $\log h/(nh^{3+3/(m-1)}) \to 0$ where $m \geqslant 3$ as $n \to \infty$. If the smoothing parameter is taken to be $h = c/n^{1/p}$ where c is a positive constant with $c = 1$ and $p = 7$, the resulting h would satisfy the requirements of both Klein and Spady [41] and Ichimura [35] estimators. Hence, in the simulations, this smoothing parameter is used for both of these estimators. For the new estimator, however, asymptotic results are not complete and hence a set of conditions for the smoothing parameter similar to Klein and Spady and Ichimura estimators do not exist. So, besides using this constant smoothing parameter, the Härdle et al. [25] idea is applied as well to quasi log-likelihood function of the proposed estimator and hence the objective function is optimized with respect to both b and h.[20] This is the first study

[19] The difficulty in single-index models stems from, as Horowitz [32, p.50] explains, "...in semiparametric single-index models, the asymptotic distribution of $[\sqrt{n}(\hat{\tilde{b}} - \tilde{b}_0)]$ does not depend on the bandwidth h. Therefore, bandwidth selection must be based on a higher-order approximation to the distribution of $[\sqrt{n}(\hat{\tilde{b}} - \tilde{b}_0)]$."

[20] The asymptotic optimality (in the sense defined above) of h obtained this way is established

which uses this idea in practice other than the original Härdle et al. [25] paper.[21] Note that in equation (2.7) observation i is excluded so in a way the objective function is "cross-validated".

2.5.3. Results

Table 2.1 has the simulation results for bias and root mean squared error (RMSE) of β_2 over 500 simulations for a sample size of 100 and 1000. In the table, KS is the quasi log-likelihood estimator of Klein and Spady [41] and PGSIM is the parametrically-guided single-index model estimator. PGSIM$_p$ results are obtained with a constant smoothing parameter h used for Klein and Spady and Ichimura estimators and PGSIM$_h$ results are obtained by the Härdle et al. idea. For identification purposes, in single-index models, β_0 is not estimated and β_1 is set equal to 1. To avoid local optima, in each simulation, different starting values (probit estimate, 0, 1, 1.5) are tried for the semiparametric estimators. A logistic cdf is the parametric guide in the PGSIM. For all of the three semiparametric estimators, a normal density function is used as the kernel. Note that in practice if, for some z_j, $G(z_j)$ is zero or near zero while $G(z_i)$ is not then the ratio $G(z_i)/G(z_j)$ blows up in PGSIM. Following Hjort and Glad [27] and Glad [17], trimming is conducted below 0.1 and above 10.

When the error distribution is standard normal (error 1), the parametric guide (logistic distribution function) is quite close to the true link function and PGSIM achieves significant bias reduction and efficiency gain with respect to semiparametric estimators of Klein and Spady [41] and Ichimura [35].

Under errors 2 and 3, the parametric guide for the link function is not correct but the estimator still achieves significant bias reduction and efficiency gain compared to the other two semiparametric estimators and the parametric probit. In fact PGSIM

only for the Ichimura [35] estimator. There is no study which finds a similar result to Härdle et al. [25] for the quasi log-likelihood functions.

[21] Of the applied papers that are reviewed in section 2.4, some use fixed h values and others use CV.

is the most efficient under these two error distributions when $n = 100$. PGSIM is still the most efficient when $n = 1000$ under error 3.

Under the bimodal distribution, PGSIM is the most efficient with both sample sizes and particularly so when $n = 100$; it achieves very significant bias reduction compared to other estimators.

When the error distribution is heteroscedastic, probit maximum likelihood estimators are inconsistent (Yatchew and Griliches [68]). Even though the estimator achieves significant bias reduction and efficiency gain against the parametric probit, the Ichimura estimator is more efficient when $n = 100$; the efficiency loss against Ichimura estimator is 16%. When $n = 1000$, however, PGSIM is the most efficient albeit by only a small difference in RMSEs.

Comparing Klein and Spady and Ichimura estimators, not surprisingly, the simulation results indicate that they perform very similarly especially as the sample size gets larger. If one has a significantly large sample, however, Klein and Spady estimator may be more attractive since it is asymptotically efficient (assuming homoscedasticity). Also, calculating standard errors for Klein and Spady is more straightforward, especially with a standard maximum likelihood routine, since all that is needed is the Hessian and the outer product of the gradient matrices.

		Error 1		Error 2		Error 3		Bimodal		Heteroscedastic	
	Estimator	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$n = 100$	Probit	0.0678	0.2384	-0.2931	0.3485	-0.3392	0.3890	-0.8753	0.8869	-0.4460	0.4769
	KS	0.0634	0.2408	0.0256	0.3860	0.0119	0.3954	-0.4687	0.9585	-0.0066	0.2972
	Ichimura	0.0634	0.2566	0.0709	0.3689	0.0458	0.3597	-0.4327	0.9452	0.0584	0.2382
	PGSIM_p	0.0460	0.2816	0.0266	0.3926	0.0248	0.4441	-0.3804	0.9246	0.0665	0.3835
	PGSIM_h	0.0477	0.2342	0.0571	0.3220	0.0352	0.3202	-0.1395	0.6920	0.0623	0.2843
$n = 1000$	Probit	0.0081	0.0697	-0.3535	0.3577	-0.3929	0.3966	-0.8763	0.8772	-0.5341	0.5363
	KS	0.0162	0.0698	0.0016	0.0923	0.0019	0.0963	-0.0043	0.4979	0.0067	0.0576
	Ichimura	0.0097	0.0697	0.0090	0.0904	0.0116	0.0965	0.0219	0.5000	0.0064	0.0573
	PGSIM_p	0.0008	0.0758	-0.0026	0.0968	-0.0017	0.0957	0.0305	0.5192	0.0031	0.0562
	PGSIM_h	0.0077	0.0732	0.0150	0.0957	0.0072	0.0927	0.0420	0.4506	-0.0042	0.0570

TABLE 2.1. Simulation Results - $\hat{\beta}_2$

		$\hat{\beta}_2$		$\hat{h}$		
		Bias	RMSE	mean	sd	median
$n = 100$	Error 1	0.0477	0.2342	0.6220	0.3302	0.5912
	Error 2	0.0571	0.3220	0.8174	0.4828	0.7177
	Error 3	0.0352	0.3202	0.7924	0.4388	0.7318
	Bimodal	-0.1395	0.6920	0.5061	0.2238	0.4952
	Heteroscedastic	0.0623	0.2843	0.6268	0.3342	0.5533
$n = 1000$	Error 1	0.0077	0.0732	0.4318	0.1225	0.4453
	Error 2	0.0150	0.0957	0.6541	0.2871	0.6058
	Error 3	0.0072	0.0927	0.5457	0.1724	0.5490
	Bimodal	0.0420	0.4506	0.3309	0.1029	0.3457
	Heteroscedastic	-0.0042	0.0570	0.3090	0.0614	0.3115

TABLE 2.2. Simulation Results (PGSIM) - $\hat{\beta}_2$ and $\hat{h}$

Table 2.2 has simulation results for $\hat{\beta}_2$ and $\hat{h}$ for the proposed estimator. Compared to the constant smoothing parameter that is used (0.5179), except for the bimodal distribution, the estimator significantly oversmooths and on average uses a bigger smoothing parameter when $n = 100$. For the bimodal distribution, on average, the estimator slightly undersmooths. For $n = 1000$, the estimator again significantly oversmooths on average compared to the constant smoothing parameter except for the bimodal and heteroscedastic distributions. As the sample size increases from 100 to 1000, the smoothing parameter converges to zero as expected and so does its standard deviation. In almost all cases, choosing the smoothing parameter in the optimization via Härdle et al. (PGSIM$_h$) rather than a constant h (PGSIM$_p$) gives better results. When this is true, usually, there is an increase in bias and a decrease in RMSE. Bias-variance tradeoff is a well known phenomenon in nonparametrics. Here, however, this probably is a finite sample issue as the asymptotic distribution of coefficient estimates in single-index models do not depend on the smoothing parameter h and thus there is no bias-variance tradeoff in the asymptotic sense. Of course the asymptotic distribution result is not provided here so this last point is more of a conjecture.

The performances of single-index models are also compared with respect to their ability to estimate probabilities ($\hat{F}$). The proposed estimator is expected to perform better here than coefficient estimates since the bias reduction idea only indirectly affects coefficient estimates whereas here it directly affects probability estimates. Table 2.3 has the results from this comparison. In the table, L_1 (L_2) is the difference of $\hat{F}$ from F measured by L_1-norm (L_2-norm).[22] Note that for single-index models $\hat{F}$ estimates converge slower (approximately proportional to $n^{1/3}$) than coefficient estimates. Hence, when comparing semiparametric estimators to parametric probit, to get a comparison of probabilities as accurate as coefficients with $n = 100$, one needs to look at results for $n = 1000$.

Probit performs very poorly in estimating probabilities, especially under bimodal and heteroscedastic errors. In fact under these error designs, even with $n = 100$, probit is the least efficient. Under error 3, probit performs competitively compared to KS and Ichimura but the new estimator with an estimated smoothing parameter performs better than probit. A small increase in sample size, however, probably to around 1500, would result in probit performing worse than all semiparametric estimators, not just PGSIM$_h$. As expected as in table 2.1, KS and Ichimura estimators perform similarly in estimating probabilities.

Even though the parametric start is not correct in any of the simulation designs (the closest it comes is under error 1), the new estimator performs better than KS and Ichimura estimators in all simulations except the heteroscedastic design when $n = 1000$. This is to be expected since as mentioned above, all that is required from the parametric start is to smooth the object to be estimated nonparametrically and not that the parametric start be correct all the time. This is the strength of the estimator and shows its usefulness in many situations.

[22]In general, L_p-norm is defined as $(E(|x|^p))^{1/p}$.

2.6. Conclusions

In this chapter, a new semiparametric estimator for binary-choice single-index models is proposed which uses parametric information in the form of a known (parametric) link function and nonparametrically corrects it. It is also shown that the estimator is consistent. An extensive simulation study is conducted and the new estimator is compared with the semiparametric estimators of Klein and Spady [41] and Ichimura [35]. The performance of the estimator is robust to the correctness of the parametric guide since all that is required from the guide is to "smooth" the function to be estimated nonparametrically so as to achieve bias reduction, not that it be correct all the time.

As for the comparison of Klein and Spady [41] and Ichimura [35] estimators, not surprisingly, they perform very similarly as the only difference is their loss functions. If one has a significantly large sample, however, Klein and Spady [41] may be a better choice since it is asymptotically efficient.[23] Also calculation of the covariance matrix is easier for Klein and Spady [41] thanks to the wide availability of maximum likelihood and Hessian and outer product of gradient matrices routines. This is the first study that compares finite sample performance of these two semiparametric estimators which are frequently used in applied work.

The method of Härdle et al. [25] is followed regarding the choice of the smoothing parameter for the proposed estimator besides using a fixed smoothing parameter and the objective function is optimized with respect to the bandwidth as well as the unknown coefficients. Other than the original Härdle et al. [25] paper, this study is the first to utilize this idea. Results show that it works well and gives quite reasonable results thus can be employed by the applied researchers especially if time is a constraint in which case a data-driven procedure, e.g., cross-validation, may not be feasible.

[23] As mentioned above, however, asymptotic efficiency result assumes independence of the errors and the regressors which rules out heteroscedasticity.

		Error 1		Error 2		Error 3		Bimodal		Heteroscedastic	
	Estimator	L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2
$n = 100$	Probit	0.0	0.0	0.0588	0.0616	0.0771	0.0858	0.2173	0.2577	0.2088	0.2367
	KS	0.0311	0.0507	0.0792	0.1064	0.1123	0.1499	0.1644	0.2269	0.1900	0.2490
	Ichimura	0.0314	0.0507	0.0774	0.1050	0.1106	0.1485	0.1615	0.2229	0.1853	0.2447
	PGSIM_p	0.0306	0.0457	0.0730	0.0930	0.1053	0.1337	0.1377	0.1884	0.1735	0.2234
	PGSIM_h	0.0371	0.0558	0.0709	0.0905	0.1011	0.1311	0.1416	0.1857	0.1788	0.2242
$n = 1000$	KS	0.0139	0.0225	0.0407	0.0571	0.0872	0.1153	0.0835	0.1241	0.0805	0.1188
	Ichimura	0.0137	0.0222	0.0405	0.0568	0.0870	0.1150	0.0832	0.1224	0.0806	0.1189
	PGSIM_p	0.0116	0.0166	0.0345	0.0451	0.0783	0.1035	0.0706	0.0927	0.0910	0.1324
	PGSIM_h	0.0136	0.0190	0.0307	0.0384	0.0741	0.1010	0.0724	0.0966	0.0964	0.1420

TABLE 2.3. Simulation Results - $\hat{F}$

3. On the Revelation of Private Information in the U.S. Crop Insurance Program

3.1. Introduction

In the U.S. crop insurance program, unlike the crop insurance programs in other countries, three rather than two economic interests are served: the federal government through the United States Department of Agriculture's Risk Management Agency (RMA); the farmers; and the insurance companies. Very little has appeared in the literature on insurance companies and their involvement in the crop insurance program, arguably one of the cornerstone programs of U.S. farm policy (for exceptions see Miranda and Glauber [48], Ker [38], Ker and McGowan [40], and Ker [39]). Figure 3.1 illustrates the breakdown of government program costs or outlays since 1981 into producer subsidies, indemnities less premium, administrative & operating (A&O) expense reimbursement to insurance companies, and underwriting gains accrued by insurance companies. There are a number of interesting features. Producer subsidies increased dramatically in 1995 as a result of the 1994 Federal Crop Insurance Act and again in 2001 as a result of the 2000 Agricultural Risk Protection Act. Indemnities less premium are quite volatile. Insurance companies' A&O has increased with increases in total premium. Finally, underwriting gains accruing to insurance companies have increased dramatically since 1994. Note that not only have total government costs increased dramatically but payments to insurance companies have tended to increase at a higher rate suggesting that they have been successful at accruing public rents. In fact, monies accruing to insurance companies are close to rivaling those accruing to producers.

