AN ABSTRACT OF THE

DISSERTATION OF

Feng Yao for the degree of Doctor of Philosophy in Economics presented on April 23, 2004.

Title: Three Essays on Nonparametric and Semiparametric Regression Models.

Abstract Approved: Redacted for privacy

Carlos Martins-Filho

This dissertation contains three essays on nonprametric and semiparametric regression models.

In the first essay, we propose an estimation procedure for *value at risk* (VaR) and *expected shortfall* (TailVaR) for conditional distributions of a time series of returns on a financial asset. Our approach combines a local polynomial estimator of conditional mean and volatility functions in a conditional heterocedastic autoregressive nonlinear (CHARN) model with Extreme Value Theory for estimating quantiles of the conditional distribution. We investigate the finite sample properties of our method and contrast them with alternatives, including the method recently proposed by McNeil and Frey(2000), in an extensive Monte Carlo study. The method we propose outperforms the estimators currently available in the literature.

In the second essay, we propose a nonparametric regression frontier model that assumes no specific parametric family of densities for the unobserved stochastic component that represents efficiency in the model. Nonparametric estimation of the regression frontier is obtained using a local linear estimator that is shown to be consistent and $\sqrt{nh_n}$ asymptotically normal under standard assumptions. The estimator we propose envelops the data but is not inherently

biased as Free Disposal Hull - FDH or Data Envelopment Analysis - DEA estimators. It is also more robust to extreme values than the aforementioned estimators. A Monte Carlo study is performed to provide preliminary evidence on the estimator's finite sample properties and to compare its performance to a bias corrected FDH estimator.

In the third essay, we establish the $\sqrt{n}$ asymptotic equivalence of V and U statistics when the statistic kernel depends on n. Combined with a lemma of Lee (1988) this result provides conditions under which U statistics projections (Hoeffding, 1961) and V statistics are $\sqrt{n}$ asymptotically equivalent. The use of this equivalence in nonparametric regression models is illustrated with two examples. The estimation of conditional variances and construction of nonparametric R-square.

Three Essays on Nonparametric and Semiparametric Regression Models

by
Feng Yao

A Dissertation

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented April 23, 2004
Commencement June 2004

Doctor of Philosophy Dissertation of Feng Yao
presented on April 23, 2004

APPROVED:

Redacted for privacy

Major Professor, representing Economics

Redacted for privacy

Chair of the Department of Economics

Redacted for privacy

Dean of Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for privacy

Feng Yao, Author

ACKNOWLEDGEMENTS

The author expresses profound indebtedness and genuine appreciation to Dr. Carlos Martins-Filho, who guided me through the whole process of my dissertation at Oregon State University. It is my honor and luck to have him as my advisor and I would like to thank him for bringing me into the door of econometrics, giving me rigorous training in Econometrics, inspiring me to think further and deeper over the topics, supporting me to work consistently, and letting me experience the excitement in Econometrics. His tremendous effort is crucial to my completion of this dissertation.

I extend my sincere gratitude to Dr. Shawna Grosskopf, Dr. Victor J. Tremblay, Dr. David S. Birkes and Dr. James R. Lundy for their kind guidance, truthful care and invaluable advice. I am also grateful to the faculty, staff and graduate students in Economics Department for their enduring support and hearty friendship, which make my study in OSU an intellectually and personally rewarding journey.

Finally I am grateful to my wife, Yin Chong, my parents, and my wife's parents, who have always been with me along the way. Their endless love and encouragement wiil always be remembered.

CONTRIBUTION of AUTHORS

Dr. Carlos Martins-Filho co-authored Chapter 2, 3 and 4.

TABLE OF CONTENTS

TABLE OF CONTENTS(Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES(Continued)

Essays on Nonparametric and Semiparametric Regression Models

Chapter 1

Introduction

Nonparametric and semiparametric regression methods offer insight to researchers as they are less dependent on the functional form assumptions when providing estimators and inference. As a consequence, they are adaptive to many unknown features of the data. They fit nicely as complement to the traditional parametric regression methods in exploring the underlying data pattern, delivering evidence on the choice of functional form through test(See Yatchew 1998). In this dissertation, three essays are provided in the nonparametric and semiparametric regression models and below is an overview.

Chapter 2 proposes a two stage estimation procedure for *value at risk* (VAR) and *expected shortfall* (TailVaR) for the conditional distributions of a time series of returns on a financial asset. The return is assumed to come from a stochastic process consisting of conditional mean, volatility and residual. In the first stage, the nonparametric local polynomial regression estimator is used to estimate the conditional mean and volatility, which allows for nonlinearity and asymmetry in the specification. Extreme Value Theory and L-moments techniques are utilized in the second stage on the normalized return by first stage results to estimate the quantiles of the conditional distribution. The VAR and TailVaR are recovered by the two stage estimators. Since the asymptotic properties of the VAR and TailVaR estimators are not available at present, we investigate the finite sample performances of our estimator by an extensive Monte Carlos study and compare it with other alternatives. Our method outperforms the other estimators currently available in the literature.

In chapter 3, we propose a nonparametric regression frontier model that assumes no specific parametric family of densities for the unobserved stochastic component that represents efficiency in the model. Nonparametric estimation of the regression frontier is obtained using a local linear estimator that is shown to be consistent and $\sqrt{nh_n}$ asymptotically normal under standard assumptions. The estimator we propose envelops the data but is not inherently biased as Free Disposal Hull - FDH or Data Envelopment Analysis - DEA estimators. It is also more robust to extreme values than the aforementioned estimators. A Monte Carlo study is performed to provide preliminary evidence on the estimator's finite sample properties and to compare its performance to a bias corrected FDH estimator.

When dealing with kernel based estimators in nonparametric and semiparametric regression models, we frequently encounter k-dimensional sums, which is basically a V-statistics. In chapter 4, we establish the $\sqrt{n}$ asymptotic equivalence of V and U statistics when the statistic kernel depends on n. Combined with a lemma of Lee (1988) this result provides conditions under which U statistics projections (Hoeffding, 1961) and V statistics are $\sqrt{n}$ asymptotically equivalent. So we have a handy tool to analyze the k-dimensional sums by U statistics projections. The use of this equivalence in nonparametric regression models is illustrated with two examples. The estimation of conditional variances and construction of nonparametric R-square. It is clear that our result can be used in a much broader context.

Chapter 2

Estimation of Value-at-risk and Expected Shortfall Based on Nonlinear Models of Return Dynamics and Extreme Value Theory

2.1 Introduction

The measurement of market risk to which financial institutions are exposed has become an important instrument for market regulators, portfolio managers and for internal risk control. As evidence of this growing importance, the Bank of International Settlements (Basel Committee,1996) has imposed capital adequacy requirements on financial institutions that are based on measurements of market risk. Furthermore, there has been a proliferation of risk measurement tools and methodologies in financial markets (Risk,1999). Two quantitative and synthetic measures of market risk have emerged in the financial literature, Value-at-Risk or VaR (RiskMetrics,1995) and Expected Shortfall or TailVaR (Artzner et al.,1999). From a statistical perspective these risk measures have straightforward definitions. Let $\{Y_t\}$ be a stochastic process representing returns on a given portfolio, stock, bond or market index, where t indexes a discrete measure of time and F_t denotes either the marginal or the conditional distribution (normally conditioned on the lag history $\{Y_{t-k}\}_{M \geq k \geq 1}$, for some $M = 1, 2, ...$) of Y_t. For $0 < \alpha < 1$, the α-VaR of Y_t is simply the α-quantile associated with F_t. [1] Expected shortfall is defined as $E_{F_t^y}(Y_t)$ where the expectation is taken with respect to F_t^y, the truncated distribution associated with $Y_t > y$ where y is a specified threshold level. When the threshold y is taken to be α-VaR, then we refer to α-TailVaR.

Accurate estimation of VaR and TailVaR depends crucially on the ability to estimate the tails of the probability density function f_t associated with F_t. Conceptually, this can be ac-

[1] We will assume throughout this paper that F_t is absolutely continuous.

complished in two distinct ways: a) direct estimation of f_t, or b) indirectly through a suitably defined (parametric) model for the tails of f_t. Unless estimation is based on a correct specification of f_t (up to a finite set of parameters), direct estimation will most likely provide a poor fit for its tails, since most observed data will likely take values away from the tail region of f_t (Diebold et al., 1998). As a result, a series of indirect estimation methods based on Extreme Value Theory (EVT) has recently emerged, including Embrechts et al.(1999), Longin(2000) and McNeil and Frey(2000). These indirect methods are based on approximating *only* the tails of f_t by an appropriately defined parametric density function.

In the case where f_t is a conditional density associated with a stochastic process of returns on a financial asset, a particularly promising approach is the two stage estimation procedure for conditional VaR and TailVaR suggested by McNeil and Frey. They envision a stochastic process whose evolution can be described as,

$$Y_t = \mu_t + \sigma_t \epsilon_t \text{ for } t = 1, 2, ..., \tag{1}$$

where μ_t is the conditional mean, σ_t is the square root of the conditional variance (volatility) and $\{\epsilon_t\}$ is an independent, identically distributed process with mean zero, variance one and marginal distribution F_ϵ. Based on a sample $\{y_t\}_{t=1}^n$, the first stage of the estimation produces $\hat{\mu}_t$ and $\hat{\sigma}_t$. In the second stage, $e_t = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t}$ for $t = 1, ..., n$ are used to estimate a generalized pareto density approximation for the tails of f_t, which in turn produce VaR and TailVaR sequences for the conditional distribution of Y_t. Backtesting of their method (using various financial return series) against some widely used *direct* estimation methods that assume a specific form for the distribution of ϵ_t (gaussian, student-t) have produced favorable, albeit specific results. Although encouraging, the backtesting results are specific to the series analyzed and uninformative regarding the statistical properties of the two stage estimators for VaR and

TailVaR. Furthermore, since various first and second stage estimators can be proposed, the important question on how to best implement the two stage procedure remains unexplored.

In this paper we make two contributions to the growing literature on VaR and TailVaR estimation. First, we note that the first stage of the estimation can play a crucial role on conditional VaR and TailVaR estimation on the second stage. In particular, if the assumptions on the (parametric) structure of μ_t, σ_t and ϵ_t are not sufficiently general, there is a distinct possibility that the resulting sequence of residuals will be inadequate for the EVT based approach that follows in the second stage. Given that popular parametric models of conditional mean and volatility of financial returns (GARCH,ARCH and their many relatives) can be rather restrictive, specifically with regards to volatility asymmetry, we consider a nonparametric markov chain model (Härdle and Tsybakov,1997 and Hafner,1998) for $\{Y_t\}$ dynamics, as well as an improved nonparametric estimation procedure for the conditional volatility of the returns (Fan and Yao,1998). The objective is to have in place a first stage estimator that is general enough to accommodate nonlinearities that have been regularly verified in empirical work (Andreou et al.,2001, Hafner,1998, Patton,2001 and Tauchen,2001). We also propose an alternative estimator for the EVT inspired parametric tail model in the second stage. The estimation we propose, which derives from L-Moment Theory (Hosking,1987), is easier and faster to implement, and in finite samples outperforms, the constrained maximum likelihood estimation methods that have prevailed in the empirical finance literature.

The second contribution we make to this literature is in the form of a Monte Carlo study. As noted previously, the statistical properties of the VaR and TailVaR estimators that result from the two stage procedure discussed above are unknown both in finite samples and asymptotically. In addition, since it is possible to envision various alternative estimators for the first and second stages of the procedure, a number of final estimators of VaR and TailVaR emerge.

To assess their properties as well as to shed some light on their relative performance we design an extensive Monte Carlo study. Our study considers various data generating processes (DGPs) that mimic the empirical regularities of financial time series, including asymmetric conditional volatility, leptokurdicity, infinite past memory and asymmetry of conditional return distributions. The ultimate goal here is to provide empirical researchers with some guidance on how to choose between a set of VaR and TailVaR estimators. Our simulations indicate that the estimation strategy we propose outperforms, as measured by the estimators' mean squared error, the method proposed by McNeil and Frey. Besides this introduction, the paper has four additional sections. Section 2 discusses the stochastic model and proposed estimation. Section 3 describes the Monte Carlo design and section 4 summarizes the results. Section 5 is a brief conclusion.

2.2 Stochastic Properties of $\{Y_t\}$ and Estimation

The practical use of discrete time stochastic processes such as (1) to model asset price returns has proceeded by making specific assumptions on μ_t, σ_t and ϵ_t. Generally, it is assumed that up to a finite set of parameters the functional specifications for μ_t and σ_t are known and that the conditioning set over which these expectation are taken depends only on past realizations of the process.[2] Furthermore, to facilitate estimation, specific distributional assumptions are normally made on ϵ_t. ARCH, GARCH, EGARCH, IGARCH and many other variants can be grouped according to this general description. Their success in fitting observed data, producing accurate forecasts, and the ease with which they can be estimated depends largely on these assumptions. In fact, the great profusion of ARCH type models is the result

[2] But see Shepherd(1996) for alternative modeling strategies.

of an attempt to accommodate empirical regularities that have been repeatedly observed in financial return series. More recently, a number of papers have emerged (Carroll, Härdle and Mammen,2002, Härdle and Tsybakov,1997, Masry and Tjøstheim,1995) which provide a less restrictive nonparametric or semiparametric modeling of μ_t, σ_t, as well as ϵ_t. These models are in general more difficult to estimate than their parametric counterparts, but there can be substantial inferential gains if the parametric models are misspecified or unduly restrictive.

Our interest is in obtaining estimates for α-VaR and α-TailVaR associated with the conditional density f_t, where in general conditioning is on the filtration $M_{t-1} = \sigma(\{Y_s : M < s \leq t-1\})$, where $-\infty \leq M \leq t-1$. We denote such conditional densities by $f(Y_t|M_{t-1})$ for $t = 2, 3, \ldots$. Letting $q(\alpha) = F_\epsilon^{-1}(\alpha)$ be the quantile of F_ϵ and given that $F_\epsilon(x) = F\left(\mu_t + \sigma_t x | M_{t-1}\right)$ we have that α-VaR for $f(y|M_{t-1})$ is given by,

$$F^{-1}(\alpha|M_{t-1}) = \mu_t + \sigma_t q(\alpha). \tag{2}$$

Similarly, α-TailVaR for $f(y|M_{t-1})$ is given by,

$$E\left(Y_t|Y_t > F^{-1}(\alpha|M_{t-1}), M_{t-1}\right) = \mu_t + \sigma_t E(\epsilon_t|\epsilon_t > q(\alpha)). \tag{3}$$

Hence, the estimation of α-VaR and α-TailVaR can be viewed as a process of estimation for the unknown functionals in (2) and (3). We start by considering the following nonparametric specifications for μ_t, σ_t and the process Y_t. Assume that $\{(Y_t, Y_{t-1})'\}$ is a two dimensional strictly stationary process with conditional mean function $E(Y_t|Y_{t-1} = x) = m(x)$ and conditional variance $E\left((Y_t - m(x))^2|Y_{t-1} = x\right) = \sigma^2(x) > 0$. The process is described by the following markov chain of order 1,

$$Y_t = m(Y_{t-1}) + \sigma(Y_{t-1})\epsilon_t \text{ for } t = 1, 2, \ldots, \tag{4}$$

where ϵ_t is an independent strictly stationary process with unknown marginal distribution F_ϵ that is absolutely continuous with mean zero and unit variance. Note that the conditional skewness, $\alpha_3(x)$ and kurtosis, $\alpha_4(x)$ of the conditional density of Y_t given $Y_{t-1} = x$ are given by, $\alpha_3(x) = E(\epsilon_t^3)$ and $\alpha_4(x) = E(\epsilon_t^4)$. We assume that such moments exist and are continuous and that $m(x)$ and $\sigma^2(x)$ have uniformly continuous second derivatives on an open set containing x.

Recursion (4) is the conditional heterocedastic autoregressive nonlinear (CHARN) model of Diebolt and Guègan(1993), Härdle and Tsybakov(1997) and Hafner(1998). It is a special case (one lag) of the nonlinear-ARCH model treated by Masry and Tjøstheim(1995). The CHARN model provides a generalization for the popular GARCH(1,1) model in that $m(x)$ is a nonparametric function, and most importantly $\sigma^2(x)$ is not a linear function of Y_{t-1}^2. The symmetry in Y_{t-1} of the conditional variance in GARCH models is a particularly undesirable restriction when modeling financial time series due to the empirically well documented *leverage effect* (Chen, 2001, Ding, Engle and Granger,1993, Hafner,1998 and Patton,2001). However, (4) is more restrictive than traditional GARCH models in that its markov property restricts its ability to effectively model the longer memory that is commonly observed in return processes. [3] Estimation of the CHARN model is relatively simple and provides much of its appeal in our context.

2.2.1 First Stage Estimation

The estimation of $m(x)$ and $\sigma^2(x)$ in (4) was considered by Härdle and Tsybakov(1997).

[3] The model of Masry and Tjøstheim(1995) and equations (2) and (3) in Carrol, Härdle and Mammen(2002) provide a full nonparametric generalization of ARCH and GARCH(1,1) models. However, nonparametric estimators for the latter model are unavailable and for the former, convergence of the proposed estimators for m and σ^2 is extremely slow as the number of lags in the conditioning set increases (curse of dimensionality).

Unfortunately, their procedure for estimating the conditional variance $\sigma^2(x)$ suffers from significant bias and does not produce estimators that are constrained to be positive. Furthermore, the estimator is not asymptotically design adaptive to the estimation of m, i.e., the asymptotic properties of their estimator for conditional volatility is sensitive to how well m is estimated. We therefore consider an alternative estimation procedure due to Fan and Yao(1998), which is described as follows. First, we estimate $m(x)$ using the local linear estimator of Fan(1992). Let $W(x), K(x) : \Re \rightarrow \Re$ be symmetric kernel functions, y in the support of the conditional density of Y_t and $h(n)$, $h_1(n)$ be sequences of positive real numbers - bandwidths - such that $h(n), h_1(n) \rightarrow 0$ as $n \rightarrow \infty$. Let

$$(\hat{\beta}, \hat{\beta}_1)' = argmin_{\beta,\beta_1} \sum_{t=2}^{n} (y_t - \beta - \beta_1(y_{t-1} - y))^2 K\left(\frac{y_{t-1} - y}{h(n)}\right),$$

then the local linear estimator of $m(y)$ is $\hat{m}(y) = \hat{\beta}(y)$. Second, let $\hat{r}_t = (y_t - \hat{m}(y_{t-1}))^2$, and define,

$$(\hat{\eta}, \hat{\eta}_1)' = argmin_{\eta,\eta_1} \sum_{t=2}^{n} (\hat{r}_t - \eta - \eta_1(y_{t-1} - y))^2 W\left(\frac{y_{t-1} - y}{h_1(n)}\right),$$

then the local linear estimator of $\sigma^2(y)$ is $\hat{\sigma}^2(y) = \hat{\eta}(y)$.

It is clear that an important element of the nonparametric estimation of m and σ^2 is the selection of the sequence of bandwidths $h(n)$ and $h_1(n)$. We select the bandwidths using the data driven plug-in method of Ruppert, Sheather and Wand(1995) and denote them by $\hat{h}(n)$ and $\hat{h}_1(n)$. $\hat{h}(n)$ and $\hat{h}_1(n)$ are obtained based on the following regressand-regressor sequences $\{(y_t, y_{t-1})\}_{t=2}^{n}$ and $\{(\hat{r}_t, y_{t-1})\}_{t=2}^{n}$, respectively. This bandwidth selection method is theoretically superior to the popular cross-validation method and is a consistent estimator of the (optimal) bandwidth sequence that minimizes the asymptotic mean integrated squared error of $\hat{m}$ and $\hat{\sigma}^2$.[4] We chose a common kernel function (gaussian) in implementing our estimators. In the context of the CHARN model the first stage estimators for μ_t and σ_t^2 in (2) and (3) are respectively $\hat{m}(y_{t-1})$ and $\hat{\sigma}^2(y_{t-1})$.

2.2.2 Second Stage Estimation

In the second stage of the estimation we obtain estimators for $q(\alpha)$ and $E(\epsilon_t|\epsilon_t > q(\alpha))$. The estimation is based on a fundamental result from extreme value theory. To state it, we need some basic definitions. Let $q \in \Re$ be such that $q < \epsilon_F$, where $\epsilon_F = sup\{x \in \Re : F_\epsilon(x) < 1\}$ is the right endpoint of F_ϵ, the marginal distribution of the iid process $\left\{\epsilon_t = \frac{Y_t - \mu_t}{\sigma_t}\right\}$. Define the distribution of the excesses of ϵ over u, $Z = \epsilon - u$, for a nonstochastic u as,

$$\Gamma(x, u) = P(Z \leq x|\epsilon > u),$$

and the generalized pareto distribution - GPD (with location parameter equal to zero) by,

$$G(x; \beta, \psi) = 1 - \left(1 + \psi \frac{x}{\beta}\right)^{-1/\psi}, x \in D$$

where $D = [0, \infty)$ if $\psi \geq 0$ and $D = [0, -\beta/\psi]$ if $\psi < 0$. By a theorem of Pickands(1975) if $F_\epsilon \in H$ then for some positive function $\beta(u)$,

$$lim_{u \to \epsilon_F} sup_{0 < x < \epsilon_F - u} |\Gamma(x, u) - G(x; \beta(u), \psi)| = 0. \tag{5}$$

The class H is the maximum domain of attraction of the generalized extreme value distribution (Leadbetter et al., 1983). Equation (5) provides a parametric approximation for the

[4] See Ruppert,Sheather and Wand(1995) and Fan and Yao(1998). For an alternative estimator of $\sigma^2(x)$ see Ziegelmann(2002).

distribution of the excesses of ϵ over u whose validity and precision depends on whether $F_\epsilon \in H$ and u is large enough. The question of what distribution functions belong to H has been addressed by Haan(1976) and for our purposes it suffices to observe that it comprises most of the absolutely continuous distributions used in statistics.[5]

First stage estimators $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ can be used to produce a sequence of standardized residuals $\left\{e_t = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t}\right\}_{t=1}^n$ which can be used to estimate the tails of f_ϵ based on (5). For this purpose we order the residuals such that $e_{j:n}$ is the j^{th} largest residual, i.e., $e_{1:n} \geq e_{2:n} \geq ... \geq e_{n:n}$ and obtain $k < n$ excesses over $e_{k+1:n}$ given by $\{e_{j:n} - e_{k+1:n}\}_{j=1}^k$, which will be used for estimation of a GPD. By fixing k we in effect determine the residuals that are used for tail estimation and randomly select the threshold. It is easy to show that for $\alpha > 1 - k/n$ and estimates $\hat{\beta}$ and $\hat{\psi}$, $q(\alpha)$ and $E(\epsilon_t|\epsilon_t > q(\alpha))$ can be estimated by,

$$\widehat{q(\alpha)} = e_{k+1:n} + \frac{\hat{\beta}}{\hat{\psi}}\left(\left(\frac{1-\alpha}{k/n}\right)^{-\hat{\psi}} - 1\right) \tag{6}$$

and for $\psi < 1$

$$\hat{E}(\epsilon_t|\epsilon_t > q(\alpha)) = \widehat{q(\alpha)}\left(\frac{1}{1-\hat{\psi}} + \frac{\hat{\beta} - \hat{\psi}e_{k+1:n}}{(1-\hat{\psi})\widehat{q(\alpha)}}\right). \tag{7}$$

It is clear that these estimators and their properties depend on the choice of k. This question has been studied by McNeil and Frey (2000) and is also addressed in the Monte Carlo study in section 4 of this paper. Combining the estimators in (6) and (7) with first stage estimators, and using (2) and (3) gives estimators for $\alpha - VaR$ and $\alpha - TailVaR$.

We now discuss the estimation of β and ψ. Given the results in Smith(1984,1987), estimation of the GPD parameters has normally been done by constrained maximum likelihood (ML). Here we propose an alternative estimator. Traditionally, raw moments have been used to

[5] See Embrechts et al.(1997) for a complete characterization of H.

describe the location and shape of distribution functions. An alternative approach is based on L-moments (Hosking, 1990). Here, we provide a brief justification for its use and direct the reader to Hosking's paper for a more thorough understanding. Let F_ϵ be a distribution function associated with a random variable ϵ and $q(u) : (0,1) \rightarrow \Re$ its quantile. The r^{th} L-moment of ϵ is given by,

$$\lambda_r = \int_0^1 q(u) P_{r-1}(u) du \text{ for } r = 1, 2, ... \tag{8}$$

where $P_r(u) = \sum_{k=0}^{r} p_{r,k} u^k$ and $p_{r,k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}$, which contrasts with raw moments $\mu_r = \int_0^1 q(u)^r du$. Theorem 1 in Hosking(1990) gives the following justification for using L-moments: a) μ_1 is finite if and only if λ_r exist for all r; b) a distribution F_ϵ with finite μ_1 is uniquely characterized by λ_r for all r.

L-moments can be used to estimate a finite number of parameters $\theta \in \Theta$ that identify a member of a family of distributions. Suppose $\{F_\epsilon(\theta) : \theta \in \Theta \subset \Re^p\}$, p a natural number, is a family of distributions which is known up to θ. A sample $\{\epsilon_t\}_{t=1}^T$ is available and the objective is to estimate θ. Since, λ_r, $r = 1, 2, 3...$ uniquely characterizes F, θ may be expressed as a function of λ_r. Hence, if estimators $\hat{\lambda}_r$ are available, we may obtain $\hat{\theta}(\hat{\lambda}_1, \hat{\lambda}_2, ...)$. From (8), $\lambda_r = \sum_{k=0}^{r} p_{r-1,k}\beta_k$ where $\beta_k = \int_0^1 q(u)u^k du$. Given the sample, we define $\epsilon_{k,T}$ to be the k^{th} smallest element of the sample, such that $\epsilon_{1,T} \leq \epsilon_{2,T} \leq ... \leq \epsilon_{T,T}$. An unbiased estimator of β_k is

$$\hat{\beta}_k = T^{-1} \sum_{j=k+1}^{T} \frac{(j-1)(j-2)...(j-r)}{(T-1)(T-2)...(T-r)} \epsilon_{j,T}$$

and we define $\hat{\lambda}_r = \sum_{k=0}^{r} p_{r-1,k}\hat{\beta}_k$. If F_ϵ is a generalized pareto distribution with $\theta = (\mu, \beta, \psi)$, it can be shown that $\mu = \lambda_1 - (2+\psi)\lambda_2$, $\beta = (1+\psi)(2+\psi)\lambda_2$, $\psi = \frac{1-3(\lambda_3/\lambda_2)}{1+(\lambda_3/\lambda_2)}$. In our case, where $\mu = 0$, $\beta = (1+\psi)\lambda_1$, $\psi = \lambda_1/\lambda_2 - 2$ we define the following L-moment estimators for ψ

and β,

$$\hat{\psi} = \frac{\hat{\lambda}_1}{\hat{\lambda}_2} - 2 \text{ and } \hat{\beta} = (1 + \hat{\psi})\hat{\lambda}_1.$$

Similar to ML estimators, these L-moment estimators are $\sqrt{T}$-asymptotically normal for $\psi > -0.5$. They are much easier to compute than ML estimators as no numerical optimization is necessary[6]. Furthermore, although asymptotically inefficient relative to ML estimators they can outperform these estimators in finite samples as indicated by Hosking(1987). Our Monte Carlo results confirm these earlier indications on their finite sample behavior.

2.2.3 Alternative First Stage Estimators

Here we define three alternatives to the first stage estimation discussed above. The first is the quasi maximum likelihood estimation (QMLE) method used by McNeil and Frey. In essence it involves estimating by maximum likelihood the following regression model,

$$Y_t = \theta_1 Y_{t-1} + \sigma_t \epsilon_t \text{ for } t = 1, 2, ... \tag{9}$$

where $\epsilon_t \sim NIID(0, 1)$ and $\sigma_t^2 = \gamma_0 + \gamma_1 (Y_{t-1} - \theta_1 Y_{t-2})^2 + \gamma_2 \sigma_{t-1}^2$. We will refer to this procedure as GARCH-N. The second alternative estimator we consider is identical to the first procedure but assumes that the ϵ_t are iid with a standardized Student-t density, denoted by $f_s(\nu)$, where $\nu > 2$ is a parameter to be estimated together $\theta_1, \gamma_0, \gamma_1, \gamma_2$ by maximum likelihood. This estimator has gained popularity in that the Y_t inherits the leptokurdicity of ϵ_t, a characteristic of financial asset returns that have been abundantly reported in the literature. We refer to this

[6] For theoretical and practical advantage of L-moments over ordinary moments, please see Hosking(1990) and Hosking and Wallis(1997).

procedure as GARCH-T.

The third estimator we consider is based on the Student-t autoregressive (StAR(1)) model in Spanos(2002). This model can be viewed as a special case of the CHARN model. $\{(Y_t, Y_{t-1})'\}$ is taken to be a strictly stationary two dimensional vector process with Markov property of order 1. In addition, it is assumed that the marginal distribution of the process is given by a bivariate Student-t with ν degrees of freedom. This assumption is represented by,

$$(Y_t, Y_{t-1})' \sim St\left(\left(\begin{array}{c}\mu\\ \mu\end{array}\right), \left(\begin{array}{cc} v & c\\ c & v\end{array}\right); \nu\right) \quad (10)$$

Under this specification it is easy to prove that, $E(Y_t|Y_{t-1} = y_{t-1}) = \theta_0 + \theta_1 y_{t-1}$ and

$$E[illegible] - E(Y_t|Y_{t-1} = y_{t-1}))^2|Y_{t-1} = y_{t-1}) = \gamma_0 \frac{\nu}{\nu - 1}\left(1 + \gamma_1 (y_{t-1} - \mu)^2\right)$$

where [illegible] $-\theta_1)\mu$, $\theta_1 = c/v$, $\gamma_0 = v - (c^2/v)$ and $\gamma_1 = 1/(\nu v)$. Furthermore, the conditional [illegible] Y_t given Y_{t-1} is a Student-t density with $\nu + 1$ degrees of freedom. The parameters [illegible] StAR(1) model are estimated by constrained maximum likelihood.

Obviously, a number of other first stage estimators can be considered.[7] Our choice of alternative estimators to be considered in the Monte Carlo study that follows was mostly guided not by the desire to be exhaustive, but rather an attempt to represent what is commonly used both in the empirical finance literature and in practice. The StAR model, although not commonly used in empirical research, is useful in our study because it represents an estimator that is misspecified both *via* the volatility and the markov property it possesses.

2.3 Monte Carlo Design

[7] For a list of many alternatives see Gourieroux(1997).

In this section we describe and justify the design of the Monte Carlo study. Our study has two main objectives. First, to provide evidence on the finite sample distribution of the various estimators proposed in the previous section and second, to evaluate the relative performance of the estimators. This is, to our knowledge, the first evidence on the finite sample performance of the two-stage estimation procedure described above. Second, to provide applied researchers with some guidance over which estimators to use when estimating VaR and TailVaR.

In designing our Monte Carlo experiments we had two goals. First, the data generating process (DGP) had to be flexible enough to capture the salient characteristics of time series on asset returns. Second, to reduce specificity problems, we had to investigate the behavior of the estimators over a number of relevant values and specifications of the design parameter and functions in the Monte Carlo DGP.