While the current approach to decreasing total uninsured losses is greater subsidization at higher coverage levels (as evidenced in figure 3.1), in 1980 insurance companies were solicited for assistance. It was believed correctly that decreased producer transaction costs via better established delivery channels would lead to significantly higher participation. Intermediaries are often used to carry out public policy when gains in efficiency are expected. In the crop insurance setting, efficiency gains were expected through two avenues. First, the better established delivery channels of insurance companies could reach a greater number of producers for a given cost. Second, the exploitation of information by insurance companies could increase the accuracy of rates thereby decreasing adverse selection activities. However, intermediaries also represent a new group of rent seekers that have the potential to decrease overall program efficiency.

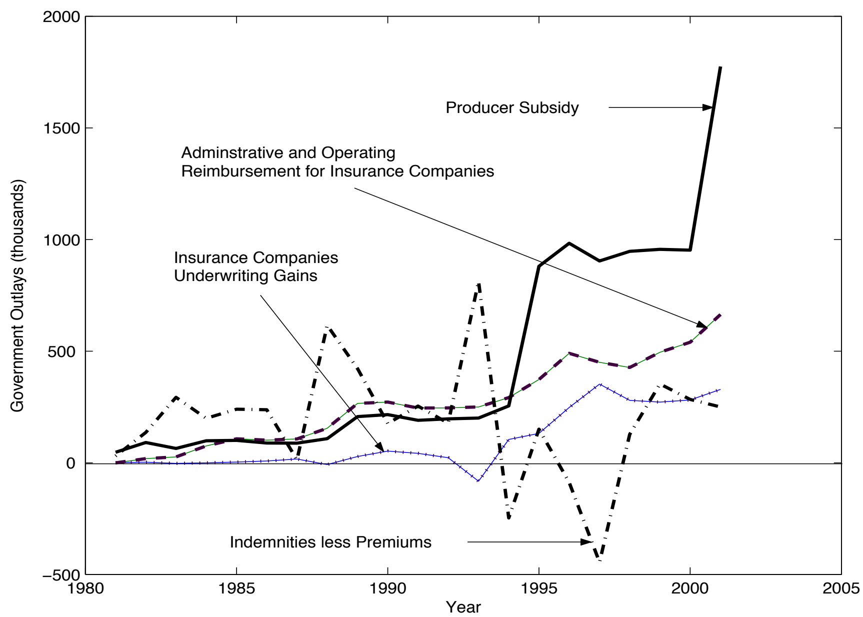

FIGURE 3.1. Government Outlays for U.S. Crop Insurance Program

Interesting policy questions arise: (i) do insurance companies reveal private information to RMA via their contract allocation decisions (explained below); (ii) are insurance companies efficiently allocating (retain and cede decisions) their book of business with respect to the SRA; (iii) would producer demand be increased more through government delivery and increased subsidies; (iv) is monitoring of insurance companies rather than sharing the underwriting gains/losses of the program more efficient; (v) is the political equilibrium stable such that future gains may not be recoverable from revealed private information; and (vi) does this degree of support to the insurance companies represent in itself a political equilibrium. There is much room for research on these important issues.

In this chapter, the focus is on the first policy question. In light of ARPA and the new subsidy structure, producer participation will continue to increase markedly and shift to higher coverage levels. As a result, total premium dollars will also continue to increase significantly and hence premium dollars diverted to insurance companies will dramatically increase. As a result, research on these important policy questions is particularly pertinent in the context of current farm policy which continues to increase monetary resources for the crop insurance program.

To test revelation of private information by the insurance companies, semiparametric as well as parametric methods are used to estimate the contract profitability decisions of insurance companies (whether a contract returns a profit or not) using a data set aggregated to crop-county-year combinations. Using semiparametric methods that avoid distributional assumptions proves to be useful as the parametric method is rejected. Out-of-sample test results show that insurance companies do possess relevant and statistically significant private information and that they strategically reinsure.

The remainder of this chapter proceeds as follows. The second section reviews the U.S. crop insurance program and ARPA as a backdrop. The third section details the Standard Reinsurance Agreement (SRA). The fourth section discusses the data and outlines the econometric methods. The fifth section presents the results while the final section focuses on the corresponding policy implications.

3.2. U.S. Crop Insurance Program

Federally regulated crop insurance programs have been a prominent part of U.S. agricultural policy since the 1930s. In 2002, the estimated number of crop insurance policies exceeded 1.25 million with total liabilities exceeding \$37 billion. Traditional crop insurance schemes offered farmers the opportunity to insure against yield losses resulting from nearly all risks, including such things as drought, fire, flood, hail, and pests. For example, if the farmer's expected wheat yield is 30 bushels per acre ($y^e = 30$), a policy purchased at the 70% coverage level ($\lambda = 0.7$) insures against a realization below 21 bushels per acre (0.7×30 bushels per acre $= 21$ bushels per acre). If the farmer realized a yield of 16 bushels per acre, they would receive an indemnity payment for the insured value of 5 bushels per acre.

A variety of crop insurance plans and a number of new pilot programs are currently under development. Standard crop yield insurance, termed 'Multiple Peril Crop Insurance', pays an indemnity at a predetermined price to replace yield losses. Group-risk yield insurance, termed 'Group Risk Plan', is based upon the county's yield. Insured farmers collect an indemnity when their county's average yield falls below a yield guarantee, regardless of the farmers' actual yields. Three farm-level revenue insurance programs are available for a limited number of crops and regions: 'Crop Revenue Coverage'; 'Income Protection'; and 'Revenue Assurance'. These programs guarantee a minimum level of crop revenue and pay an indemnity if revenues fall beneath the guarantee. The recently developed 'Group Risk Income Plan', a variation of the Group Risk Plan, insures county revenues rather than yields.

Figure 3.1 illustrates that companies are a major participant in the U.S. crop insurance program and warrant attention. ARPA increases the prominence of the crop insurance program in farm policy. The additional cost of this legislation is estimated to be \$8.2 billion over a 5-year period approximately doubling the federal budget on crop insurance programs to \$16.1 billion. ARPA has mandated the expansion of crop insurance in three important dimensions: expanded product coverage including, for example, livestock products; expanded geographical availability for existing crops; and increasing producer demand by doubling subsidies from approximately 30% to 60% of the estimated actuarially fair premium rate. Finally, the current form of the SRA will remain in effect through the 2004 reinsurance year. All legislative actions suggest that crop insurance will remain one of the predominant policy instruments to funnel resources to agricultural producers. As a result, significant public resources will flow to insurance companies. The program clearly deserves close analysis.

3.3. The Standard Reinsurance Agreement (SRA)

The involvement of the insurance companies in the U.S. crop insurance program is defined by the SRA. The insurance companies sell policies and conduct claim adjustments and in return, RMA compensates them for these administrative and operating expenses. The underwriting gain/loss, which is defined as total premiums less total claims or indemnity payments, are shared, asymmetrically, between the insurance companies and the RMA. Both the provisions by which the underwriting gains and losses are shared and the reimbursement for A&O expenses are set out in the SRA.[1]

3.3.1. Provisions of Sharing the Underwriting Gains/Losses

Section II.A.2 of the 1998 SRA states that an insurance company *"...must offer all approved plans of insurance for all approved crops in any State in which it writes an eligible crop insurance contract and must accept and approve all applications from all eligible producers."* An eligible farmer will not be denied access to an available, federally subsidized, crop insurance product. Therefore, an insurance company wishing to conduct business in a state cannot discriminate among farmers, crops, or insurance

[1] See Ker [39] and Skees [62] for a discussion of the A&O expense reimbursement and it's economic implications.

products in that state. An unusual situation arises; the responsibility for pricing the crop policies lies with the RMA but the insurance company must accept some liability for each policy they write and cannot choose which policy they will or will not write.

To elicit the participation of insurance companies, two mechanisms are required that emulate a private market. First, given that insurance companies do not set premium rates, there needs to be a mechanism by which they can cede the liability, or the majority thereof, of an undesirable policy. In a private market, the insurance company would not write a policy deemed undesirable. Second, a mechanism providing an adequate return to the insurance company's capital and a level of protection against ruin (bankruptcy) is needed. Premium rates in a private market reflect a return to capital and a loading factor guarding against ruin. The premium rates set by RMA do not reflect a return to capital but include a loading factor. The SRA provides these two mechanisms which, in effect, emulate a private market from the perspective of the insurance company. In so doing, the SRA also provides a vehicle by which an insurance company can exploit information by strategically reinsuring its book of business.

Under the SRA, insurance companies cannot cede or retain the total underwriting gain/loss of a policy but must place each into one of three funds: assigned risk, developmental, or commercial. For each state in which the insurance company does business, there is a separate assigned risk fund, developmental fund, and commercial fund. The structure of the risk sharing is identical but the parameters that dictate the amount of sharing vary greatly across funds. For each fund, the underwriting gain/loss the insurance company retains is equal to the total underwriting gain/loss for the fund multiplied by two parameters. Formally,

$$\Omega_{IC}^{k} = \Omega^{k} \cdot \mu_1^{k} \cdot \mu_2^{k}$$

where Ω_{IC}^{k} denotes the underwriting gain/loss retained by the insurance company for fund k, Ω^{k} denotes the underwriting gain/loss for fund k, μ_1^{k} is the first parameter for fund k, and μ_2^k is the second parameter for fund k. The underwriting gain/loss retained by the RMA may be defined as:

$$\Omega_{RMA}^k = \Omega^k \cdot (1 - \mu_1^k \cdot \mu_2^k)$$

where Ω_{RMA}^k denotes the underwriting gain/loss retained by the RMA.

The first parameter, μ_1^k, represents an *ex ante* choice variable for the insurance company with respect to the commercial and developmental funds. For $k \equiv$ assigned risk fund, $\mu_1^k = 0.2$. For $k \equiv$ developmental fund, $\mu_1^k \in [0.35, 1.0]$ while for $k \equiv$ commercial fund, $\mu_1^k \in [0.5, 1.0]$. They must choose μ_1^k by July 1 of the preceding crop year.

The second parameter, μ_2^k, is not a fixed scalar but a function of the fund loss ratio. Figure 3.2 illustrates the relationship between the fund loss ratio and the percent of premium retained. The fund loss ratio is defined as the ratio of total claims to total premiums. A loss ratio greater than one results when total claims exceed total premiums and thus corresponds to an underwriting loss. Conversely, a loss ratio less than one results when total premiums exceed total claims and thus corresponds to an underwriting gain.

Figure 3.2 illustrates aspects of the three funds (for the boundary values of μ_1) that pertain directly to the empirical analysis: (i) insurance companies minimize their exposure to the underwriting gains and losses for those policies in the assigned risk fund and (ii) insurance companies maximize their exposure to the underwriting gains and losses for those policies in the commercial fund. Therefore, rational behavior would indicate that policies the insurance companies expect to yield underwriting gains would be placed in the commercial fund while policies the insurance companies expect to yield underwriting losses would be placed in the assigned risk fund. Very little can be determined about policies placed in the developmental fund. If $\mu_1^k = 0.35$ (minimum) this fund resembles the assigned risk fund while if $\mu_1^k = 1.00$ (maximum) this fund resembles the commercial fund. Therefore, without knowledge of μ_1^k for

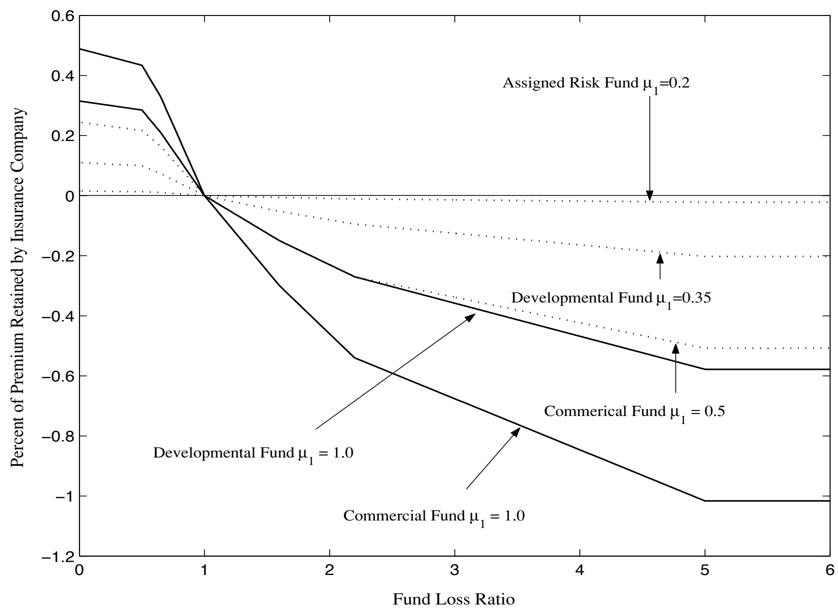

FIGURE 3.2. Percent of Premium Retained by Insurance Company Relative to Fund Loss Ratio

the developmental fund, the expectations of the insurance companies regarding those policies can not be determined.

Two final points of significance need discussed. First, there exists separate developmental and commercial funds for "catastrophic policies", "revenue policies", and "other policies". This latter category is comprised of MPCI/APH policies and GRP policies (GRP policies make up a negligible fraction of the total policies). The empirical analysis considers only the three fund allocations for the "other policies" because insurance companies have significantly less experience and historical information with the "revenue policies" and "catastrophic policies" and thus their fund allocations may not be as efficient. Also note that while these funds (except assigned risk) are not aggregated across types of policies, they are aggregated across crops. Second, insurance companies face a constraint, at the state level, on the maximum percent of premium in their book of business that can be placed in the assigned risk fund. These maximums, which vary quite significantly by state, are located in Table 3.1. While this may inhibit the insurance companies' ability to cede unwanted policies, by choosing $\mu_1^k = 0.35$ for the developmental fund, they can make it resemble the assigned risk fund and there does not exist such restrictions on the developmental fund.

State	Percent	State	Percent	State	Percent
Alabama	50%	Louisiana	50%	Ohio	25%
Alaska	75%	Maine	75%	Oklahoma	50%
Arizona	55%	Maryland	20%	Oregon	30%
Arkansas	50%	Massachusetts	45%	Pennsylvania	25%
California	20%	Michigan	50%	Rhode Island	75%
Colorado	20%	Minnesota	20%	South Carolina	55%
Connecticut	35%	Mississippi	50%	South Dakota	30%
Delaware	30%	Missouri	20%	Tennessee	35%
Florida	40%	Montana	75%	Texas	75%
Georgia	75%	Nebraska	20%	Utah	75%
Hawaii	10%	Nevada	75%	Vermont	15%
Idaho	45%	New Hampshire	10%	Virginia	30%
Illinois	20%	New Jersey	50%	Washington	30%
Indiana	20%	New Mexico	55%	West Virginia	75%
Iowa	15%	New York	40%	Wisconsin	35%
Kansas	20%	North Carolina	20%	Wyoming	35%
Kentucky	25%	North Dakota	45%		

Table 3.1. Maximum Percent of Premium in Assigned Risk Fund by State

While the involvement of the insurance companies was initially justified to increase participation through better established delivery channels, this does not necessarily explain why insurance companies share in the underwriting gains and losses rather than just receive an A&O reimbursement. There are three possible reasons. The first is RMA wishes to share the risk with the private market. This is unlikely because RMA can self-insure without cost while in order to share underwriting gains and losses with the private market, they must pay a risk premium.[2] The second reason is so that

[2]The government can absorb that risk since it is the unique agent that can diversify over time and across space more than any other agent in the economy.

the insurance companies are incentive compatible with RMA when conducting claim adjustments. That is, because the insurance companies must share, to some extent, the underwriting losses, there is less exposure to fraudulent claims. However, in so doing, RMA must offer a vehicle for companies to adversely select. In essence, RMA has traded moral hazard for adverse selection with respect to insurance companies. It is unlikely that the cost of monitoring insurance companies' claim adjustments would be in the order of magnitude of the necessary risk premium that RMA pays to insurance companies. The final reason for having insurance companies involved in the underwriting gains and losses is to design a contract that would reveal relevant private information to RMA regarding premium rates. This information could then be used to improve the accuracy of premium rates. The allocation of the policies to the three funds does reveal the expectations of the insurance companies with respect to the profitability of those policies. The key question, and the one studied here, is whether those expectations/allocations reveal unknown information.

3.4. Data and Methodology

Recall that the purpose here is to test whether relevant private information is revealed with respect to rating policies (premium rates) in the fund allocations of insurance companies. This hypothesis can be tested by predicting whether policies are profitable or not. If a policy is correctly expected to be profitable (premium exceeds expected indemnities), this would suggest that the premium rate needs to decrease. Conversely, if a policy is correctly expected to be unprofitable (premium less than expected indemnities), this would suggest that the premium rate needs to increase. Specifically, one can test whether the percent of correct predictions increases significantly when the insurance companies' fund allocations are included as explanatory variables.

The dependent variable is whether a set of policies returned a profit or not. If the total premium is greater than indemnities, $y = 1$. Conversely, if total premium is less than indemnities, $y = 0$. The first model is:

$$y = F(v\beta) + \epsilon \tag{3.1}$$

where v is information available to RMA such as historical loss ratio, crop dummies, state maximums on the assigned risk fund, and liability changes. $F(\cdot)$ is the link function and $v\beta$ is the index. The second model is:

$$y = F(v\beta + \text{ insurance company fund allocations} * \gamma) + \epsilon \tag{3.2}$$

where the set of explanatory variables now includes the insurance company fund allocations.[3]

3.4.1. The Data

The data is comprised of the total premium, indemnities, liability, and number of policies in each of the three funds by crop-county-year combination on corn, cotton, soybeans, and wheat for the reinsurance years 1998, 1999, 2000, and 2001. Combinations with less than $500,000 in liability are removed leaving 7,600 crop-county-year combinations.

There are three caveats regarding the data that require discussion. First, the data is aggregated to the county level; policy specific fund allocation decisions were not

[3] The dependent variable is based on whether a set of policies returned a profit or not rather than the level of profit. As argued in Ker and McGowan [40] (independent strategy), this is the decision that the insurance company faces. Under their "independent strategy" which assumes the loss ratio of a given policy is independent of the loss ratio for that fund, the insurance company only needs to consider whether the policy is expected to return an underwriting gain or loss. Recall, insurance companies maximize their share of the underwriting gains/losses with the commercial fund and minimizes their share with the assigned risk fund. If the loss ratio of the policy is independent of the fund loss ratio, then the insurance company maximizes their total underwriting gain/loss by maximizing the size of the commercial fund. This of course is maximized by only allocating those policies with expected underwriting gain in the commercial fund. Those with expected underwriting loss are allocated to the assigned risk fund. Therefore, given this allocation rule the most that can be ascertained from their actual allocations is whether a policy is expected, by the insurance company, to return an underwriting gain or an underwriting loss, not the expected magnitude.

available. This is not problematic, however, because anecdotal evidence (discussion with companies as well as looking at their fund allocations) suggests that insurance companies tend to allocate by crop-county combinations rather than individual policies. Second, the data is aggregated across coverage level. Again, insurance companies tend to allocate by crop-county combinations and do not consider coverage levels because of the lack of information at coverage levels differing from the 65% coverage level.[4] Third, the data is aggregated across insurance companies. While company specific fund allocations would be preferred, it was only possible to obtain aggregated data.

3.4.2. Econometric Methodology

For the estimation of (3.1) and (3.2), three methods are considered: parametric probit and two semiparametric estimators, namely, SLS estimator of Ichimura [35] and PGSIM. These semiparametric estimators have been discussed in detail in chapter two. The smoothing parameter h for these semiparametric estimators is chosen by applying the Härdle et al. [25] idea (see section 2.5.2). In estimations, a normal density function truncated at plus and minus 3 standard deviations is used as the kernel for both of the single-index model estimators. The parametric start for PGSIM is a probit model, i.e., normal cdf. No trimming of the data was conducted for either of the semiparametric estimators.[5] Note that, in chapter two, a consistent covariance matrix estimator is not provided for PGSIM. In absence of this estimator, it would be ideal to bootstrap the standard errors. Unfortunately this is not feasible due to the sample size and the time that is required for optimization.[6] For this reason, initially,

[4]One may have more concern here in that premium rates at higher coverage levels tend to be biased upwards in high premium rate areas. However, it is likely that by eliminating those combinations with less than $500,000 in liability, the problem have been mitigated; participation in high rate areas tends to be weak.

[5]As Horowitz [31, p.53] explains "...this amounts to assuming that the support of [the index] is larger than that observed in the data."

[6]Note that during the optimization, at each iteration, a nonparametric estimation is required.

it was experimented with the robust covariance matrix estimator of White [66] since the problem is likelihood based. This estimator requires second derivatives, however, and due to highly nonlinear nature of the problem, the numerical second derivatives from the optimizer were not reliable at all. Thus, the so-called BHHH estimator using the analytical gradient were calculated. However, one should note that the empirical tests are based on out-of-sample forecast performance rather than in-sample standard errors.

3.5. Estimation Results

To test the hypothesis that no private information is revealed in the fund allocation decisions of insurance companies, the sample is randomly split into an estimation sample and a prediction sample. The hypothesis is evaluated using out-of-sample rather than in-sample methods because (i) the insurance companies must make their allocation decisions out-of-sample; and (ii) out-of-sample tests minimize spurious results from over-fitting the data (particularly concerning for nonparametric methods which, if applied inappropriately, can be made to over fit the data).

The explanatory variables used in the analysis are crop dummies for cotton, soybeans, and wheat, historical loss ratios (from 1981 to year prior to the corresponding crop year), ratio of current liability to the previous year liability, the maximum percent of premium allowed in the assigned risk fund for that state, percent of premium placed in the commercial fund, and the percent of premium placed in the assigned risk fund.

3.5.1. Revelation of Private Information

To test the hypothesis about the revelation of private information, two sets of models are estimated. The difference between the first and second set of models is that the fund allocation explanatory variables are only included in the second set of models.

The estimation results and predictive performances for the models without and with the fund allocation data are located in tables 3.2 and 3.3 respectively (standard errors are in parentheses).

Parameter Estimate	Probit	Ichimura	PGSIM
intercept	1.6902	0.0000*	0.0000*
	(0.0683)	n/a	n/a
d_{cotton}	-0.2374	-5.1377	3.9632
	(0.0881)	(0.0754)	(0.0146)
$d_{soybeans}$	-0.1077	3.1030	4.1033
	(0.0587)	(0.0630)	(0.0337)
d_{wheat}	-0.1067	-2.7825	0.8989
	(0.0715)	(0.0911)	(0.0224)
liability ratio	-0.0020	-0.0687	-0.2558
	(0.0078)	(0.0067)	(0.0013)
state risk	-1.6868	-6.8218	-2.1699
	(0.1533)	(0.1150)	(0.0013)
historical LR	-0.2521	-0.2521*	-0.2521*
	(0.0475)	n/a	n/a
h	n/a	0.3264	0.1089
Predictive Performance	74.66%	77.84%	78.34%

* - parameter is restricted as necessitated by estimation procedure

TABLE 3.2. Estimation Results and Predictive Performance without Fund Allocation Data

Note that d_{cotton} is the dummy variable for cotton, $d_{soybeans}$ is the dummy variable for soybeans, d_{wheat} is the dummy variable for wheat, liability ratio is the ratio of current year's liability to the previous year's liability, state risk is the percent of premium in the insurance companies book of business that is allowed in the assigned risk fund, historical LR is the historical loss ratio up to but not including that years insurance experience, commercial is the percent of premium placed in the commercial fund, and assigned is the percent of premium placed in the assigned risk fund. The percent of premium placed in the developmental fund is not included as that would result in a singularity problem as the three percentages in the three funds always

Parameter Estimate	Probit	Ichimura	PGSIM
intercept	1.3813	0.0000*	0.0000*
	(0.1426)	n/a	n/a
d_{cotton}	-0.2129	-0.1836	-0.2954
	(0.0885)	(0.0631)	(0.0286)
$d_{soybeans}$	-0.1083	-0.0728	1.6128
	(0.0589)	(0.0404)	(0.0183)
d_{wheat}	-0.0880	-0.1330	-0.1540
	(0.0718)	(0.0527)	(0.0256)
liability ratio	-0.0028	-0.0022	-0.0141
	(0.0078)	(0.0095)	(0.0040)
state risk	-1.6507	-2.9153	-1.9087
	(0.1545)	(0.1102)	(0.0382)
commercial	0.2854	0.1448	0.5151
	(0.1236)	(0.0445)	(0.0171)
assigned	-0.3957	-0.7199	-0.8445
	(0.2141)	(0.1147)	(0.0417)
historical LR	-0.1738	-0.1738*	-0.1738*
	(0.0523)	n/a	n/a
h	n/a	0.1025	0.0484
Predictive Performance	75.24%	79.63%	79.18%

* - parameter is restricted as necessitated by estimation procedure

TABLE 3.3. Estimation Results and Predictive Performance with Fund Allocation Data

equals one. Recall the dependent variable is set equal to 1 if the set of policies resulted in a profit (premium greater than indemnities) and 0 if the set of policies resulted in a loss (premium less than indemnities). Finally, for the semiparametric and nonparametric estimators, the intercept is restricted to 0 and the parameter estimate on the historical loss ratio is restricted to the probit estimate as is commonly done.[7]

There are no a priori expectations about the signs of the dummy variables whereas one might have expectations about the signs of the other parameter estimates. First, the sign of liability ratio is negative as expected. If liability increases (decreases)

[7]This parameter can be set to any finite constant.

significantly from one year to the next, this may suggest that producers perceive their return to that insurance contract to have increased (decreased) and thus the expected return for the insurance company may decrease (increase). The parameter estimate on state risk is negative (as expected) and significant. This indicates, quite interestingly, that policies in those states with higher bounds on the percent of premium allowed in the assigned risk fund are less likely to be profitable. It is interesting from a political economy perspective that these parameters were negotiated in the 1980s and yet they still provide an indicator as to the profitability of a current crop insurance contract. The parameter estimate on the historical loss ratio in the probit models are negative and significant as expected; the higher the loss ratio the less likely the policies are profitable. The parameter on the percent of premium in the commercial fund is positive as expected. This suggests that policies the insurance company places in the commercial fund are more likely to be profitable. This is statistically significant in both the Probit and semiparametric models. Finally, the parameter on the assigned variable is negative as expected suggesting that policies the insurance company places in the assigned risk fund are less likely to be profitable.