2.3.1 The base DGP

The main DGP we consider is a nonparametric GARCH model first proposed by Hafner(1998) and later studied by Carroll, Härdle and Mammen(2002). We assume that $\{Y_t\}$ is a stochastic process representing the log-returns on a financial asset with $E(Y_t|M_{t-1}) = 0$ and $E(Y_t^2|M_{t-1}) = \sigma_t^2$, where $M_{t-1} = \sigma(\{Y_s : M < s \leq t-1\})$, where $-\infty \leq M \leq t-1$. We assume that the process evolves as,

$$Y_t = \sigma_t \epsilon_t \text{ for } t = 1, 2, ..., \text{ where} \tag{11}$$

$$\sigma_t^2 = g(Y_{t-1}) + \gamma \sigma_{t-1}^2 \tag{12}$$

where $g(x)$ is a positive, twice continuously differentiable function and $0 < \gamma < 1$ is a parameter. $\{\epsilon_t\}$ is assumed to be a sequence of independent and identically distributed random variables with the skewed Student-t density function. This density was proposed by Hansen(1994) and is given by

$$f(x;v,\lambda)=\begin{cases} bc\left(1+\frac{1}{v-2}\left(\frac{bx+a}{1+\lambda}\right)^2\right)^{-(v+1)/2} & \text{for } x\geq -a/b \\ bc\left(1+\frac{1}{v-2}\left(\frac{bx+a}{1-\lambda}\right)^2\right)^{-(v+1)/2} & \text{for } x\leq -a/b \end{cases}$$

where $c\equiv\frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{\pi(v-2)}}$, $a\equiv 4\lambda c\left(\frac{v-2}{v-1}\right)$, $b\equiv\sqrt{1+3\lambda^2-a^2}$. Hansen proved that $E(\epsilon_t)=0$ and $V(\epsilon_t)=1$. The following lemma gives expressions for the skewness and kurtosis of the asymmetric Student-t density.

Lemma 1: *Let $f(x;v,\lambda)$ be the skewed t-Student density function of Hansen(1994). Let κ_i for $i=1,2,3,4$ be as defined in* **Proposition 1** *in the appendix, then the skewness α_3 of the density is given by,*

$$\alpha_3=\frac{8\kappa_3}{b^3}(\lambda^3+\lambda)-\frac{6\kappa_2 a}{b^3}(\lambda^3+3\lambda)+\frac{12a^2\kappa_1}{b^3}\lambda-\frac{a^3+3a(1-\lambda)^3}{b^3},$$

and its kurtosis is given by,

$$\begin{aligned}\alpha_4 &= \frac{\kappa_4}{b^4}(2\lambda^5+20\lambda^3+10\lambda)-\frac{32a\kappa_3}{b^4}(\lambda^3+\lambda)\\ &+\frac{12a^2\kappa_2}{b^4}(\lambda^3+3\lambda)-\frac{16a^3\lambda\kappa_1}{b^4}+\frac{1}{b^4}\left(a^4+\frac{3(1-\lambda)^5(v-2)}{v-4}+6a^2(1-\lambda)^3\right).\end{aligned}$$

It is clear from these expressions that skewness and kurtosis are controlled by the parameters λ and v. When $\lambda=0$ the distribution is a symmetric standardized Student-t. The α-VaR for ϵ_t was obtained by Patton(2001) and is given by,

$$\alpha-VaR=\begin{cases} \frac{1-\lambda}{b}\sqrt{\frac{v-2}{v}}F_s^{-1}\left(\frac{\alpha}{1-\lambda};v\right)-\frac{a}{b} & \text{for } 0<\alpha<\frac{1-\lambda}{2} \\ \frac{1+\lambda}{b}\sqrt{\frac{v-2}{v}}F_s^{-1}\left(0.5+\frac{1}{1+\lambda}\left(\alpha-\frac{1-\lambda}{2}\right);v\right)-\frac{a}{b} & \text{for } \frac{1-\lambda}{2}\leq\alpha<1 \end{cases},$$

where F_s is the cumulative distribution of a random variable with Student-t density and v degrees of freedom. In the following lemma we obtain α-TailVaR for ϵ_t when $\alpha \geq -\frac{a}{b}$.

Lemma 2: *Let X be a random variable with density function given by an asymmetric Student-t and define the truncated density,*

$$f_{X>z}(x; v, \lambda) = \frac{f(x; v, \lambda)}{1 - F(z)} \text{ for } z \geq -a/b$$

where F is the distribution function of X. Then, the expected shortfall of X, $E(X|X > z) = \int_z^{\infty} x f_{X>z}(x; v, \lambda)dx$ is given by,

$$\begin{aligned} & E(X|X > z) \\ = & \ (1 - F(z))^{-1} \left(\frac{c(1+\lambda)^2}{b} \left(\frac{v-2}{v-1} \right) \beta^{(v-1)/2} - \frac{(1+\lambda)a}{b} \left(1 - F_s \left(\frac{bz+a}{1+\lambda} \sqrt{\frac{v}{v-2}} \right) \right) \right) \end{aligned}$$

where $\beta = \left(\cos \left(\arctan \left(\frac{bz+a}{(1+\lambda)\sqrt{v-2}} \right) \right) \right)^2$ and F_s is the cumulative distribution of a random variable with Student-t density and v degrees of freedom.

Under (11),(12) and the assumptions on ϵ_t, it is easy to verify that the conditional skewness $\alpha_3(Y_t|M_{t-1}) = E(\epsilon_t^3)$ and the conditional kurtosis $\alpha_4(Y_t|M_{t-1}) = E(\epsilon_t^4)$.

This DGP incorporates many of the empirically verified regularities normally ascribed to returns on financial assets: (1) asymmetric conditional variance with higher volatility for large negative returns and smaller volatility for positive returns (Hafner,1998); (2) conditional skewness (Aït-Sahalia and Brandt,2001, Chen,2001, Patton,2001); (3) Leptokurdicity (Tauchen,2001, Andreou et al., 2001); and (4) nonlinear temporal dependence. Our objective, of course, was to provide a DGP design that is flexible enough to accommodate these empirical regularities and to mimic the properties observed in return time series.

We designed a total of 144 experiments over the base DGP described above. Table 1 provides a numbering scheme for the experiments that is used in the description of the Monte

Carlo results. In summary, there are two sample sizes considered for the first stage of the estimation $n_S = \{500, 1000\}$, three values for γ, $n_\gamma = \{0.3, 0.6, 0.9\}$, three values for λ, $n_\lambda = \{0, -0.25, -0.5\}$, two values for α, $n_\alpha = \{0.95, 0.99\}$, two values for the number of observations used in the second stage of the estimation k, $n_k = \{80, 100\}$, and two functional forms for $g(x)$, which we denote by $g_1(x) = \frac{exp(-x)}{1+exp(-x)}$ and $g_2(x) = 1 - 0.9exp(-2x^2)$. Figures 2.1 and 2.2 provide the general shape for these volatility specifications together with that implied by GARCH type models. In the skewed student-t density distribution for $\{\epsilon_t\}$, the parameter v is fixed at 5. Let α_3 and α_4 be the skewness and kurtosis of $\{\epsilon_t\}$. Then by lemma 1, for $\lambda = 0$, $\alpha_3 = 0$, $\alpha_4 = 9$; for $\lambda = -0.25$, $\alpha_3 = -1.049$, $\alpha_4 = 11.073$; for $\lambda = -0.5$, $\alpha_3 = -1.84$, $\alpha_4 = 15.65$. Each of the experiments is based on 1000 repetitions. The set of ϵ's are generated independently for each experiment and repetition according to the specification on λ, α, n, γ, k and $g(\cdot)$. We now turn to the results of our Monte Carlo.

FIGURE 2.1 CONDITIONAL VOLATILITY BASED ON $g_1(x)$ AND GARCH MODEL

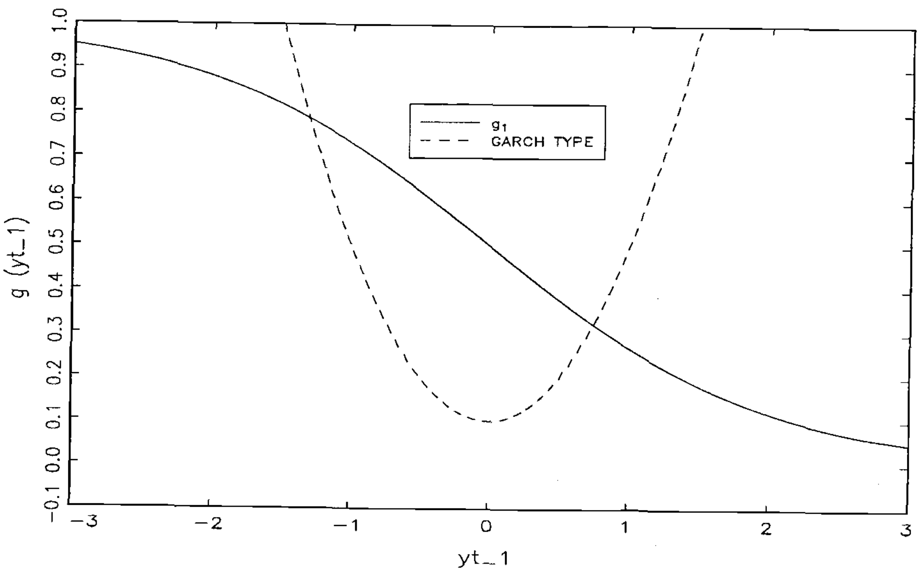
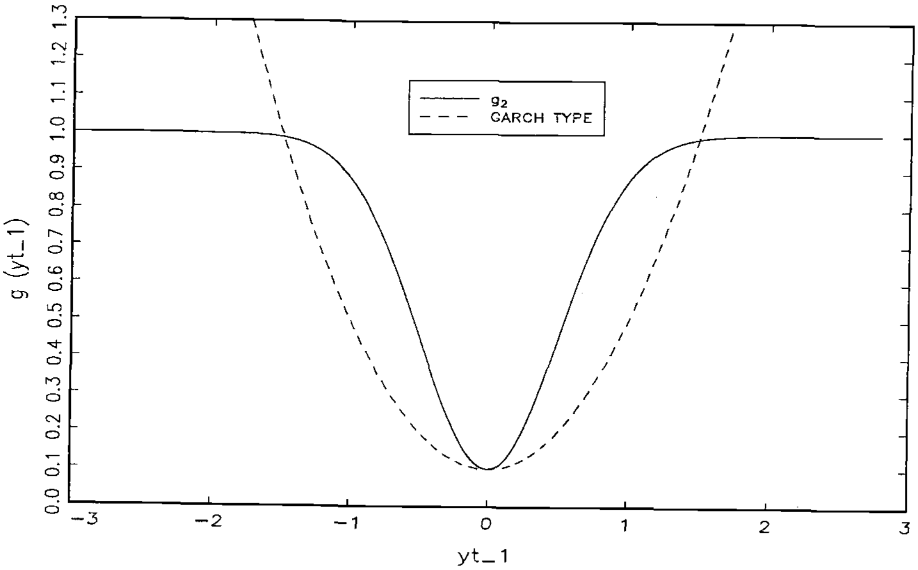


FIGURE 2.2 CONDITIONAL VOLATILITY BASED ON $g_2(x)$ AND GARCH MODEL



2.4 Results

We considered a total of nine estimators for VaR and TailVaR. There are four first stage estimators: the nonparametric method we propose, GARCH-N, GARCH-T, StAR; and two second stage estimators: the L-moments based estimator we propose and the ML estimator. In addition we consider a direct estimation method that estimates VaR and TailVaR assuming that ϵ_t in (11) is indeed distributed as an asymmetric Student-t density. In this case, all parameters are estimated *via* maximum likelihood. Since this direct ML estimator is based on a correct specification of the conditional density we expect it to outperform all other methods. Our focus is on the remaining estimators, which are all based on stochastic models that are misspecified relative to the DGPs considered. Specifically, the nonparametric and StAR estimators are based on models in which the volatility function is assumed to depend only on Y_{t-1} rather than the entire history of time series (markov property of order 1). The GARCH type models and StAR are misspecified in that g in our DGPs is not linear in Y_{t-1}. A summary of simulation results is provided in Appendix 2.

General Results on Relative Performance : Tables 2A-2F provide the ratio between an estimator's mean squared error(MSE) and the MSE for the direct estimation method for $n = 1000$, which we call relative MSE.[8] As expected, in virtually all experiments, relative MSE>1 indicating that the direct estimation method (correctly specified DGP) outperforms all other estimators for VaR and TailVaR. More interestingly, however, is that a series of very general conclusions can be reached regarding the relative performance of the other estimators:[9]

[8]Tables and graphs omit the results for the StAR method. This estimation procedure is consistently outperformed by all other methods. This result is not surprising, since the estimator is based on a stochastic model that is misspecified in **two** important ways relative to the DGP, i.e., it is markov of order 1 and linear on Y_{t-1}.

[9]Unless explicitly noted conclusions are also valid for the cases when $n = 500$.

1) for both VaR and TailVaR estimation and all estimators considered, second stage estimation based on L-moments produces smaller MSE than when based on ML. This conclusion holds for virtually all experiments, volatilities and λ. The performance of ML based estimators seems to improve with $k = 100$ and n, but not significantly. 2) For DGPs based on g_2, VaR and TailVaR estimators based on the nonparametric method produces lower MSE for virtually all experimental designs. Estimators based on GARCH-t are consistently the second best option. For DGPs based on g_1, nonparametric based estimators of VaR and TailVaR outperform GARCH type methods in virtually all cases where $\gamma = 0.6, 0.9$. We detect no decisive impact of λ,n and other design parameters on the relative performance of the estimators.

These results generally indicate that the combined nonparametric/L-moment estimation procedure we propose is superior to GARCH/ML (and StAR) type estimators in virtually all experimental designs. Furthermore, since our estimator assumes that $\{Y_t\}$ is markov of order 1, contrary to GARCH type models, our results reveal that nonlinearities in volatility may be more important to VaR and TailVaR estimation performance then accounting for the non-markov property of the series. In fact, given that $g_1(x)$ produces conditional volatilities that are similar to those empirically obtained in Hafner(1998), it seems warranted to conclude that this would indeed be the case for some typical time series of asset returns.[10] We now turn to a discussion of the bias and MSE of the estimators.

Results on MSE and Bias : 1) Tables 3A-F show that MSE for VaR and TailVaR across all estimators falls with increased n for virtually all design parameters and volatilities. There is,

[10]We found that in finite samples, VaR and TailVaR estimators based on more sophisticated nonparametric estimators in the first stage, such as that proposed by Carrol, Härdle and Mammen(2002), did not outperform ours in finite samples.

however, no distinguishable impact of increasing k on MSE for any of the estimators.[11] This is most likely due to the range of k we have used. 2) MSE for VaR and TailVaR estimators increases with the quantile across all design parameters and volatilities. 3) The impact of γ on MSE is ambiguous for DGPs based on g_2 and positive for g_1 across all estimators and design parameters. 4) MSE for VaR and TailVaR estimators decrease significantly with λ across all parameter designs and volatilities, specially when considering the nonparametric procedure. This is illustrated in Figures 2.3 and 2.4, where the ratio $R = \frac{MSE(\lambda=-0.5)}{MSE(\lambda=0)} < 1$. Since in our DGPs $\lambda \leq 0$ the data is skewed towards the positive quadrant. For second stage estimation we are selecting data that are larger than the k^{th} order statistic and considering the 95 and 99 percent quantiles. As such, more representative data of tail behavior is used when λ decreases. 5) The MSE across all design parameters, volatilities and estimators is significantly larger when estimating TailVaR compared to VaR. Hence, our simulations seem to indicate that the estimation of TailVaR is more difficult than VaR, at least as measured by MSE. The result is largely due to increased variance in TailVaR estimation rather than bias.

[11] Insensitivity of the GARCH-N/ML VaR estimator to changes in k in this range was also obtained in a simulation study by McNeil and Frey.

FIGURE 2.3 R=$\frac{\text{MSE}(\lambda=-0.5)}{\text{MSE}(\lambda=0)}$ ON TailVaR USING L-MOMENTS WITH $n = 1000$, VOLATILITY BASED ON $g_1(x)$ FOR GARCH-T, NONPARAMETRIC

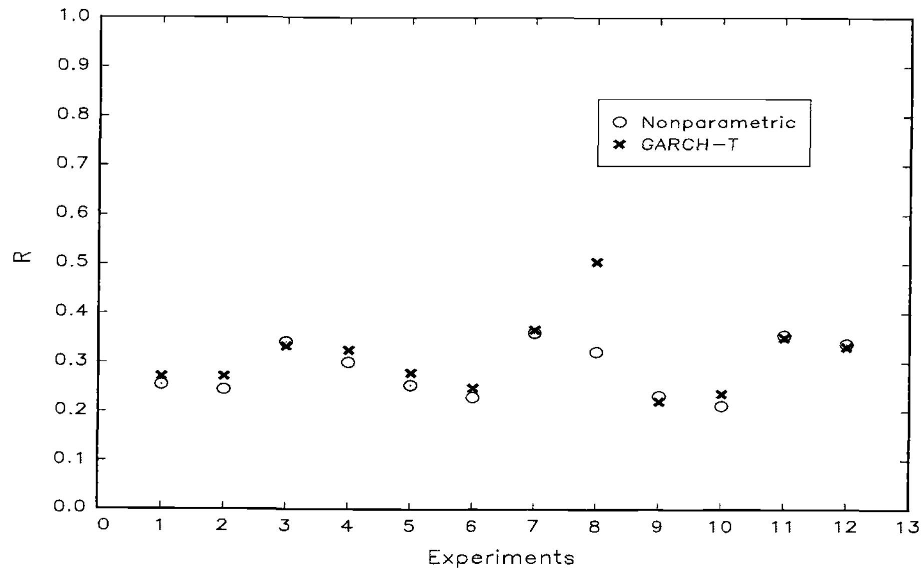
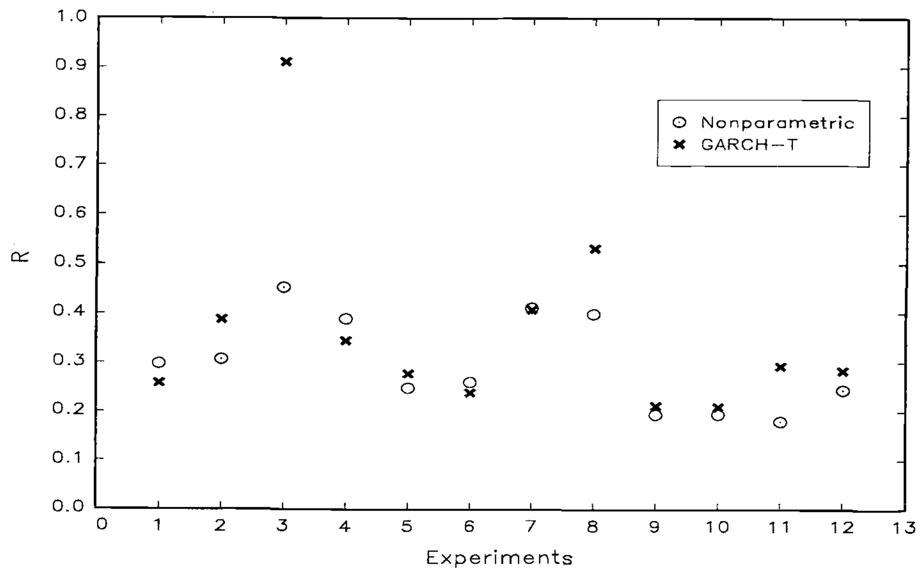


FIGURE 2.4 R=$\frac{\text{MSE}(\lambda=-0.5)}{\text{MSE}(\lambda=0)}$ ON TailVaR USING L-MOMENTS WITH $n = 1000$, VOLATILITY BASED ON $g_2(x)$ FOR GARCH-T, NONPARAMETRIC



The impact of various design parameters on the bias in VaR and TailVaR estimation across procedures, design parameters and volatilities is much less clear and definitive than that on MSE. In particular, as the sample size n or k increases there is no clear impact on bias. Hence, the reduction on MSE when n increases observed above is largely due to a reduction on variance. The most definite result regarding estimation bias we could infer from the simulations is that all estimation procedures seem to have a positive bias in the estimation of VaR and TailVaR. Figures 2.5, 2.6, 2.7, 2.8 illustrate this conclusion for g_1, g_2, $\lambda = -0.25$, $n = 1000$ for the best estimation methods, i.e., GARCH-T and the nonparametric method together with L-moment estimation. The exceptions seem to occur when $\gamma = 0.9$, in which case there is negative bias.[12] Although there are exceptions, in most DGPs the nonparametric method seems to have a slightly smaller bias than the other estimators.

[12]These conclusion are valid for DGPs where $\lambda = 0, -0.25$, $n = 500$, as well as for the other estimators considered.

FIGURE 2.5 BIAS ×100 ON VaR USING L-MOMENTS WITH $n = 1000$, $\lambda = -0.25$, VOLATILITY BASED ON $g_1(x)$ FOR GARCH-T, NONPARAMETRIC

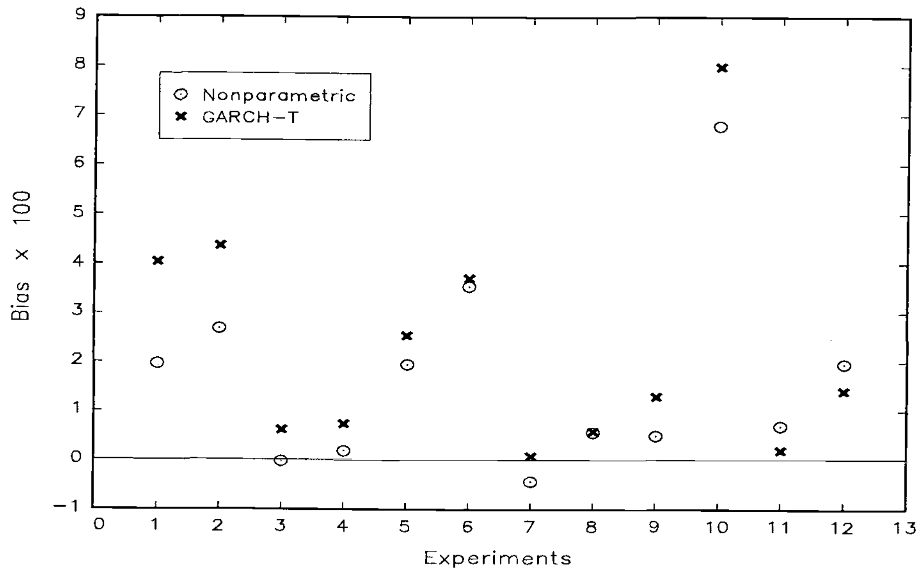


FIGURE 2.6 BIAS ×100 ON TailVaR USING L-MOMENTS WITH $n = 1000$, $\lambda = -0.25$, VOLATILITY BASED ON $g_1(x)$ FOR GARCH-T, NONPARAMETRIC

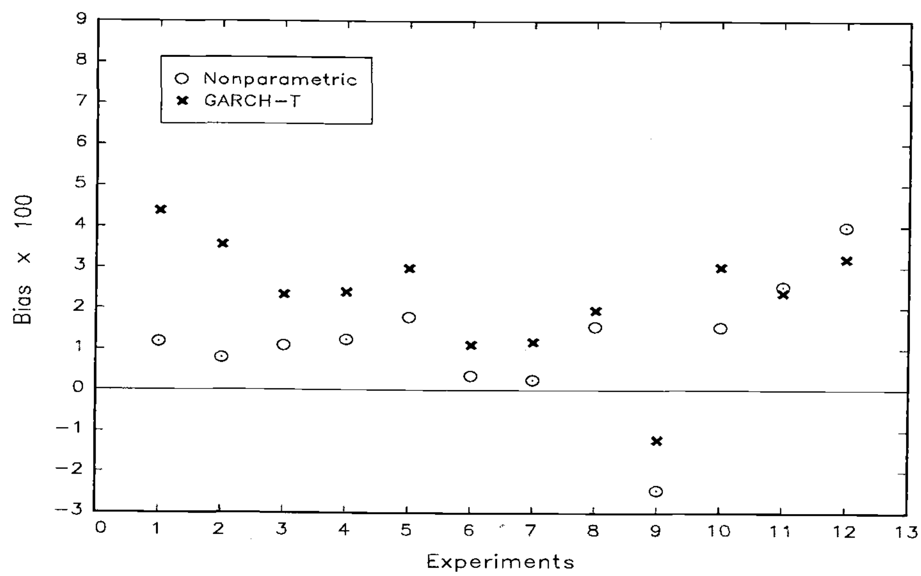


FIGURE 2.7 BIAS ×100 ON VaR USING L-MOMENTS WITH $n = 1000$, $\lambda = -0.25$, VOLATILITY BASED ON $g_2(x)$ FOR GARCH-T, NONPARAMETRIC

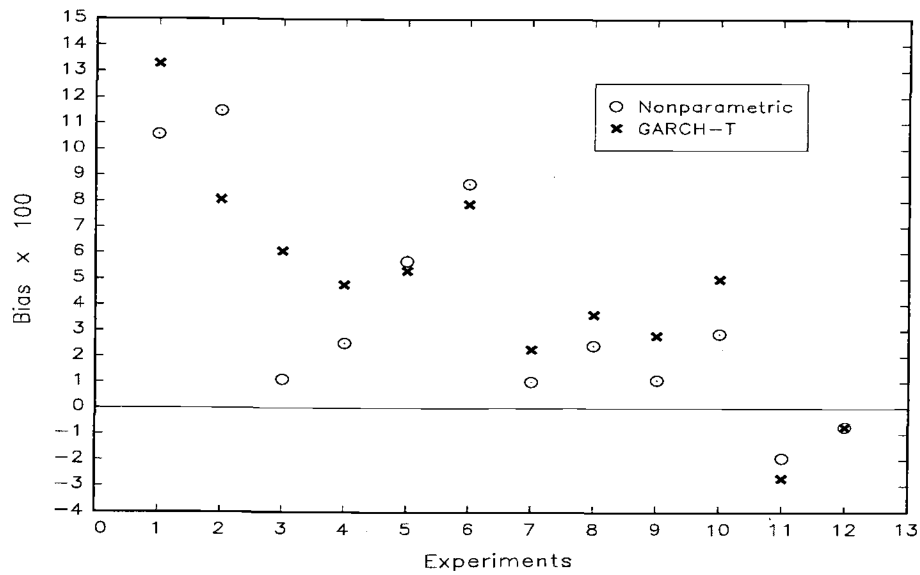
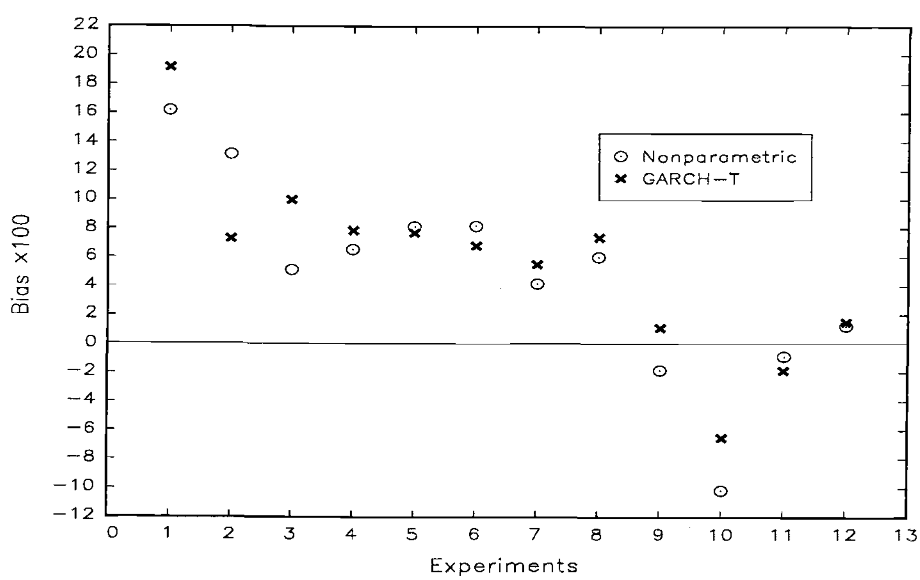


FIGURE 2.8 BIAS ×100 ON TailVaR USING L-MOMENTS WITH $n = 1000$, $\lambda = -0.25$, VOLATILITY BASED ON $g_2(x)$ FOR GARCH-T, NONPARAMETRIC



2.5 Conclusion

In this paper we have proposed a novel way to estimate VaR and TailVaR, two measures of risk that have become extremely popular in the empirical, as well as theoretical finance literature. Our procedure combines the methodology originally proposed by McNeil and Frey(2000) with nonparametric models of volatility dynamics and L-moment estimation. A Monte Carlo study that is based on a DGP that incorporates empirical regularities of returns on financial time series reveals that our estimation method outperforms the methodology put forth by McNeil and Frey. To the best of our knowledge, this is the first evidence on the finite sample performance of VaR and TailVaR estimators for conditional densities. It is important at this point to highlight the fact that asymptotic results for these types estimators are currently unavailable. In addition, results from our simulations seem to indicate that nonlinearities in volatility dynamics may be very important in accurately estimating measures of risk. In fact, our simulations indicate that accounting for nonlinearities may be more important than richer modeling of dependency.