The null hypothesis is that no private information is revealed in the fund allocation decisions. To test this, the percent of policies correctly predicted with and without the fund allocation explanatory variables are compared. Specifically, the percent of policies correctly predicted should increase significantly when the fund allocation explanatory variables are included in the model. The test may be formally written as:

$$H_o : \rho_f - \rho_{nf} = 0 \text{ versus } H_a : \rho_f - \rho_{nf} > 0$$

where ρ_f corresponds to the percent of correct predictions from the model that includes the two fund variables while ρ_{nf} corresponds to the percent of correct predictions from the model that does not include the fund variables. Table 3.4 summarizes the empirical tests. Standard errors are calculated by bootstrapping the prediction sample and recovering the difference in the percent of correct predictions (500 bootstraps are used).

Test	Test Statistic	Standard Error
Private Information Tests		
Model 2 less Model 1 with Probit	0.0058	0.00169
Model 2 less Model 1 with Ichimura	0.0179	0.00576
Model 2 less Model 1 with PGSIM	0.0084	0.00592
Probit versus Nonparametric Tests		
Ichimura less Probit - Model 1	0.0318	0.00557
PGSIM less Probit - Model 1	0.0368	0.00506
Ichimura less Probit - Model 2	0.0439	0.00504
PGSIM less Probit - Model 2	0.0394	0.00498

Table 3.4. Hypothesis Test Results

The out-of-sample test results reveal—exception being the PGSIM—that predictive performance increases significantly when the fund allocations are included as explanatory variables indicating that there exists relevant private information. This coincides with the in-sample results which suggested that the fund allocations were significant at explaining profitable and nonprofitable sets of policies. Therefore, the null hypothesis that no relevant private information is revealed in the fund allocation data is rejected.

3.5.2. Testing the Parametric Probit Model

Three tests are undertaken for the parametric probit model. The first is the so-called HH test of Horowitz and Härdle [33]. This test is motivated by conditional moment tests. Horowitz and Härdle [33] replace the parametric alternative model with a semiparametric one. The advantage of this test relative to tests with arbitrary nonparametric alternatives is that as long as only the shape of the link function, and not the single-index structure, is the issue, the HH test will be more powerful as the latter tests suffer from curse of dimensionality.[8] The HH test, on the other hand, assumes that the conditional expectation of the dependent variable depends on the regressors only through the index not only in the null but in the alternative as well and thus avoids curse of dimensionality. The HH test statistic is

$$T_n = \sqrt{h}\sum_{i=1}^{n} w(\hat{z}_i)[y_i - F(\hat{z}_i)][\hat{F}(\hat{z}_i) - F(\hat{z}_i)]$$

where $\hat{z}_i = v_i\hat{\beta}_{probit}$ is the estimated index from the parametric probit model, w is a nonnegative weight function which can be chosen to be an indicator variable of an interval that contains 95-99% of $\hat{z}$, and F is the normal distribution function. For $\hat{F}$, they use the jackknife-like method of Schucany and Sommers [59] to achieve asymptotic unbiasedness. Formally,

$$\hat{F}(l) = [\hat{F}_h(l) - (h/s)^r\hat{F}_s(l)]/[1 - (h/s)^r]$$

and

$$\hat{F}_t(l) = \sum_{j\neq i} y_j K\left(\frac{l-\hat{z}_j}{t}\right) \Big/ \sum_{j\neq i} K\left(\frac{l-\hat{z}_j}{t}\right) \quad \text{for} \quad t = h, s$$

where $h = cn^{-1/(2r+1)}$, $s = c'n^{-\delta/(2r+1)}$ with $c, c' > 0$, $0 < \delta < 1$, and K is a kernel of order $r \geqslant 2$. Horowitz and Härdle [33] show that T_n is asymptotically distributed as $N(0, \sigma_T^2)$ where

$$\sigma_T^2 = 2C_K \int_{-\infty}^{\infty} w(l)^2[\sigma^2(l)]^2 dl. \tag{3.3}$$

In (3.3), $C_K = \int_{-\infty}^{\infty} K(u)^2 du$ and $\sigma^2(l) = Var(y|z = l)$. This test is conducted for both models 1 and 2 using a standard normal density as the kernel ($r = 2$) and w was taken to be the indicator variable which equals 1 on an interval containing 98% of $\hat{z}$ and 0 elsewhere. There is no optimal way of choosing h and s. Following Härdle

[8] As the number of regressors increases, estimation precision declines rapidly. This phenomenon is known as curse of dimensionality.

et al. [23], s is determined according to $s = hn^{(1-\delta)/5}$ with $\delta = 0.1$.[9] For h, several values were used which were found after a graphical inspection of $\hat{F}$. Based on those values, $T_n/\hat{\sigma}_T$ was in the range 6.66-7.63 for model 1 and 6.81-7.27 for model 2. Thus, for both models, the probit is rejected.

The second test calculates the difference in the predictive performance of the semiparametric methods versus the probit for both models 1 and 2 (see Table 3.4). These test results reject the probit model in favor of the semiparametric methods. The test results could not reject either of the semiparametric model in favor of the other.

A third less formal but pictorially pleasing test follows the graphical approach of Horowitz [32, p.53]. Figures 3.3 and 3.4 show nonparametric kernel estimates of $dF/d\hat{z}$, pointwise 95% bootstrap confidence interval, and the normal density function. Note that for a probit model, $dF/d\hat{z}$ would be the normal density function. In these nonparametric estimations, the standard normal density is used as the kernel. For bandwidth selection, initially, cross-validation for derivative estimation (see Härdle [22, pp.160-161]) was tried, however, numerical minimization of this objective function was not successful for the most part so after experimenting with CV, the bandwidths are chosen accordingly. In both graphs, the derivative of the link function is clearly left skewed that can not be accommodated by the symmetric normal density. Pointwise confidence intervals are represented by the dotted lines. In both figures, the derivatives are bimodal which suggests that the true data generating processes may possibly be a mixture of two populations. Using a parametric probit model clearly misses these features of the data.

[9] Härdle et al. [23] suggest using bootstrap instead of normal approximation to calculate critical values and show that bootstrap yields better approximations to the critical values in a simulation study with $n = 200$. Here, however, the sample size is relatively large ($n = 3,800$) and thus normal approximation is used.

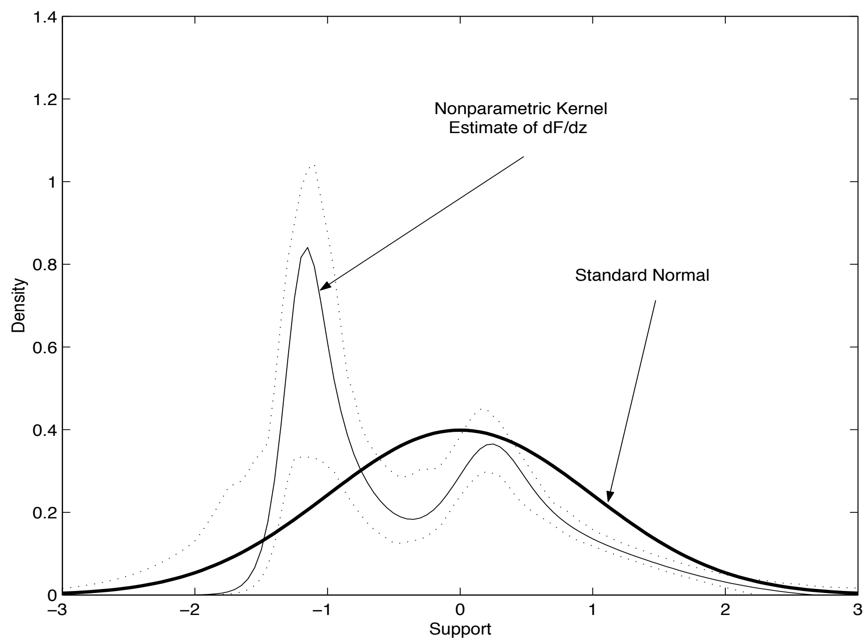

FIGURE 3.3. Test of Probit for the Data not Including Fund Allocations

3.6. Conclusions and Policy Implications

Although the crop insurance program has garnered significant attention in the academic literature, surprisingly little has focused on the insurance companies and in particular the SRA. However, the rents obtained by the insurance companies in return for their involvement are close to rivaling those obtained by producers. Consequently, more research is needed, both theoretically and empirically, focusing on the involvement of insurance companies.

This chapter focused on whether insurance companies reveal private information through their reinsurance (fund allocation) decisions. Using both parametric and semiparametric estimators, out-of-sample tests have been conducted and it is shown that insurance companies do possess statistically significant private information that may warrant their involvement in the crop insurance program. However, the percentage increase in predictive performance is rather marginal in that the increase ranged

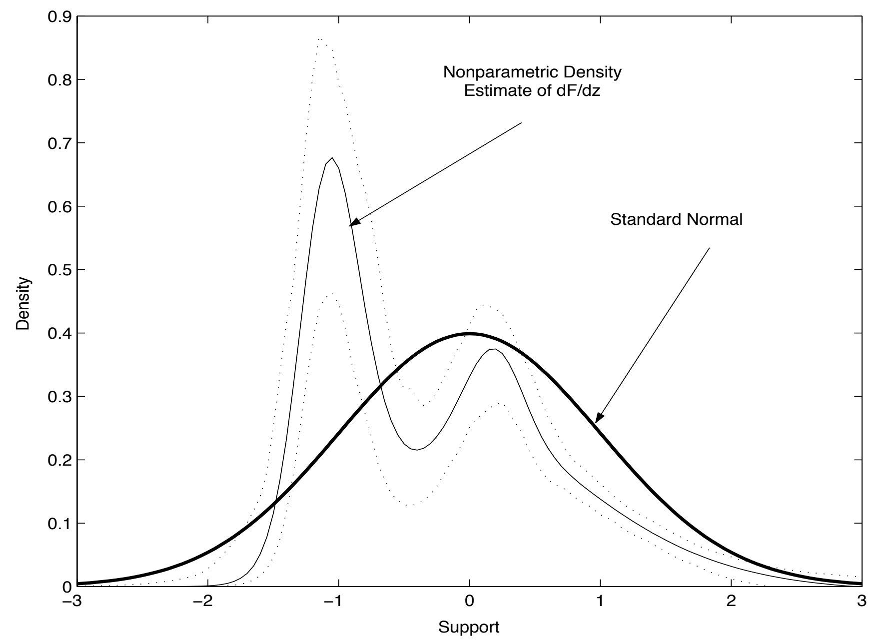

FIGURE 3.4. Test of Probit for the Data Including Fund Allocations

from 0.58 percentage points with the probit model to 1.79 percentage points with the Ichimura model.

Recall the arguments for involving the insurance companies in the crop insurance program are: (i) lower delivery costs; (ii) increased efficiency due to the revelation of private information; and (iii) risk sharing. First, Ker [39] conjectures that delivery costs are not lower with insurance companies. Second, even though the insurance companies reveal private information through their reinsurance decisions, it is not clear whether the RMA adjusts premium rates given the revelation of new information. Third, the government need not be risk sharing with insurance companies as it can self-insure at no cost.

Why then are insurance companies involved in this crop insurance program? One may conjecture that they were initially involved to increase demand through lower transaction costs incurred by producers by the better established delivery channels of insurance companies. However, this increase in demand was obviously not sufficient to ward off ad-hoc disaster aid. In addition, one can argue that the savings brought about by a government delivered insurance program, if funneled back to producers through subsidies, would have a substantially greater impact on demand, particularly now that producers can sign-up for their insurance electronically. It appears that the insurance companies have been active and successful in a political economy sense to not only survive but obtain significant rents. There is ample evidence of this in ARPA. For example, premium subsidies have caused a pronounced shift in demand to higher coverage levels. Therefore, the amount of A&O reimbursement has increased according to the increases in premium at these higher coverage levels. The percentage rate of A&O reimbursement should decline with the coverage level as there is both a fixed and variable cost component with A&O activities. However, this has not happened.

4. Trading Collar, Intraday Periodicity, and Stock Market Volatility

4.1. Introduction

On October 19, 1987, the Dow Jones Industrial Average (DJIA) plunged 508 points, approximately 25% from the previous day's close. Portfolio insurance and especially program trading (explained below) were blamed for the excessive volatility and this crash. To date, in the academic world and daily press, there seems to be an even split on whether program trading is really to blame for volatility.

New York Stock Exchange (NYSE) releases data about program trading each week and this information can be used by the public to identify the most active program traders on the NYSE. For a short while, the NYSE stopped releasing this information but immediately reinstated the practice in response to public requests and expressed concerns by the House telecommunications and finance subcommittee (WSJ [64]). In the past, some big institutional investors have announced they would end their business with brokerage firms that do program trading (Reibstein and Friday [56]). Needless to say, most of those firms' customers are small investors who "get scared by volatility".

Program trading is basically the simultaneous purchase or sale of a basket of 15 or more stocks with a total value of $1 million or more. The most famous form of program trading is a derivative product-related strategy called index arbitrage. Index arbitrage is the purchase or sale of a basket of stocks in conjunction with the sale or purchase of a derivative product, such as index futures, in order to profit from the price difference between the basket and the derivative product. In other words, as Booth [7] explains, "...index arbitrage kicks in when the price discrepancy between an index future and the underlying stocks grows large enough that it is possible to lock in a profit by selling whichever of the two is over-priced and buying the other". Using the illustration in Hogan et al. [28, p.736], consider the no-artibrage condition (ignoring transaction costs)

$$\log F_{t,T} = \log S_t + \log(1 + r_{t,T}) - \log D_{t,T}$$

where $F_{t,T}$ is the price of an index futures at time t that expires at time T. S_t denotes the stock index price at time t and $r_{t,T}$ is the appropriate interest rate. $D_{t,T}$ is the future value of dividends between t and T on a portfolio that mimics the index. If, for instance,

$$\log F_{t,T} > \log S_t + \log(1 + r_{t,T}) - \log D_{t,T}$$

then the index arbitrage opportunity arises and the investor purchases the portfolio that mimics the index while selling the index futures.

Different studies were conducted and recommendations were made in response to the October 1987 crash and the mini-crash of October 1989. Indeed, the NYSE established a set of rules and regulations to avoid excessive market volatility and to regain (especially small) investor confidence.[1] The most famous of these rules is Rule 80A Collar.

The collar, formally known as Rule 80A, was instituted on August 1, 1990 and basically restricts (explained below) index arbitrage form of program trading in component stocks of the S&P500 stock price index. Originally, the collar was set at 50 points, i.e., Rule 80A restrictions on index arbitrage trading in component stocks of the S&P500 index were to be imposed when the DJIA was above or below its closing value on the previous trading day by 50 points or more. The restrictions were to be removed when the DJIA returned to within 25 points of previous day's closing value. In February 1999, percentage levels were implemented and since then have been adjusted quarterly; the level for the collar is calculated as 2 percent of the av-

[1] See Greenwald and Stein [21] for recommendations by the Presidential Task Force on Market Mechanisms and Lindsey and Pecora [44] for details of regulatory developments.

erage closing value of the DJIA for the last month of the previous quarter and the collar is removed when the DJIA advances or retreats from the prior day's close to less than or equal to half of the 2 percent value. The collar restriction in component stocks of the S&P500 index works in the following way: If the DJIA declines (with third quarter of 2001 values) 210 points or more then the index arbitrage sell orders of the S&P500 stocks have to be stabilizing, i.e., sell plus.[2] If the DJIA advances 210 points or more then the index arbitrage buy orders of the S&P500 stocks have to be stabilizing, i.e., buy minus[3] (see NYSE [50] for details of the definitions). In simpler words, Rule80A forces sell (buy) orders to be done at a higher (lower) price when the market is declining (rising).

Generalized autoregressive conditional heteroscedasticity (GARCH) models have become almost a standard for modeling market volatility. But a vast majority of market volatility models have not taken the trading collars into consideration. The presence of trading collars could alter volatility dynamics and volatility models should account for this. Forecasts from models that do not incorporate important institutional details may not be as accurate. In this study, GARCH models are estimated that explicitly account for the NYSE's trading collar rules using intraday data. The motivation and contribution is to explain what is happening to market volatility when Rule 80A is in effect. That is, during trade, if the collar restrictions are imposed, are there any significant changes to market volatility? Does the volatility react to shocks in the same magnitude when the rule is in effect as when it is not? The chapter also provides a descriptive analysis of the collar including the percentage of observations during the collar regime and what percent of the time the rule was in effect due to a bull and a bear market.

Financial markets exhibit strong periodic dependencies across the trading day—

[2]An order to sell "plus" is an order to sell a stated amount of a stock provided that the price to be obtained is not lower than the last sale if the last sale was a "plus" tick (see NYSE [50]).

[3]An order to buy "minus" is an order to buy a stated amount of a stock provided that the price to be obtained is not higher than the last sale if the last sale was a "minus" tick (see NYSE [50]).

typically volatility is highest at the open and toward the close of the day—and failure to account for this may seriously distort the inferences made from the models (Bollerslev [6]). Two approaches have been used in the literature to capture intraday seasonal patterns in volatility in the context of GARCH models: use of dummy variables in the conditional variance equation (e.g., Baillie and Bollerslev [3] and Ederington and Lee [11]) and use of Flexible Fourier forms (e.g., Andersen and Bollerslev [1] and Martens [47]).

In this study, a polynomial function is used to capture systematic intraday periodicities in volatility and the seasonal components are estimated simultaneously with the rest of the model. The polynomial functional form, like the Fourier form introduced by Andersen and Bollerslev [1], can be viewed as a flexible form for approximating the true, unknown seasonal pattern. By increasing the number of terms in the polynomial, the function can be made arbitrarily close to the true seasonal pattern. Also, using simple parametric restrictions, the function can be made continuous and smooth as it cycles from one day to the next.

Andersen and Bollerslev [1] first estimate seasonality and then use this to deseasonalize the returns. Then they fit GARCH models to the deseasonalized data. In this study, the seasonal components and the GARCH parameters are estimated simultaneously. By estimating seasonality simultaneously with the GARCH parameters, this approach avoids the shortfalls of the two-step procedure with estimated data (Murphy and Topel [49]). This is the first study to examine market volatility in the context of trading collars while simultaneously accounting for intraday seasonality.

The polynomial form advocated in this chapter is parsimonious in parameters and is easy to estimate. For high frequency data, the dummy variable approach requires too many parameters to completely specify the intraday seasonality. For example, for the five-minute interval data that is used here, it would take as many as 78 parameters using the dummy variables approach to capture the time of the day effects on conditional volatility. With a polynomial specification, a sufficiently flexible seasonal pattern can be estimated often times using just 4 or 5 parameters.

The organization of the chapter is as follows: In the next section, the small literature on the NYSE Rules 80A and 80B is reviewed. Section 4.3 discusses the data and creation of certain variables in detail. Then the model is discussed in section 4.4. Estimation results are in section 4.5. Section 4.6 concludes.

4.2. Literature Review

In this section, the few studies on the NYSE Rules 80A and 80B are reviewed although the data and the methodologies that they use are not as detailed and thorough as in this study.

In [8], Booth and Broussard use extreme value statistical theory to determine the probability that a particular extreme negative return (which could trigger the circuit breaker[4]) will occur sometime during the day. Their finding is that the NYSE fixed-point circuit breaker is an inflexible tool that may lead to unwanted triggering of the mechanism. Since April 1998, fixed-point circuit breaker has been replaced with percentage levels as the authors suggest.

Kuserk et al. [43] look at whether the triggering of Rule 80A is effective in delinking of the futures and cash markets (namely S&P500 index futures and the cash index) and whether the triggering causes increased volatility in the (S&P500 index) futures market. They don't find conclusive evidence to support either one.

[4]Circuit breaker (Rule 80B) points were originally adopted on October 19, 1988 and represent the threshold values for the DJIA at which trading on NYSE is halted for single day declines (below its closing values on the previous trading day) in the DJIA. Until April 1998 these declines were point level declines. In April 1998 percentage levels were implemented and since then have been adjusted quarterly. Since their adoption, circuit breakers have been triggered only once on October 27, 1997.

Two important differences between Rule 80A and Rule 80B are worth emphasizing: Rule 80A collar does not stop program trading, "...it just throws sand in the gears" and "...forces the arbitragers to trade against the trend" (Power [55]) whereas Rule 80B circuit breakers halt all trading including program trading. The second difference is that Rule 80A kicks in when the DJIA is off (up or down) by certain points from the previous day's close whereas Rule 80B halts trading only when the DJIA is below (by certain points) its closing value on the previous trading day.

In [58], Santoni and Liu look at whether the adoption of Rules 80A and 80B reduce the volatility of stock prices in the cash market. They use daily closing values of the S&P500 composite index to estimate a GARCH(1,1) model with several dummy variables in the conditional variance equation that correspond to periods following the adoption and revision of the rules by the NYSE and Chicago Mercantile Exchange (CME)[5] to test for structural breaks. They find that the adoption of the rules and their revisions had no appreciable effect on volatility and conclude that the daily data is not consistent with the hypothesis that these rules reduce volatility. Furthermore, they also analyze intraday data on days when the Rule 80A collar which restricts index arbitrage was triggered. For intraday data, they focus on the unconditional variance and even though they find some evidence suggesting that the variance is lower following a trigger point, they conclude that the decline in volatility is not immediately associated with the trigger.

Kuhn et al. [42] look at whether circuit breakers[6] moderated (S&P500 and Major Market Index(MMI)) cash and index futures price volatility on October 13 and 16, 1989.[7] They calculate the standard deviation of the log price change, average absolute log price change, and range of price change calculated as the log of the high price divided by the low price, all over one-minute intervals. They find no evidence, neither in cash nor in index futures markets, to support the hypothesis that circuit breakers moderate volatility.

The relation between volume and volatility in financial markets is well documented. Hogan et al. [28] use a bivariate error-correction GARCH model to examine the relationship between program and non-program trading volume and volatility in the S&P500 cash and futures markets. Because much of program trading (especially

[5]After the October 1987 crash, besides NYSE, CME adopted new rules as well and imposed opening and intraday daily price limits on the S&P500 futures contract.

[6]In their definition, circuit breakers include downside price limits, trading halts, and the restrictions on certain types of trading.

[7]On October 13, 1989, the DJIA plunged 191 points, a 6.9% decline.

index arbitrage) exploits the price differences between the cash and futures markets, one would expect that program trading would affect both markets. The conditional variances are modeled as GARCH(1,1) with volume variables (several models are estimated with different combinations of program trading volume, non-program trading volume, and buy and sell program trade volumes in the conditional variance equations). They find that program trading leads to a higher market volatility whereas non-program trading is only weakly related to volatility. Furthermore, they also examine whether sell-program trades and buy-program trades have different effects and find that sell-program trades are associated with higher market volatility than buy-program trades.

4.3. The Data

The data is obtained from Tick Data Inc. for the period 4/5/1993 to 8/31/2001. Unfortunately the intraday data for pre-1987 or at least pre-1990 (the collar was instituted in August, 1990) could not be obtained. In this case, obviously, it is not possible to examine volatility by means of a pre-collar vs. post-collar analysis.

Using five-minute DJIA data, a dummy variable D is created which takes the value of one when the collar is in effect and zero when it is not. Another variable, Y, is also created which is the current value of the Dow minus the previous day's close. This dummy is created in the following way: Since it was instituted, the value of the collar has been changed several times after percentage levels were implemented in February 1999. The collar values since the rule's adoption are given in table 4.1. In the table the collar pair x-z is that if the current value of the Dow is above or below the previous day's close by x points or more then the collar is triggered. Once the collar is in effect, the value of the Dow should go back within z points or less of the previous day's close for the rule to be lifted. Otherwise the collar remains in effect until the end of the day. No matter what the situation was at the previous day's close

	Collar Levels
August 1, 1990 - February 15, 1999	50-25
February 16, 1999 - March 31, 1999	180-90
April 1, 1999 - June 30, 1999	190-90
July 1, 1999 - December 31, 1999	210-100
January 3, 2000 - March 31, 2000	220-110
April 3, 2000 - June 30, 2000	200-100
July 3, 2000 - March 30, 2001	210-100
April 2, 2001 - June 29, 2001	200-100
July 2, 2001 - September 28, 2001	210-100

TABLE 4.1. Rule 80A Trading-Collar Levels

(whether the rule was in effect or not) the beginning of each trading day is a new start, that is trade does not open with the rule in effect. So the dummy variable D is created considering these differing values of the collar for different time periods and that the first observation of a day should be treated as a no collar observation even if the previous observation, i.e., the previous day's close, may be a collar observation. In other words, the Y variable for the first observation of a day should be compared with x even if the previous observation may be compared with z. Two more dummy variables, D^b and D^a, are created as well. D^b is the dummy variable which takes a value of one if the collar is triggred from below, that is if the rule is in effect due to decreases. In other words D^b takes a value of one if $D = 1$ and $Y < 0$. D^a is just the opposite. Clearly $D^a + D^b = D$.

Note that the collar values are the same until February 1999 and the bull market of 1990s means that from the day it was instituted until February 1999, the likelihood that the collar was triggered and stayed in effect increases in an almost artificial way because a 50 point collar in early 1990s is higher percentage wise than in late 1990s. For instance, 50-25 point collar pair corresponds to 1.48% - 0.74% based on closing value on 4/5/1993 whereas it corresponds to 0.54% - 0.27% based on closing value on 2/12/1999.

Table 4.2 shows the mean values for the dummy variables for different time periods. With the exception of 1995, the percentage of the collar observations increases steadily as expected until February 16, 1999. Since this date on, the value of the collar has been changed quarterly so it can be argued that this period is not "flawed" like the other periods. For this last period, the mean of the dummy variable D is 0.0600.

			D	D^b	D^a
4/6/1993	-	12/31/1993	0.0022	0.0016	0.0006
1/3/1994	-	12/30/1994	0.0374	0.0189	0.0185
1/3/1995	-	12/29/1995	0.0336	0.0188	0.0148
1/2/1996	-	12/31/1996	0.1574	0.0773	0.0801
1/2/1997	-	12/31/1997	0.4479	0.1986	0.2493
1/2/1998	-	2/12/1999	0.5645	0.2720	0.2925
2/16/1999	-	8/31/2001	0.0600	0.0355	0.0245

TABLE 4.2. Mean Values of Dummy Variables for Different Time Periods

In other words 6% of the observations in this period were observed when the collar was in effect.[8] The means of D^a and D^b are, respectively, 0.0245 and 0.0355. This suggests that, during this period, Rule 80A was in effect due to increases in the Dow 41% of the time and due to decreases 59% of the time. Several important numbers to better characterize the collar are calculated as well for this last period and they will be discussed here. During this period, which is a little more than two and a half years, the collar was triggered 99 times. Once it was triggered, the collar stayed in effect, on average, 29.72 observations. Because five-minute data is used in this study, this corresponds to an average of 2 hours and 29 minutes. That is, once the collar was triggered, on average, it stayed in effect for almost two and a half hours.