Table 2.1 Numbering of Experiments
$\lambda = 0,\ -0.25,\ -0.5$; $n = 500,\ 1000$;Volatility based on $g_1(x),\ g_2(x)$; Number of repetitions=1000

Exp	γ	α	k
1	0.3	0.99	60
2	0.3	0.99	100
3	0.3	0.95	60
4	0.3	0.95	100
5	0.6	0.99	60
6	0.6	0.99	100
7	0.6	0.95	60
8	0.6	0.95	100
9	0.9	0.99	60
10	0.9	0.99	100
11	0.9	0.95	60
12	0.9	0.95	100

Table 2.2 Relative MSE for $n = 1000$, $\lambda = 0$,
Volatility based on $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.074	1.113	1.213	1.328	1.130	1.187	1.271	1.383	1.148	1.184	1.278	1.385
2	1.051	1.056	1.253	1.253	1.014	1.026	1.254	1.256	1.022	1.032	1.229	1.234
3	1.032	1.037	1.024	1.040	1.017	1.022	1.014	1.045	1.070	1.074	1.049	1.086
4	1.127	1.126	1.094	1.115	1.000	1.000	1.000	1.009	1.078	1.080	1.049	1.062
5	1.086	1.132	1.247	1.402	1.047	1.107	1.219	1.382	1.142	1.192	1.275	1.441
6	1.023	1.036	1.188	1.138	1.000	1.014	1.169	1.110	1.123	1.143	1.208	1.195
7	1.020	1.021	1.045	1.080	1.044	1.053	1.059	1.081	1.102	1.096	1.099	1.123
8	1.136	1.133	1.110	1.107	1.001	1.000	1.001	1.000	1.092	1.101	1.098	1.127
9	1.098	1.195	1.448	1.558	1.138	1.225	1.488	1.572	1.228	1.361	1.528	1.676
10	1.157	1.161	1.502	1.504	1.065	1.106	1.460	1.490	1.306	1.350	1.612	1.647
11	1.035	1.039	1.066	1.148	1.177	1.228	1.155	1.221	1.257	1.264	1.318	1.390
12	1.031	1.032	1.074	1.090	1.128	1.131	1.118	1.133	1.181	1.193	1.175	1.222

Table 2.3 Relative MSE for $n = 1000$, $\lambda = -0.25$, Volatility based on $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.056	1.109	1.215	1.326	1.498	1.555	1.568	1.654	1.166	1.204	1.285	1.378
2	1.042	1.040	1.179	1.175	1.001	1.000	1.153	1.143	1.038	1.038	1.164	1.155
3	1.018	1.015	1.025	1.031	1.000	1.001	1.000	1.005	1.082	1.084	1.080	1.085
4	1.008	1.001	1.018	1.034	1.007	1.000	1.000	1.012	1.037	1.032	1.025	1.039
5	1.084	1.170	1.338	1.441	1.141	1.207	1.367	1.453	1.169	1.242	1.371	1.466
6	1.088	1.083	1.234	1.231	1.072	1.071	1.258	1.271	1.155	1.153	1.326	1.333
7	1.006	1.006	1.056	1.089	1.112	1.110	1.147	1.175	1.176	1.178	1.229	1.278
8	1.013	1.015	1.058	1.082	1.051	1.046	1.081	1.093	1.098	1.099	1.121	1.152
9	1.116	1.211	1.451	1.602	1.200	1.310	1.510	1.681	1.794	1.950	2.134	2.366
10	1.117	1.112	1.456	1.450	1.246	1.240	1.584	1.564	1.441	1.430	1.701	1.676
11	1.007	1.020	1.000	1.035	1.116	1.133	1.066	1.127	1.351	1.359	1.304	1.319
12	1.019	1.064	1.048	1.176	1.074	1.094	1.114	1.166	1.077	1.124	1.131	1.229

Table 2.4 Relative MSE for $n = 1000$, $\lambda = -0.5$, Volatility based on $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.007	1.055	1.137	1.216	1.143	1.176	1.271	1.326	1.149	1.194	1.248	1.320
2	1.040	1.058	1.149	1.151	1.191	1.207	1.270	1.267	1.358	1.377	1.432	1.425
3	1.021	1.024	1.023	1.046	1.000	1.004	1.000	1.027	1.058	1.060	1.070	1.109
4	1.035	1.039	1.028	1.049	1.000	1.007	1.000	1.017	1.022	1.026	1.020	1.028
5	1.017	1.032	1.142	1.196	1.158	1.201	1.233	1.323	1.280	1.357	1.314	1.460
6	1.046	1.058	1.142	1.159	1.164	1.165	1.223	1.220	1.401	1.399	1.419	1.403
7	1.000	1.000	1.011	1.033	1.031	1.026	1.043	1.056	1.394	1.389	1.433	1.469
8	1.030	1.031	1.014	1.020	1.233	1.238	1.444	1.457	3.214	3.210	3.760	3.764
9	1.123	1.217	1.378	1.602	1.116	1.233	1.348	1.578	1.469	1.581	1.613	1.866
10	1.144	1.172	1.426	1.479	1.365	1.441	1.548	1.648	2.167	2.164	2.108	2.078
11	1.000	1.009	1.044	1.129	1.111	1.120	1.118	1.199	1.741	1.791	1.675	1.751
12	1.007	1.000	1.004	1.010	1.059	1.065	1.021	1.055	1.513	1.528	1.445	1.485

Table 2.5 Relative MSE for $n = 1000$, $\lambda = 0$, Volatility based on $g_2(x)$												
Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.057	1.063	1.124	1.146	2.534	2.516	2.683	2.557	2.670	2.612	2.799	2.664
2	1.035	1.037	1.138	1.139	1.717	1.717	1.729	1.725	1.894	1.895	1.839	1.837
3	1.010	1.009	1.035	1.049	2.118	2.128	2.062	2.074	2.370	2.389	2.293	2.303
4	1.000	1.006	1.024	1.024	2.357	2.379	2.477	2.484	2.552	2.572	2.642	2.648
5	1.057	1.075	1.219	1.262	1.571	1.597	1.711	1.772	1.928	1.955	2.050	2.088
6	1.058	1.068	1.153	1.150	1.262	1.279	1.338	1.363	1.219	1.228	1.280	1.280
7	1.012	1.016	1.025	1.044	1.234	1.234	1.269	1.291	1.390	1.394	1.407	1.427
8	1.014	1.017	1.043	1.063	1.674	1.681	1.804	1.801	1.936	1.949	2.007	2.013
9	1.192	1.360	1.637	1.884	1.297	1.471	1.693	1.922	1.623	1.815	1.881	2.118
10	1.113	1.117	1.523	1.575	1.149	1.212	1.508	1.577	1.253	1.259	1.580	1.597
11	1.441	1.448	1.520	1.601	1.000	1.005	1.000	1.101	1.863	1.869	1.790	1.880
12	1.077	1.090	1.106	1.179	1.269	1.262	1.184	1.197	1.471	1.464	1.372	1.386

Table 2.6 Relative MSE for $n = 1000$, $\lambda = -0.25$, Volatility based on $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.060	1.068	1.079	1.092	2.868	2.868	2.811	2.803	3.207	3.209	3.029	3.039
2	1.056	1.077	1.047	1.073	1.787	1.797	1.638	1.623	2.098	2.101	1.861	1.834
3	1.011	1.015	1.031	1.052	3.344	3.354	3.494	3.497	4.235	4.243	4.377	4.384
4	1.013	1.014	1.023	1.029	2.354	2.374	2.324	2.321	2.757	2.780	2.691	2.689
5	1.066	1.111	1.171	1.283	1.137	1.167	1.248	1.316	1.519	1.552	1.549	1.631
6	1.075	1.096	1.194	1.210	1.241	1.253	1.341	1.338	1.330	1.353	1.378	1.392
7	1.000	1.006	1.019	1.067	1.510	1.508	1.496	1.484	1.890	1.884	1.843	1.808
8	1.034	1.028	1.021	1.017	1.413	1.410	1.389	1.386	2.135	2.136	2.049	2.049
9	1.127	1.334	1.632	2.011	1.545	1.760	1.859	2.253	1.520	1.817	1.850	2.317
10	1.130	1.207	1.538	1.608	1.675	1.742	1.794	1.831	1.650	1.752	1.752	1.835
11	1.000	1.008	1.126	1.272	1.036	1.040	1.000	1.082	1.212	1.210	1.161	1.283
12	1.008	1.000	1.000	1.021	1.302	1.313	1.203	1.240	2.117	2.131	1.808	1.873

TABLE 2.7 RELATIVE MSE FOR $n = 1000$, $\lambda = -0.5$, VOLATILITY BASED ON $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.001	1.024	1.000	1.033	2.111	2.129	2.071	2.084	2.253	2.268	2.218	2.228
2	1.007	1.006	1.000	1.002	1.916	1.899	1.916	1.891	2.287	2.294	2.268	2.263
3	1.023	1.025	1.000	1.008	4.212	4.214	4.025	4.036	3.320	3.326	3.137	3.144
4	1.001	1.008	1.000	1.011	2.203	2.213	2.145	2.139	2.505	2.510	2.365	2.362
5	1.013	1.058	1.001	1.069	1.626	1.693	1.568	1.658	1.994	2.051	1.840	1.914
6	1.011	1.047	1.000	1.040	1.089	1.106	1.067	1.090	1.349	1.371	1.287	1.310
7	1.000	1.002	1.010	1.021	1.255	1.249	1.243	1.240	2.432	2.441	2.355	2.362
8	1.005	1.007	1.004	1.007	2.405	2.396	2.325	2.338	4.646	4.458	4.327	4.041
9	1.103	1.309	1.557	1.875	1.545	1.703	1.759	2.018	3.699	3.845	3.092	3.326
10	1.109	1.179	1.525	1.540	1.218	1.262	1.630	1.622	1.602	1.675	1.787	1.843
11	1.014	1.000	1.043	1.087	1.142	1.147	1.110	1.198	1.530	1.552	1.456	1.576
12	1.000	1.016	1.023	1.102	1.368	1.388	1.283	1.320	3.412	3.395	2.911	2.976

TABLE 2.8 MSE FOR $\lambda = 0$,
VOLATILITY BASED ON $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.140	0.109	0.387	0.259	0.137	0.114	0.396	0.272	0.146	0.116	0.384	0.273
2	0.158	0.117	0.408	0.290	0.143	0.113	0.380	0.290	0.205	0.114	0.506	0.284
3	0.043	0.036	0.103	0.079	0.043	0.035	0.102	0.078	0.047	0.037	0.112	0.081
4	0.040	0.040	0.094	0.090	0.039	0.036	0.093	0.083	0.042	0.039	0.096	0.087
5	0.189	0.145	0.575	0.371	0.191	0.140	0.578	0.362	0.201	0.152	0.590	0.379
6	0.207	0.144	0.604	0.406	0.209	0.141	0.618	0.399	0.218	0.158	0.623	0.413
7	0.050	0.042	0.126	0.100	0.057	0.043	0.137	0.101	0.071	0.046	0.166	0.105
8	0.053	0.055	0.129	0.121	0.054	0.048	0.130	0.109	0.059	0.053	0.140	0.119
9	0.573	0.386	2.251	1.256	0.592	0.400	2.290	1.291	0.700	0.431	2.640	1.325
10	0.592	0.417	1.992	1.293	0.614	0.384	1.941	1.257	0.649	0.471	1.999	1.389
11	0.125	0.094	0.357	0.237	0.148	0.107	0.385	0.257	0.149	0.114	0.398	0.293
12	0.126	0.093	0.360	0.231	0.140	0.102	0.384	0.241	0.148	0.107	0.411	0.253

TABLE 2.9 MSE FOR $\lambda = -0.25$,
VOLATILITY BASED ON $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.073	0.068	0.182	0.148	0.075	0.097	0.186	0.191	0.084	0.075	0.199	0.157
2	0.082	0.070	0.196	0.148	0.081	0.067	0.190	0.145	0.086	0.070	0.195	0.146
3	0.029	0.027	0.059	0.052	0.029	0.026	0.058	0.051	0.031	0.028	0.063	0.055
4	0.029	0.025	0.062	0.050	0.028	0.025	0.060	0.049	0.036	0.026	0.073	0.050
5	0.107	0.085	0.291	0.214	0.135	0.089	0.336	0.219	0.147	0.092	0.357	0.219
6	0.119	0.088	0.307	0.200	0.135	0.086	0.330	0.204	0.144	0.093	0.332	0.215
7	0.038	0.033	0.082	0.067	0.042	0.036	0.090	0.073	0.053	0.038	0.107	0.078
8	0.037	0.032	0.081	0.064	0.041	0.033	0.086	0.066	0.044	0.035	0.090	0.068
9	0.314	0.217	1.015	0.637	0.329	0.233	1.021	0.663	0.366	0.348	1.055	0.937
10	0.310	0.202	0.976	0.591	0.298	0.225	0.921	0.643	0.315	0.261	0.938	0.691
11	0.082	0.070	0.194	0.146	0.098	0.078	0.220	0.156	0.122	0.094	0.252	0.191
12	0.087	0.065	0.215	0.138	0.108	0.069	0.249	0.147	0.138	0.069	0.296	0.149

TABLE 2.10 MSE FOR $\lambda = -0.5$, VOLATILITY BASED ON $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.040	0.035	0.087	0.066	0.041	0.039	0.087	0.074	0.047	0.040	0.095	0.072
2	0.045	0.039	0.085	0.071	0.045	0.044	0.082	0.079	0.053	0.050	0.089	0.089
3	0.019	0.017	0.032	0.027	0.022	0.016	0.039	0.026	0.030	0.017	0.050	0.028
4	0.019	0.017	0.031	0.027	0.020	0.016	0.032	0.027	0.022	0.016	0.035	0.027
5	0.059	0.048	0.126	0.094	0.063	0.054	0.133	0.101	0.070	0.060	0.142	0.108
6	0.059	0.049	0.123	0.093	0.059	0.054	0.119	0.099	0.065	0.065	0.125	0.115
7	0.025	0.022	0.045	0.036	0.028	0.022	0.048	0.037	0.035	0.030	0.060	0.051
8	0.037	0.029	0.059	0.039	0.035	0.034	0.057	0.055	0.035	0.090	0.057	0.144
9	0.130	0.124	0.378	0.291	0.161	0.123	0.449	0.285	0.250	0.162	0.598	0.341
10	0.143	0.122	0.348	0.274	0.153	0.145	0.362	0.297	0.494	0.230	0.768	0.405
11	0.050	0.046	0.100	0.084	0.059	0.052	0.109	0.090	0.162	0.081	0.244	0.135
12	0.053	0.045	0.101	0.078	0.061	0.047	0.108	0.080	0.067	0.067	0.117	0.113

TABLE 2.11 MSE FOR $\lambda = 0$,
VOLATILITY BASED ON $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.378	0.361	0.830	0.730	0.690	0.865	1.333	1.742	0.734	0.911	1.349	1.817
2	0.403	0.355	0.804	0.727	0.723	0.590	1.329	1.104	0.757	0.651	1.373	1.174
3	0.136	0.122	0.300	0.263	0.269	0.256	0.601	0.524	0.301	0.286	0.656	0.582
4	0.128	0.128	0.279	0.275	0.440	0.301	1.054	0.664	0.432	0.326	0.863	0.709
5	0.428	0.348	1.146	0.797	0.453	0.518	1.179	1.118	0.507	0.635	1.228	1.339
6	0.435	0.381	1.033	0.842	0.980	0.455	2.298	0.977	1.261	0.439	2.778	0.934
7	0.113	0.109	0.257	0.236	0.159	0.134	0.356	0.292	0.219	0.150	0.476	0.324
8	0.117	0.114	0.288	0.251	0.165	0.188	0.413	0.434	0.182	0.218	0.415	0.483
9	0.707	0.403	3.101	1.627	0.793	0.438	3.218	1.682	1.126	0.548	4.006	1.869
10	0.757	0.397	2.763	1.567	0.843	0.410	2.868	1.551	0.895	0.447	2.958	1.626
11	0.175	0.155	0.476	0.415	0.208	0.108	0.519	0.273	0.288	0.201	0.707	0.489
12	0.145	0.100	0.473	0.283	0.163	0.118	0.464	0.303	0.222	0.136	0.606	0.351

Table 2.12 MSE for $\lambda = -0.25$, Volatility based on $g_2(x)$												
Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.273	0.225	0.524	0.407	0.526	0.608	0.977	1.061	0.539	0.680	0.939	1.143
2	0.282	0.213	0.529	0.386	0.712	0.361	1.259	0.604	0.897	0.424	1.555	0.686
3	0.094	0.089	0.187	0.169	0.185	0.296	0.357	0.574	0.206	0.374	0.406	0.719
4	0.097	0.092	0.191	0.173	0.225	0.214	0.423	0.393	0.270	0.250	0.497	0.455
5	0.251	0.205	0.570	0.417	0.412	0.219	0.837	0.445	0.521	0.293	0.989	0.552
6	0.242	0.207	0.508	0.429	0.330	0.238	0.624	0.482	0.366	0.255	0.668	0.495
7	0.083	0.074	0.168	0.146	0.102	0.112	0.210	0.215	0.158	0.140	0.320	0.265
8	0.090	0.083	0.178	0.157	0.126	0.113	0.231	0.213	0.180	0.171	0.331	0.315
9	0.379	0.232	1.520	0.784	0.444	0.318	1.622	0.893	0.526	0.313	1.771	0.889
10	0.428	0.223	1.414	0.771	0.494	0.330	1.468	0.899	0.590	0.325	1.594	0.878
11	0.094	0.062	0.247	0.169	0.114	0.065	0.265	0.150	0.139	0.076	0.315	0.175
12	0.085	0.058	0.236	0.142	0.229	0.075	0.449	0.171	0.276	0.121	0.563	0.257

TABLE 2.13 MSE FOR $\lambda = -0.5$,
VOLATILITY BASED ON $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.149	0.137	0.245	0.217	0.558	0.289	0.844	0.450	0.685	0.309	0.992	0.482
2	0.146	0.145	0.218	0.223	0.333	0.275	0.443	0.428	0.369	0.328	0.494	0.506
3	0.072	0.072	0.117	0.119	0.512	0.296	0.843	0.478	0.758	0.234	1.131	0.372
4	0.071	0.064	0.117	0.107	0.161	0.141	0.249	0.229	0.190	0.161	0.289	0.252
5	0.132	0.118	0.247	0.198	0.193	0.189	0.339	0.311	0.319	0.232	0.502	0.365
6	0.128	0.134	0.211	0.219	0.217	0.144	0.315	0.234	0.241	0.179	0.336	0.282
7	0.059	0.059	0.097	0.097	0.083	0.074	0.136	0.119	0.135	0.143	0.216	0.226
8	0.068	0.060	0.112	0.100	0.135	0.144	0.209	0.231	0.355	0.278	0.516	0.430
9	0.161	0.103	0.535	0.316	0.203	0.145	0.595	0.357	0.411	0.346	0.819	0.627
10	0.175	0.098	0.530	0.305	0.218	0.108	0.582	0.326	0.607	0.142	1.092	0.357
11	0.067	0.039	0.131	0.075	0.072	0.044	0.135	0.080	0.119	0.058	0.206	0.105
12	0.049	0.035	0.104	0.069	0.071	0.049	0.130	0.086	0.108	0.121	0.184	0.195

Chapter 3

A Nonparametric Model of Frontiers

3.1 Introduction

The specification and estimation of production frontiers and as the measurement of the associated efficiency level of production units has been the subject of a vast and growing literature since the seminal work of Farrell(1957). The main objective of this literature can be stated simply. Consider $(y, x) \in \Re_+ \times \Re_+^K$ where y describes the output of a production unit and x describes the K inputs used in production. The production technology is given by the set $T = \{(y, x) \in \Re_+ \times \Re_+^K : x \text{ can produce } y\}$ and the production function or frontier associated with T is $\rho(x) = sup\{y \in \Re_+ : (y, x) \in T\}$ for all $x \in \Re_+^K$. Let $(y_0, x_0) \in T$ characterize the performance of a production unit and define $0 \leq R_0 \equiv \frac{y_0}{\rho(x_0)} \leq 1$ to be this unit's (inverse) Farrell output efficiency measure. The main objective in production and efficiency analysis is, given a random sample of production units $\{(Y_t, X_t)\}_{t=1}^n$ that share a technology T, to obtain estimates of $\rho(\cdot)$ and by extension $R_t = \frac{Y_t}{\rho(X_t)}$ for $t = 1, \cdots, n$. Secondary objectives, such as efficiency rankings and relative performance of production units, can be subsequently obtained.

There exists in the current literature two main approaches for the estimation of $\rho(\cdot)$. The deterministic approach, represented by Charnes et al.(1978) data envelopment analysis (DEA) and Deprins et al.(1984) free disposal hull (FDH) estimators, is based on the assumption that all observed data lies in the technology set T, i.e., $P((Y_t, X_t) \in T) = 1$ for all t. The stochastic approach, pioneered by Aigner, Lovell and Schmidt(1977) and Meeusen and van den Broeck(1977), allows for random shocks in the production process and consequently $P((Y_t, X_t) \notin T) > 0$. Although more appealing from an econometric perspective, it is unfortunate that identification of stochastic frontier models requires strong parametric assumptions on the joint distribution of (Y_t, X_t) and/or $\rho(\cdot)$. These parametric assumptions may lead to misspecification of $\rho(\cdot)$ and invalidate any optimal derived properties of the proposed estimators (generally maximum likelihood) and consequently lead to erroneous inference. In addition, as recently pointed out by Baccouche and Kouki(2003), estimated inefficiency levels and firm efficiency rankings are sensitive to the specification of the joint density of (Y_t, X_t). Hence, different density specifications can lead to different conclusions regarding technology and efficiency from the same random sample. Such deficiencies of stochastic frontier models have contributed to the popularity of deterministic frontiers.[13]

Deterministic frontier estimators, such as DEA and FDH, have gained popularity among applied researchers because their construction relies on very mild assumptions on the technology T. Specifically, there is no need to assume any restrictive parametric structure on $\rho(\cdot)$ or the joint density of (Y_t, X_t). In addition to the flexible nonparametric structure, the appeal of these estimators has increased since Gijbels et al.(1999) and Park, Simar and Weiner(2000) have obtained their asymptotic distributions under some fairly reasonable assumptions.[14] Although much progress has been made in both estimation and inference in the deterministic frontier literature, we believe that alternatives to DEA and FDH estimators may be desirable. Recently, Cazals et al.(2002) have proposed a new estimator based on the joint survivor function that is more robust to extreme values and outliers than DEA and FDH estimators and does not suffer from their inherent biasedness.[15]

[13]see Seifford(1996) for an extensive literature review that illustrates the widespread use of deterministic frontiers.

[14]See the earlier work of Banker(1993) and Korostelev, Simar and Tsybakov(1995) for some preliminary asymptotic results.

[15]Bias corrected FDH and DEA estimators are available but their asymptotic distributions are not known. Again, see Gijbels et al.(1999) and Park, Simar and Weiner(2000)

In this paper we propose a new deterministic production frontier regression model and estimator that can be viewed as an alternative to the methodologies currently available, including DEA and FDH estimators and the estimator of Cazals et al.(2002). Our frontier model shares the flexible nonparametric structure that characterizes the data generating processes (DGP) underlying the results in Gijbels et al.(1999) and Park, Simar and Weiner(2000) but in addition our estimation procedure has some general properties that can prove desirable *vis a vis* DEA and FDH. First, as in Cazals et al.(2002), the estimator we propose is more robust to extreme values and outliers; second, our frontier estimator is a smooth function of input usage, not a discontinuous or piecewise linear function (as in the case of FDH and DEA estimators, respectively); third, the construction of our estimator is fairly simple as it is in essence a local linear kernel estimator; and fourth, although our estimator envelops the data, it is not intrinsically biased and therefore no bias correction is necessary. In addition to these general properties we are able to establish the asymptotic distribution and consistency of the production frontier and efficiency estimators under assumptions that are fairly standard in the nonparametric statistics literature. We view our proposed estimator not necessarily as a substitute to estimators that are currently available but rather as an alternative that can prove more adequate in some contexts.

In addition to this introduction, this paper has five more sections. Section 2 describes the model in detail, contrasts its assumptions with those in the past literature and describes the estimation procedure. Section 3 provides supporting lemmas and the main theorems establishing the asymptotic behavior of our estimators. Section 4 contains a Monte Carlo study that implements the estimator, sheds some light on its finite sample properties and compares its performance to the bias corrected FDH estimator of Park, Simar and Weiner(2000). Section 5 provides a conclusion and some directions for future work.

3.2 A Nonparametric Frontier Model

The construction of our frontier regression model is inspired by data generating processes for multiplicative regression. Hence, rather placing primitive assumptions directly on (Y_t, X_t) as it is common in the deterministic frontier literature, we place primitive assumptions on (X_t, R_t) and obtain the properties of Y_t by assuming a suitable regression function. We assume that $Z_t \equiv (X_t, R_t)'$ is a $K+1$-dimensional random vector with common density g for all $t \in \{1, 2, ...\}$ and that $\{Z_t\}$ forms an independently distributed sequence. We assume there are observations on a random variable Y_t described by

$$Y_t = \sigma(X_t)\frac{R_t}{\sigma_R}. \tag{13}$$

R_t is an unobserved random variable, X_t is an observed random vector taking values in $\Re_+^K$, $\sigma(\cdot) : \Re_+^K \rightarrow (0, \infty)$ is a measurable function and σ_R is an unknown parameter. In the case of production frontiers we interpret Y_t as output, $\rho(\cdot) \equiv \frac{\sigma(\cdot)}{\sigma_R}$ as the production frontier with inputs X_t and R_t as efficiency with values in $[0, 1]$. R_t has the effect of contracting output from optimal levels that lie on the production frontier. The larger R_t the more efficient the production unit because the closer the realized output is to that on the production frontier. In section 3 we provide a detailed list of assumptions that is used in obtaining the asymptotic properties of our estimator, however in defining the elements of the model and the estimator, two important conditional moment restrictions on R_t must be assumed; $E(R_t|X_t = x) \equiv \mu_R$ where $0 < \mu_R < 1$ and $V(R_t|X_t = x) \equiv \sigma_R^2$. It should be noted that by construction $0 < \sigma_R^2 < \mu_R < 1$. The parameter μ_R is interpreted as a mean efficiency given input usage and the common technology T and σ_R is a scale parameter for the conditional distribution of R_t that also locates the production frontier. These conditional moment restrictions together with equation (13) imply that $E(Y_t|X_t = x) = \frac{\mu_R}{\sigma_R}\sigma(x)$ and $V(Y_t|X_t = x) = \sigma^2(x)$. The model can therefore be rewritten as,

$$Y_t = \sigma(X_t)\frac{R_t}{\sigma_R} = b\sigma(X_t) + \sigma(X_t)\frac{(R_t - \mu_R)}{\sigma_R} = m(X_t) + \sigma(X_t)\epsilon_t \tag{14}$$

where $b = \frac{\mu_R}{\sigma_R}$, $\epsilon_t = \frac{R_t - \mu_R}{\sigma_R}$, $m(X_t) = b\sigma(X_t)$, $E(\epsilon_t|X_t = x) = 0$ and $V(\epsilon_t|X_t = x) = 1$.[16]

The frontier model described in (14) has a number of desirable properties. First, the frontier $\rho(\cdot) \equiv \frac{\sigma(\cdot)}{\sigma_R}$ is not restricted to belong to a known parametric family of functions and therefore there is no *a priori* undue restriction on the technology T. Second, although the existence of conditional moments are assumed for R_t, no specific parametric family of densities is assumed, therefore bypassing a number of potential problems arising from misspecification. Third, the model allows for conditional heteroscedasticity of Y_t as has been argued for in previous work (Caudill et al., 1995 and Hadri, 1999) on production frontiers. Finally, the structure of (14) is similar to regression models studied by Fan and Yao(1998), therefore lending itself to similar estimation *via* kernel procedures. This similarity motivates the estimation procedure that is described below.

The nonparametric local linear frontier estimation we propose can be obtained in three easily implementable stages. For any $x \in \Re^K_+$ we first obtain $\hat{m}(x) \equiv \hat{\alpha}$ where

$$(\hat{\alpha}, \hat{\beta}) = argmin_{\alpha,\beta} \sum_{t=1}^{n} (Y_t - \alpha - \beta(X_t - x))^2 K\left(\frac{X_t - x}{h_n}\right)$$

$K(\cdot) : \Re^K \rightarrow \Re$ is a density function and $0 < h_n \rightarrow 0$ as $n \rightarrow \infty$ is a bandwidth. This is the local linear kernel estimator of Fan(1992) with regressand Y_t and regressors X_t. In the second stage, we follow Hall and Carroll(1989) and Fan and Yao(1998) by defining $e_t \equiv (Y_t - \hat{m}(X_t))^2$

[16] For simplicity in notation, we will henceforth write $E(\cdot|X_t = x)$ or $V(\cdot|X_t = x)$ simply as $E(\cdot|X_t)$ or $V(\cdot|X_t)$.

to obtain $\hat{\alpha}_1 \equiv \hat{\sigma}^2(x)$, where

$$(\hat{\alpha}_1, \hat{\beta}_1) = argmin_{\alpha_1,\beta_1} \sum_{t=1}^{n} (e_t - \alpha_1 - \beta_1(X_t - x))^2 K\left(\frac{X_t - x}{h_n}\right)$$

which provides an estimator $\hat{\sigma}(x) = \left(\hat{\sigma}^2(x)\right)^{1/2}$. In the third stage an estimator for σ_R,

$$s_R = \left(max_{1\leq t\leq n}\frac{Y_t}{\hat{\sigma}(X_t)}\right)^{-1}$$

is obtained. Hence, a production frontier estimator at $x \in \Re^K$ is given by $\hat{\rho}(x) = \frac{\hat{\sigma}(x)}{s_R}$. We note that by construction, provided that the chosen kernel K is smooth, $\hat{\rho}(x)$ is a smooth estimator that envelops the data (no observed pair Y_t lies above $\hat{\rho}(X_t)$) but may lie above or below the true frontier $\rho(X_t)$.

In our model, the parameter σ_R provides the location of the production frontier, whereas its shape is provided by $\sigma(\cdot)$. Since besides the conditional moment restrictions on R_t there are no other restrictions other than $R_t \in [0,1]$, the observed data $\{(Y_t, X_t)\}_{t=1}^n$ may or may not be dispersed close to the frontier, hence the estimation of σ_R requires an additional normalization assumption. We assume that there exists one observed production unit that is efficient, in that the forecasted value for R_t associated with this unit is identically one. This normalization provides the motivation for the above definition of s_R. The problem of locating the production frontier is also inherent in obtaining DEA and FDH estimators. The normalization in these cases involves a number of production units being forced by construction to be efficient, i.e., lie on the frontier. This results from the fact that these estimators are defined to be minimal functions (with some stated properties, e.g., convexity and monotonicity) that envelope the data. Hence, if the stochastic process that generates the data is such that (Y_t, X_t) lie away from the true frontier, e.g., μ_R and σ_R are small, DEA and FDH will provide a downwardly biased location for the frontier. It is this dependency on boundary data points that makes these estimators highly susceptible to extreme values. This is in contrast with the estimator we propose which by construction is not a minimal enveloping function of the data. Furthermore, we note that although the location of the frontier in our model depends on the estimator s_R and its inherent normalization, if estimated efficiency levels are defined as $\hat{R}_t = \frac{s_R Y_t}{\hat{\sigma}(X_t)}$, the efficiency ranking of firms, as well as their estimated relative efficiency $\frac{\hat{R}_t}{\hat{R}_\tau}$ for $t, \tau = 1, 2, \cdots, n$ are entirely independent of the estimator s_R. In the next section we investigate the asymptotic properties of our estimators.

3.3 Asymptotic Characterization of the Estimators

In this section we establish the asymptotic properties of the frontier estimator described above. We first provide a sufficient set of assumptions for the results we prove below and provide some contrast with the assumptions made in Gijbels et al. (1999) and Park, Simar and Wiener(2000) to obtain the asymptotic distribution of DEA and FDH estimators.

ASSUMPTION A1. 1. $Z_t = (X_t, R_t)'$ for $t = 1, 2, \cdots, n$ is an independent and identically distributed sequence of random vectors with density g. We denote by $g_X(x)$ and $g_R(r)$ the common marginal densities of X_t and R_t respectively, and by $g_{R|X}(r; X)$ the common conditional density of R_t given X. 2. $0 < \underline{B}_{gx} \leq g_X(x) \leq \bar{B}_{gx} < \infty$ for all $x \in G$, G a compact subset of $\Theta = \times_{t=1}^{K}(0, \infty)$, which denotes the Cartesian product of the intervals $(0, \infty)$.

ASSUMPTION A2. 1. $Y_t = \sigma(X_t)\frac{R_t}{\sigma_R}$; 2. $R_t \in [0, 1]$, $X_t \in \Theta$; 3. $E(R_t|X_t) = \mu_R$, $V(R_t|X_t) = \sigma_R^2$; 4. $0 < \underline{B}_\sigma \leq \sigma(x) \leq \bar{B}_\sigma < \infty$ for all $x \in \Theta$; 5. $\sigma^2(\cdot) : \Theta \to \Re$ is a measurable twice continuously differentiable function in Θ; 6. $|\sigma^{2(2)}(x)| < \bar{B}_{2\sigma}$ for all $x \in \Theta$

Assumptions A1.1 and A2 imply that $\{(Y_t, X_t)\}_{t=1}^n$ forms an iid sequence of random variables with some joint density $\phi(y, x)$. This corresponds to assumption AI in Park, Simar and Wiener(2000) and is also assumed in Gijbels et al.(1999). Given that $0 < \sigma_R < 1$, A2.4 and A2.5 are implied by assumption AIII in Park, Simar and Wiener(2000). A2.6 is implied by A2 in Gijbels et al.(1999) and AIII in Park, Simar and Wiener(2000). The following assumption A3 is standard in nonparametric estimation and involves only the kernel $K(\cdot)$. We observe that A3 is satisfied by commonly used kernels such as Epanechnikov, Biweight and others.

Assumption A3. $K(x) : \Re \rightarrow \Re$ is a symmetric density function with bounded support $S_K \subset \Re^K$ satisfying: 1. $\int xK(x)dx = 0$; 2. $\int x^2 K(x)dx = \sigma_K^2$; 3. for all $x \in \Re^K$, $|K(x)| < B_K < \infty$; 4. for all $x, x' \in \Re^K$, $|K(x) - K(x')| < m||x - x'||$ for some $0 < m < \infty$;

Assumption A4. For all $x, x' \in \Theta$, $|g_X(x) - g_X(x')| < m_g||x - x'||$ for some $0 < m_g < \infty$.