Table 4.3 has the descriptive statistics for the continuously compounded return

[8]Because five-minute data is used, clearly, it is assumed that in the five-minute interval between two observations, there is no data which would change the status of the collar. The same dummy variables D, D^a, and D^b for this period are created with one-minute data as well with mean values of 0.0621, 0.0261, and 0.0360 respectively which are quite close to the numbers in table 4.2. To keep the sample size manageable, five-minute returns were used instead of one-minute returns data

on the Dow times 100 for the period 2/16/1999 - 8/31/2001 using five-minute data giving a total of 49,001 observations. In the table, B-J is Bera-Jarque test statistic for

	Mean	Standard Deviation	Skewness	Excess Kurtosis	B-J	ARCH
Return	0.0000854	0.1182	0.0591	19.11	745,644.28	2,072.23

TABLE 4.3. Descriptive Statistics of five-minute DJIA Returns times 100

the null hypothesis of normality which is distributed as chi-squared with two degrees of freedom. ARCH is LM test for 12th order ARCH effects for the null hypothesis of no ARCH and is distributed as chi-squared with 12 degrees of freedom. Overnight returns are excluded so it is exclusively five-minute returns in the data set. Normality is strongly rejected as Bera-Jarque statistic is highly significant. Since the returns are highly leptokurtic, a distribution, like student-t, which has thicker tails than the normal would be appropriate. Also there is strong evidence for ARCH effects as the LM statistic for up to 12th order ARCH effects is highly significant. The sample mean and standard deviation for five-minute returns for the DJIA are 8.540×10^{-7} and 1.182×10^{-3}. Assuming returns are uncorrelated, the standard error for the mean equals $1.182 \times 10^{-3}/\sqrt{49,001} = 5.340 \times 10^{-6}$ making the mean indistinguishable from zero at standard significance levels.

The intraday seasonality of volatility in financial markets is well documented (see Goodhart and O'Hara [18] and Bollerslev [6]). Market volatility follows roughly a U-shaped pattern in a trading day: typically highest at the open and towards the close of trade but at relatively low levels in the midst of the trade. To capture this characteristic of the intraday data, a variable S is created. For a given trading day which has n five-minute returns, first observation of variable S for that day is $1/n$, second observation is $2/n, \ldots$ and the last observation of variable S for that day is $n/n = 1$. In the next section it is explained how this variable is used.

4.4. The Model

Consider the following GARCH model

$$r_t = \mu + \varepsilon_t \tag{4.1}$$

$$\varepsilon_t \mid \psi_{t-1} \sim \phi(0, h_t, \nu) \tag{4.2}$$

where $r_t = 100 \cdot \log(DJIA_t/DJIA_{t-1})$, ϕ specifies a student-t distribution with zero mean, variance h_t, and ν degrees of freedom. ψ_{t-1} is the information set that the analyst conditions on that is available at the end of period $t-1$, or equivalently, at the beginning of period t. To capture volatility dynamics during the periods of trading collars, following specifications for the conditional variance h_t are proposed and estimated:

$$h_t = (1+\gamma D_t)\omega_t + (1+\gamma D_t)\alpha\varepsilon_{t-1}^2 + (1+\gamma D_t)\beta h_{t-1} \tag{4.3a}$$

$$h_t = (1+\gamma^a D_t^a + \gamma^b D_t^b)\omega_t + (1+\gamma^a D_t^a + \gamma^b D_t^b)\alpha\varepsilon_{t-1}^2 + (1+\gamma^a D_t^a + \gamma^b D_t^b)\beta h_{t-1} \tag{4.3b}$$

$$h_t = (1+\gamma_0 D_t)\omega_t + (1+\gamma_1 D_t)\alpha\varepsilon_{t-1}^2 + (1+\gamma_2 D_t)\beta h_{t-1} \tag{4.3c}$$

$$h_t = (1+\gamma_0^a D_t^a + \gamma_0^b D_t^b)\omega_t + (1+\gamma_1^a D_t^a + \gamma_1^b D_t^b)\alpha\varepsilon_{t-1}^2 + (1+\gamma_2^a D_t^a + \gamma_2^b D_t^b)\beta h_{t-1} \tag{4.3d}$$

where $\omega_t = \exp(a_0 + a_1 S_t + a_2 S_t^2 + \cdots + a_p S_t^p)$. The log-likelihood function is

$$\log L = T \log\left[\frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]\sqrt{\pi(\nu-2)}}\right] - \frac{1}{2}\sum_{t=1}^{T}\log h_t - \left[\frac{\nu+1}{2}\right]\sum_{t=1}^{T}\log\left[1+\frac{\varepsilon_t^2}{h_t(\nu-2)}\right]$$

where $\Gamma[\cdot]$ is the gamma function, ν is the degrees of freedom, and $\varepsilon_t = r_t - \mu$. For the conditional variance h_t, (4.3a)-(4.3d) are used. This log-likelihood can be numerically maximized subject to the constraint $\nu > 2$.

Equations (4.3a)-(4.3d) will be referred to as models A to D respectively. Intraday seasonality of volatility is captured by the pth degree polynomial in the ω_t term and this term will be referred to as the seasonality term. The seasonality term can be made continuous by restricting $\omega_t|_{S_t=0} = \omega_t|_{S_t=1}$, i.e., $\sum_{i=1}^{p} a_i = 0$.[9] Note that models A through D reduce to conventional GARCH models with no intraday seasonality when the seasonality parameters $a_1, \ldots, a_p$ are restricted to zero. Day-of-the-week dummies were included in the preliminary estimations as well in the conditional variance specifications above but they turned out to be insignificant and those results are not reported here. Use of daily data instead of intraday data would probably reveal significant coefficients for those dummies.

Models A and B allow the trading collar to influence conditional variance by allowing GARCH parameters to vary by a constant proportion when the trading collar is in effect. Models C and D are more flexible and nest models A and B respectively. Models C and D also allow conditional variance dynamics to be different during the trading collars but do not restrict GARCH parameters to vary by the same constant proportion. Models A and C treat up and down markets the same while models B and D allow different volatility dynamics during up and down markets. Model D is the most general model and nests the other three models in it while model A is the most restrictive and is nested in the other three.

4.5. Estimation Results

Maximum likelihood estimates of the parameters for the four models are given in table 4.4. In the table, Akaike Information Criteria (AIC) is $2(\log L - k)$ where k is the number of estimated parameters and the numbers in parentheses are the asymptotic t-ratios.

One over the degrees of freedom parameter is significantly different from zero for all four models indicating that the t distribution is indeed a better choice than the normal as the conditional distribution. Estimated values for this parameter are quite close and give a degrees of freedom of around 7.18.

[9] In addition to continuity, the seasonality term can also be made smooth as it cycles from one

Estimated GARCH coefficients α and β are highly significant and similar across the models. Also the sum of these coefficients is close to one indicating persistence of memory which is to be expected in high frequency data.

In model A, γ is positive and significant. An estimated value of 0.02 indicates that market volatility is 2.0% higher during the periods of trading collar. However, model A does not distinguish rising and declining markets. To distinguish the effects of trading collar in rising and declining markets, the results from model B can be looked at. The γ^a coefficient is insignificant indicating that in rising markets, the presence of the collar has no bearing on market volatility. In contrast, γ^b is significant indicating that market volatility is appreciably higher in declining markets. A value of 0.034 for γ^b indicates that market volatility is 3.4% higher when the collar is in effect in declining markets. All three trading collar coefficients $\gamma_0, \gamma_1, \gamma_2$ are significant in model C implying that volatility dynamics are different during trading collar regimes. Results from model D indicate that volatility dynamics are not affected by the presence of trading collar during rising markets. However, volatility dynamics are significantly affected by the presence of the trading collar in declining markets. The Akaike Information Criteria indicates that model D is the preferred model. Thus, empirical results suggest that volatility dynamics are significantly different when trading collar is imposed and that these dynamics are not identical during up and down markets.

In the seasonality term ω_t, a cubic polynomial was found to be satisfactory ($p = 3$ and $a_3 = -a_1 - a_2$). Estimation results for all four models reveal that the incorporation of 3rd degree polynomial indeed captures the intraday seasonality as the polynomial coefficients are highly significant. Also these seasonality parameters are markedly similar across the models indicating that the estimated intraday seasonality polynomial is robust across the four model specifications.

Figure 4.1 is a graphical depiction of the estimated intraday seasonal pattern in

day to the next by imposing $d\omega_t/dS_t\,|_{S_t=0} = d\omega_t/dS_t|_{S_t=1}$, i.e., $\sum_{i=2}^{p} i \cdot a_i = 0$.

volatility for model D. Seasonal patterns for other three models are similar. The figure indicates that the DJIA is about six times more volatile at the open and the end of the day than at midday. The estimated seasonality polynomial supports a U-shaped volatility pattern reported by earlier studies for other equity markets (e.g., Andersen and Bollerslev [1]).

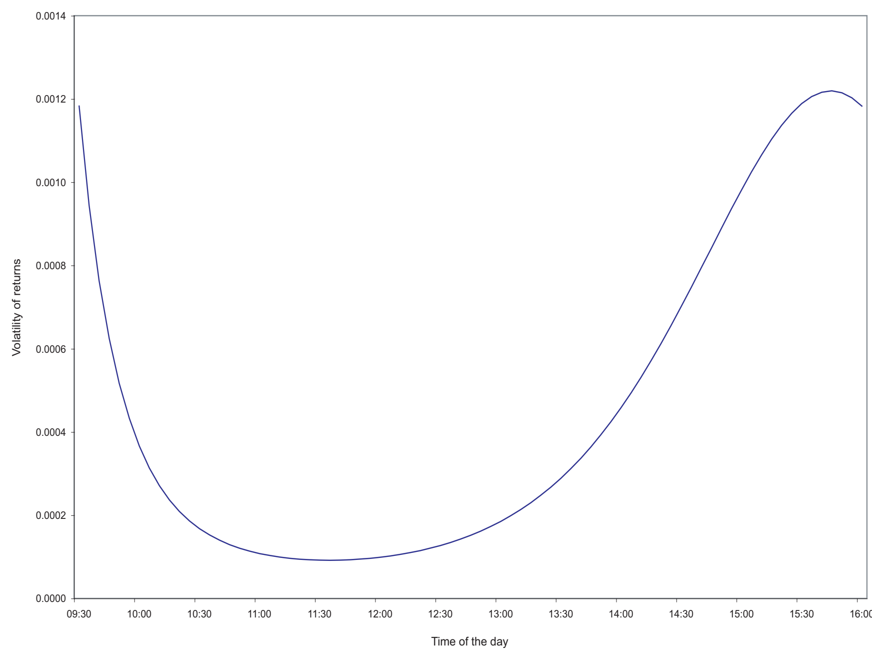

FIGURE 4.1. Estimated Intraday Seasonal Patterns in Volatility of the DJIA

4.6. Conclusions

In this chapter, using intraday data, volatility models for the DJIA are estimated that account for the presence of Rule 80A restrictions on index arbitrage. It is shown that the volatility does not react to shocks in the same magnitude when the rule is in effect as when it is not and when the rule is in effect due to a market increase or a decrease. In doing this, the models are also able to capture the well documented U-shaped pattern of intraday volatility using a polynomial specification.

Specifically, the results indicate that: (i) market volatility is 2% higher when trading collars are in effect; (ii) when it is differentiated between up and down markets, volatility is higher by 3.4% if the collar restrictions are in effect during a down market and during an up market, the collar has no effect on the volatility; (iii) from April 1993, which is the start of the data set, to February 1999, the percentage of observations when the rule was in effect increases steadily. This is due to constant collar values and the bull market of 1990s. Since February 1999, the collar values have been adjusted quarterly and this leads to a sharp decrease in the number of times the collar has been triggered and stayed in effect. In the estimation period, which is from February 16, 1999 to August 31, 2001, 6% of the observations were observed when the rule was in effect and 41% of the time this was due to increases in the Dow and 59% of the time due to decreases.

	Model A	Model B	Model C	Model D
μ	0.00066	0.00065	0.00064	0.00062
	(1.71)	(1.67)	(1.65)	(1.60)
a_0	-6.75	-6.74	-6.82	-6.74
	(-78.65)	(-92.61)	(-81.23)	(-99.75)
a_1	-18.26	-18.11	-18.35	-18.02
	(-13.49)	(-19.57)	(-13.43)	(-22.05)
a_2	38.52	38.12	38.97	37.62
	(10.94)	(15.87)	(11.00)	(17.99)
α	0.0740	0.0732	0.0701	0.0709
	(23.05)	(24.98)	(22.03)	(24.93)
β	0.888	0.888	0.893	0.891
	(187.94)	(232.74)	(199.93)	(252.98)
γ	0.020			
	(5.24)			
γ^a		-0.0016		
		(-0.27)		
γ^b		0.034		
		(7.63)		
γ_0			1.228	
			(2.76)	
γ_1			0.843	
			(3.16)	
γ_2			-0.060	
			(-3.14)	
γ_0^a				0.752
				(3.02)
γ_0^b				1.156
				(2.46)
γ_1^a				-0.013
				(-0.06)
γ_1^b				0.807
				(2.94)
γ_2^a				-0.026
				(-1.35)
γ_2^b				-0.037
				(-2.16)
$1/\nu$	0.1395	0.1395	0.1393	0.1393
	(34.21)	(34.26)	(34.21)	(34.26)
$\log L$	71,169.33	71,182.55	71,192.63	71,205.94
AIC	142,322.66	142,347.10	142,365.26	142,385.88

TABLE 4.4. Estimation Results

5. Dissertation Conclusions

It is expected that the methodological contribution of this dissertation in chapter two and the empirical findings of chapters three and four will benefit future research on semiparametric estimation, insurance company involvement in and efficiency of U.S. crop insurance program, and volatility modeling in equity markets where there are implicit or explicit limits on the movement of prices.