A Lipschitz condition such as A4 is also assumed in Park, Simar and Wiener(2000). We note that consistency and asymptotic normality of the DEA and FDH estimators for the production frontier and associated firm efficiency depends crucially on the assumption (AII in Park, Simar and Wiener(2000)) that the joint density $\phi(y, x)$ of (Y_t, X_t) is positive at the frontier.[17] At an intuitive level this means that the data generating process (DGP)cannot be one that repeatedly gives observations that are bounded away from the frontier. In reality, there might be situations in which this assumption can be too strong. Consider, for example, the analysis of efficiency in a domestic industry or sector of an economy that is institutionally protected from foreign - potentially more efficient - competition. Unless there is an institutional change (open markets)

[17] By consequence this assumption is also crucial in obtaining the asymptotic distribution of the estimator proposed by Cazals et al.(2002).

it seems unreasonable to assume that the DGP is one that would produce efficient production units. In contrast, we assume that R_t takes values in the entire interval $[0,1]$, but there is no need for the density of the data to be positive at the frontier to obtain consistency or asymptotic normality of the frontier estimator. However, asymptotic normality of the frontier, as is made explicit in Theorem 2 requires a particular assumption on the speed of convergence of $max_{1\leq t\leq n}R_t$ to 1 as $n \to \infty$, which clearly implies some restriction on the shape of g_R.

Lastly, we make some general comments on our assumptions. As alluded to before the assumption that Z_t are iid does not prevent the model from allowing for conditional heteroscedasticity. Also, we do not assume that X_t and R_t are contemporaneously independent as it is usually done in stochastic frontier models. All that is assumed here is that conditional first and second centered moments are independent of input usage.

The main difficulties in obtaining the asymptotic properties of $\hat{\sigma}$ and by consequence those of $\frac{\hat{\sigma}}{s_R}$ derive from the fact that $\hat{\sigma}$ is based on regressands that are themselves residuals from a first stage nonparametric regression. This problem is in great part handled by the use of Lemma 3 on U statistics that appears in the appendix. Although we need only deal with U-statistics of dimension 2, Lemma 3 generalizes to $k \leq n$ Lemma 3.1 in Powell et al.(1989) where the case for $k = 2$ is proven. This lemma is of general interest and can be used whenever there is a need to analyze some specific linear combinations of nonparametric kernel estimators. For simplicity, but without loss of generality, all of our proofs are for $K = 1$. For $K > 1$ all of the results hold with appropriate adjustments on the relative speed of n and h_n^K.[18]

Lemma 1 below establishes the order in probability of certain linear combinations of kernel functions that appear repeatedly in component expressions of our estimators. The proof of the lemmas and theorems that follow depend on the repeated application of a version of Lebesgue's

[18]If different bandwidths $h_1, \cdots, h_K$ are used, a more extensive adjustment of the relative speed assumptions of n and h_i are necessary, but with no qualitative consequence to the results obtained.

dominated convergence theorem which can be found in Pagan and Ullah(1999, p.362) and Prakasa-Rao (1983, p.35). Henceforth, we refer to this result as *the proposition of Prakasa-Rao.*

Lemma 1 Assume A1, A2, A3 and suppose that $f(x,r) : (0,\infty) \times [0,1] \rightarrow \Re$ is a continuous function in G a compact subset of $(0,\infty)$ with $|f(x,r)| < B_f < \infty$. Let

$$s_j(x) = (nh_n)^{-1} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^j f(X_t, R_t) \text{ with } j = 0, 1, 2.$$

a) If $nh_n^2 \rightarrow \infty$ then $sup_{x\in G}|s_j(x) - E(s_j(x))| = O_p\left(\frac{ln(n)}{nh_n}\right)$.

b) If $nh_n^{2p+1}(ln(h_n))^{-1} \rightarrow \infty$ for $p > 0$, then $sup_{x\in G}|s_j(x) - E(s_j(x))| = o_p(h_n^p)$

Proof [Lemma 1] a) We prove the case where $j = 0$. Similar arguments can be used for $j = 1, 2$. Let $B(x_0, r) = \{x \in \Re : |x - x_0| < r\}$ for $r \in \Re^+$. G compact implies that there exists $x_0 \in G$ such that $G \subseteq B(x_0, r)$. Therefore for all $x, x' \in G$ $|x - x'| < 2r$. Let $h_n > 0$ be a sequence such that $h_n \rightarrow 0$ as $n \rightarrow \infty$ where $n \in \{1, 2, 3 \cdots\}$. For any n, by the Heine-Borel theorem there exists a finite collection of sets $\left\{B\left(x_k, \left(\frac{n}{h_n^2}\right)^{-1/2}\right)\right\}_{k=1}^{l_n}$ such that $G \subset \cup_{k=1}^{l_n} B\left(x_k, \left(\frac{n}{h_n^2}\right)^{-1/2}\right)$ for $x_k \in G$ with $l_n < \left(\frac{n}{h_n^2}\right)^{1/2} r$. For $x \in B\left(x_k, \left(\frac{n}{h_n^2}\right)^{-1/2}\right)$,

$$|s_0(x) - s_0(x_k)| \leq (nh_n)^{-1} \sum_{t=1}^{n} m|h_n^{-1}(x_k - x)|B_f < B_f m(nh_n^2)^{-1/2} \text{ and}$$

$$|E(s_0(x_k)) - E(s_0(x))| < B_f m(nh_n^2)^{-1/2}.$$

Hence,

$$|s_0(x) - E(s_0(x))| \leq |s_0(x_k) - E(s_0(x_k))| + 2B_f m(nh_n^2)^{-1/2} \text{ and}$$

$$sup_{x\in G}|s_0(x) - E(s_0(x))| \leq max_{1\leq k\leq l_n}|s_0(x_k) - E(s(x_k))| + 2B_f m(nh_n^2)^{-1/2}.$$

If $nh_n^2 \rightarrow \infty$, then to prove a) it suffices to show that there exists a constant $\Delta > 0$ such that for all $\epsilon > 0$ there exists N such that for all $n > N$,

$$P\left(\frac{nh_n}{ln(n)} max_{1\leq k\leq l_n}|s_0(x) - E(s_0(x))| \geq \Delta\right) \leq \epsilon.$$

Let $\varepsilon_n = \frac{ln(n)}{nh_n}\Delta$. Then, for every n,

$$P\left(max_{1\leq k\leq l_n}|s_0(x_k) - E(s_0(x_k))| \geq \varepsilon_n\right) \leq \sum_{k=1}^{l_n} P\left(|s_0(x_k) - E(s_0(x_k))| \geq \varepsilon_n\right)$$

But $|s_0(x_k) - E(s_0(x_k))| = |\frac{1}{n}\sum_{t=1}^{n} W_{tn}|$ where

$W_{tn} = \frac{1}{h_n}K(\frac{X_t - x_k}{h_n})f(X_t, R_t) - \frac{1}{h_n}E\left(K(\frac{X_t - x_k}{h_n})f(X_t, R_t)\right)$ with $E(W_{tn}) = 0$ and $|W_{tn}| \leq \frac{2B_K B_f}{h_n} = \frac{B_W}{h_n}$. Since $\{W_{tn}\}_{t=1}^{n}$is an independent sequence, by Bernstein's inequality

$$P\left(|s_0(x_k) - E(s_0(x_k))| \geq \varepsilon_n\right) < 2exp\left(\frac{-nh_n\varepsilon_n^2}{2h_n\bar{\sigma}^2 + \frac{2B_W\varepsilon_n}{3})}\right)$$

where $\bar{\sigma}^2$

$= n^{-1}\sum_{t=1}^{n} V(W_{tn}) = h_n^{-2}E\left(K^2\left(\frac{X_t - x_k}{h_n}\right)f^2(X_t, R_t)\right) - \left(h_n^{-1}E\left(K\left(\frac{X_t - x_k}{h_n}\right)f(X_t, R_t)\right)\right)^2$.

Under assumptions A1 and A3 and the fact that $f(x, r)$ and $g(x, r)$ are continuous in G we have that $h_n\bar{\sigma}^2 \to B_{\bar{\sigma}^2}$ by the proposition of Prakasa-Rao. Hence, for any $n > N$ there exists a constant $B_c > 0$ such that, $2h_n\bar{\sigma}^2 + \frac{2}{3}B_W\varepsilon_n \leq B_c\varepsilon_n$. Hence, $\frac{-nh_n\varepsilon_n^2}{2h_n\bar{\sigma}^2 + \frac{2B_W\varepsilon_n}{3}} \leq \frac{-nh_n\varepsilon_n^2}{B_c\varepsilon_n} = \frac{-\Delta ln(n)}{B_c}$. Hence, for any $\epsilon > 0$ there exists N such that for all $n > N$,

$P\left(max_{1\leq k\leq l_n}|s_0(x_k) - E(s_0(x_k))| \geq \varepsilon_n\right) < 2l_n n^{-\Delta/B_c}$

$< 2\left(\frac{n}{h_n^2}\right)^{1/2} rn^{-\Delta/B_c} < 2\left(nh_n^2\right)^{-1/2} r < \epsilon$ provided $\Delta > B_c$.

b) As in part a) define a collection of sets $\{B(x_k, h_n^a)\}_{k=1}^{l_n}$ such that $G \subset \cup_{k=1}^{l_n} B(x_k, h_n^a)$ for $x_k \in G$ with $l_n < h_n^{-a}r$ for $a \in (0, \infty)$. By assumption $|s_0(x) - s_0(x_k)| \leq (nh_n)^{-1}\sum_{t=1}^{n} m|h_n^{-1}(x_k - x)|B_f < B_f m h_n^{a-2}$ for $x \in B(x_k, h_n^a)$. Similarly, $|E(s_0(x_k)) - E(s_0(x))| < B_f m h_n^{a-2}$ for $x \in B(x_k, h_n^a)$. Hence, $|s_0(x) - E(s_0(x))| \leq |s_0(x_k) - E(s_0(x_k))| + 2B_f m h_n^{a-2}$ for $x \in B(x_k, h_n^a)$ and

$$sup_{x\in G}|s_0(x) - E(s_0(x))| \leq max_{1\leq k\leq l_n}|s_0(x_k) - E(s(x_k))| + 2B_f m h_n^{a-2}.$$

To show that $lim_{n\to\infty}P\left(sup_{x\in G}|s_0(x) - E(s_0(x))| \geq h_n^p\epsilon\right) = 0$ for $p > 0$ we need $h_n^{a-p-2} \to 0$ as $n \to \infty$ and $lim_{n\to\infty}P\left(max_{1\leq k\leq l_n}|s_0(x) - E(s_0(x))| \geq h_n^p\epsilon\right) = 0$. But

$$P\left(max_{1\leq k\leq l_n}|s_0(x_k) - E(s_0(x_k))| \geq h_n^p\epsilon\right) \leq \sum_{k=1}^{l_n} P\left(|s_0(x_k) - E(s_0(x_k))| \geq h_n^p\epsilon\right)$$

Using Bernstein's inequality as in a), we have

$$P\left(|s_0(x_k) - E(s_0(x_k))| \geq h_n^p \epsilon\right) < 2exp\left(\frac{-nh_n^{2p}\epsilon^2}{2(\bar{\sigma}^2 + B_W \frac{h_n^p \epsilon}{3})}\right).$$

Hence for the desired result the righthand side of the inequality must approach zero as $n \to \infty$. It suffices to show that $\frac{nh_n^{2p}\epsilon^2}{2\bar{\sigma}^2 + 2/3 B_W h_n^p \epsilon} + aln(h_n) \to \infty$, which given that $\bar{\sigma}^2 = O(1)$ will result if $\frac{nh_n^{2p+1}}{ln(h_n)} \to \infty$.

Comment. An important special case of part b) in Lemma 1 occurs when $f(x,r) \equiv 1$ for all x, r and $p = 1$. In this case we have

$$sup_{x \in G} \frac{1}{h_n} \left| (nh_n)^{-1} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^j - E(s_j(x)) \right| = o_p(1)$$

for $j = 0, 1, 2$. This result in combination with assumption A4 can be used to show that $s_0(x) - g_X(x) = O_p(h_n)$, $s_1(x) = O_p(h_n)$ and $s_2(x) - g_X(x)\sigma_K^2 = O_p(h_n)$ uniformly in G. These uniform boundedness results are used to prove Lemma 2.

Lemma 2 Assume A1, A2, A3 and A4. If $h_n \to 0$ and $\frac{nh_n^3}{ln(h_n)} \to \infty$, then for every $x \in G$ the compact set described Lemma 1, we have

$$\hat{\sigma}^2(x) - \sigma^2(x) = \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\hat{r}_t - \sigma^2(x) - \sigma^{2(1)}(x)(X_t - x)\right) + O_p(R_{n,1}(x))$$

uniformly in G, where $\hat{r}_t = \sigma^2(X_t)\epsilon_t^2 + (m(X_t) - \hat{m}(X_t))^2 + 2(m(X_t) - \hat{m}(X_t))\sigma(X_t)\epsilon_t$, $\sigma^{2(1)}(x)$ is the first derivative of $\sigma^2(x)$,

$R_{n,1}(x) = n^{-1}\left(\left|\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) r_t^*\right| + \left|\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right) r_t^*\right|\right)$ and

$r_t^* = \hat{r}_t - \sigma^2(x) - \sigma^{2(1)}(x)(X_t - x)$.

Proof [Lemma 2] Let $R_n \equiv (1_n, \vec{x} - 1_n x)$, $P_n \equiv diag\left\{K\left(\frac{X_t - x}{h_n}\right)\right\}_{t=1}^{n}$, $\hat{r}' = (\hat{r}_1, ..., \hat{r}_n)$ with $\hat{r}_t = \sigma^2(X_t)\epsilon_t^2 + (m(X_t) - \hat{m}(X_t))^2 + 2(m(X_t) - \hat{m}(X_t))\sigma(X_t)\epsilon_t$, $1_n = (1, \cdots, 1)'$, $\vec{x} =$

$(X_1, \cdots, X_n)'$,

$$S_n(x) = (nh_n)^{-1} \begin{pmatrix} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) & \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\frac{X_t - x}{h_n} \\ \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\frac{X_t - x}{h_n} & \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2 \end{pmatrix}$$

and $S(x) = \begin{pmatrix} g_X(x) & 0 \\ 0 & g_X(x)\sigma_K^2 \end{pmatrix}$.

Then, $\hat{\sigma}^2(x) - \sigma^2(x) = \frac{1}{nh_n}\sum_{t=1}^n W_n\left(\frac{X_t - x}{h_n}, x\right) r_t^*$ where $W_n(z, x) = (1, 0)S_n^{-1}(x)(1, z)'K(z)$ and $r_t^* = \hat{r}_t - \sigma^2(x) - \sigma^{2(1)}(x)(X_t - x)$. Let $A_n(x) \equiv \hat{\sigma}^2(x) - \sigma^2(x) - \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^*$, then

$$\begin{aligned} |A_n| &= \frac{1}{nh_n}\left|\sum_{t=1}^n \left(W_n\left(\frac{X_t - x}{h_n}, x\right) - \frac{1}{g_X(x)}K\left(\frac{X_t - x}{h_n}\right)\right) r_t^*\right| \\ &= \frac{1}{nh_n}\left|(1, 0)(S_n^{-1}(x) - S^{-1}(x)) \begin{pmatrix} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^* \\ \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\frac{X_t - x}{h_n} r_t^* \end{pmatrix}\right| \\ &\leq \frac{1}{h_n}\left((1, 0)(S_n^{-1}(x) - S^{-1}(x))^2(1, 0)'\right)^{1/2} \\ &\quad \frac{1}{n}\left(\left|\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^*\right| + \left|\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\frac{X_t - x}{h_n} r_t^*\right|\right) \end{aligned}$$

By the comment following Lemma 1, $B_n(x) \equiv \frac{1}{h_n}\left((1, 0)(S_n^{-1}(x) - S^{-1}(x))^2(1, 0)'\right)^{1/2} = O_p(1)$ uniformly in G. Hence if we put

$R_{n,1}(x) \equiv n^{-1}\left(\left|\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) r_t^*\right| + \left|\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\frac{X_t - x}{h_n} r_t^*\right|\right)$ and the proof is complete.

Comment. Similar arguments can be used to prove that,

$$\hat{m}(x) - m(x) = \frac{1}{nh_n g_X(x)}\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\left(Y_t - m(x) - m^{(1)}(x)(X_t - x)\right) + O_p(R_{n,2}(x))$$

where $R_{n,2}(x) = n^{-1}\left(\left|\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) Y_t^*\right| + \left|\sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right) Y_t^*\right|\right)$ and

$Y_t^* = Y_t - m(x) - m^{(1)}(x)(X_t - x)$.

Lemmas 2 and 3 are used to prove Theorem 1, which is the basis for establishing uniform consistency and asymptotic normality of the frontier estimator. Theorem 1 contains two results.

The first (a) shows that the difference $\hat{\sigma}^2(x)-\sigma^2(x)$ is $h_n^{3-\delta}$ uniformly asymptotically equivalent to $\frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right) r_t^*$ in G for $\delta > 0$. Hence, we can investigate the asymptotic properties of $\hat{\sigma}^2(x)-\sigma^2(x)$ by restricting attention to $\frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right) r_t^*$. The second (b) establishes the asymptotic normality of a suitable normalization of $\frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right) r_t^*$.

Some of the assumptions in the following theorems are made for convenience on ϵ_t rather than R_t. Since $\epsilon_t = \frac{R_t-\mu_R}{\sigma_R}$ these assumptions have a direct counterpart for R_t. Specifically we have $E(\epsilon_t^4|X_t) = \mu_4(X_t) \Rightarrow E(R_t^4|X_t)$ exists as a function of X_t and $E(|\epsilon_t||X_t) = \mu_1 \Rightarrow E(|R_t-\mu_R||X_t)$ exists as a function of X_t.

Theorem 1 Suppose that assumptions A1, A2, A3 and A4 are holding. In addition assume that $E(|\epsilon_t||X_t) = \mu_1$ for $X_t \in G$ a compact subset of $(0,\infty)$. If $h_n \to 0$, $\frac{nh_n^3}{ln(h_n)} \to \infty$ and $\frac{nh_n^3}{ln(n)} \to C$ where C is a constant, then for every $x \in G$

a) $sup_{x\in G}\left|\hat{\sigma}^2(x)-\sigma^2(x)-\frac{1}{nhg_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right) r_t^*\right| = O_p(h_n^3)$

b) if in addition we assume that 1. $E(\epsilon_t^4|X_t=x)=\mu_4(x)$ is continuous in $(0,\infty)$ then

$$\sqrt{nh_n}\left(\hat{\sigma}^2(x)-\sigma^2(x)-B_{0n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^4(x)}{g_X(x)}(\mu_4(x)-1)\int K^2(y)dy\right),$$

for all $x \in G$ where $B_{0n} = \frac{h_n^2\sigma_K^2}{2}\sigma^{2(2)}(x) + o_p(h_n^2)$.

Proof [Theorem 1] (a) Given the upperbound $\bar{B}_{g_X}$ and Lemma 2

$$\begin{aligned}
&\left|\hat{\sigma}^2(x)-\sigma^2(x)-\frac{1}{nh_n g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right) r_t^*\right| \\
\leq\ & \bar{B}_{g_X} B_n(x) h_n \left(\frac{1}{nh_n g_X(x)}\left(\left|\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right) r_t^*\right| + \left|\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)\left(\frac{X_t-x}{h_n}\right) r_t^*\right|\right)\right) \\
=\ & \bar{B}_{g_X} B_n(x) h_n \left(|c_1(x)|+|c_2(x)|\right).
\end{aligned}$$

Since $B_n(x) = O_p(1)$ uniformly in G from the comment following Lemma 1, it suffices to investigate the order in probability of $|c_1(x)|$ and $|c_2(x)|$. Here, we establish the order of $c_1(x)$

noting that the proof for $c_2(x)$ follows a similar argument given assumption A3. We write $c_1(x) = I_{1n} + I_{2n} - I_{3n} + I_{4n}$ where

$$\begin{aligned} I_{1n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)(\sigma^2(X_t) - \sigma^2(x) - \sigma^{2(1)}(X_t - x)) \\ I_{2n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2 - 1) \\ I_{3n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\sigma(X_t)\epsilon_t(\hat{m}(X_t) - m(X_t)) \\ I_{4n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)(m(X_t) - \hat{m}(X_t))^2 \end{aligned}$$

and examine each term separately. $I_{1n}(x)$: by Taylor's theorem there exists $X_{tb} = \lambda X_t + (1-\lambda)x$ for some $\lambda \in [0,1]$ such that $I_{1n} = \frac{h_n}{2ng_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2 \sigma^{2(2)}(X_{tb})$. Given A1.2 and A2.6 we have

$$\begin{aligned} & sup_{x\in G}|I_{1n}(x)| \\ \leq\ & \frac{\bar{B}_{2\sigma}\underline{B}_{gX}^{-1}}{2}\left(h_n^2 sup_{x\in G}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2 \right.\right. \\ & \left.\left. - E\left(\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2\right)\right| + h_n^2 sup_{x\in G} E\left(\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2\right)\right) \\ =\ & \frac{\bar{B}_{2\sigma}\underline{B}_{gX}^{-1}}{2}\left(h_n^3 o_p(1) + h_n^2 O(1)\right) = O_p(h_n^2) \text{ by part b) of Lemma 1 with } p = 1. \end{aligned}$$

$I_{2n}(x)$: Note that by assumption A1.2

$$\begin{aligned} |I_{2n}(x)| &\leq \underline{B}_{gX}^{-1} sup_{x\in G}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2 - 1)\right| \text{ and} \\ sup_{x\in G}|I_{2n}(x)| &\leq \underline{B}_{gX}^{-1} sup_{x\in G}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{1}{h_n}K\left(\frac{X_t - x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2 - 1)\right| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right) \end{aligned}$$

where the last equality follows from part a) in Lemma 1 with $f(X_t, R_t) = \sigma^2(X_t)(\epsilon_t^2 - 1)$, which is bounded in G by assumptions A2.2 and A2.4.

$I_{3n}(x)$: From the comment following Lemma 2 and by Taylor's theorem there exists $X_{kt} = \lambda X_k + (1-\lambda)X_t$ for some $\lambda \in [0,1]$ such that $I_{3n}(x) = I_{31n}(x) + I_{32n}(x) + I_{33n}(x)$, where

$$\begin{aligned} I_{31n}(x) &= \frac{2}{n^2h_n^2 g_X(x)}\sum_{t=1}^{n}\sum_{k=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)K\left(\frac{X_k-X_t}{h_n}\right)\sigma(X_t)\sigma(X_k)\epsilon_t\epsilon_k \\ I_{32n}(x) &= \frac{h_n^2}{n^2h_n^2 g_X(x)}\sum_{t=1}^{n}\sum_{k=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)K\left(\frac{X_k-X_t}{h_n}\right)\left(\frac{X_k-X_t}{h_n}\right)^2 \times \\ &\quad \sigma(X_t)\epsilon_t m^{(2)}(X_{kt}) \\ I_{33n}(x) &= \frac{2}{nh_n g_X(x)}\sum_{t=1}^{n}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)\epsilon_t \times \\ &\quad \left(\hat{m}(X_t) - m(X_t) - \frac{1}{nh_n g_X(X_t)}\sum_{t=1}^{n}K\left(\frac{X_k-X_t}{h_n}\right)Y_k^*\right) \end{aligned}$$

We now examine each of these terms separately. Note that,

$$|I_{31n}(x)| \leq 2\underline{B}_{gX}^{-1}\frac{1}{nh_n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t| sup_{x\in G}\frac{1}{nh_n}\left|\sum_{k=1}^{n}K\left(\frac{X_k-x}{h_n}\right)\sigma(X_k)\epsilon_k\right|$$

Since $|\sigma(X_k)\epsilon_k| < C$ for a generic constant C. If $nh_n^2 \to \infty$ we have by part a) of Lemma 1,

$$sup_{x\in G}\frac{1}{nh_n}\left|\sum_{k=1}^{n}K\left(\frac{X_k-x}{h_n}\right)\sigma(X_k)\epsilon_k\right| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right).$$

Therefore,

$$sup_{x\in G}|I_{31n}(x)| \leq 2\underline{B}_{gX}^{-1}O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right) sup_{x\in G}\left|\frac{1}{nh_n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right|$$

Since $\left|\frac{\sigma(X_t)|\epsilon_t|}{g_X(X_t)}\right| < C$,

$$\begin{aligned} & sup_{x\in G}\left|\frac{1}{nh_n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right| \\ \leq\ & sup_{x\in G}\left|\frac{1}{nh_n}\sum_{t=1}^{n}\left(\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right.\right. \\ & \left.\left. - E\left(\frac{1}{g_X(X_t)}h_n^{-1}K(\frac{X_t-x}{h_n})\sigma(X_t)|\epsilon_t|\right)\right)\right| \\ & + sup_{x\in G}E\left(\frac{1}{g_X(X_t)}\frac{1}{h_n}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right) \\ =\ & O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right) + \frac{1}{h_n}sup_{x\in G}E\left(\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right), \end{aligned}$$

by part a) of Lemma 1. Now, $\frac{1}{h_n}E\left(\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|\right) = \int K(\phi)\sigma(x+h_n\phi)\mu_1 d\phi$ and by the proposition in Prakasa-Rao,

$$\frac{1}{h_n}sup_{x\in G}E\left(\frac{1}{g_X(X_t)}K(\frac{X_t-x}{h_n})\sigma(X_t)|\epsilon_t|\right) \leq \mu_1 \int K(\phi)d\phi sup_{x\in G}\sigma(x) \leq C$$

given assumption A2.4 and $E(|\epsilon_t||X_t) = \mu_1$. Therefore,
$\frac{1}{h_n}sup_{x\in G}E\left(\frac{1}{g_X(X_t)}K(\frac{X_t-x}{h_n})\sigma(X_t)|\epsilon_t|\right) = O(1)$ and consequently
$sup_{x\in G}|I_{31n}(x)| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right)$.

Now,by assumptions A2.1 and A2.6

$$\begin{aligned} |I_{32n}(x)| &\leq \underline{B}_{gX}^{-1}bB_{2\sigma}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t| \\ &\quad \times sup_{x\in G}\left|\frac{1}{n}\sum_{k=1}^{n}K\left(\frac{X_k-x}{h_n}\right)\left(\frac{X_k-x}{h_n}\right)^2\right|. \end{aligned}$$

From the analysis of I_{1n}, $sup_{x\in G}\left|\frac{1}{n}\sum_{t=1}^{n}K\left(\frac{X_t-x}{h_n}\right)\left(\frac{X_t-x}{h_n}\right)^2\right| = O_p(h_n)$ and by using part b) of Lemma 1 $sup_{x\in G}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{g_X(X_t)}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t| = O_p(h_n)$, which gives $sup_{x\in G}|I_{32n}| = O_p(h_n^2)$. From the comments following Lemma 2

$$D_n(X_t) \equiv \left|\hat{m}(X_t) - m(X_t) - \frac{1}{nh_n g_X(x)}\sum_{k=1}^{n}K\left(\frac{X_t-x}{h_n}\right)Y_k^*\right| \leq B_n(X_t)R_{n,2}(X_t),$$

hence $|I_{33n}(x)| \leq O_p(1)\frac{2}{nh_n g_X(x)}\sum_{k=1}^{n}K\left(\frac{X_t-x}{h_n}\right)\sigma(X_t)|\epsilon_t|R_{n,2}(X_t)$. Now, we can write

$$R_{n,2}(X_t) \leq |R_{11}(X_t)| + |R_{12}(X_t)| + |R_{21}(X_t)| + |R_{22}(X_t)|,$$

where $R_{11}(X_t) = \frac{1}{n}\sum_{k=1}^{n}K\left(\frac{X_k-X_t}{h_n}\right)\sigma(X_k)\epsilon_k$,
$R_{12}(X_t) = \frac{1}{2n}\sum_{k=1}^{n}K\left(\frac{X_k-X_t}{h_n}\right)(X_k-X_t)^2 m^{(2)}(X_{kt})$,
$R_{21}(X_t) = \frac{1}{n}\sum_{k=1}^{n}K\left(\frac{X_k-X_t}{h_n}\right)\left(\frac{X_k-X_t}{h_n}\right)^2\sigma(X_k)\epsilon_k$ and
$R_{22}(X_t) = \frac{h_n^2}{2n}\sum_{k=1}^{n}K\left(\frac{X_k-X_t}{h_n}\right)\left(\frac{X_k-X_t}{h_n}\right)^3 m^{(2)}(X_{kt})$.

By part b) of Lemma 1 $sup_{X_t\in G}|R_{11}(X_t)| = o_p(h_n^2)$ and by the analysis of I_{32n} we have that $sup_{X_t\in G}|R_{12}(X_t)| = O_p(h_n^3)$. Again by Lemma 1 and the fact that $E(\epsilon_t|X_t) = 0$ we have that

$sup_{X_t \in G}|R_{21}(X_t)| = o_p(h_n^2)$. Finally, given that K is defined on a bounded support, by Lemma 1 and A2.6 we obtain $sup_{X_t \in G}|R_{22}(X_t)| = O_p(h_n^3)$. Hence, $sup_{X_t \in G} R_{n,2}(X_t) = o_p(h_n^2)$ and

$$|I_{33n}(x)| \leq 2\underline{\mathrm{B}}_{g_x}^{-1} O_p(1) o_p(h_n^2) \frac{1}{nh_n} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \sigma(X_t)|\epsilon_t| = 2\underline{\mathrm{B}}_{g_x}^{-1} o_p(h_n^2) I_{331n}.$$

By Lemma 1, $sup_{x \in G} I_{331n} = o_p(h_n) + O(1)$ and therefore $sup_{x \in G}|I_{33n}| = o_p(h_n^2)$. Combining all results we have $sup_{x \in G}|I_{3n}| = O_p(h_n^2) + O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right)$.