The semiparametric approach in chapter two shows that in binary-choice models where one might have a reliable prior expectation about the shape of the unknown link function, it is quite straightforward to incorporate this information into the estimation process. More importantly, as shown via simulations, one's initial guess about the shape of the function need only be a rough guess, not an exact one, to achieve significant bias reduction and efficiency gain. Hence this parametric start need not be viewed as an additional nuisance parameter whose choice might otherwise require a data-driven, e.g., cross-validation type procedure. Furthermore, the usefulness of this approach is not limited to estimation of coefficients. In fact, this is where the approach has only an indirect effect. As shown in simulations (chapter two) and an application (chapter three), this semiparametric approach can achieve more significant efficiency gains in the estimation of probabilities where the bias reduction idea has direct effect. Future research will involve looking at asymptotic and finite sample properties of this semiparametric approach in estimation of marginal effects, i.e., derivatives of the unknown link function with respect to covariates.

This approach is certainly applicable to single-index models in general as in Ichimura [35], not just binary-choice models as in Klein and Spady [41], however, in the latter, it is easier to come up with an initial guess for the unknown function, e.g., a distribution function.

This new semiparametric approach has been successfully applied to insurance data along with competing parametric and semiparametric estimators in chapter three. This chapter stands almost alone in the literature as the overwhelming majority of the research analyses producer involvement in the U.S. crop insurance program even though the rents obtained by the private insurance companies have increased dramatically over the years. Although more research is needed, the findings of this chapter show that the insurance company involvement in this program may be too costly to justify their participation. Another future research that this study stimulates is looking at efficiency of the crop insurance program in terms of premium rates and rating practices of the federal government. Although preliminary, one of the findings of this chapter is that the premium rates may not be efficient which could have serious impact on financial soundness of the program.

Modeling market volatility has been the subject of many studies especially after the seminal work of Engle [12]. Among the equity markets that have been analyzed, stock and foreign exchange markets have been the subject of overwhelming majority of this literature. Many of these stock and foreign exchange markets have a common feature which is that the equity prices may freely move only in a certain band and their move outside of that band may be restricted or even prohibited. The modeling approach in chapter four explicitly takes a similar feature of the NYSE into account which is ignored by almost all studies. It is shown in this chapter that this can be done easily by some data manipulation and can be incorporated into popular volatility models. Future research will look at commodity future prices where the "price limits" prohibit movement of prices outside the limits as opposed to NYSE Rule 80A collar which restricts price movements. Another contribution of this chapter is how the well documented intraday seasonality of market volatility is captured by a simple polynomial specification as opposed to the two-step Fourier transform procedure. A comparison of these two procedures, in terms of estimation simplicity and performance, may have important empirical relevance and will be the subject of future research.

A. Appendix

Proof of lemma 1

Our proof of lemma 1 follows closely Bierens [4], Bierens [5], and Pagan and Ullah [52, pp.36-39]. Note that $\hat{F}$ in (2.10) can be written as $\hat{F}_1/\hat{F}_2$ where

$$\hat{F}_1 = \frac{1}{(n-1)h}\sum_{j\neq i} y_j \mathbf{1}_{[x_j\in A_{nx}]}\left\{\frac{G(x_ib)}{G(x_jb)}\right\}K\left(\frac{x_ib-x_jb}{h}\right)$$
$$\hat{F}_2 = \frac{1}{(n-1)h}\sum_{j\neq i} \mathbf{1}_{[x_j\in A_{nx}]}K\left(\frac{x_ib-x_jb}{h}\right).$$

Since $\hat{F}_2$ is $\hat{F}_1$ with $y_j = 1$ and $G(\cdot)$ a constant function, we will only show uniform convergence of $\hat{F}_1$.[1] Now observe that

$$E(y|x) = F(x) = \frac{G(x)\int y\frac{1}{G(x)}f(y,x)dy}{\int f(y,x)dy}.$$

Let $g(x) = \int y\frac{1}{G(x)}f(y,x)dy/\int f(y,x)dy$ and $h(x) = \int f(y,x)dy$. So $F(x) = G(x)g(x)$ and $G(x)g(x)h(x) = G(x)\int y\frac{1}{G(x)}f(y,x)dy$. Thus $\hat{F}_1 = G(x)\hat{g}(x)\hat{h}(x)$. Notice that by assumption 5, G is uniformly bounded and we have $\sup_x |G(x)| = O(1)$. In fact a plausible start is a distribution function in which case $\sup_x |G(x)| = 1$. Hence it

[1] The derivation below assumes that (y_i, x_i) is absolutely continuous. Obviously in binary-choice models this is not true as y_i is a Bernoulli random variable. We will keep the absolute continuity interpretation as it is more general and give the necessary changes here for the binary-response case. Using a notation similar to Klein and Spady [41], let g_x be the unconditional density for x and $g_{x|y}$ be the density for x conditional on y for $y = 0, 1$. We have the following series of equalities

$$E(y|x) = F(x) = Pr(y=1|x) = \frac{Pr(y=1)g_{x|1}}{g_x} = \frac{g_{1x}}{g_x} = G(x)\frac{g_{1x}/G(x)}{g_x} = G(x)g(x)$$

where $g(x) = (g_{1x}/G(x))/g_x$. Thus $\hat{F}_1 = G(x)\hat{g}(x)\hat{g}_x$ where $\hat{g}(x)\hat{g}_x = ((n-1)h)^{-1}\sum_{j\neq i}\frac{y_j}{G(x_j)}K((x-x_j)/h)$. So there is no change from (A.1) to (A.8). In equation (A.8), we can take an iterated expectation to get $E_X\left[1/G(x_j)K((z-x_j)/h)Pr(y_j=1|x)\right] = E_X\left[K((z-x_j)/h)g(x)\right] = \int K((z-x)/h)g(x)g_x dx$. And now $1/h$ times this last term would replace equation (A.9) and we can apply Taylor expansion to $\psi(x) = g(x)g_x$.

suffices to show $\sup_x |\hat{g}(x)\hat{h}(x) - g(x)h(x)| \to 0$. So we have

$$\sup_x |\hat{g}(x)\hat{h}(x) - g(x)h(x)| \leqslant \sup_x |\hat{g}(x)\hat{h}(x) - E\hat{g}(x)\hat{h}(x)| + \sup_x |E\hat{g}(x)\hat{h}(x) - g(x)h(x)|. \tag{A.1}$$

Like Ichimura [35], we will refer to the second term of the right-hand side as bias term and show that it converges to 0 at the rate h^2. But first notice that

$$\hat{g}(x)\hat{h}(x) = \frac{1}{nh}\sum_{j=1}^{n} \frac{y_j}{G(x_j)} K\left(\frac{x - x_j}{h}\right). \tag{A.2}$$

From the inversion formula (see Fristedt and Gray [14, p.231]) and by assumption 6 we have

$$K(a) = \frac{1}{2\pi}\int \exp(-ita)\phi(t)dt \tag{A.3}$$

where $\phi(t)$ is the characteristic function of K and $i^2 = -1$. Using (A.2) and (A.3) and letting $s = t/h$ we get

$$\hat{g}(x)\hat{h}(x) = \frac{1}{2\pi}\int \left\{\frac{1}{n}\sum_{j=1}^{n} \frac{y_j}{G(x_j)} \exp(itx_j)\right\} \exp(-itx)\phi(ht)dt. \tag{A.4}$$

From (A.4) we get

$$E\hat{g}(x)\hat{h}(x) = \frac{1}{2\pi}\int E\left[\frac{y_j}{G(x_j)} \exp(itx_j)\right] \exp(-itx)\phi(ht)dt. \tag{A.5}$$

From (A.4) and (A.5) and noting that $|\exp(-itx)| = 1$

$$\begin{aligned}|\hat{g}(x)\hat{h}(x) - E\hat{g}(x)\hat{h}(x)| \leqslant &\frac{1}{2\pi}\int \left|\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j}{G(x_j)}\exp(itx_j)\right.\right.\\ &\left.\left. -E\left[\frac{y_j}{G(x_j)}\exp(itx_j)\right]\right\}\right| |\phi(ht)|dt.\end{aligned}$$

So

$$\begin{aligned}\sup_x |\hat{g}(x)\hat{h}(x) - E\hat{g}(x)\hat{h}(x)| \leqslant &\frac{1}{2\pi}\int \left|\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j}{G(x_j)}\exp(itx_j)\right.\right.\\ &\left.\left. -E\left[\frac{y_j}{G(x_j)}\exp(itx_j)\right]\right\}\right| |\phi(ht)|dt.\end{aligned}$$

Using $\exp(itx_j) = \cos(tx_j) + i\sin(txj)$ we can write

$$
\begin{aligned}
&E\left|\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j}{G(x_j)}\exp(itx_j) - E\left[\frac{y_j}{G(x_j)}\exp(itx_j)\right]\right\}\right| \\
=&E\Bigg|\underbrace{\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j}{G(x_j)}\cos tx_j - E\left[\frac{y_j}{G(x_j)}\cos tx_j\right]\right\}}_{A} \\
&+ i\underbrace{\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j}{G(x_j)}\sin tx_j - E\left[\frac{y_j}{G(x_j)}\sin tx_j\right]\right\}}_{B}\Bigg|
\end{aligned}
\tag{A.6}
$$

Note that we can write $|A+iB| = (A^2+B^2)^{1/2}$ and so $E|A+iB| = E(A^2+B^2)^{1/2} \leqslant (EA^2+EB^2)^{1/2} = (Var(A)+Var(B))^{1/2}$ where the inequality comes from Jensen's inequality and by construction $EA = 0$ and $EB = 0$. So (A.6) is

$$
\begin{aligned}
&\leqslant \Bigg\{Var\left[\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j}{G(x_j)}\cos tx_j - E\left[\frac{y_j}{G(x_j)}\cos tx_j\right]\right\}\right] \\
&\qquad +Var\left[\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j}{G(x_j)}\sin tx_j - E\left[\frac{y_j}{G(x_j)}\sin tx_j\right]\right\}\right]\Bigg\}^{1/2} \\
&= \left\{Var\left[\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j}{G(x_j)}\cos tx_j\right\}\right] + Var\left[\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{y_j}{G(x_j)}\sin tx_j\right\}\right]\right\}^{1/2} \\
&= \left\{\frac{1}{n}\left(Var\left[\frac{y_j}{G(x_j)}\cos tx_j\right] + Var\left[\frac{y_j}{G(x_j)}\sin tx_j\right]\right)\right\}^{1/2}.
\end{aligned}
$$

Note that $VarX \leqslant EX^2$ so

$$
\begin{aligned}
&\leqslant \left\{\frac{1}{n}\left(E\left[\left(\frac{y_j}{G(x_j)}\right)^2\cos^2 tx_j\right] + E\left[\left(\frac{y_j}{G(x_j)}\right)^2\sin^2 tx_j\right]\right)\right\}^{1/2} \\
&= \frac{1}{\sqrt{n}}\left\{E\left[\left(\frac{y_j}{G(x_j)}\right)^2\right]\right\}^{1/2}
\end{aligned}
$$

noting that $\cos^2 tx_j + \sin^2 tx_j = 1$. So

$$E \sup_x |\hat{g}(x)\hat{h}(x) - E\hat{g}(x)\hat{h}(x)| \leqslant \frac{1}{2\pi}\frac{1}{\sqrt{n}} \left\{ E\left[\left(\frac{y_j}{G(x_j)}\right)^2\right]\right\}^{1/2} \int |\phi(ht)|dt$$
$$= \frac{1}{2\pi}\frac{1}{h\sqrt{n}} \left\{ E\left[\left(\frac{y_j}{G(x_j)}\right)^2\right]\right\}^{1/2} \int |\phi(s)|ds \quad \text{(A.7)}$$

after a change of variables ($s = ht$) and the last term goes to zero as $h\sqrt{n} \to \infty$. Finally using Markov's inequality with (A.7) we get

$$Pr\left(\sup_x |\hat{g}(x)\hat{h}(x) - E\hat{g}(x)\hat{h}(x)| > \epsilon\right) \to 0 \quad \text{as} \quad n \to \infty.$$

Now for $|E\hat{g}(x)\hat{h}(x) - g(x)h(x)|$ note that

$$E\hat{g}(z)\hat{h}(z) = E\left[\frac{1}{nh}\sum_{j=1}^{n} y_j \frac{1}{G(x_j)} K\left(\frac{z-x_j}{h}\right)\right]$$
$$= \frac{1}{h}E\left[y_j \frac{1}{G(x_j)} K\left(\frac{z-x_j}{h}\right)\right] \quad \text{(A.8)}$$
$$= \frac{1}{h}\int K\left(\frac{z-x}{h}\right) \underbrace{\int y \frac{1}{G(x)} f(y,x)dy}_{g(x)h(x)}\, dx. \quad \text{(A.9)}$$

Now let $\psi(x) = g(x)h(x)$ and $s = (z-x)/h$ for the Taylor expansion

$$\psi(x) = \psi(z - sh) = \psi(z) - hs\psi'(z) + \frac{1}{2}h^2 s^2 \psi''(z) + o(h^2).$$

So

$$\frac{1}{h}\int \left(\psi(z) - hs\psi'(z) + \frac{1}{2}h^2 s^2 \psi''(z)\right) K(s)hds = \psi(z) + \frac{1}{2}h^2\psi''(z)\int s^2 K(s)ds.$$

Thus we can write

$$\sup_x \left|E\hat{\psi}(x) - \psi(x)\right| = \sup_x \left|\psi(x) + \frac{1}{2}h^2\psi''(x)\int s^2 K(s)ds - \psi(x)\right|$$
$$\leqslant \frac{1}{2}h^2 \sup_x |\psi''(x)| \int |s^2 K(s)|ds$$

so the last term goes to zero at the rate h^2. This completes the proof of lemma 1.

Proof of theorem 1

Note that

$$\sup_b |\hat{Q}_n(b) - Q_0(b)| \leqslant \sup_b |\hat{Q}_n(b) - Q_n(b)| + \sup_b |Q_n(b) - Q_0(b)|$$

where

$$\hat{Q}_n(b) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[x_i \in A_x]}(y_i \log[\hat{F}(x_i b)] + (1-y_i)\log[1-\hat{F}(x_i b)]),$$
$$Q_n(b) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[x_i \in A_x]}(y_i \log[F(x_i b)] + (1-y_i)\log[1-F(x_i b)]),$$
$$Q_0(b) = \frac{1}{n}\sum_{i=1}^{n} E\left[\mathbf{1}_{[x_i \in A_x]}(y_i \log[F(x_i b)] + (1-y_i)\log[1-F(x_i b)])\right].$$

Let $\hat{Q}_{1n} = n^{-1}\sum_{i=1}^{n} \mathbf{1}_{[x_i \in A_x]} y_i \log[\hat{F}(x_i b)]$ and similarly for Q_{1n}. Let $\hat{F}_i \equiv \hat{F}(x_i b)$ and similarly for F_i. $\hat{Q}_{1n}$ can be viewed as a function of $\hat{F}_i$, so from a functional mean-value expansion of $\hat{Q}_{1n}$ about F_i we get

$$\begin{aligned}|\hat{Q}_{1n} - Q_{1n}| = &\left|\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[x_i \in A_x]} y_i \log[F_i] + \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[x_i \in A_x]} y_i \frac{1}{\tilde{F}_i}(\hat{F}_i - F_i)\right. \\ &\left. - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[x_i \in A_x]} y_i \log[F_i]\right| \\ \leqslant &\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[x_i \in A_x]} y_i \frac{1}{\tilde{F}_i}|\hat{F}_i - F_i|\end{aligned}$$

where $\tilde{F}_i$ is between $\hat{F}_i$ and F_i. So we have

$$\sup_b |\hat{Q}_{1n} - Q_{1n}| \leqslant \sup_{i,b} |q_i| \frac{1}{n}\sum_{i=1}^{n} \sup_b |\hat{F}_i - F_i|$$

where $q_i = \mathbf{1}_{[x_i \in A_x]} y_i / \tilde{F}_i$. Note that $\sup_{i,b}|q_i| = O_p(1)$ and from lemma 1 $\sup_b |\hat{F}_i - F_i| \to 0$ in probability so $\sup_b |\hat{Q}_{1n} - Q_{1n}| = o_p(1)$. A similar result can be obtained for $y_i = 0$ part of the likelihood. Thus we have $\sup_b |\hat{Q}_n(b) - Q_n(b)| = o_p(1)$.

Now let $q_i(x_i, y_i, b) = \mathbf{1}_{[x_i \in A_x]}(y_i \log[F(x_i b)] + (1 - y_i) \log[1 - F(x_i b)])$. So

$$\sup_b |Q_n(b) - Q_0(b)| = \sup_b \left| \frac{1}{n} \sum_{i=1}^{n} (q_i(x_i, y_i, b) - E[q_i(x_i, y_i, b)]) \right|. \tag{A.10}$$

As Ichimura [35, p.91], we can use the uniform law of large numbers by Andrews [2] and so (A.10) goes to 0 in probability. This completes the proof of theorem 1.

References

[1] Andersen, T.G., and T. Bollerslev (1997): “Intraday periodicity and volatility persistence in financial markets,” *Journal of Empirical Finance*, 4, 115-158.

[2] Andrews, D.W.K. (1987): “Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers,” *Econometrica*, 55, 1465-1471.

[3] Baillie, R.T., and T. Bollerslev (2001): “Intraday and inter market volatility in foreign exchange rates,” *Review of Economic Studies*, 58, 565-585.

[4] Bierens, H.J. (1987): “Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions,” *Journal of the American Statistical Association*, 78, 699-707.

[5] Bierens, H.J. (1987): “Kernel Estimators of Regression Functions,” in *Advances in Econometrics, Vol. 1*, ed by. T.F. Bewley. Cambridge: Cambridge University Press, 99-144.

[6] Bollerslev, T. (2001): “Financial Econometrics: Past developments and future challenges,” *Journal of Econometrics*, 100, 41-51.

[7] Booth, R.A. (1992): “In Defense of Program Trading,” *Wall Street Journal*, April 1.

[8] Booth, G.G., and J.P. Broussard (1998): “Setting NYSE Circuit Breaker Triggers,” *Journal of Financial Services Research*, 13, 187-204.

[9] Chen, S. (2000): “Efficient estimation of binary choice models under symmetry,” *Journal of Econometrics*, 96, 183-199.

[10] Cosslett, S.R. (1987): “Efficiency Bounds for Distribution-Free Estimators of the Binary Choice and Censored Regression Models,” *Econometrica*, 55, 559-586.

[11] Ederington, L., and J.H. Lee (2001): “Intraday volatility in interest-rate and foreign-exchange markets: ARCH, announcement, and seasonality effects,” *Journal of Futures Markets*, 21, 517-552.

[12] Engle, R.F. (1982): “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, 50, 987-1007.

[13] Fernandez, A.I., and J.M. Rodriguez-Poo (1997): “Estimation and Specification Testing in Female Labor Participation Models: Parametric and Semiparametric Methods,” *Econometric Reviews*, 16, 229-247.

[14] Fristedt, B., and L. Gray (1997): *A Modern Approach to Probability Theory.* Birkhäuser.

[15] Gabler, S., F. Laisney, and M. Lechner (1993): "Semi-nonparametric maximum likelihood estimation of binary choice models with an application to labour force participation," *Journal of Business and Economic Statistics*, 11, 61-80.

[16] Gerfin, M. (1996): "Parametric and Semiparametric Estimation of the Binary Response Model of Labour Market Participation," *Journal of Applied Econometrics*, 11, 321-339.

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

[18] Goodhart, C.A.E., and M. O'Hara (1997): "High frequency data in financial markets: Issues and applications," *Journal of Empirical Finance*, 4, 73-114.

[19] Goodwin, B.K., and M.T. Holt (2002): "Parametric and Semiparametric Modeling of the Off-Farm Labor Supply of Agrarian Households in Transition Bulgaria," *American Journal of Agricultural Economics*, 84, 184-209.

[20] Gozalo, P., and O. Linton (2000): "Local Nonlinear Least Squares: Using Parametric Information In Nonparametric Regression," *Journal of Econometrics*, 99, 63-106.

[21] Greenwald, B., and J. Stein (1988): "The Task Force Report: The Reasoning Behind the Recommendations," *Journal of Economic Perspectives*, 2, 3-23.

[22] Härdle, W. (1990): *Applied Nonparametric Regression.* Cambridge University Press.

[23] Härdle, W., E. Mammen, and I. Proença (2000): "A Bootstrap Test for Single Index Models," unpublished manuscript.

[24] Härdle, W., and J.S. Marron (1985): "Asymptotic nonequivalence of some bandwidth selectors in nonparametric regression," *Biometrika*, 72, 481-484.

[25] Härdle, W., P. Hall, and H. Ichimura (1993): "Optimal Smoothing in Single-Index Models," *Annals of Statistics*, 21, 157-178.

[26] Hjort, N.L., and M.C. Jones (1996): "Locally parametric nonparametric density estimation," *Annals of Statistics*, 24, 1619-1647.

[27] Hjort, N.L., and I.K. Glad (1995): "'Nonparametric Density Estimation With a Parametric Start," *Annals of Statistics*, 23, 882-904.

[28] Hogan, K.C., K.F. Kroner, and J. Sultan (1997): "Program Trading, Nonprogram Trading, and Market Volatility," *Journal of Futures Markets*, 17, 733-756.

[29] Horowitz, J.L. (1992): "A Smoothed Maximum Score Estimator For the Binary Response Model," *Econometrica*, 60, 505-531.

[30] Horowitz, J.L. (1993): "Semiparametric and Nonparametric Estimation of Quantal Response Models," in *Handbook of Statistics*, Vol. 11, ed. by G.S. Maddala, C.R. Rao, and H.D. Vinod. North-Holland.

[31] Horowitz, J.L. (1993): "Semiparametric estimation of a work-trip mode choice model," *Journal of Econometrics*, 58, 49-70.

[32] Horowitz, J.L. (1998): *Semiparametric Methods in Econometrics*. Springer-Verlag.

[33] Horowitz, J.L., and W. Härdle (1994): "Testing a Parametric Model Against a Semiparametric Alternative," *Econometric Theory*, 10, 821-848.

[34] Horowitz, J.L., and W. Härdle (1996): "Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates," *Journal of the American Statistical Association*, 91, 1632-1640.

[35] Ichimura, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71-120.

[36] Ichimura, H., and L.F. Lee (1991): "Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. A. Barnett, J. Powell, and G. Tauchen. Cambridge University Press.

[37] Jones, M.C., and D.F. Signorini (1997): "A Comparison of Higher-Order Bias Kernel Density Estimators," *Journal of the American Statistical Association*, 92, 1063-1073.

[38] Ker, A. (1999): "Private Insurance Companies and the U.S. Crop Insurance Program," *Choices*.

[39] Ker, A. (2001): "Private Insurance Company Involvement in the U.S. Crop Insurance Program," *Canadian Journal of Agricultural Economics*,

[40] Ker, A., and P. McGowan (2000): "Weather-Based Adverse Selection and the U.S. Crop Insurance Program: The Private Insurance Company Perspective," *Journal of Agricultural and Resource Economics*, 25, 386-410.

[41] Klein, R.W., and R.H. Spady (1993): "An Efficient Semiparametric Estimator For Binary Response Models," *Econometrica*, 61, 387-421.

[42] Kuhn, B.A., G.J. Kuserk, and P. Locke (1991): "Do Circuit Breakers Moderate Volatility? Evidence from October 1989," *Review of Futures Markets*, 10, 136-175.

[43] Kuserk, G.J., P.R. Locke, and C.L. Sayers (1992): "The Effects of Amendments to Rule 80A on Liquidity, Volatility, and Price Efficiency in the S&P500 Futures," *Journal of Futures Markets*, 12, 383-409.

[44] Lindsey, R.R., and A.P. Pecora (1998): "Ten Years After: Regulatory Developments in the Securities Markets Since the 1987 Market Break," *Journal of Financial Services Research*, 13, 283-314.

[45] Manski, C.F. (1975): "The Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205-228.

[46] Manski, C.F. (1988): "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, 729-738.

[47] Martens, M. (2001): "Forecasting daily exchange rate volatility using intraday returns," *Journal of International Money and Finance*, 20, 1-23.

[48] Miranda, M., and J. Glauber (1997): "Systematic Risk, Reinsurance, and the Failure of Crop Insurance Markets," *American Journal of Agricultural Economics*, 79, 206-215.

[49] Murphy, K.M., and R.H. Topel (1985): "Estimation and inference in two-step econometric models," *Journal of Business and Economic Statistics*, 3, 370-379.

[50] NYSE Glossary available at http://www.nyse.com.

[51] Newey, W.K., J.L. Powell, and J. R. Walker (1990): "Semiparametric Estimation of Selection Models: Some Empirical Results," *American Economic Review*, 80, 324-328.

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

[53] Powell, J.L. (1994): "Estimation of Semiparametric Models," in *Handbook of Econometrics*, Vol. 4, ed. by R.F. Engle and D.L. McFadden. North-Holland.

[54] Powell, J.L., J.H. Stock, and T.M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 474-523.

[55] Power, W. (1994): "A Chorus at Big Board of 94: Another Day, Another Collar," *Wall Street Journal*, September 22.

[56] Reibstein, L., and C. Friday (1989): "The Villain in the Volatility," *Newsweek*, November 6.

[57] Ruud, P.A. (1983): "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models," *Econometrica*, 51, 225-228.

[58] Santoni, G.J., and T. Liu (1993): "Circuit Breakers and Stock Market Volatility," *Journal of Futures Markets*, 13, 261-277.

[59] Schucany, W.R., and J.P. Sommers (1977): "Improvements of Kernel Type Density Estimators," *Journal of the American Statistical Association*, 72, 420-423.

[60] Schuster, E., and S. Yakowitz (1979): "Contributions to the Theory of Nonparametric Regression With Application to System Identification," *Annals of Statistics*, 7, 139-149.

[61] Skees, J.R. (1999): "Agricultural Risk Management or Income Enhancement?," *Regulation*.

[62] Skees, J.R. (2001): "The Bad Harvest," *Regulation*.

[63] Stoker, T.M. (1986): "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461-1481.

[64] Wall Street Journal (1994): "Big Board Reinstates Program Trade Data Relating to Volatility," March 15.

[65] Wand, M.P., and M.C. Jones (1995): *Kernel Smoothing.* Chapman & Hall.

[66] White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

[67] Wright, B.D., and J.A. Hewitt (1994): "All-Risk Crop Insurance: Lessons From Theory and Experience," in *Economics of Agricultural Crop Insurance: Theory and Evidents*, ed. by D.L. Hueth and W.H. Furtan. Boston: Kluwer Academic Publishers.

[68] Yatchew, A., and Z. Griliches (1985): "Specification Error in Probit Models," *Review of Economics and Statistics*, 67, 134-139.