$I_{4n}(x)$: We write $I_{4n} = I_{41n}(x) + I_{42n}(x) + I_{43n}(x) + I_{44n}(x) + I_{45n}(x) + I_{46n}(x)$ where

$$\begin{aligned}
I_{41n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \frac{1}{n^2 h_n^2 g_X^2(X_t)} \sum_{k=1}^{n} \sum_{l=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) K\left(\frac{X_l - X_t}{h_n}\right) \\
&\quad \times \sigma(X_t)\sigma(X_l)\epsilon_k \epsilon_l \\
I_{42n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \frac{1}{4n^2 h_n^2 g_X^2(X_t)} \sum_{k=1}^{n} \sum_{l=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) (X_k - X_t)^2 \\
&\quad \times K\left(\frac{X_l - X_t}{h_n}\right) (X_l - X_t)^2 m^{(2)}(X_{kt}) m^{(2)}(X_{lt}) \\
I_{43n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) D_n^2(X_t) \\
I_{44n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \frac{1}{n^2 h_n^2 g_X^2(X_t)} \sum_{k=1}^{n} \sum_{l=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) K\left(\frac{X_l - X_t}{h_n}\right) \\
&\quad \times (X_l - X_t)^2 m^{(2)}(X_{lt}) \sigma(X_k)\epsilon_k \\
I_{45n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \frac{2D_n(X_t)}{nh_n g_X(X_t)} \sum_{k=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) \sigma(X_k)\epsilon_k \\
I_{46n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \frac{D_n(X_t)}{nh_n g_X(X_t)} \sum_{k=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) m^{(2)}(X_{kt})(X_k - X_t)^2
\end{aligned}$$

where $D_n(X_t) = \hat{m}(X_t) - m(X_t) - \frac{1}{nh_n g_X(X_t)} \sum_{k=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) Y_k^*$ where Y_k^* is defined as in the comment following Lemma 2. We now examine each term separately. First,

$$\begin{aligned}
I_{41n}(x) &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \left(\frac{1}{nh g_X(X_t)} \sum_{l=1}^{n} K\left(\frac{X_l - X_t}{h_n}\right) \sigma(X_l)\epsilon_l\right)^2 \\
&= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) (I_{411}(X_t))^2
\end{aligned}$$

where $I_{411}(X_t) = \frac{1}{nhg_X(X_t)} \sum_{l=1}^{n} K\left(\frac{X_l - X_t}{h_n}\right) \sigma(X_l)\epsilon_l$. But,

$$\begin{aligned} sup_{X_t \in G}|I_{411}(X_t)| &\leq \underline{B}_{gX}^{-1} h_n \frac{1}{h_n} sup_{X_t \in G} \left| \frac{1}{nh_n} \sum_{l=1}^{n} K\left(\frac{X_l - x}{h_n}\right) \sigma(X_l)\epsilon_l \right| \\ &= \underline{B}_{gX}^{-1} h_n o_p(1) \text{ by part b) of Lemma 1.} \end{aligned}$$

Hence, $sup_{X_t \in G}|I_{411}(X_t)| = o_p(h_n)$ and $sup_{X_t \in G}(I_{411})^2 = o_p(h_n^2)$ and

$$sup_{X_t \in G}|I_{41n}(x)| \leq o_p(h_n^2) sup_{x \in G} \left| \frac{1}{nh_n} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \right| = o_p(h_n^2).$$

Now,

$$\begin{aligned} |I_{42n}(x)| &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \\ &\quad \times \left(\frac{1}{2nh_n g_X(X_t)} \sum_{t=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) m^{(2)}(X_{kt})(X_k - X_t)^2 \right)^2 \\ &= \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) (I_{421}(X_t))^2 \end{aligned}$$

where $I_{421}(X_t) = \frac{1}{2nh_n g_X(X_t)} \sum_{t=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) m^{(2)}(X_{kt})(X_k - X_t)^2$. But $|I_{421}(X_t)| \leq \underline{B}_{gX}^{-1} h_n^{-1} |R_{12}(X_t)|$ and since $sup_{X_t \in G}|R_{12}(X_t)| = O_p(h_n^3)$ from above, we have that $sup_{X_t \in G}(I_{421}(X_t))^2 = O_p(h_n^4)$. Since $\frac{1}{nh_n} \left|\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\right| = O_p(1)$ we have $sup_{x \in G}|I_{42n}| = O_p(h_n^4)$.

For the $I_{43n}(x)$ we first observe that from our analysis of I_{33n} we have that $sup_{X_t \in G}|D_n(X_t)| = o_p(h_n^2)$ hence $|I_{43n}(x)| \leq \underline{B}_{g_x}^{-1} o_p(h_n^4) \left|\frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\right|$ and consequently $sup_{x \in G}|I_{43}(x)| = o_p(h_n^4)$ since $\frac{1}{nh_n} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) = O_p(1)$ uniformly in G.

Now,

$$|I_{44n}(x)| \leq \left| \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \right| sup_{X_t \in G}|I_{441}(X_t)| sup_{X_t \in G}|I_{442}(X_t)|, \text{ where}$$

$I_{441}(X_t) = \frac{1}{nh_n g_X(X_t)} \sum_{k=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) \sigma(X_k)\epsilon_k$,

$I_{442}(X_t) = \frac{1}{nh_n g_X(X_t)} \sum_{k=1}^{n} K\left(\frac{X_k - X_t}{h_n}\right) (X_k - X_t)^2 m^{(2)}(X_{kt})$. But given that

$sup_{X_t \in G} I_{441}(X_t) \leq \underline{B}_{gX}^{-1} h_n^{-1} sup_{X_t \in G} |R_{11}(X_t)| = o_p(h_n)$ and

$$sup_{X_t \in G} I_{442}(X_t) \leq 2\underline{B}_{gX}^{-1} h_n^{-1} sup_{X_t \in G} |R_{12}(X_t)| = O_p(h_n^2),$$

we have $sup_{x \in G} I_{44n}(x) = o_p(h_n^3)$. Finally,

$$|I_{45n}(x)| \leq \frac{2}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) sup_{X_t \in G} |D_n(X_t)| sup_{X_t \in G} |I_{441}(X_t)|$$

which implies from above that $sup_{x \in G} |I_{45n}(x)| = o_p(h_n^3)$ and

$$|I_{46n}(x)| \leq \frac{2}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) sup_{X_t \in G} |D_n(X_t)| sup_{X_t \in G} |I_{421}(X_t)|$$

which from above gives $sup_{x \in G} |I_{46n}(x)| = o_p(h_n^4)$, hence $sup_{x \in G} |I_{4n}| = o_p(h_n^2)$. Combining all terms we have that $sup_{x \in G} |c_1(x)| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right) + O_p(h_n^2)$ and also $sup_{x \in G} |c_2(x)| = O_p\left(\left(\frac{nh_n}{ln(n)}\right)^{-1}\right) + O_p(h_n^2)$. Provided that $\frac{nh_n^3}{ln(n)} \to C$ for some constant C we have that $sup_{x \in G} |c_1(x)|, sup_{x \in G} |c_2(x)| = O_p(h_n^2)$. Consequently,

$$\left| \hat{\sigma}^2(x) - \sigma^2(x) - \frac{1}{nh_n g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) r_t^* \right| \leq O_p(h_n^3).$$

(b) From part a) $I_{1n}(x) = \frac{1}{2} \frac{h_n}{n} \frac{1}{g_X(x)} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right) \left(\frac{X_t - x}{h_n}\right)^2 \sigma^{2(2)}(X_{tb})$, and given A1,

$$E\left(\frac{I_{1n}(x)}{h_n^2}\right) = \frac{1}{2g_X(x)} \int \phi^2 K(\phi) \sigma^{2(2)}(x + h_n \theta\phi) g_X(x + h_n \phi) d\phi \text{ and}$$

$$V\left(\frac{I_{1n}(x)}{h_n^2}\right) = \frac{1}{4g_X(x)^2} \left(\frac{1}{nh_n^2} E\left(K^2\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^4 (\sigma^{2(2)}(X_{tb}))^2\right) - \right.$$

$$\left. \frac{1}{n}\left(\frac{1}{h_n} E\left(K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^2 \sigma^{2(2)}(X_{tb})\right)\right)^2\right)$$

for $|\theta| \leq 1$. Given assumptions A1,A2.5 and A3 and by the proposition of Prakasa-Rao,

$$E\left(\frac{I_{1n}(x)}{h_n^2}\right) \to \frac{1}{2}\sigma^{2(2)}(x)\sigma_K^2 \text{ and } V\left(\frac{I_{1n}(x)}{h_n^2}\right) \to 0$$

hence by Chebyshev's inequality $\frac{I_{1n}(x)}{h_n^2} - \frac{1}{2}\sigma^{2(2)}(x)\sigma_K^2 = o_p(1)$.

We now establish that $\sqrt{nh}I_{2n} \xrightarrow{d} N\left(0, \frac{\sigma^4(x)}{g_X(x)}(\mu_4(x)-1)\int K^2(y)dy\right)$. To this end, note that

$$\sqrt{nh}I_{2n} = \sum_{t=1}^{n} \frac{1}{\sqrt{nh}g_X(x)} K\left(\frac{X_t-x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2-1) = \sum_{t=1}^{n} Z_{tn}$$

where $\{Z_{tn} : t=1,\cdots,n; n=1,2,\cdots\}$ forms an independent triangular array with $E(Z_{tn})=0$ and

$$\begin{aligned} s_n^2 = \sum_{t=1}^{n} E(Z_{tn}^2) &= \frac{1}{nh_n g_X^2(x)} \sum_{t=1}^{n} E\left(K^2\left(\frac{X_t-x}{h_n}\right)\sigma^4(X_t)(\epsilon_t^2-1)^2\right) \\ &= \frac{1}{h_n g_X^2(x)} E\left(K^2\left(\frac{X_t-x}{h_n}\right)\sigma^4(X_t)(\mu_4(X_t)-1)\right) \end{aligned}$$

where $\mu_4(X_t) = E\left(\epsilon_t^4|X_t\right)$. By the proposition of Prakasa-Rao and the continuity of $\mu_4(X_t)$, $s_n^2 \to \frac{\sigma^4(x)}{g_X(x)}(\mu_4(x)-1)\int K^2(\phi)d\phi$. By Liapounov's central limit theorem $\sum_{t=1}^{n}\frac{Z_{tn}}{s_n} \xrightarrow{d} N(0,1)$ provided that $lim_{n\to\infty}\sum_{t=1}^{n} E\left|\frac{Z_{tn}}{s_n}\right|^{2+\delta} = 0$ for some $\delta > 0$. Now,

$$\begin{aligned} \sum_{t=1}^{n} E\left|\frac{Z_{tn}}{s_n}\right|^{2+\delta} &= (s_n^2)^{-1-\delta/2}\sum_{t=1}^{n} E\left|Z_{tn}\right|^{2+\delta} \\ &= (s_n^2)^{-1-\delta/2}\frac{g_X(x)^{-2-\delta}}{(nh_n)^{\delta/2}}\frac{1}{h_n}E\left|K\left(\frac{X_t-x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2-1)\right|^{2+\delta} \end{aligned}$$

But,

$$\begin{aligned} &\frac{1}{h_n}E\left|K\left(\frac{X_t-x}{h_n}\right)\sigma^2(X_t)(\epsilon_t^2-1)\right|^{2+\delta} \\ =\ &\frac{1}{h_n}E\left(K^{2+\delta}\left(\frac{X_t-x}{h_n}\right)(\sigma^2(X_t))^{2+\delta}E(|\epsilon_t^2-1|^{2+\delta}|X_t)\right) \\ \leq\ &\frac{C}{h_n}E\left(K^{2+\delta}\left(\frac{X_t-x}{h_n}\right)\right) \to Cg_X(x)\int K^{2+\delta}(x)dx \end{aligned}$$

where the inequality follows from the existence of $\mu_4(X_t)$, A1, A2.4 and A3.

We now examine $I_{3n}(x)$. As in part a) we write $I_{3n}(x) = I_{31n}(x) + I_{32n}(x) + I_{33n}(x)$ and look at each term separately. Using the notation of Lemma 3 in the appendix,

$$I_{31n}(x) = \frac{2K(0)}{n^2h_n^2 g_X(x)}\sum_{t=1}^{n} K\left(\frac{X_t-x}{h_n}\right)\sigma^2(X_t)\frac{\epsilon_t^2}{g_X(X_t)} + \frac{n-1}{n}\left(\begin{array}{c} n \\ 2 \end{array}\right)^{-1}\sum_{t<k}\psi_n(Z_t,Z_k)$$

$$= \quad I_{311} + \frac{n-1}{n} I_{312}$$

where, $\psi_n(Z_t, Z_k) = h_{tk} + h_{kt}$, $h_{tk} = \frac{1}{g_X(x)h_n^2}\frac{1}{g_X(X_t)}K\left(\frac{X_t - x}{h_n}\right)K\left(\frac{X_k - X_t}{h_n}\right)\sigma(X_t)\sigma(X_k)\epsilon_t\epsilon_k$, $Z_t = (X_t, \epsilon_t)$. Given our assumptions,

$$E\left(\sqrt{nh_n} I_{311}\right) = \frac{2K(0)}{\sqrt{nh_n} g_X(x)} \frac{1}{h_n} \int K\left(\frac{z-x}{h_n}\right) \frac{\sigma^2(z)}{g_X(z)} g_X(z) dz,$$

$$E_{gx}\left(V\left(\sqrt{nh_n} I_{311} | \vec{x}\right)\right) = \frac{4K(0)}{n^2 h_n^2 g_X^2(x)} \frac{1}{h_n} \int K^2\left(\frac{z-x}{h_n}\right) \frac{\sigma^4(z)}{g_X^2(z)} (\mu^4(z) - 1) g_X(z) dz$$

and

$$\begin{aligned} V_{gx}\left(E\left(\sqrt{nh_n} I_{311} | \vec{x}\right)\right) &= \frac{4K^2(0)}{nh_n g_X^2(x)} \left(\frac{1}{nh_n} \frac{1}{h_n} \int K^2\left(\frac{z-x}{h_n}\right) \frac{\sigma^4(z)}{g_X^2(z)} g_X(z) dz \right. \\ &- \left. \frac{1}{n}\left(\frac{1}{h_n} \int K\left(\frac{z-x}{h_n}\right) \frac{\sigma^2(z)}{g_X(z)} g_X(z) dz\right)^2\right). \end{aligned}$$

Since, $V\left(\sqrt{nh_n} I_{311}\right) = E_{gx}\left(V\left(\sqrt{nh_n} I_{311} | \vec{x}\right)\right) + V_{gx}\left(E\left(\sqrt{nh_n} I_{311} | \vec{x}\right)\right)$, provided that $nh_n \to \infty$ a direct application of the proposition of Prakasa-Rao gives, $E\left(\sqrt{nh_n} I_{311}\right), V\left(\sqrt{nh_n} I_{311}\right) \to 0$ and consequently by Chebyshev's inequality we have $I_{311} = o_p((nh_n)^{-1/2})$. Given our assumptions it is easily verified that $E(\psi_n(Z_t, Z_k)) = 0$ and $\psi_{1n}(Z_t) = 0$. Hence, by direct use of Lemma 3, we have $\sqrt{n} I_{312} = o_p(1)$ provided that $E(\psi_n^2(Z_t, Z_k)) = o(n)$. We now turn to verifying that $E(\psi_n^2(Z_t, Z_k)) = o(n)$. Note that,

$$\begin{aligned} &\frac{1}{n} E(\psi_n^2(Z_t, Z_k)) \\ &= \frac{1}{n g_X^2(x) h_n^4} E\left(K^2\left(\frac{X_t - X_j}{h_n}\right) K^2\left(\frac{X_t - x}{h_n}\right) \sigma^2(X_t)\sigma^2(X_j)\epsilon_t^2\epsilon_j^2 \frac{1}{g_X^2(X_t)}\right) \\ &+ \frac{1}{n g_X^2(x) h_n^4} E\left(K^2\left(\frac{X_t - X_j}{h_n}\right) K^2\left(\frac{X_j - x}{h_n}\right) \sigma^2(X_t)\sigma^2(X_j)\epsilon_t^2\epsilon_j^2 \frac{1}{g_X^2(X_j)}\right) \\ &+ \frac{2}{n g_X^2(x) h_n^4} E\left(K^2\left(\frac{X_t - X_j}{h_n}\right) K\left(\frac{X_t - x}{h_n}\right) K\left(\frac{X_j - x}{h_n}\right) \sigma^2(X_t)\sigma^2(X_j)\epsilon_t^2\epsilon_j^2 \right. \\ &\quad \times \left. \frac{1}{g_X(X_j) g_X(X_t)}\right) \\ &= U_1 + U_2 + U_3 \end{aligned}$$

We focus on the first term - U_1. Since, $t \neq j$ we have that,

$$E(U_1|\vec{x}) = \frac{1}{ng_X^2(x)h_n^4} K^2\left(\frac{X_t - X_j}{h_n}\right)\sigma^2(X_t)\sigma^2(X_j)K^2\left(\frac{X_t - x}{h_n}\right)\frac{1}{g_X^2(X_t)} \text{ and}$$

$$\begin{aligned} E(U_1) &= \frac{1}{ng_X^2(x)h_n^4}\int\int K^2\left(\frac{X_t - X_j}{h_n}\right)\sigma^2(X_t)\sigma^2(X_j)K^2\left(\frac{X_t - x}{h_n}\right)\frac{1}{g_X^2(X_t)} \\ &\quad \times g_X(X_t)g_X(X_j)dX_t dX_j. \end{aligned}$$

Given our assumptions, if $nh_n^2 \to \infty$, by Lebesgue's Dominated Convergence theorem we have $E(U_1) \to 0$. We omit the analysis of U_2 and U_3 which can be treated similarly. Hence, combining the results on I_{311} and I_{312} we have that $\sqrt{nh_n}I_{31n} = o_p(1)$. Now we turn to the analysis of $I_{32n}(x)$. Using the notation of Lemma 3 we have
$I_{32n}(x) = \frac{n-1}{2n}\frac{1}{g_X(x)}\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\sum_{t<k}\psi_n(Z_t, Z_k)$ where $\psi_n(Z_t, Z_k) = h_{tk} + h_{kt}$ and

$$h_{tk} = K\left(\frac{X_t - x}{h_n}\right)K\left(\frac{X_t - X_k}{h_n}\right)\left(\frac{X_t - X_k}{h_n}\right)^2 m^{(2)}(X_{tk})\frac{\sigma(X_t)\epsilon_t}{g_X(X_t)}$$

and $Z_t = (X_t, \epsilon_t)$. Given our assumptions $E(\psi_n(Z_t, Z_k)) = 0$ and

$$\psi_{1n}(Z_t) = K\left(\frac{X_t - x}{h_n}\right)\frac{\sigma(X_t)\epsilon_t}{g_X(X_t)}E\left(K\left(\frac{X_t - X_k}{h_n}\right)\left(\frac{X_t - X_k}{h_n}\right)^2 m^{(2)}(X_{tk})|Z_t\right)$$

Hence, using the notation in Lemma 3, $\sqrt{n}\hat{u}_n = \frac{2}{\sqrt{n}}\sum_{t=1}^n \phi_{1n}(Z_t)$, with $E(\sqrt{n}\hat{u}_n) = 0$ and

$$\begin{aligned} V(\sqrt{n}\hat{u}_n) = 4E\Bigg(&K^2\left(\frac{X_t - x}{h_n}\right)K\left(\frac{X_t - X_k}{h_n}\right)\left(\frac{X_t - X_k}{h_n}\right)^2\frac{\sigma^2(X_t)}{g_X^2(X_t)}m^{(2)}(X_{tk}) \times \\ &K\left(\frac{X_t - X_l}{h_n}\right)\left(\frac{X_t - X_l}{h_n}\right)^2 m^{(2)}(X_{tl})\Bigg) \end{aligned}$$

Using the proposition of Prakasa-Rao we have $V(\sqrt{n}\hat{u}_n) \to 0$ and consequently by Lemma 3, $\sqrt{n}I_{32n} = o_p(1)$ provided that $E(\psi_n^2(Z_t, Z_k)) = o(n)$. Now,

$$\frac{1}{n}E(\psi_n^2(Z_t, Z_j))$$

$$
\begin{aligned}
&= \frac{1}{4nh_n^4}\int\int K^2\left(\frac{X_t-X_j}{h_n}\right)K^2\left(\frac{X_t-x}{h_n}\right)(X_t-X_j)^4\frac{\sigma^2(X_t)\epsilon_t^2 m^{(2)2}(X_{tj})}{g_X^2(X_t)} \\
&\quad \times g_X(X_t)g_X(X_j)dX_t dX_j \\
&+ \frac{1}{4nh_n^4}\int\int K^2\left(\frac{X_t-X_j}{h_n}\right)K^2\left(\frac{X_j-x}{h_n}\right)(X_t-X_j)^4\frac{\sigma^2(X_j)\epsilon_j^2 m^{(2)2}(X_{tj})}{g_X^2(X_j)} \\
&\quad g_X(X_t)g_X(X_j)dX_t dX_j \\
&+ \frac{2}{4nh_n^4}\int\int K^2\left(\frac{X_t-X_j}{h_n}\right)K\left(\frac{X_j-x}{h_n}\right)K\left(\frac{X_t-x}{h_n}\right)(X_t-X_j)^4 \\
&\quad \times\frac{\sigma(X_t)\sigma(X_j)\epsilon_t\epsilon_j m^{(2)}(X_{tj})m^{(2)}(X_{tj})}{g_X(X_t)g_X(X_j)}g_X(X_t)g_X(X_j)dX_t dX_j \\
&= U_1+U_2+U_3
\end{aligned}
$$

Given our assumptions, a direct application of Lebesgue's dominated convergence theorem gives $U_1, U_2, U_3 \to 0$. Since from part a) $I_{33n} = o_p(h_n^2)$ we have that by combining all terms $I_{3n}(x) = o_p(n^{-1/2}) + o_p(h_n^2)$. Finally, since we have already established in part a) that $I_{4n}(x) = o_p(h_n^2)$, combining all convergence results for $I_{1n}(x), I_{2n}(x), I_{3n}(x)$ and $I_{4n}(x)$ we have,

$$
\sqrt{nh_n}\left(\hat{\sigma}^2(x)-\sigma^2(x)-B_{0n}\right)\xrightarrow{d}N\left(0,\frac{\sigma^4(x)}{g_X(x)}(\mu_4(x)-1)\int K^2(y)dy\right),
$$

for all $x \in G$ where $B_{0n} = \frac{h_n^2\sigma_K^2}{2}\sigma^{2(2)}(x) + o_p(h_n^2)$, which completes the proof.

It is a direct consequence of part a) in Theorem 1 that $sup_{x\in G}\left|\hat{\sigma}^2(x)-\sigma^2(x)\right| = O_p(h_n^2)$ which implies that $\hat{\sigma}^2(x)-\sigma^2(x)\xrightarrow{p}0$ uniformly G. We now use the former result to show that $\hat{\sigma}(x)-\sigma(x) = O_p(h_n^2)$ uniformly in G and consequently obtain $\hat{\sigma}(x)-\sigma(x) = o_p(1)$ uniformly in G.

Corollary 1 Assume A1, A2, A3, A4 and that $h_n \to 0$, $\frac{nh^3}{ln(h_n)} \to \infty$ and $\hat{\sigma}^2(x)-\sigma^2(x) = O_p(h_n^2)$ uniformly in G a compact subset of $(0,\infty)$. Then, for all $\epsilon, \delta > 0$ there exists $N_{\epsilon,\delta}$ such that for $n > N_{\epsilon,\delta}$, $P\left(\{\underline{\mathrm{B}}_\sigma^2 > inf_{x\in G}|\hat{\sigma}^2(x)|\}\right) < \delta$ and $\hat{\sigma}(x)-\sigma(x) = O_p(h_n^2)$ uniformly in G.

Proof [Corollary 1] Fix $\epsilon, \delta > 0$. Then for all $x \in G$ $|\hat{\sigma}^2(x)| \leq |\hat{\sigma}^2(x) - \sigma^2(x)| + \bar{B}_\sigma$. Therefore, $sup_G|\hat{\sigma}^2(x)| \leq sup_G|\hat{\sigma}^2(x) - \sigma^2(x)| + \bar{B}_\sigma$ and
$P\left(\epsilon + \bar{B}_\sigma < sup_G|\hat{\sigma}^2(x)|\right) \leq P\left(sup_G|\hat{\sigma}^2(x) - \sigma^2(x)| > \epsilon\right) < \delta$ for $n > N_{\epsilon,\delta}$. Also, for all $x \in G$ $\underline{\mathbf{B}}_\sigma^2 - |\hat{\sigma}^2(x)| \leq |\sigma^2(x) - \hat{\sigma}^2(x)|$ and $\underline{\mathbf{B}}_\sigma^2 - inf_G|\hat{\sigma}^2(x)| \leq sup_G|\sigma^2(x) - \hat{\sigma}^2(x)|$ which gives

$$P\left(\left\{inf_G|\hat{\sigma}^2(x)| < \underline{\mathbf{B}}_\sigma^2 - \epsilon\right\}\right) \leq P\left(\left\{sup_G|\sigma^2(x) - \hat{\sigma}^2(x)| > \epsilon\right\}\right) < \delta.$$

for $n > N_{\epsilon,\delta}$. By the mean value theorem and A2, there exists $\sigma_b^2(x) = \theta\sigma^2(x) + (1-\theta)\hat{\sigma}^2(x)$ for some $0 \leq \theta \leq 1$ and $\forall x \in G$ such that

$$\begin{aligned} sup_G|\hat{\sigma}(x) - \sigma(x)| &= \frac{1}{2} sup_G \left|\frac{1}{\sqrt{\sigma_b^2(x)}}\right| sup_G|\hat{\sigma}^2(x) - \sigma^2(x)| \\ &= \frac{1}{2}\left(inf_G|\sigma_b^2(x)|\right)^{-1/2} sup_G|\hat{\sigma}^2(x) - \sigma^2(x)| \end{aligned}$$

Note that $inf_G|\sigma_b^2(x)| \geq \theta\underline{\mathbf{B}}_\sigma^2 + (1-\theta)inf_G|\hat{\sigma}^2(x)|$ and therefore

$$P\left(\left\{\underline{\mathbf{B}}_\sigma^2 > inf_G|\sigma_b^2(x)|\right\}\right) \leq P\left(\left\{\underline{\mathbf{B}}_\sigma^2 > inf_G|\hat{\sigma}^2(x)|\right\}\right) < \delta$$

for $n > N_{\epsilon,\delta}$. Hence, $\sigma_b^2(x)^{-1/2} = O_p(1)$ uniformly in G which combines with $\hat{\sigma}^2(x) - \sigma^2(x) = O_p(h_n^2)$ uniformly in G from the comment following Theorem 1 to give $\hat{\sigma}(x) - \sigma(x) = O_p(h_n^2)$ uniformly in G.

The asymptotic normality of $\hat{\sigma}(x)$ is easily obtained from part b) of Theorem 1 by noting that

$$\begin{aligned} &\sqrt{nh_n}\left(\hat{\sigma}(x) - \sigma(x) - \frac{1}{2\sigma(x)}B_{0n} + \left(\frac{1}{2\sigma(x)} - \frac{1}{2\sigma_b(x)}\right)B_{0n}\right) \\ = &\frac{1}{2\sqrt{\sigma_b^2(x)}}\sqrt{nh_n}(\hat{\sigma}^2(x) - \sigma^2(x) - B_{0n}) \end{aligned}$$

and given the uniform consistency of $\hat{\sigma}(x)$ in G from the corollary we have that,

$$\sqrt{nh_n}\left(\hat{\sigma}(x) - \sigma(x) - \frac{1}{2\sigma(x)}B_{0n}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{4g_X(x)}(\mu_4(x) - 1)\int K^2(y)dy\right)$$

The results in Theorem 1 and its corollary refer to the estimator $\hat{\sigma}(x)$, but since our main interest lies on $\hat{\rho}(x) \equiv \frac{\hat{\sigma}}{s_R}$, a complete characterization of the asymptotic behavior of the frontier estimator requires that we provide convergence results on s_R. Theorem 2 below shows that given that $\hat{\sigma}(x) - \sigma(x) = O_p(h_n^2)$, we are able to show that $s_R - \sigma_R = O_p(h_n^2)$ provided that $max_{1\leq t\leq n} R_t$ converges to 1 sufficiently fast. It should be noted that the required speed of convergence on $max_{1\leq t\leq n} R_t$ is not necessary to establish the consistency of s_R, which results directly from $\hat{\sigma}(x) - \sigma(x) = O_p(h_n^2)$. As made explicit below, its use is necessary only in obtaining asymptotic distributional results on $\hat{\rho}(x)$.

Theorem 2 Suppose that (1) $\hat{\sigma}(x) - \sigma(x) = O_p(h_n^2)$ uniformly in G and that (2) for all $\delta > 0$ there exists a constant $\Delta > 0$ such that for all $n > N_\delta$ we have that
$P\left(max_{1\leq t\leq n} R_t > 1 - h_n^2\Delta\right) > 1-\delta$. Then, $s_R - \sigma_R = O_p(h_n^2)$.

Proof [Theorem 2] We start by noting that $|s_R - \sigma_R| = s_R\sigma_R|s_R^{-1} - \sigma_R^{-1}|$. By Corollary 1 $\left(sup_{X_t\in G}\hat{\sigma}(X_t)\right)^{-1} = O_p(1)$, hence by definition $s_R \leq \left(sup_{X_t\in G}\hat{\sigma}(X_t)\right)^{-1}\left(max_{1\leq t\leq n} Y_t\right)^{-1} = O_p(1)$. Hence, to obtain the desired result it suffices to show that $s_R^{-1} - \sigma_R^{-1} = O_p(h_n^2)$. Since, $|s_R^{-1} - \sigma_R^{-1}| = \sigma_R^{-1}|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1|$ we need only show that $max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1 = O_p(h_n^2)$. Note that for some $\Delta', \Delta > 0$,

$$P\left(h_n^{-2} sup_{X_t\in G}\left|\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} - 1\right| < \Delta\right) \geq P\left(h_n^{-2} sup_{X_t\in G}\left|\sigma(X_t) - \hat{\sigma}(X_t)\right| < \Delta'\right).$$

Therefore, given supposition (1) in the statement of the theorem, for all $\delta > 0$ there exists $\Delta > 0$ such that for all $n > N_\delta$,

$$P\left(h_n^{-2} sup_{X_t\in G}\left|\frac{\sigma(X_t)}{\hat{\sigma}(X_t)} - 1\right| < \Delta\right) > 1-\delta. \tag{15}$$

Now suppose that $max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1 \geq 0$. Then,

$$|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1| = max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)} - 1$$

$$\leq \ max_{1\leq t\leq n}R_t sup_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)}-1$$
$$\leq \ sup_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)}-1$$

Hence, $h_n^{-2}|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)}-1|\leq h_n^{-2}|sup_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)}-1|$ and by inequality (15)

$$P\left(h_n^{-2}|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)}-1|<\Delta\right)>1-\delta$$

Now suppose that $max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)}-1<0$. Then,

$$|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)}-1| \ = \ 1-max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)}$$
$$\leq \ max_{1\leq t\leq n}R_t inf_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)}$$

and

$$P\left(h_n^{-2}|max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)}-1|<\Delta\right)\geq P\left(max_{1\leq t\leq n}R_t inf_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)}>1-h_n^2\Delta\right).$$

By inequality (15) and assumption (2) in the statement of the theorem, for all $\delta>0$ there is some $\Delta_1,\Delta>0$ such that whenever $n>N_\delta$, $P\left(inf_{X_t\in G}\frac{\sigma(X_t)}{\hat{\sigma}(X_t)}>1-h_n^2\Delta\right)>1-\delta$ and $P\left(max_{1\leq t\leq n}R_t>1-h_n^2\Delta_1\right)>1-\delta$. Hence, for all $\delta>0$ there is some $\Delta_2>0$ such that whenever $n>N_\delta$

$$P\left(max_{1\leq t\leq n}\frac{\sigma(X_t)R_t}{\hat{\sigma}(X_t)}>1-h_n^2\Delta_2\right)>1-\delta.$$

which completes the proof.

Since, $\frac{\hat{\sigma}(x)}{s_R}-\frac{\sigma(x)}{\sigma_R}=\hat{\sigma}(x)\left(\frac{1}{s_R}-\frac{1}{\sigma_R}\right)+\frac{1}{\sigma_R}\left(\hat{\sigma}(x)-\sigma(x)\right)$ and $\hat{\sigma}(x)=O_p(1)$ an immediate consequence of Theorem 2 is that $\frac{\hat{\sigma}(x)}{s_R}-\frac{\sigma(x)}{\sigma_R}=o_p(1)$, establishing consistency of the frontier estimator. The asymptotic distribution of $\hat{\rho}(x)$ can be easily obtained from Theorems 1 and 2 by first noting that from Theorem 1 we have that,

$$\sqrt{nh_n}\left(\frac{\hat{\sigma}(x)}{\sigma_R}-\frac{\sigma(x)}{\sigma_R}-\frac{1}{2\sigma(x)\sigma_R}B_{0n}\right)\xrightarrow{d}N\left(0,\frac{\sigma^2(x)}{4\sigma_R^2 g_X(x)}(\mu_4(x)-1)\int K^2(y)dy\right).$$

Also, since

$$\begin{aligned}&\sqrt{nh_n}\left(\frac{\hat{\sigma}(x)}{\sigma_R}-\frac{\sigma(x)}{\sigma_R}-\frac{1}{2\sigma(x)\sigma_R}B_{0n}\right)\\ \equiv\ &\sqrt{nh_n}\left(\frac{\hat{\sigma}(x)}{s_R}-\frac{\sigma(x)}{\sigma_R}-\hat{\sigma}(x)\left(s_R^{-1}-\sigma_R^{-1}\right)-\frac{1}{2\sigma(x)\sigma_R}B_{0n}\right)\end{aligned}$$

we have by Theorem 2 that $\hat{\sigma}(x)\left(s_R^{-1}-\sigma_R^{-1}\right)=O_p(h_n^2)$ and consequently we can write,

$$\sqrt{nh_n}\left(\frac{\hat{\sigma}(x)}{s_R}-\frac{\sigma(x)}{\sigma_R}-\frac{1}{2\sigma(x)\sigma_R}B_{1n}\right)\xrightarrow{d}N\left(0,\frac{\sigma^2(x)}{4\sigma_R^2 g_X(x)}(\mu_4(x)-1)\int K^2(y)dy\right) \quad (16)$$

where $B_{1n}=\frac{h_n^2\sigma_K^2\sigma^{(2)2}(x)}{4\sigma_R\sigma(x)}+O_p(h_n^2)$. The asymptotic properties of the frontier estimator can be used directly to obtain the properties of the implied inverse Farrell efficiency. If (y_0,x_0) is a production plan with $x_0\in G$, then $\hat{R}_0-R_0=o_p(1)$ and

$$\sqrt{nh_n}\left(\hat{R}_0-R_0+B_{2n}\right)\xrightarrow{d}N\left(0,\frac{R_0^2}{4g_X(x_0)}(\mu_4(x_0)-1)\int K^2(y)dy\right) \quad (17)$$

where $B_{2n}=\frac{h_n^2\sigma_K^2R_0\sigma^{2(2)}(x_0)}{4\sigma^2(x_0)}+O_p(h_n^2)$.

The importance of Theorem 2, and in particular its assumption (2), in establishing the asymptotic normality of the frontier and efficiency estimators lies in establishing that the term $\hat{\sigma}(x)\left(s_R^{-1}-\sigma_R^{-1}\right)$ is of the same order as B_{0n}. This allows us to combine the asymptotic biases introduced by the local linear nonparametric estimation and the term introduced by the estimation of σ_R. Assumption (2) places an additional constraint on the DGP that goes beyond those in A1, A2 and A4. Informally, the assumption can be interpreted as a shape restriction on the marginal distribution - $F_R(r)$ of R_t that guarantees that for all $\epsilon>0$ as $n\to\infty$, $F_R^n(1-\epsilon)\to 0$ sufficiently fast.

Given the results described in theorems 1 and 2, standard bandwidth selection methods (Fan and Gijbels, 1995; Ruppert, Sheather and Wand, 1995) can be used to obtain a data driven h_n. These data driven bandwidth selection methods are asymptotically equivalent to an optimal bandwidth which is $O(n^{-1/5})$. In addition, as is typical in nonparametric regression, if there is undersmoothing the bias terms vanish asymptotically. In the next section we perform a simulation study that sheds some light on the estimators finite sample performance and compares it to the bias corrected FDH estimator of Park et al.(2000).

3.4 Monte Carlo Study

In this section we investigate some of the finite sample properties of our estimator, henceforth referred to as NP *via* a Monte Carlo. For comparison purposes, we also include in the study the bias corrected FDH estimator described in Park, Simar and Weiner (2000). Our simulations are based on model (13), i.e.,

$$Y_t = \frac{\sigma(X_t)}{\sigma_R} R_t \text{ with } K = 1.$$

We generate data with the following characteristics. The X_t are pseudo random variables from a uniform distribution with support given by $[10, 100]$. $R_t = exp(-Z_t)$ where Z_t are pseudo random variables from an exponential distribution with parameter $\beta > 0$, therefore R_t has support in $(0, 1]$. We consider two specifications for $\sigma(\cdot)$: $\sigma_1(x) = \sqrt{x}$ and $\sigma_2(x) = 0.0015x^2$, which are associated with production functions that admit decreasing and increasing returns to scale respectively. Three parameters for the exponential distribution were considered: $\beta_1 = 3$, $\beta_2 = 1$, $\beta_3 = 1/3$. These choices of parameters produce, respectively, the following values for the parameters of $g_{R|X}$: $(\mu_R, \sigma_R^2) = (0.25, 0.08), (0.5, 0.08)$, and $(0.75, 0.04)$. Three sample size $n = 100, 300, 600$ are considered and 1000 repetitions are performed for each alternative experimental design. We evaluate the frontiers and construct confidence intervals for efficiency at $(y_0, x_0) = (10, 32.5), (10, 55), (10, 77.5)$ for $\sigma_1(x)$ and at $(y_0, x_0) = (2.5, 32.5), (2.5, 55), (2.5, 77.5)$ for $\sigma_2(x)$. The values of X correspond to the 25^{th},50^{th}

and 75^{th} percentile of its support and the values of Y are arbitrarily chosen output levels below the frontier.

Given the convergence in (17) asymptotic confidence intervals for efficiency R_0 can be constructed. To construct a $1-\alpha$ confidence interval for R_0, we obtain a bandwidth h_n for $\hat{\sigma}(x)$ such that $nh_n^5 \to 0$ as $n \to \infty$ (undersmoothing) which eliminates the asymptotic bias. Hence, for quantiles $Z_{\frac{\alpha}{2}}$ and $Z_{1-\frac{\alpha}{2}}$ of a standard normal distribution we have

$$\lim_{n\to\infty}\{P(\hat{R}_0 - (\sqrt{nh})^{-1}\hat{\sigma}_0(x_0,R_0)Z_{1-\frac{\alpha}{2}} \leq R_0 \leq \hat{R}_0 - (\sqrt{nh})^{-1}\hat{\sigma}_0(x_0,R_0)Z_{\frac{\alpha}{2}})\} = 1-\alpha$$

where $\hat{\sigma_0}^2(x_0,R_0) = \frac{\hat{R}_0^2}{4\hat{g}_X(x_0)}(\hat{\mu}_4(x_0)-1)\int K^2(y)dy$, $K(\cdot)$ is the Epanechnikov kernel, $\hat{R}_0 = \frac{y_0}{\hat{\sigma}(x_0)}s_R$, $\hat{g}_X(x_0)$ is the Rosenblatt kernel density estimator and $\hat{\mu}_4 = \frac{1}{n}\sum_{t=1}^n\left(\frac{Y_t}{\hat{\sigma}(X_t)} - \hat{b}\right)^4$. The estimator $\hat{\mu}_4$ depends on an estimator for b which we define as $\hat{b} = \frac{\sum_{t=1}^n \hat{\sigma}(X_t)Y_t}{\sum_{t=1}^n \hat{\sigma}^2(X_t)}$. Consistency of this estimator is proved in Lemma 4 in the appendix.[19]

Confidence intervals for R_0 using the bias corrected FDH estimator are given in Park, Simar and Wiener(2000). We follow their suggestion and choose their constant C to be 1 and select their bandwidth (ξ) to be proportional to $n^{-1/3}$.

The evaluation of the overall performance of the efficiency estimator was based on three different measures. First, we consider the correlation between the efficiency rankings produced by the estimator and the true efficiency rankings:

$$R_{rank} = \frac{cov(rank(\hat{R}_t), rank(R_t))}{\sqrt{var(rank(\hat{R}_t))\, var(rank(R_t))}}$$
$$= \frac{\sum_{t=1}^n (rank(\hat{R}_t) - \overline{rank(\hat{R}_t)})(rank(R_t) - \overline{rank(R_t)})}{\sqrt{\sum_{t=1}^n (rank(\hat{R}_t) - \overline{rank(\hat{R}_t)})^2 \sum_{t=1}^n (rank(R_t) - \overline{rank(R_t)})^2}}$$

where $rank(R_t)$ gives the ranking index according to the magnitude of R_t and $\overline{rank(R_t)}$ is the mean of $rank(R_t)$. The closer R_{rank} for $\hat{R}_t$ is to 1, the higher the correlation between the

[19] Note that together, the consistency of s_R from Theorem 2 and Lemma 4 can be used to define a consistent estimator for μ_R, $\hat{\mu}_R = \hat{b}s_R$.

true R_t and $\hat{R}_t$, thus the better the estimator $\hat{R}_t$. The second measure we consider is

$$R_{mag} = \frac{1}{n}\sum_{t=1}^{n}(\hat{R}_t - R_t)^2$$

which is simply the squared Euclidean distance between the estimated vector of efficiencies and the true vector of efficiencies. The third measure we use is,

$$R_{rel} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{R}_t}{\hat{R}_i} - \frac{R_t}{R_i}\right|,$$

where i is the position index for $R_i = \max_{1\leq t\leq n} R_t$, and $\hat{R}_i$ is the i^{th} corresponding element in $\{\hat{R}_t\}_{t=1}^{n}$, which may or may not be the maximum of $\hat{R}_t$. Hence R_{rank}, R_{mag} summarize the performance of the estimator $\hat{R}_t$ in ranking and calculating the magnitude of efficiency. R_{rel} captures the relative efficiency. In our simulations we consider estimates $\hat{R}_t$ based on both our estimator and the bias corrected FDH estimator.

The results of our simulations are summarized in Tables 3.1, 3.2, 3.3 and 3.4. Table 3.1 provides the bias and mean squared error - MSE of s_R and $\hat{\sigma}(x)$ at three different values of x. Table 3.2 gives the bias and MSE of our estimator (NP) as well as those of the bias corrected FDH frontier estimator. Table 3.3 gives the empirical coverage probability (the frequency that the estimated confidence interval contains the true efficiency in 1000 repetitions) for efficiency for both estimators and Table 3.4 gives the overall performance of the efficiency estimators according to the measures described above. We first identify some general regularities on estimation performance.

General Regularities. As expected from the asymptotic results of section 3, as the sample size n increases, the bias and the MSE for s_R, $\hat{\sigma}(x)$, and the frontier estimator based on NP generally decrease, with some exceptions when it comes to the bias. The frontier estimator based on the bias corrected FDH also exhibits decreasing MSE and bias, with a number of exceptions in the latter case. We observe that the empirical coverage probability for NP is close to the true 95% and generally approaches 95% as n increases with exceptions for small μ_R, while that for FDH is usually below 95% and there is no clear evidence that they get closer to 95% as n increases. The asymptotics of both estimators seem to be confirmed in general terms as their performances improve with large n.

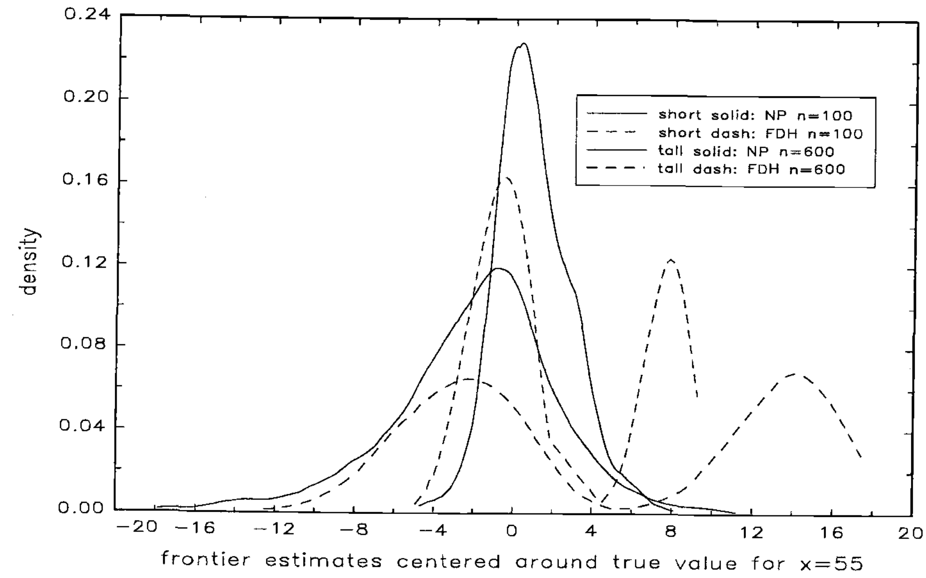
We now turn to the impact of different values of μ_R on the performance of NP and FDH. As μ_R increases, the bias and MSE of s_R increase, with the bias being generally negative except for small μ_R and small sample($n = 100$). The bias of $\hat{\sigma}(x)$, which is negative for $\sigma_1(x)$ and mostly positive for $\sigma_2(x)$, doesn't seem to be impacted by μ_R. Note that the sign of these biases is in accordance to what the asymptotic results predict due to the presence of $\sigma^{2(2)}(x)$ in the bias term. Also, in accordance to the asymptotic results derived in section 3, the MSE for $\hat{\sigma}(x)$ oscillates with μ_R, which reflects the fact that the variance of $\hat{\sigma}(x)$ depends on μ_R in a nonlinear fashion, as indicated by Theorem 1. The bias of the NP frontier estimator is generally positive, except for small μ_R and $n = 100$, and generally increases with μ_R except for the case where $n = 100$, whereas its MSE oscillates with μ_R. In general, the FDH frontier estimator has a positive bias, which together with MSE decreases with μ_R in most experiments, exceptions occuring when $\sigma(x) = 0.0015x^2$. No clear pattern is discerned from the impact of larger μ_R on the empirical coverage probability for NP, but there is weak evidence that FDH is improved. Regarding the measures of overall performance for the efficient estimator described above, the NP estimator seems to perform worse when μ_R is larger for R_{rank}, R_{mag} and R_{rel}. The FDH estimator performs worse when μ_R is larger and the performance measured considered is R_{rank}, while in the case of R_{mag} and R_{rel}, FDH performs better as μ_R increases for the case of $\sigma_1(x)$, but the performance oscillates when we consider $\sigma_2(x)$.

Lastly, as one would expect from the NP estimation procedure, the experimental results indicate that as measured by bias and MSE, the estimation of the NP frontier is less accurate and precise than that of $\sigma(x)$, since the NP frontier estimator involves the estimation of both $\sigma(x)$ and σ_R.

Relative Performance of Estimators. On estimating the production frontier (Table 3.2) there seems to be evidence that NP dominates FDH in terms of bias and MSE when $\mu_R = 0.25$ and $\mu_R = 0.5$, with exceptions in cases where $\sigma(x) = 0.0015x^2$, while FDH is better with $\mu_R = 0.75$. Regarding the empirical coverage probabilities (Table 3.3), the NP estimator is superior in most experiments, i.e., NP estimates are much closer to the intended probability $1 - \alpha = 95\%$. When the different measures of overall performance we considered are analyzed (Table 3.4), we observe that the NP estimator outperforms FDH in terms of R_{rank} and R_{rel}, except when $\mu_R = 0.75$ and $\sigma(x) = \sqrt{x}$. In terms of R_{mag}, NP generally outperforms FDH when $\mu_R = 0.25, 0.5$, while FDH is better when $\mu_R = 0.75$. Based on these results, it seems reasonable to conclude that when we are dealing with DGPs that produce inefficient and mediocre firms with large probability, then the fact that the NP estimator is impacted to a lesser degree by extreme values results in better performance *vis a vis* the FDH estimator, whose construction depends heavily on boundary points. This improved performance is easily perceived in Figure 3.1. The figure shows kernel density estimates for the frontier around the true value evaluated at $x = 55$ for NP $\left(\frac{\hat{\sigma}(x)}{s_R} - \frac{\sigma(x)}{\sigma_R}\right)$ and FDH $\left(\hat{\rho}_{FDH}(x) - \frac{\sigma(x)}{\sigma_R}\right)$ based on 1000 simulations, $\mu_R = 0.25$ and $\sigma(x) = \sqrt{x}$, for $n = 100$ and 600. The kernel density estimates were calculated using an Epanechnikov kernel and bandwidths were selected using the *rule-of-thumb* of Silverman(1986). We observe that the NP estimator is more tightly centered around the true frontier and shows the familiar symmetric bell shape, while that of FDH is generally bimodal with greater variability. Figure 1 also shows that the estimated densities become tighter with more acute spikes as the sample size increases, as expected from the available asymptotic results.[20]

[20] Similar graphs but with less dramatic differences between the NP and FDH estimators are obtained when

FIGURE 3.1 DENSITY ESTIMATES FOR NP AND FDH ESTIMATORS



$\mu_R = 0.5$.

3.5 Conclusion

In this paper we proposed a new nonparametric frontier model together with estimators for the frontier and associated efficiency levels of production units or plans. Our estimator can be viewed as an alternative to DEA and FDH estimators that are popular and have been widely used in the empirical literature. The estimator is easily implementable, as it is in essence a local linear kernel estimator, and we show that it is consistent and asymptotically normal when suitably normalized. Efficiency rankings and relative efficiency of firms are estimated based only on some rather parsimonious restrictions on conditional moments. The assumptions required to obtain the asymptotic properties of the estimator are standard in nonparametric statistics and are flexible enough to preserve the desirable generality that has characterized nonparametric deterministic frontier estimators. In contrast to DEA and FDH estimators, our estimator is not intrinsically biased but it does envelop the data, in the sense that no observation can lie above the estimated frontier. The small Monte Carlo study we perform seems to confirm the asymptotic results we have obtained and also seems to indicate that for a number of DGPs our proposed estimator can outperform bias corrected FDH according to various performance measures.

Our estimator together with DEA, FDH and the recently proposed estimator of Cazals et al. (2002) forms a set of procedures that can be used for estimating nonparametric deterministic frontiers and for which asymptotic distributional results are available. Future research on the relative performance of all of these alternatives under various DGPs would certainly be desirable from a theoretical and practical viewpoints. Furthermore, extensions of all such models and estimators to accommodate stochastic frontiers with minimal additional assumptions that result in identification is also desirable. Lastly, with regards to our estimator, an extension to the case of multiple outputs should be accomplished.

Acknowledgments. We thank R. Färe and S. Grosskopf for helpful discussions.

Table 3.1: Bias and MSE for S_R and $\hat{\sigma}(x)$

$\sigma_1(x)=\sqrt{x}$	n	S_R bias	S_R MSE($\times 10^{-1}$)	$\hat{\sigma}(x_1): x_1=32.5$ bias	$\hat{\sigma}(x_1): x_1=32.5$ MSE	$\hat{\sigma}(x_2): x_2=55$ bias	$\hat{\sigma}(x_2): x_2=55$ MSE	$\hat{\sigma}(x_3): x_3=77.5$ bias	$\hat{\sigma}(x_3): x_3=77.5$ MSE
$\mu_R=0.25$	100	0.010	0.005	-0.237	0.963	-0.322	1.608	-0.490	2.440
	300	-0.010	0.004	-0.080	0.281	-0.113	0.469	-0.137	0.698
	600	-0.012	0.003	-0.032	0.151	-0.057	0.240	-0.077	0.350
$\mu_R=0.5$	100	-0.026	0.012	-0.205	0.422	-0.284	0.761	-0.334	1.075
	300	-0.018	0.005	-0.070	0.139	-0.115	0.211	-0.078	0.295
	600	-0.014	0.003	-0.042	0.065	-0.064	0.110	-0.061	0.154
$\mu_R=0.75$	100	-0.046	0.026	-0.230	0.974	-0.304	1.601	-0.355	2.420
	300	-0.027	0.009	-0.124	0.334	-0.127	0.531	-0.115	0.758
	600	-0.019	0.005	-0.029	0.171	-0.044	0.247	-0.066	0.380

$\sigma_2(x)=0.0015x^2$	n	S_R bias	S_R MSE($\times 10^{-1}$)	$\hat{\sigma}(x_1): x_1=32.5$ bias	$\hat{\sigma}(x_1): x_1=32.5$ MSE	$\hat{\sigma}(x_2): x_2=55$ bias	$\hat{\sigma}(x_2): x_2=55$ MSE	$\hat{\sigma}(x_3): x_3=77.5$ bias	$\hat{\sigma}(x_3): x_3=77.5$ MSE
$\mu_R=0.25$	100	0.016	0.008	0.027	0.104	-0.121	0.722	-0.402	2.643
	300	-0.005	0.003	0.065	0.036	0.006	0.227	-0.070	0.849
	600	-0.010	0.003	0.058	0.020	0.028	0.123	-0.016	0.451
$\mu_R=0.5$	100	-0.024	0.014	0.036	0.046	-0.057	0.307	-0.272	1.386
	300	-0.015	0.004	0.057	0.017	0.001	0.102	-0.031	0.373
	600	-0.013	0.003	0.045	0.010	0.025	0.052	-0.025	0.213
$\mu_R=0.75$	100	-0.050	0.031	0.036	0.106	-0.166	0.782	-0.503	2.781
	300	-0.026	0.009	0.082	0.041	0.007	0.226	-0.108	0.899
	600	-0.020	0.005	0.057	0.022	0.028	0.128	-0.029	0.488

TABLE 3.2: BIAS AND MSE OF NONPARAMETRIC AND FDH FRONTIER ESTIMATORS

			$x_1 = 32.5$		$x_2 = 55$		$x_3 = 77.5$	
$\sigma_1(x) = \sqrt{x}$	n		NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	bias	-1.422	4.446	-1.909	4.673	-2.639	4.246
		MSE	12.895	89.907	20.873	95.434	32.012	90.600
	300	bias	0.486	2.939	0.604	2.998	0.713	3.225
		MSE	3.916	39.069	6.751	39.532	10.751	40.645
	600	bias	0.797	2.174	0.984	2.314	1.127	2.197
		MSE	2.854	21.894	4.650	23.276	6.273	22.135
$\mu_R = 0.5$	100	bias	1.290	3.432	1.611	3.522	1.929	3.046
		MSE	8.920	62.745	15.201	62.717	22.018	56.067
	300	bias	1.047	1.689	1.282	1.624	1.736	1.848
		MSE	3.243	20.337	5.218	19.569	8.239	21.934
	600	bias	0.835	0.999	1.052	1.257	1.303	1.094
		MSE	1.594	8.937	2.720	11.349	3.857	9.727
$\mu_R = 0.75$	100	bias	8.552	3.030	10.899	2.834	12.826	2.932
		MSE	255.763	53.209	325.684	50.862	349.823	50.477
	300	bias	4.126	1.577	5.633	1.397	6.881	1.362
		MSE	30.484	17.467	59.484	15.665	85.023	15.498
	600	bias	3.075	0.884	3.978	0.766	4.639	0.839
		MSE	16.200	7.512	26.989	6.595	38.007	7.520

			$x_1 = 32.5$		$x_2 = 55$		$x_3 = 77.5$	
$\sigma_2(x) = 0.0015x^2$	n		NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	bias	-0.184	0.535	-1.214	-1.561	-2.967	-4.712
		MSE	1.361	6.059	9.930	11.365	37.537	41.254
	300	bias	0.353	0.537	0.356	-0.810	0.407	-2.649
		MSE	0.604	2.780	3.327	4.591	11.454	15.303
	600	bias	0.429	0.456	0.729	-0.389	1.170	-1.983
		MSE	0.449	1.615	2.429	2.275	7.705	8.328
$\mu_R = 0.5$	100	bias	0.695	1.093	1.359	0.191	2.078	-1.566
		MSE	1.617	6.090	9.213	6.484	34.783	12.922
	300	bias	0.527	0.649	0.904	0.434	1.668	-0.762
		MSE	0.507	2.496	2.353	2.767	8.032	4.871
	600	bias	0.425	0.594	0.843	0.355	1.393	-0.450
		MSE	0.299	1.592	1.498	1.842	4.792	2.680
$\mu_R = 0.75$	100	bias	3.378	1.148	7.713	0.434	14.211	-1.152
		MSE	20.865	5.903	128.855	6.508	456.643	11.019
	300	bias	1.832	0.768	3.879	0.514	6.943	-0.496
		MSE	5.067	2.563	26.146	2.572	89.554	3.999
	600	bias	1.278	0.539	2.883	0.465	5.230	-0.125
		MSE	2.387	1.410	13.516	1.721	47.090	2.072

TABLE 3.3: EMPIRICAL COVERAGE PROBABILITY FOR $\hat{R}$ BY NONPARAMETRIC AND FDH FOR $1-\alpha = 95\%$

		$x_1 = 32.5, y_1 = 10$		$x_2 = 55, y_2 = 10$		$x_3 = 77.5, y_3 = 10$	
$\sigma_1(x) = \sqrt{x}$	n	NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	0.958	0.748	0.957	0.748	0.965	0.727
	300	0.984	0.776	0.981	0.771	0.976	0.792
	600	0.964	0.787	0.970	0.779	0.972	0.785
$\mu_R = 0.5$	100	0.994	0.810	0.987	0.825	0.996	0.801
	300	0.964	0.830	0.967	0.812	0.955	0.831
	600	0.946	0.827	0.952	0.846	0.951	0.839
$\mu_R = 0.75$	100	0.999	0.836	1.000	0.811	1.000	0.845
	300	0.979	0.855	0.966	0.843	0.968	0.836
	600	0.939	0.837	0.939	0.832	0.932	0.834

		$x_1 = 32.5, y_1 = 2.5$		$x_2 = 55, y_2 = 2.5$		$x_3 = 77.5, y_3 = 2.5$	
$\sigma_2(x) = 0.0015x^2$	n	NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	0.925	0.616	0.932	0.763	0.945	0.490
	300	0.921	0.680	0.970	0.766	0.975	0.557
	600	0.911	0.691	0.947	0.800	0.965	0.561
$\mu_R = 0.5$	100	0.957	0.735	0.985	0.757	0.993	0.773
	300	0.903	0.712	0.961	0.777	0.980	0.773
	600	0.879	0.756	0.944	0.730	0.956	0.782
$\mu_R = 0.75$	100	0.995	0.756	0.999	0.782	0.999	0.785
	300	0.945	0.777	0.984	0.767	0.979	0.774
	600	0.918	0.780	0.953	0.753	0.954	0.785

TABLE 3.4: OVERALL MEASURES OF EFFICIENCY ESTIMATORS BY NONPARAMETRIC AND FDH

		R_{rank}		R_{mag}		R_{rel}	
$\sigma_1(x) = \sqrt{x}$	n	NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	0.990	0.966	0.014	0.014	0.054	0.200
	300	0.997	0.986	0.002	0.006	0.026	0.120
	600	0.999	0.992	0.001	0.004	0.019	0.078
$\mu_R = 0.5$	100	0.966	0.934	0.034	0.012	0.074	0.200
	300	0.990	0.973	0.003	0.005	0.036	0.101
	600	0.996	0.986	0.001	0.003	0.024	0.067
$\mu_R = 0.75$	100	0.785	0.893	0.148	0.008	0.161	0.133
	300	0.893	0.962	0.017	0.002	0.086	0.059
	600	0.938	0.981	0.009	0.001	0.059	0.039

		R_{rank}		R_{mag}		R_{rel}	
$\sigma_2(x) = 0.0015x^2$	n	NP	FDH	NP	FDH	NP	FDH
$\mu_R = 0.25$	100	0.987	0.944	0.014	0.026	0.059	0.314
	300	0.996	0.976	0.008	0.013	0.033	0.202
	600	0.998	0.987	0.006	0.008	0.026	0.148
$\mu_R = 0.5$	100	0.956	0.830	0.072	0.033	0.091	0.427
	300	0.983	0.919	0.011	0.016	0.052	0.262
	600	0.990	0.951	0.009	0.009	0.039	0.189
$\mu_R = 0.75$	100	0.747	0.641	0.097	0.029	0.185	0.332
	300	0.863	0.841	0.056	0.011	0.111	0.176
	600	0.906	0.912	0.042	0.005	0.080	0.126

Chapter 4

A Note on the Use of V and U Statistics in Nonparametric Models of Regression

4.1 Introduction

The use of U-statistics (Hoeffding, 1948) as a means of obtaining asymptotic properties of kernel based estimators in semiparametric and nonparametric regression models is now common practice in econometrics. For example, Powell, Stock and Stoker (1989) use U-statistics to obtain the asymptotic properties of their semiparametric index model estimator. Ahn and Powell (1993) use U-statistics to investigate the properties of a two stage semiparametric estimation of censored selection models where the first stage estimator involves a nonparametric regression estimator for the selection variable. Fan and Li (1996) use U-statistics to consider a set of specification tests for nonparametric regression models and Zheng (1998) uses U-statistics to provide a nonparametric kernel based test for parametric quantile regression models.[21]

The appeal for the use of U-statistics derives from the fact that when estimators of interest are expressed as linear combinations of semiparametric and/or nonparametric estimators, their component parts (see section 3 for examples) can normally be expressed as k-dimensional sums. Thus, obtaining the asymptotic characteristics of these component parts normally involves a two-stage strategy. First, the k-dimensional sum is broken up into two parts, the first involving a single indexed sum and the second takes the form of a U-statistic. The first term is normally handled by a suitable central limit theorem or law of large numbers. The U-statistic component is often studied using Hoeffding's(1961) H-decomposition which breaks it into an average of independent and identically distributed terms (projection) and a remainder term

[21] See also Kemp(2000) and D'Amico(2003) for more recent uses.

that is orthogonal to the space in which the projection lies and is of order smaller than that of the projection (see Lee, 1990 and Serfling, 1980).

In this note we show that a convenient approach to establishing the asymptotic properties of these estimators is to consider the k-dimensional component sums directly via *von Mises'* V-statistics. To this end we show that under suitable conditions V-statistics are $\sqrt{n}$ asymptotically equivalent to U-statistics. Combined with the use of Hoeffding's H-decomposition our results establish the $\sqrt{n}$ asymptotic equivalence of V-statistics and the corresponding U-statistic projection. The remainder of the paper is structured as follows. We introduce the asymptotic equivalence result of V and U statistics in section 2. Applications to the estimation of conditional variance in nonparametric regression models and constructing nonparametric R-square are provided in section 3. A brief conclusion is given in section 4.

4.2 Asymptotic Equivalence of U and V Statistics

Let $\{Z_i\}_{i=1}^n$ be a sequence of i.i.d. random variables and $\psi_n(Z_1, \cdots, Z_k)$ be a symmetric function with $k \leq n$. We call $\psi_n(Z_1, \cdots, Z_k)$ a kernel function and a k-dimensional U-statistic will be denoted by u_n which is defined as,

$$u_n = \binom{n}{k}^{-1} \sum_{(n,k)} \psi_n(Z_{i_1}, \cdots, Z_{i_k})$$

where $\sum_{(n,k)}$ denotes a sum over all subsets $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ of $\{1, 2, \cdots, n\}$. A k-dimensional V-statistic, denoted by v_n is defined as,

$$v_n = n^{-k} \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \psi_n(Z_{i_1}, \cdots, Z_{i_k}).$$

The following Theorem establishes the $\sqrt{n}$ asymptotic equivalence of U and V-statistics under suitable conditions.

Theorem 1 Let $\{Z_i\}_{i=1}^n$ be a sequence of i.i.d. random variables and u_n and v_n be U and V statistics with kernel function $\psi_n(Z_1, \cdots, Z_k)$. If $E\left(\psi_n^2(Z_{i_1}, \cdots, Z_{i_k})\right) = o(n)$, for all $1 \leq i_1, ..., i_k \leq n$, $k \leq n$, then $u_n - v_n = o_p(n^{-1/2})$.

Proof Let $\sum_{c'}$ denote the sum over all $(i_1, ..., i_k)$ subsets of $\{1, 2, ..., n\}$ such that at least one index $i_j = i_l$ for $j \neq l$ and $P_k^n = n!/(n-k)!$. Simple algebra gives $u_n - v_n = L_n(u_n - w_n)$ where $L_n = 1 - P_k^n/n^k$ and $w_n = (n^k - P_k^n)^{-1} \sum_{c'} \psi_n(Z_{i_1}, ..., Z_{i_k})$. We start by showing that $\sum_{c'} \psi_n(Z_{i_1}, ..., Z_{i_k})$ is a U-statistic. First, we write

$$\sum_{c'} \psi_n(Z_{i_1}, ..., Z_{i_k}) = \sum_{c_0} \psi_n(Z_{i_1}, ..., Z_{i_k}) + \sum_{c_1} \psi_n(Z_{i_1}, ..., Z_{i_k}) + \cdots + \sum_{c_{k-2}} \psi_n(Z_{i_1}, ..., Z_{i_k})$$

where $\sum_{c_r}$ is the sum over all $1 \leq i_1, i_2, ..., i_k \leq n$ such that r indices differ from the others. Note that if $\psi_n(Z_{i_1}, ..., Z_{i_{k-1}})$ is symmetric and $2 \leq k < n$

$$\sum_{i_1=1}^{n-(k-2)} \sum_{i_2=i_1+1}^{n-(k-3)} \cdots \sum_{i_{k-1}=i_{k-2}+1}^{n} \psi_n(Z_{i_1}, ..., Z_{i_{k-1}}) = (n-k+1)^{-1} \sum_{(n,k)} \psi_n'(Z_{i_1}, ..., Z_{i_k}) \qquad (18)$$

where $\psi_n'(Z_{i_1}, ..., Z_{i_k}) = \psi_n(Z_{i_1}, ..., Z_{i_{k-1}}) + \psi_n(Z_{i_1}, ..., Z_{i_{k-2}}, Z_{i_k}) + \cdots + \psi_n(Z_{i_2}, ..., Z_{i_k})$. Hence, by repeated use of (18) we have,

$$\begin{aligned}
\sum_{c_0} \psi_n(Z_{i_1}, ..., Z_{i_k}) &= (n-1)^{-1} \sum_{(n,2)} \left(\psi_n(Z_{i_1}, ..., Z_{i_1}) + \psi_n(Z_{i_2}, ..., Z_{i_2})\right) \\
&= (n-1)^{-1} \sum_{(n,2)} \phi_{2,n}^{(0)}(Z_{i_1}, Z_{i_2}) \\
\text{where} \quad & \phi_{2,n}^{(0)}(Z_{i_1}, Z_{i_2}) = \psi_n(Z_{i_1}, ..., Z_{i_1}) + \psi_n(Z_{i_2}, ..., Z_{i_2}), \text{ similarly below.} \\
&= ((n-1)(n-2))^{-1} \sum_{(n,3)} (\phi_{2,n}^{(0)}(Z_{i_1}, Z_{i_2}) + \phi_{2,n}^{(0)}(Z_{i_1}, Z_{i_3}) + \phi_{2,n}^{(0)}(Z_{i_2}, Z_{i_3})) \\
&= ((n-1)(n-2))^{-1} \sum_{(n,3)} \phi_{3,n}^{(0)}(Z_{i_1}, Z_{i_2}, Z_{i_3}) \\
&\ \vdots \\
&= ((n-1)(n-2) \cdots (n-(k-1)))^{-1} \sum_{(n,k)} \phi_{k,n}^{(0)}(Z_{i_1}, Z_{i_2}, Z_{i_3}, \cdots, Z_{i_k})
\end{aligned}$$

$$\begin{aligned}
\sum_{c_1}\psi_n(Z_{i_1},...,Z_{i_k}) &= \binom{k}{1}\sum_{(n,2)}(\psi_n(Z_{i_1},Z_{i_2}...,Z_{i_2})+\psi_n(Z_{i_2},Z_{i_1},...,Z_{i_1}))\\
&= \binom{k}{1}\sum_{(n,2)}\phi^{(1)}_{2,n}(Z_{i_1},Z_{i_2})\\
&= \binom{k}{1}(n-2)^{-1}\sum_{(n,3)}(\phi^{(1)}_{2,n}(Z_{i_1},Z_{i_2})+\phi^{(1)}_{2,n}(Z_{i_1},Z_{i_3})+\phi^{(1)}_{2,n}(Z_{i_2},Z_{i_3}))\\
&= \binom{k}{1}(n-2)^{-1}\sum_{(n,3)}\phi^{(1)}_{3,n}(Z_{i_1},Z_{i_2},Z_{i_3})\\
&\vdots\\
&= \binom{k}{1}((n-2)\cdots(n-(k-1)))^{-1}\sum_{(n,k)}\phi^{(1)}_{k,n}(Z_{i_1},Z_{i_2},Z_{i_3},\cdots,Z_{i_k})
\end{aligned}$$

$$\begin{aligned}
\sum_{c_2}\psi_n(Z_{i_1},...,Z_{i_k}) &= \binom{k}{2}\sum_{(n,3)}(\psi_n(Z_{i_1},Z_{i_2},Z_{i_3},...,Z_{i_3})+\psi_n(Z_{i_2},Z_{i_1},...,Z_{i_3})\\
&+ \psi_n(Z_{i_1},Z_{i_3},...,Z_{i_2})+\psi_n(Z_{i_3},Z_{i_1},Z_{i_2},...,Z_{i_2})\\
&+ \psi_n(Z_{i_2},Z_{i_3},...,Z_{i_1})+\psi_n(Z_{i_3},Z_{i_2},...,Z_{i_1}))\\
&= \binom{k}{2}\sum_{(n,3)}\phi^{(2)}_{3,n}(Z_{i_1},Z_{i_2},Z_{i_3})\\
&= \binom{k}{2}(n-3)^{-1}\sum_{(n,4)}\phi^{(2)}_{4,n}(Z_{i_1},Z_{i_2},Z_{i_3},Z_{i_4})\\
&\vdots\\
&= \binom{k}{2}((n-3)\cdots(n-(k-1)))^{-1}\sum_{(n,k)}\phi^{(2)}_{k,n}(Z_{i_1},Z_{i_2},Z_{i_3},\cdots,Z_{i_k})
\end{aligned}$$

Expressing each sum in a similar way we finally have

$$\sum_{c_{k-2}}\psi_n(Z_{i_1},...,Z_{i_k})=\binom{k}{k-2}(n-(k-1))^{-1}\sum_{(n,k)}\phi^{(k-2)}_{k,n}(Z_{i_1},Z_{i_2},Z_{i_3},\cdots,Z_{i_k})$$

and therefore

$$\begin{aligned}
\sum_{c'}\psi_n(Z_{i_1},...,Z_{i_k}) &= \binom{n}{k}^{-1}\sum_{(n,k)}\sum_{r=0}^{k-2}\frac{n(n-1)\cdots(n-r)\binom{k}{r}}{k!}\phi^{(r)}_n(Z_{i_1},\cdots,Z_{i_k})\\
&= \binom{n}{k}^{-1}\sum_{(n,k)}\phi'_n(Z_{i_1},\cdots,Z_{i_k})
\end{aligned}$$

where $\phi'_n(Z_{i_1},\cdots,Z_{i_k})$ is symmetric by construction. Hence,

$$\begin{aligned} u_n - v_n &= L_n\left(u_n - (n^k - P_k^n)^{-1}\binom{n}{k}^{-1}\sum_{(n,k)}\phi'_n(Z_{i_1},\cdots,Z_{i_k})\right) \\ &= L_n\left(\binom{n}{k}^{-1}\sum_{(n,k)}\left(\psi_n(Z_{i_1},\cdots,Z_{i_k}) - (n^k-P_k^n)^{-1}\phi'_n(Z_{i_1},\cdots,Z_{i_k})\right)\right) \\ &= L_n\binom{n}{k}^{-1}\sum_{(n,k)}H'_n(Z_{i_1},\cdots,Z_{i_k}) = L_n\frac{1}{n!}\sum_p W(Z_{i_1},\cdots,Z_{i_n}) \end{aligned}$$

where $W(Z_{i_1},\cdots,Z_{i_n}) = \gamma^{-1}(H'_n(Z_1,\cdots,Z_k)+H'_n(Z_{k+1},\cdots,Z_{2k})+\cdots+H'_n(Z_{\gamma k-k+1},\cdots,Z_{\gamma k}))$, $\gamma=[n/k]$ is the greatest integer $\leq n/k$, $\sum_p$ denotes the sum over all permutations $(i_1,...,i_n)$ of $\{1,2,...,n\}$. Now, $E(u_n-v_n) = L_nE(u_n-w_n)$ and $E(u_n-v_n)^2 = L_n^2E\left(\frac{1}{n!}\sum_p W(Z_{i_1},\cdots,Z_{i_n})\right)^2$. By Minkowski's inequality,

$$E(u_n-v_n)^2 \leq L_n^2 n!^{-2}\left(\sum_p (E|W(Z_{i_1},\cdots,Z_{i_n})|^2)^{1/2}\right)^2 = L_n^2E|W(Z_{i_1},\cdots,Z_{i_n})|^2$$

and by Chebyshev's inequality, for all $\epsilon>0$, $P(n^{1/2}|u_n-v_n|>\epsilon) \leq \frac{nL_n^2E|W(Z_{i_1},\cdots,Z_{i_n})|^2}{\epsilon^2}$. Hence, to conclude the proof it suffices to show that $nL_n^2E|W(Z_{i_1},\cdots,Z_{i_n})|^2 = o(1)$. Now, $W(Z_{i_1},\cdots,Z_{i_n}) = \gamma^{-1}\sum_{j=1}^{\gamma}H'_n(Z_{i_{jk-k+1}},...,Z_{i_{jk}})$ where $\{H'_n(Z_{i_{jk-k+1}},...,Z_{i_{jk}})\}_{j=1}^{\gamma}$ is an iid sequence with zero expectation. Since $n^k-P_k^n = O(n^{k-1})$ and $\frac{n(n-1)(n-2)...(n-r)}{n^{k-1}} = O(n^{r-k+2})$, $max_rO(n^{r-k+2}) = O(1)$ for $r=0,...,k-2$, . Hence,

$$(n^k-P_k^n)^{-1}\phi'_n(Z_{i_1},\cdots,Z_{i_k}) = O(max_r|\psi_n(Z_{i_1},\cdots,Z_{i_k})|)$$

where the maximum is taken over all $\psi_n(Z_{i_1},\cdots,Z_{i_k})$ in $max_r|\phi_n^{(r)}(Z_{i_1},\cdots,Z_{i_k})|$. Therefore, since

$$H'_n(Z_{i_1},\cdots,Z_{i_k}) = \psi_n(Z_{i_1},\cdots,Z_{i_k}) - (n^k-P_k^n)^{-1}\phi'_n(Z_{i_1},\cdots,Z_{i_k})$$

we have that $H'_n(Z_{i_1},\cdots,Z_{i_k}) = O(max\{\psi_n(Z_{i_1},\cdots,Z_{i_k}), max_r|\psi_n(Z_{i_1},\cdots,Z_{i_k})|\})$ and so there exists $\Delta>0$ such that for all n

$$E\left|H'_n(Z_{i_1},\cdots,Z_{i_k})\right|^2 < \Delta^2E\psi_n^2(Z_{i_1},\cdots,Z_{i_k}) = o(n)$$

where the last equality is by assumption. Now,

$$E|W(Z_{i_1},\cdots,Z_{i_n})|^2 = \frac{1}{\gamma^2}E|\sum_{j=1}^{\gamma} H_n'(Z_{i_{jk-k+1}},\cdots,Z_{jk})|^2 \leq \frac{1}{\gamma}o(n)$$

Hence,

$$nE(u_n - v_n)^2 \quad \leq \quad n^2 L_n^2 \frac{1}{\gamma} o(1) = O(1)o(1) = o(1) \text{ since } n^2 L_n^2 = O(1) \text{ and } \gamma \to \infty \text{ as } n \to \infty$$

Corollary 1 Let $\{Z_i\}_{i=1}^n$ be a sequence of i.i.d. random variables and u_n and v_n be U and V statistics with kernel function $\psi_n(Z_1,\cdots,Z_k)$. In addition, let $\hat{u}_n = \frac{k}{n}\sum_{i=1}^n(\psi_{1n}(Z_i) - \theta_n) + \theta_n$, where $\psi_{1n}(Z_i) = E(\psi_n(Z_1,\cdots,Z_k)|Z_i)$ and $\theta_n = E(\psi_n(Z_1,\cdots,Z_k))$.
If $E(\psi_n^2(Z_1,\cdots,Z_k)) = o(n)$ then $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$.

Proof From **Theorem 1** we have that $\sqrt{n}(v_n - u_n) = o_p(1)$ and from Lemma 2.1 in Lee(1988) we have that $\sqrt{n}(u_n - \hat{u}_n) = o_p(1)$. Hence, $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$.

4.3 Some Applications in Regression Model

We provide applications of **Theorem 1** and its corollary around the following regression model

$$y_t = m(x_t) + \sigma(x_t)\epsilon_t \tag{19}$$

where $\{y_t, x_t\}_{t=1}^n$ is a sequence of iid random variables $y_t, x_t \in \Re$ with joint density $q(y,x)$ for all t, $E(\epsilon_t|x_t) = 0$, and $Var(\epsilon_t|x_t) = 1$. This is similar to the regression model considered by Fan and Yao(1998), with the exception that here the observations are iid rather than strictly stationary. Our applications all involve establishing the asymptotic distribution of statistics that are constructed as linear combinations of a first stage nonparametric estimator of $m(x)$. As the first stage estimator for $m(x)$ we consider the local linear estimator of Fan(1992), i.e., for $x \in \Re$, we obtain $\hat{m}(x) = \hat{\alpha}$ where

$$(\hat{\alpha}, \hat{\beta}) = argmin_{\alpha,\beta} \sum_{t=1}^{n} (Y_t - \alpha - \beta(X_t - x))^2 K\left(\frac{X_t - x}{h_n}\right)$$

where $K(\cdot)$ is a density function and $0 < h_n \to 0$ as $n \to \infty$ is a bandwidth. We make the following assumptions.

A1. $0 < \underline{\mathbf{B}}_g \leq g(x) \leq \bar{B}_g < \infty$ for all $x \in G$, G a compact subset of $\Re$, where g is the common marginal density of x_t.

A2. 1. $0 < \underline{\mathbf{B}}_\sigma \leq \sigma(x) \leq \bar{B}_\sigma < \infty$ for all $x \in \Re$, $\sigma^2(\cdot) : \Re \to \Re$ is a measurable twice continuously differentiable function in $\Re$ with $|\sigma^{2(2)}(x)| < \bar{B}_{2\sigma}$ for all $x \in \Re$. 2. $0 < \underline{\mathbf{B}}_m \leq m(x) \leq \bar{B}_m < \infty$ for all $x \in \Re$, $m(\cdot) : \Re \to \Re$ is a measurable twice continuously differentiable function in $\Re$ with $|m^{2(2)}(x)| < \bar{B}_{2m}$ for all $x \in \Re$.

A3. $K(\cdot) : \Re \to \Re$ is a symmetric density function with bounded support $S_K \in \Re^K$ satisfying 1. $\int xK(x)dx = 0$; 2. $\int x^2 K(x)dx = \sigma_K^2$; 3. for all $x \in \Re^K$, $|K(x)| < B_K < \infty$; 4. for all $x, x' \in \Re^K$, $|K(x) - K(x')| < m_K|x - x'|$ for some $0 < m_K < \infty$;

A4. For all $x, x' \in \Re$, $|g(x) - g(x')| < m_g|x - x'|$ for some $0 < m_g < \infty$.

A5. $nh_n^2 \to \infty$, and $nh_n^3(|ln(h_n)|)^{-1} \to \infty$.

The following two lemmas provide auxiliary results that are useful in the applications that follow. We note that the proofs of these lemmas are themselves facilitated by using **Theorem 1** and its corollary. In addition, the proofs rely on repeated use of a version of Lebesgue's dominated convergence theorem which can be found Prakasa-Rao(1983, p.35). Henceforth, we refer to this result as *the proposition of Prakasa-Rao.* In addition, the proofs rely on Lemma 2 and Theorem 1 in Martins-Filho and Yao(2003).

Lemma 1 Let $f(x_t, y_t)$ be continuous at x_t and assume that $E(f^2(x_t, y_t)), E(y_t^2 f^2(x_t, y_t)) < \infty$. Let $I_n = \frac{1}{n}\sum_{t=1}^{n}(\hat{m}(x_t) - m(x_t))f(x_t, y_t)$, then given equation (19) and assumptions A1 - A5 we have that

$$I_n = \frac{1}{n}\sum_{t=1}^{n}\sigma(x_t)\epsilon_t \int f(x_t, y)g(y|x_t)dy + \frac{1}{2}h_n^2\sigma_K^2 E(m^{(2)}(x_t)f(x_t, y_t)) + o_p(n^{-\frac{1}{2}}) + o_p(h_n^2).$$

Proof Note that,

$$\begin{aligned}
&\hat{m}(x_t) - m(x_t) \\
=& \frac{1}{nhg(x_t)}\sum_{k=1}^{n}K(\frac{x_k - x_t}{h_n})\sigma(x_k)\epsilon_k + \frac{1}{2nhg(x_t)}\sum_{k=1}^{n}K(\frac{x_k - x_t}{h_n})m^{(2)}(x_{kt})(x_k - x_t)^2 + w(x_t)
\end{aligned}$$

where $w(x_t) = \hat{m}(x_t) - m(x_t) - \frac{1}{nhg(x_t)}\sum_{k=1}^{n}K(\frac{x_k - x_t}{h_n})y_k^*$ and $y_k^* = \sigma(x_k)\epsilon_k + \frac{1}{2}m^{(2)}(x_{kt})(x_k - x_t)^2$ with $x_{kt} = \lambda x_t + (1-\lambda)x_k$ for some $\lambda \in (0,1)$. Hence, $I_n = I_{1n} + I_{2n} + I_{3n}$, where

$$\begin{aligned}
I_{1n} &= \frac{1}{n^2 h_n}\sum_{t=1}^{n}\sum_{k=1}^{n}K(\frac{x_k - x_t}{h_n})\sigma(x_k)\epsilon_k f(x_t, y_t)\frac{1}{g(x_t)} \\
I_{2n} &= \frac{h_n}{2n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}K(\frac{x_k - x_t}{h_n})(\frac{x_k - x_t}{h_n})^2 m^{(2)}(x_{kt})f(x_t, y_t)\frac{1}{g(x_t)} \\
I_{3n} &= \frac{1}{n}\sum_{t=1}^{n}w(x_t)f(x_t, y_t)
\end{aligned}$$

We treat each term separately. Observe that I_{1n} can be written as

$$\begin{aligned}
I_{1n} &= \frac{1}{2n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\left(\frac{1}{h_n}K(\frac{x_k - x_t}{h_n})\sigma(x_k)\epsilon_k f(x_t, y_t)\frac{1}{g(x_t)}\right. \\
&\quad \left. + \frac{1}{h_n}K(\frac{x_t - x_k}{h_n})\sigma(x_t)\epsilon_t f(x_k, y_k)\frac{1}{g(x_k)}\right) \\
&= \frac{1}{2n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}(h_{tk} + h_{kt}) = \frac{1}{2n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\psi_n(Z_t, Z_k) = \frac{1}{2}v_n
\end{aligned}$$

where v_n is a V-statistic with $Z_t = (x_t, y_t)$ and $\psi_n(Z_t, Z_k)$ a symmetric kernel. Hence by the corollary to **Theorem 1** if $\frac{1}{n}E\left(\psi_n^2(Z_t, Z_k)\right) = o(1)$, then $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$. First, note that $\frac{1}{n}E\left(\psi_n^2(Z_t, Z_k)\right) = \frac{2}{n}E(h_{tk}^2) + \frac{2}{n}E(h_{tk}h_{kt})$ and since $E(\epsilon_k^2|x_1, \cdots, x_n) = 1$, we have from the proposition of Prakasa-Rao that

$$\frac{1}{n}E(h_{tk}^2) = \frac{1}{nh_n^2}E\left(K^2(\frac{x_k - x_t}{h_n})\sigma^2(x_k)E(f^2(x_t, y_t)|x_1, \cdots, x_n)\frac{1}{g^2(x_t)}\right) \to 0,$$

provided that $nh_n \to \infty$. Similarly it can be shown that $E(\frac{1}{n}h_{tk}h_{kt}) = o(1)$ and therefore $\frac{1}{n}E(\psi_n^2(Z_t, Z_k)) = o(1)$. Now, since $E(\epsilon_t|x_1, \cdots, x_n) = 0$, $\theta_n = E\left(\psi_n(Z_t, Z_k)\right) = 0$ and $\hat{u}_n = \frac{2}{n}\sum_{t=1}^{n}\psi_{1n}(Z_t)$. Therefore,

$$\begin{aligned}\psi_{1n}(Z_t) &= \frac{1}{h_n}\sigma(x_t)\epsilon_t \int\int K(\frac{x_k - x_t}{h_n})f(x_k, y_k)g(y_k|x_k)dx_k dy_k \\ &= \sigma(x_t)\epsilon_t \int f(x_t, \alpha)g(\alpha|x_t)d\alpha + o(1),\end{aligned}$$

where $g(y|x) = \frac{g(y,x)}{g(x)}$. Consequently,

$$I_{1n} = \frac{1}{n}\sum_{t=1}^{n}\sigma(x_t)\epsilon_t \int f(x_t, y)g(y|x_t)dy + o_p(n^{-\frac{1}{2}}) \qquad (20)$$

given that $\frac{1}{n}\sum_{t=1}^{n}\sigma(x_t)\epsilon_t = O_p(n^{-\frac{1}{2}})$.

Now, note that $I_{2n} = \frac{1}{4n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}(h_{tk} + h_{kt}) = \frac{1}{4n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\psi_n(Z_t, Z_k) = \frac{1}{4}v_n$ where $h_{tk} = h_n K(\frac{x_k - x_t}{h_n})(\frac{x_k - x_t}{h_n})^2 m^{(2)}(x_{kt})f(x_t, y_t)\frac{1}{g(x_t)}$. Again by the Corollary if $E(\psi_n^2(Z_t, Z_k)) = o(1)$, then $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$. Hence we focus on $\hat{u}_n$. Note that, $\hat{u}_n = \frac{2}{n}\sum_{t=1}^{n}\psi_{1n}(Z_t) - \theta_n$, where $\psi_{1n}(Z_t) = E(h_{tk}|Z_t) + E(h_{kt}|Z_t)$ and $E(\hat{u}_n) = \theta_n = 2E(h_{tk})$. Hence, by the proposition of Prakasa-Rao,

$$\begin{aligned}E\left(\frac{\hat{u}_n}{h_n^2}\right) &= \frac{2}{h_n}\int\int\int K\left(\frac{x_k - x_t}{h_n}\right)\left(\frac{x_k - x_t}{h_n}\right)^2 m^{(2)}(x_{kt})f(x_t, y_t)g(x_t, y_t)dx_k dx_t dy_t \\ &\to \sigma_K^2 E(m^{(2)}(x_t)f(x_t, y_t)).\end{aligned}$$

In addition, note that

$$\begin{aligned}V\left(\frac{1}{h_n^2}\hat{u}_n\right) &= \frac{4}{n^2h_n^4}nV(E(h_{tk}|Z_t) + E(h_{kt}|Z_t)) \\ &= \frac{4}{nh_n^4}\left(E(E(h_{tk}|Z_t))^2 + E(E(h_{kt}|Z_t))^2 + 2E(E(h_{tk}|Z_t)E(h_{kt}|Z_k)) - \theta_n^2\right)\end{aligned}$$

Under our assumptions it is straightforward to show that

$$\frac{1}{nh_n^4}E(E(h_{tk}|Z_t))^2 = \frac{1}{nh_n^2}E\left(f(x_t, y_t)\frac{1}{g^2(x_t)}E^2(K(\frac{x_k - x_t}{h_n})(\frac{x_k - x_t}{h_n})^2 m^{(2)}(x_{kt})|Z_t)\right) \to 0$$

as $n \to \infty$. Similarly, $\frac{1}{nh_n^4}E(E(h_{kt}|Z_t))^2 \to 0$ and $\frac{1}{nh_n^4}E(E(h_{tk}|Z_t)E(h_{kt}|Z_t)) \to 0$ as $n \to \infty$. Hence, $V\left(\frac{1}{h_n^2}\hat{u}_n\right) \to 0$ and by Markov's inequality we conclude that $\hat{u}_n = 2h_n^2\sigma_K^2 E(m^{(2)}(x_t)f(x_t,y_t)) + o_p(h_n^2)$ and by the corollary we have

$$I_{2n} = \frac{1}{2}\sigma_K^2 E(m^{(2)}(x_t)f(x_t,y_t)) + o_p(h_n^2). \tag{21}$$

Finally, verification that $E(\psi_n^2(Z_t,Z_k)) = 2E(h_{tk}^2) + 2E(h_{tk}h_{kt}) = o(n)$ follows directly from our assumptions by using the proposition of Prakasa-Rao. From Martins-Filho and Yao's(2003) Lemma 2 and Theorem 1

$$|w(x_t)| = |\hat{m}(x_t) - m(x_t) - \frac{1}{nhg(x_t)}\sum_{k=1}^{n} K(\frac{x_k - x_t}{h_n})y_k^*| = O_p(R_{n,1}(x_t))$$

and $\sup_{x_t\in G}|R_{n,1}(x_t)| = o_p(h_n^2)$, hence $\sup_{x_t\in G}|w(x_t)| = o_p(h_n^2)$. As a result we have that $|I_{3n}| \le o_p(h_n^2)\frac{1}{n}\sum_{t=1}^{n}|f(x_t,y_t)| = o_p(h_n^2)$, since $E(f^2(x_t,y_t)) < \infty$. Therefore, $I_{3n} = \frac{1}{n}\sum_{t=1}^{n} w(x_t)f(x_t,y_t) = o_p(h_n^2)$. Combining this result with (20) and (21) proves the lemma.

Lemma 2 Let $f(x_t,y_t)$ be continuous at x_t and assume that $E(f^2(x_t,y_t)), E(y_t^4 f^2(x_t,y_t)) < \infty$. Let $I_n = \frac{1}{n}\sum_{t=1}^{n}(\hat{m}(x_t) - m(x_t))^2 f(x_t,y_t)$, then given equation (19) and assumptions A1 - A5 we have that

$$\begin{aligned} I_n &= \frac{\sigma_K^2 h_n^2}{n}\sum_{t=1}^{n}\sigma(x_t)\epsilon_t m^{(2)}(x_t)E(f(x_t,y)|x_1,\cdots,x_n) \\ &+\frac{1}{4}h_n^4\sigma_K^4 E(m^{(2)}(x_t)f(x_t,y_t)) + o_p(n^{-\frac{1}{2}}) + o_p(h_n^3). \end{aligned}$$

Proof Note that

$$\begin{aligned} I_n &= \frac{1}{n^3h_n^2}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} f(x_t,y_t)\frac{1}{g^2(x_t)}K(\frac{x_k - x_t}{h_n})K(\frac{x_l - x_t}{h_n})\sigma(x_k)\sigma(x_l)\epsilon_k\epsilon_l \\ &+ \frac{h_n^2}{4n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{1}{g^2(x_t)}f(x_t,y_t)K(\frac{x_k - x_t}{h_n})(\frac{x_k - x_t}{h_n})^2 K(\frac{x_l - x_t}{h_n})(\frac{x_l - x_t}{h_n})^2 \end{aligned}$$

$$
\begin{aligned}
& \times m^{(2)}(x_{kt})m^{(2)}(x_{lt}) + \frac{1}{n}\sum_{t=1}^{n} w^2(x_t)f(x_t,y_t) \\
+ & \frac{1}{n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{1}{g^2(x_t)}f(x_t,y_t)K(\frac{x_k-x_t}{h_n})K(\frac{x_l-x_t}{h_n})(\frac{x_l-x_t}{h_n})^2 m^{(2)}(x_{lt})\sigma(x_k)\epsilon_k \\
+ & \frac{2}{n^2h_n}\sum_{t=1}^{n}\sum_{k=1}^{n} w(x_t)\frac{1}{g(x_t)}f(x_t,y_t)K(\frac{x_k-x_t}{h_n})\sigma(x_k)\epsilon_k \\
+ & \frac{h_n}{n^2}\sum_{t=1}^{n}\sum_{k=1}^{n} w(x_t)\frac{1}{g(x_t)}f(x_t,y_t)K(\frac{x_k-x_t}{h_n})(\frac{x_k-x_t}{h_n})^2 m^{(2)}(x_{kt}) \\
= & I_{1n} + I_{2n} + I_{3n} + I_{4n} + I_{5n} + I_{6n}
\end{aligned}
$$

We examine each term separately. $I_1 = \frac{1}{6}\frac{1}{n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}(h_{tkl}+h_{tlk}+h_{ktl}+h_{klt}+h_{ltk}+h_{lkt}) = \frac{1}{6}\frac{1}{n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\psi_n(Z_t,Z_k,Z_l) = \frac{1}{6}v_n$, where $h_{tkl} = \frac{1}{h_n^2}K(\frac{x_k-x_t}{h_n})K(\frac{x_l-x_t}{h_n})\sigma(x_k)\sigma(x_l)\epsilon_k\epsilon_l f(x_t,y_t)\frac{1}{g^2(x_t)}$. By the corollary to **Theorem 1** if $E(\psi_n^2(Z_t,Z_k,Z_l)) = o(n)$ then $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$. Since $\psi_{1n}(Z_t) = 0$ and $\theta_n = 0$, we have that $\hat{u}_n = 0$ and therefore $\sqrt{n}I_{1n} = o_p(1)$. To verify that $\frac{1}{n}E(\psi_n^2(Z_t,Z_k,Z_l)) = o(1)$ we note $\frac{1}{n}E(\psi_n^2(Z_t,Z_k,Z_l)) = \frac{4}{n}(3E(h_{tkl}^2) + 2E(h_{tkl}h_{ktl}) + 2E(h_{tkl}h_{ltk}) + 2E(h_{ktl}h_{ltk}))$. Under A1-A5 repeated application of the proposition of Prakasa-Rao each of these expectations can be shown to approach zero as $n \to \infty$.

$$
\begin{aligned}
I_{2n} &= \frac{1}{24n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}[h_{tkl}+h_{tlk}+h_{ktl}+h_{klt}+h_{ltk}+h_{lkt}] \\
&= \frac{1}{24n^3}\sum_{t=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\psi_n(Z_t,Z_k,Z_l) = \frac{1}{24}v_n,
\end{aligned}
$$

where $h_{tkl} = h_n^2 K(\frac{x_k-x_t}{h_n})(\frac{x_k-x_t}{h_n})^2 K(\frac{x_l-x_t}{h_n})(\frac{x_l-x_t}{h_n})^2 m^{(2)}(x_{kt})m^{(2)}(x_{lt})\frac{1}{g^2(x_t)}f(x_t,y_t)$. By the corollary to Theorem 1 $\sqrt{n}(v_n - \hat{u}_n^2) = o_p(1)$ provided that $E(\psi_n(Z_t,Z_k,Z_l)) = o(n)$. As above, verification of this condition is straightforward given our assumptions. Note that $\hat{u}_n = \frac{3}{n}\sum_{t=1}^{n}\psi_{1n}(Z_t) - 2\theta_n$ with $E(\hat{u}_n) = \theta_n$. Hence,

$$\frac{1}{h_n^4}E(\hat{u}_n) = \frac{6}{h_n^2}E(K(\frac{x_k - x_t}{h_n})(\frac{x_k - x_t}{h_n})^2 K(\frac{x_l - x_t}{h_n})(\frac{x_l - x_t}{h_n})^2 m^{(2)}(x_{kt})m^{(2)}(x_{lt})\frac{1}{g^2(x_t)}f(x_t, y_t)$$

and by the proposition of Prakasa-Rao we have,

$\frac{1}{h_n^4}E(\hat{u}_n) \to 6\sigma_K^4 E(m^{(2)2}(x_t)f(x_t, y_t))$ and $V(\frac{1}{h_n^4}\hat{u}_n) \to 0$. By Markov's inequality we conclude that $I_{2n} = \frac{1}{4}h_n^4\sigma_K^4 E(m^{(2)2}(x_t)f(x_t, y_t)) + o_p(h_n^4)$. $I_{3n} = \frac{1}{n}\sum_{t=1}^n w^2(x_t)f(x_t, y_t) = o_p(h_n^4)$ follows directly from the analysis of the I_{3n} term in Lemma 1. Now,

$$\begin{aligned} I_{4n} &= \frac{1}{6n^3}\sum_{t=1}^n\sum_{k=1}^n\sum_{l=1}^n h_{tkl} + h_{tlk} + h_{ktl} + h_{klt} + h_{ltk} + h_{lkt} \\ &= \frac{1}{6n^3}\sum_{t=1}^n\sum_{k=1}^n\sum_{l=1}^n \psi_n(Z_t, Z_k, Z_l) = \frac{1}{6}v_n \end{aligned}$$

where $h_{tkl} = K(\frac{x_k - x_t}{h_n})K(\frac{x_l - x_t}{h_n})(\frac{x_l - x_t}{h_n})^2 m^{(2)}(x_{kt})\sigma(x_k)\epsilon_k\frac{1}{g^2(x_t)}f(x_t, y_t)$. Once again if $E(\psi_n^2(Z_t, Z_k, Z_l)) = o(n)$ we have $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$, where $\hat{u}_n = \frac{3}{n}\sum_{t=1}^n \psi_{1n}(Z_t) - 2\theta_n$. Given that in this case $\theta_n = 0$ we have

$$\hat{u}_n = \frac{6}{n}\sum_{t=1}^n \sigma(x_t)\epsilon_t E(K(\frac{x_k - x_t}{h_n})K(\frac{x_l - x_t}{h_n})(\frac{x_l - x_t}{h_n})^2 m^{(2)}(x_{kt})\sigma(x_k)\epsilon_k\frac{1}{g^2(x_t)}f(x_t, y_t)|Z_t),$$

Under A1-A5 the proposition of Prakasa Rao gives,

$$\begin{aligned} &E(\frac{1}{h_n^2}K(\frac{x_k - x_t}{h_n})K(\frac{x_l - x_t}{h_n})(\frac{x_l - x_t}{h_n})^2 m^{(2)}(x_{kt})\sigma(x_k)\epsilon_k\frac{1}{g^2(x_t)}f(x_t, y_t)|Z_t) \\ &\to \sigma_K^2 m^{(2)}(x_t)E(f(x_t, y|x_1, \cdots, x_n)). \end{aligned}$$

Hence, $\hat{u}_n = \frac{6h_n^2}{n}\sum_{t=1}^n \sigma(x_t)\epsilon_t m^{(2)}(x_t)\sigma_K^2 E(f(x_t, y)|x_1, \cdots, x_n) + o_p(n^{-\frac{1}{2}})$,

given that $\frac{6}{n}\sum_{t=1}^n \sigma(x_t)\epsilon_t = O_p(n^{-\frac{1}{2}})$. Consequently,

$I_{4n} = \frac{\sigma_K^2 h_n^2}{n}\sum_{t=1}^n \sigma(x_t)\epsilon_t m^{(2)}(x_t)E(f(x_t, y)|x_1, \cdots, x_n) + o_p(n^{-\frac{1}{2}})$.

Now note that $|I_{5n}| = \left|\frac{2}{n}\sum_{t=1}^n w(x_t)\left(\frac{1}{nh_n g(x_t)}\sum_{k=1}^n K(\frac{x_k - x_t}{h_n})\sigma(x_k)\epsilon_k\right)f(x_t, y_t)\right|$. From Martins-Filho and Yao (2003) Theorem 1 we have $\sup_{x_t \in G}|w(x_t)| = o_p(h_n^2)$ and

$\sup_{x_t \in G} |\frac{1}{nh_n g(x_t)} \sum_{k=1}^{n} K(\frac{x_k - x_t}{h_n})\sigma(x_k)\epsilon_k| = o_p(h_n)$.

Hence, $|I_{5n}| \leq \frac{2}{n}\sum_{t=1}^{n} |f(x_t, y_t)| o_p(h_n^3) = O_p(1)o_p(h_n^3) = o_p(h_n^3)$

since $\frac{2}{n}\sum_{t=1}^{n} |f(x_t, y_t)| = O_p(1)$, which gives $I_{5n} = o_p(h_n^3)$.

Finally, $I_{6n} = \frac{1}{n}\sum_{t=1}^{n} w(x_t) f(x_t, y_t)(\frac{h_n}{ng(x)t)} \sum_{k=1}^{n} K(\frac{x_k - x_t}{h_n})(\frac{x_k - x_t}{h_n})^2 m^{(2)}(x_{kt}))$. Again, from Martins-Filho and Yao (2003) Theorem 1 we have

$$\sup_{x_t \in G} |\frac{h_n}{ng(x)t)} \sum_{k=1}^{n} K(\frac{x_k - x_t}{h_n})(\frac{x_k - x_t}{h_n})^2 m^{(2)}(x_{kt})| = O_p(h_n^2)$$

and $|I_{6n}| \leq \frac{1}{n}\sum_{t=1}^{n} |f(x_t, y_t)| o_p(h_n^2) O_p(h_n^2) = O_p(1)o_p(h_n^4) = o_p(h_n^4)$. Combining the results for each term proves the Lemma.

We now use the lemmas given above to help establish the asymptotic distribution of some statistics of interest.

4.3.1 Estimating Conditional Variances

Consider first a special case of model (19), where ϵ_t is conditionally $N(0,1)$, with $V(y_t|x_t) = \sigma^2$. Then a natural estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\sigma^2\epsilon_t^2 + 2(m(x_t) - \hat{m}(x_t))\sigma\epsilon_t + (m(x_t) - \hat{m}(x_t))^2\right)$$

The difficulty in dealing with such expressions lies in the average terms involving $(m(x_t) - \hat{m}(x_t))$ and $(m(x_t) - \hat{m}(x_t))^2$, since the first term is an average of an iid sequence. By using the corollary to **Theorem 1** and the lemmas we have a convenient way to establish their asymptotic properties. This is shown in the next theorem.

Theorem 2 Assume that in model (19) ϵ_t is conditionally $N(0,1)$, with $V(y_t|x_t) = \sigma^2$, and $E(y_t^4) < \infty$. Under assumptions A1-A5 we have $\sqrt{n}(\hat{\sigma}^2 - \sigma^2 - b_{1n}) \xrightarrow{d} N(0, 2\sigma^4)$, where

$b_{1n} = o_p(h_n^2)$.

Proof Note that $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\sigma^2\epsilon_t^2 + 2(m(x_t) - \hat{m}(x_t))\sigma\epsilon_t + (m(x_t) - \hat{m}(x_t))^2\right)$. Now, letting $f(x_t, y_t) = -2\sigma\epsilon_t$ in Lemma 1 we have $\frac{1}{n}\sum_{t=1}^{n}\sigma\epsilon_t\int f(x_t, y_t)q(y_t, x_t)dy_t = 0$, since $E(\epsilon_t|x_1, \cdots, x_n) = 0$. Also,

$$\frac{1}{2}h_n^2\sigma_K^2 E(m^{(2)2})f(x, y)) = 0,$$

hence $\frac{1}{n}\sum_{t=1}^{n} 2(m(x_t) - \hat{m}(x_t))\sigma\epsilon_t = o_p(n^{-\frac{1}{2}}) + o_p(h_n^2)$. By Lemma 2 $\frac{1}{n}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))^2$ and $f = 1$ we have

$$\frac{1}{4}h_n^4\sigma_K^4 E(m^{(2)2}(x_t)f(x_t, y_t)) = O_p(h_n^4),$$

$\frac{\sigma_K^2 h_n^2}{n}\sum_{t=1}^{n}\sigma\epsilon_t m^{(2)2}(x_t)E(f(x_t, y)|x_1, \cdots, x_n)) = \frac{\sigma_K^2 h_n^2}{n}\sum_{t=1}^{n}\sigma\epsilon_t m^{(2)2}(x_t) = O_p(n^{-\frac{1}{2}}h_n^2) = o_p(n^{-\frac{1}{2}})$ by the Central Limit Theorem for iid sequences. Hence, $\frac{1}{n}\sum_{t=1}^{n}(m(x_t) - \hat{m}(x_t))^2 = o_p(n^{-\frac{1}{2}}) + o_p(h_n^3)$ provided $E(y_t^4) < \infty$. Finally, note that $\frac{1}{n}\sum_{t=1}^{n}\sigma^2(\epsilon_t^2 - 1) = \frac{1}{n}\sum_{t=1}^{n}\zeta_t$, then $E(\zeta_t) = 0$, $V(\zeta_t) = E(\zeta_t^2) = \sigma^4(E(\epsilon_t^4) - 1) = 2\sigma^4$, as $E(\epsilon_t^4) = 3$. Hence, $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma^2(\epsilon_t^2 - 1) \xrightarrow{d} N(0, 2\sigma^4)$. Combining the results for each term proves the theorem.

Theorem 2 can be generalized by relaxing the conditional normality assumption and allowing $V(y_t|x_t) = \sigma^2(x_t)$. This generalization was done by Martins-Filho and Yao(2003) but their proof can be conveniently simplified by using Theorem 1 and its corollary.

4.3.2 Estimating nonparametric R^2

In regression analysis we are usually interested in Pearson's correlation ratio, i.e.,

$$\eta^2 = \frac{V(E(y|x))}{V(y)} = \frac{V(m(x))}{V(y)},$$

where y is a regressand and x is the regressor in the context of model (19). Since $Var(y) = Var(E(y|x))) + E(Var(y|x)) = V(m(x)) + E(\sigma^2(x))$, η^2 gives the fraction of the variability of y which is explained with the best predictor based on x, $m(x)$, and can be interpreted as a nonparametric coefficient of determination or nonparametric R^2.

Estimation of nonparametric R^2 has been studied by Doksum and Samarov (1995) using Nadaraya-Watson estimator. Similar topics, such as estimation of noise to signal ratios has been considered by Yao and Tong (1996). Here, given that $Var(m(x)) = Var(y) - E(Y - m(x))^2$ we have $\eta^2 = 1 - \frac{E(y-m(x))^2}{Var(y)}$ and we consider

$$\hat{\eta}^2 = 1 - \frac{\frac{1}{n}\sum_{t=1}^n (y_t - \hat{m}(x_t))^2}{\frac{1}{n}\sum_{t=1}^n (y_t - \bar{y})^2} \quad \text{where } \bar{y} = \frac{1}{n}\sum_{t=1}^n y_t.$$

The following theorem gives the asymptotic characterization of our proposed nonparametric R^2.

Theorem 3 Under assumptions A1-A5, $E(y_t^4) < \infty$ and model (19) we have

$$\sqrt{n}(\hat{\eta}^2 - \eta^2 - b_{2n}) \xrightarrow{d} N(0, V(\zeta)) \text{ where } \zeta = \frac{1}{Var(y_t)}(\sigma^2(x_t)\epsilon^2 - (1-\eta^2)(y_t - E(y_t))^2)$$

where $b_{2n} = o_p(h_n^2)$.

Proof $\hat{\eta}^2 - \eta^2 = -\frac{1}{\frac{1}{n}\sum_{t=1}^n (y_t - \bar{y})^2}[\beta_1 - \beta_2 \frac{E(y_t - m(x_t))^2}{Var(y_t)}]$ where $\beta_1 = \frac{1}{n}\sum_{t=1}^n (y_t - \hat{m}(x_t))^2 - E(y_t - m(x_t))^2$, $\beta_2 = \frac{1}{n}\sum_{t=1}^n (y_t - \bar{y})^2 - Var(y_t)$. First, note that by the law of large numbers $\frac{1}{n}\sum_{t=1}^n y_t^2 \xrightarrow{p} E(y_t^2)$ and $\bar{y}^2 \xrightarrow{p} E(y_t^2)$, hence

$$\left(\frac{1}{n}\sum_{t=1}^n (y_t - \bar{y})^2\right)^{-1} \xrightarrow{p} V(y_t)^{-1}$$

For β_2, $\bar{y}^2 - (Ey_t)^2 = \frac{1}{2}\frac{1}{n^2}\sum_{t=1}^n \sum_{k=1}^n -E(m(x_t))^2 = \frac{1}{2}v_n - (E(m(x_t)))^2$. By the Corollary to Theorem 1 $\frac{1}{n}E(\psi_n^2(y_t, y_k)) = \frac{4}{n}E(y_t^2)^2 = o(1)$, hence $\sqrt{n}(v_n - \hat{u}_n) = o_p(1)$ where $\hat{u}_n = \frac{2}{n}\sum_{t=1}^n (\psi_{1n}(y_t)) - \theta_n = \frac{4}{n}\sum_{t=1}^n y_t E(m(x_t)) - 2E(m(x_t))^2$.

Hence, $\bar{y}^2 - E(y_t)^2 = \frac{2}{n}\sum_{t=1}^n y_t E(m(x_t)) - 2(E(m(x_t)))^2 + o_p(n^{-\frac{1}{2}})$. Hence,

$\beta_2 \frac{E(y_t - m(x_t))^2}{Var(y_t)} = \frac{1}{n}\sum_{t=1}^n \left\{ \frac{E\sigma^2(x_t))}{Var(y_t)}(y_t^2 - Ey_t^2) - 2\frac{E\sigma^2(x_t))E(m(x_t))}{Var(y_t)}(y_t - E(m(x_t)))\right\} + o_p(n^{-\frac{1}{2}})$.

For β_1,

$\frac{1}{n}\sum_{t=1}^n (y_t - \hat{m}(x_t))^2 = \frac{1}{n}\sum_{t=1}^n \left(\sigma^2(x_t)\epsilon_t^2 + 2(m(x_t) - \hat{m}(x_t))\sigma(x_t)\epsilon_t + (m(x_t) - \hat{m}(x_t))^2\right)$. Similarly to the proof of Theorem 2, we use Lemma 1 by letting $f(x_t, y_t) = -2\sigma(x_t)\epsilon_t$, and obtain $\frac{1}{n}\sum_{t=1}^n 2(m(x_t) - \hat{m}(x_t))\sigma(x_t)\epsilon_t = o_p(n^{-\frac{1}{2}}) + o_p(h_n^2)$. By Lemma 2, we let $f(x_t, y_t) = 1$ and obtain $\frac{1}{n}\sum_{t=1}^n (m(x_t) - \hat{m}(x_t))^2 = o_p(n^{-\frac{1}{2}}) + o_p(h_n^3)$. Therefore, $\frac{1}{n}\sum_{t=1}^n (y_t - \hat{m}(x_t))^2 = \frac{1}{n}\sum_{t=1}^n \sigma^2(x_t)\epsilon_t^2 + o_p(n^{-\frac{1}{2}}) + o_p(h_n^2)$. Hence,

$$\begin{aligned}
\beta_1 - \beta_2\frac{E(y_t - m(x_t))^2}{Var(y_t)} &= \frac{1}{n}\sum_{t=1}^n (\sigma^2(x_t)\epsilon_t^2 - E(\sigma^2(x_t)) - \frac{E(\sigma^2(x_t))}{V(y_t)}(y_t^2 - Ey_t^2) \\
&+ \frac{2E(\sigma^2(x_t))E(m(x_t))}{V(y_t)}(y_t - E(m(x_t))) + o_p(n^{-\frac{1}{2}}) + o_p(h_n^2) \\
&= \frac{1}{n}\sum_{t=1}^n \zeta_t + o_p(n^{-\frac{1}{2}}) + o_p(h_n^2)
\end{aligned}$$

Since $E(\zeta_t) = 0$, $V(\zeta_t) = V(\sigma^2(x_t)\epsilon_t^2 - (1-\eta^2)(y_t - E(y_t))^2) < \infty$ and ζ_t forms an iid sequence, then by the central limit theorem we have

$$\sqrt{n}\left(\beta_1 - \beta_2\frac{E(y_t - m(x_t))^2}{Var(y_t)} - o_p(h_n^2)\right) \xrightarrow{d} N(0, V(\zeta)).$$

Together with $\frac{1}{\frac{1}{n}\sum_{t=1}^n (y_t - \bar{y})^2} \xrightarrow{p} \frac{1}{V(y_t)}$ and given that
$\hat{\eta}^2 - \eta^2 = -\frac{1}{\frac{1}{n}\sum_{t=1}^n (y_t - \bar{y})^2}\left(\beta_1 - \beta_2\frac{E(y_t - m(x_t))^2}{Var(y_t)}\right)$, we obtain the desired result.

4.4 Conclusion

In this note we establish the $\sqrt{n}$ asymptotic equivalence between V and U-statistics when their kernels depend on n. We combine our result with a result of Lee(1988) to obtain the $\sqrt{n}$ asymptotic equivalence between V statistics and U-statistics projections. Some illustrations on how our results can be used in nonparametric kernel estimation are given but it seems clear that our result can be used in a much broader context.

Bibliography

1. Ahn, H. and Powell J. L., 1993, Semiparametric estimation of censored selection models with a nonparametric selection mechanism, Journal of Econometrics, 58, 3-29.
2. Aigner, D., C.A.K. Lovell and P. Schmidt, 1977, Formulation and estimation of stochastic frontiers production function models, Journal of Econometrics, 6, 21-37.
3. Aït-Sahalia, Y. and M.W. Brandt, 2001, Variable selection for portfolio choice. Journal of Finance, 56, 1297-1355.
4. Andreou, E., N. Pittis and A. Spanos, 2001, On modelling speculative prices: the empirical literature. Journal of Economic Surveys, 15, 187-220.
5. Artzner, P., F. Delbaen, J. Eber, J. and D. Heath, 1999, Coherent measures of risk. Mathematical Finance, 9, 203-228.
6. Baccouche, R. and M. Kouki, 2003, Stochastic production frontier and technical inefficiency: a sensitivity analysis, Econometric Reviews, 22, 79-91.
7. Banker, R., 1993, Maximum likelihood, consistency and data development analysis: a statistical foundation, management Science, 39, 1265-1273.
8. Basle Committee, 1996, Overview of the ammendment of the capital accord to incorporate market risk. Basle Committee on Banking Supervision.
9. Bollerslev, T., 1986, Generalized autoregressive conditional heterocedasticity. Journal of Econometrics, 31, 307-327.
10. Carroll, R.J., W. Härdle and E. Mammen, 2002, Estimation in an additive model when the components are linked parametrically. Econometric Theory, 18, 886-912.
11. Caudill, S., J. Ford, D. Gropper, 1995, Frontier estimation and firm specific inefficiency measures in the presence of heterocedasticity, Journal of Business and Economic Statistics, 13, 105-111.
12. Cazals, C., J.-P. Florens, L. Simar, 2002, Nonparametric frontier estimation: a robust approach, Journal of Econometrics, 106, 1-25.
13. Charnes, A., W. Cooper and E. Rhodes, 1978, Measuring the efficiency of decision making units, European Journal of Operational Research, 429-444.
14. Chen, Yi-Ting, 2001, Testing conditional symmetry with an application to stock returns. Working Paper, Academia Sinica.
15. Christoffersen, P., J. Hahn, and A. Inoue, 2001, Testing and comparing value-at-risk measures, Journal of Empirical Finance, 8, 325-342.
16. D'Amico, S., 2003, Density estimation and combination under model ambiguity via a pilot nonparametric estimate: an application to stock returns. Working paper, Department of Economics, Columbia University (http://www.columbia.edu/~sd445/).

17. Deprins, D., L. Simar and H. Tulkens, 1984, Measuring labor inefficiency in post offices. In M. Marchand, P. Pestiau and H. Tulkens(Eds.), The performance of public enterprises: concepts and measurements. North Holland, Amsterdam.

18. Diebold, F.X., T. Schuermann, J.D. Stroughair, 1998, Pitfalls and opportunities in the use of extreme value theory in risk management, in Advances in Computational Finance, A.-P. Refenes, J.D. Moody and A.N. Burgess eds., 2-12. Kluwer Academic: Amsterdam.

19. Ding, Z., C.W.J. Granger and R.F. Engle, 1993, A long memory property of stock market returns and a new model. Journal of Empirical Finance, 1, 83-106.

20. Diebolt, J. and D. Guégan, 1993, Tail behaviour of the stationary density of general nonlinear autoregressive processes of order 1. Journal of Applied Probability, 30, 315-329.

21. Doksum K. and A. Samarov, 1995, Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression, The Annals of Statistics, Vol. 23, No. 5, 1443-1473.

22. Embrechts, P., S. Resnick, and G. Samorodnitsky, 1999, Extreme value theory as a risk management tool. North American Actuarial Journal 3, 30-41.

23. Embrechts, P., C. Klüppelberg, T. Mikosh, 1997, Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

24. Fan, J., 1992, Design adaptive nonparametric regression. Journal of the American Statistical Association, 87, 998-1004.

25. Fan, J. and I. Gijbels, 1995, Data driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation, Journal of the Royal Statistical Society B 57, 371-394.

26. Fan, J. and Q. Yao, 1998, Efficient estimation of conditional variance functions in stochastic regression. Biometrika, 85, 645-660.

27. Fan, Y. and Q. Li, 1996, Consistent Model Specification test: ommitted variables and semiparametric functional forms, Econometrica, 64, No.4, July, 865-890.

28. Farrell, M., 1957, The measurement of productive efficiency, Journal of the Royal Statistical Society A, 120, 253-290.

29. Gijbels, I., E. Mammen, B. Park and L. Simar, 1999, On estimation of monotone and concave frontier functions, Journal of the American Statistical Association, 94, 220-228.

30. Gouriéroux, C., 1997, ARCH models and financial applications. Springer: Berlin.

31. Haan, L. de, 1976, Sample extremes: an elementary introduction. Statistica Neerlandica, 30, 161-172.

32. Hadri, K., 1999, Estimation of a doubly heterocedastic stochastic frontier cost function, Journal of Business and Economic Statistics, 17, 359-363.

33. Hafner, C., 1998, Nonlinear time series analysis with applications to foreign exchange volatility. Physica-Verlag: Heidelberg.

34. Hall, P. and R. Carroll, 1989, Variance function estimation in regression: the effect of estimation of the mean. journal of the Royal Statistical Society B, 51, 3-14.

35. Hansen, B., 1994, Autoregressive conditional density estimation. International Economic Review, 35, 705-729.

36. Härdle, W. and A. Tsybakov, 1997, Local Polynomial estimators of the volatility function in nonparametric autoregression. Journal of Econometrics, 81, 223-242.

37. Hoeffding, W., 1948, A class of statistics with asymptotically normal distribution, Annals of Mathematical Statistics, 19, 293-325.

38. Hoeffding, W., 1961, The strong law of large numbers for U-statistics, Institute of Statistics, Mimeo Series 302, University of North Carolina, Chapel Hill, NC.

39. Hols, M.C.A.B. and C.G. de Vries, 1991, The limiting distribution of extremal exchange rate returns. Journal of Applied Econometrics, 6, 287-302.

40. Hosking, J.R.M., 1990, L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society, B, 52, 105-124.

41. Hosking, J.R.M. and J. R. Wallis, 1987, Parameter and quantile estimation for the generalized pareto distribution. Technometrics, 29, 339-349.

42. Hosking, J.R.M. and J.R. Wallis, 1997, Regional frequency analysis. Cambridge University Press: Cambridge, UK.

43. Kemp, G., 2000, Semiparametric estimation of a logit model. Working paper, Department of Economics, University of Essex
(http://privatewww.essex.ac.uk/~ kempgcr/papers/logitmx.pdf).

44. Korostelev, L. Simar and Tsybakov, 1995, Efficient estimation of monotone boundaries, Annals of Statistics, 23, 476-489.

45. Leadbetter, M.R., G. Lindgren, H. Rootzén, 1983, Extremes and related properties of random sequences and processes. Springer-Verlag: Berlin.

46. Lee, A.J., 1990, U-Statistics. Marcel Dekker, New York.

47. Lee, B., 1988, Nonparametric tests using a kernel estimation method, Doctoral dissertation, Department of Economics, University of Wisconsin, Madison, WI.

48. Longin, F., 2000, From value-at-risk to stress testing, the Extreme Value Approach. Journal of Banking and Finance, 24, 1097-1130.

49. Martins-Filho, C. and F. Yao, 2003, A Nonparametric Model of Frontiers, Working Paper, Department of Economics, Oregon State University (It is also Chapter 3 of this dissertation. http://oregonstate.edu/~ martinsc/martins-filho-yao(03).pdf).

50. Masry, E., 1996, Multivariate local polynomial regression for time series: uniform strong consistency and rates. Journal of Time Series Analysis, 17, 571-599.

51. Masry, E. and D. Tjøstheim, 1995, Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality. Econometric Theory, 11, 258-289.

52. McNeil, A.J. and R. Frey, 2000, Estimation of tail-related risk measures for heterocedastic financial time series: an extreme value approach. Journal of Empirical Finance, 7, 271-300.

53. Meeusen, W. and J. van den Broeck, 1977, Efficiency estimation from cobb-douglas production functions with composed error, International Economic Review, 18, 435-444.

54. Pagan, A. and G.W. Schwert, 1990, Alternative models for conditional stochastic volatility. Journal of Econometrics, 45, 267-291.

55. Pagan, A. and A. Ullah, 1999, Nonparametric econometrics. Cambridge University Press, Cambridge, UK.

56. Park, B., L. Simar and Ch. Weiner, 2000, The FDH estimator for productivity efficient scores: asymptotic properties, Econometric Theory, 16, 855-877.

57. Patton, A.J., 2001, On the importance of skewness and asymetric dependence in stock returns for asset allocation. Working Paper, LSE.

58. Pickands, J., 1975, Statistical Inference Using Extreme Order Statistics, *The Annals of Statistics,* **3** 119-131.

59. Powell, J., J. Stock, T. Stoker, 1989, Semiparametric estimation of index coefficients, Econometrica, 57, 1403-1430.

60. Prakasa-Rao, B.L.S., 1983, Nonparametric functional estimation. John Wiley and Sons, New York.

61. Risk, 1999, Risk Software Survey. January, 67-80.

62. RiskMetrics, 1995, RiskMetrics Technical Document, 3rd Edition, JP Morgan.

63. Ruppert, Sheather, and Wand, M., 1995, An effective bandwidth selector for local least squares regression, Journal of the American Statistical Association, 90, 1257-1270.

64. Scaillet, O., 2002, Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance, forthcoming.

65. Seifford, L., 1996, Data envelopment analysis: the evolution of the state of the art(1978-1995), Journal of Productivity Analysis 7, 99-137.

66. Serfling, R. J., 1980, Approximation Theorems of Mathematical Statistics. John Wiley & Sons, Inc.

67. Smith, R.L., 1984, Threshold methods for sample extremes, in Statistical Extremes and applications, ed. Thiago de Oliveira, 621-638. Dordrecht: D. Reidel.

68. Smith, R.L., 1987, Estimating tails of probability distributions. The Annals of Statistics, 15, 1174-1207.

69. Shephard, M., 1996, Statistical aspects of ARCH and stochastic volatility models, in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, eds. Time Series Models in Econometrics, Finance and Other Fields, 1-67. London: Chapman and Hall.

70. Silverman, B.W,, 1986, Density estimation for statistics and data analysis. Chapman and Hall, London.

71. Spanos, A., 2002, Student's t autoregressive model with dynamic heterocedasticity. Working Paper, Virginia Tech.

72. Tauchen, G., 2001, Notes on Financial Econometrics. Journal of Econometrics, 100, 57-64.

73. Yao, Q. and H. Tong, 2000, Nonparametric estimation of ratios of noise to signal in stochastic regression, Statistica Sinica, 10, 751-770.

74. Yatchew, A., 1998, Nonparametric Regression Techniques in Economics. Journal of Economic Literature, XXXVI, 669-721.

75. Zheng, J. X., 1998, A consistent nonparametric test of parametric regression models under conditional quantile restrictions, Econometric Theorey, 14, 123-138.

76. Ziegelmann, F., 2002, Nonparametric estimation of volatility functions: the local exponential estimator. Econometric Theory, 18, 985-991.

Appendices

Appendix A: Proof of Proposition 1 in Chapter 2

Proposition 1 : Let $g(y;v) = c\left(1+\frac{1}{v-2}y^2\right)^m$ where $c, m \in \Re$, $v > 2$ a positive integer and $-\infty < y < \infty$. Let $\kappa_p = \int_0^\infty y^p g(y;v)dy$ for $p = 1,2,3,4$. Then,

$$\kappa_1 = \tfrac{c(v-2)}{2}\int_0^1 u^{-m-2}du,$$
$$\kappa_2 = \tfrac{c(v-2)^{3/2}}{2}\left(B[1/2,-m-3/2] - B[1/2,-m-1/2]\right),$$
$$\kappa_3 = \tfrac{c(v-2)^2}{2}\left(\int_0^1 u^{-m-3}du - \int_0^1 u^{-m-2}du\right),$$
$$\kappa_4 = \tfrac{c(v-2)^{5/2}}{2}\left(B[1/2,-m-5/2] - 2B[1/2,-m-3/2] + B[1/2,-m-1/2]\right),$$

where $B[\alpha,\beta]$ is the beta function.

Proof: Let $\theta = (v-2)^{-1/2}$, then $\kappa_1 = c(v-2)\int_0^\infty \theta(1+\theta^2)^m d\theta$. Now, put $\theta = tan(w)$ and using the fact $sin(w)^2 = 1 - cos(w)^2$ and $1 + tan^2(w) = cos(w)^{-2}$, we have

$$\kappa_1 = -c(v-2)\int_0^{\pi/2} cos(w)^{-2m-3}dcos(w) = \frac{c(v-2)}{2}\int_0^1 u^{-m-2}du.$$

For κ_2 we have, $\kappa_2 = c(v-2)^{3/2}\int_0^\infty \theta^2(1+\theta^2)^m d\theta$. Using the same transformations above, we have

$$\kappa_2 = c(v-2)^{3/2}\left(\int_0^{\pi/2} cos(w)^{-2m-4}dw - \int_0^{\pi/2} cos(w)^{-2m-2}dw\right).$$

It is easy to show that for $h \in \Re$ and $B[\alpha,\beta]$ the beta function,

$$\int_0^{\pi/2} cos(w)^{-2m-h}dw = \frac{1}{2}\int_0^1 u^{1/2-1}(1-u)^{-m-h/2-1/2}du = \frac{1}{2}B[1/2,-m-h/2+1/2],$$

which gives the desired result. For κ_3 we have, $\kappa_3 = c(v-2)^2\int_0^\infty \theta^3(1+\theta^2)^m d\theta$ and we obtain

$$\kappa_3 = c(v-2)^2\left(\int_0^{\pi/2} cos(w)^{-2m-3} dcos(w) - \int_0^{\pi/2} cos(w)^{-2m-5} dcos(w)\right)$$
$$= \frac{c(v-2)^2}{2}\left(\int_0^1 u^{-m-3}du - \int_0^1 u^{-m-2}du\right)$$

Finally, for κ_4 we have $\kappa_4 = c(v-2)^{5/2}\int_0^\infty \theta^4(1+\theta^2)^m d\theta$ and it is straightforward to show that,

$$\kappa_4 = c(v-2)^{5/2}\int_0^{\pi/2}\left(cos(w)^{-2m-6} - 2cos(w)^{-2m-4} + cos(w)^{-2m-2}\right)dw$$

Using the previous results we obtain the desired expression.

Appendix B: Proof of Lemma 1 in Chapter 2

Proof of **Lemma 1** : From Hansen(1994), $\int_{-\infty}^\infty xf(x;v,\lambda)dx = 0$ and $\int_{-\infty}^\infty x^2 f(x;v,\lambda)dx = 1$, therefore $\alpha_3 = \int_{-\infty}^\infty x^3 f(x;v,\lambda)dx$ and $\alpha_4 = \int_{-\infty}^\infty x^4 f(x;v,\lambda)dx$. First consider α_3. Note that

$$\begin{aligned}\alpha_3 \;=\; & \int_{-\infty}^{-a/b} x^3 bc\left(1+\frac{1}{v-2}\left(\frac{bx+a}{1-\lambda}\right)^2\right)^{-(v+1)/2} dx + \\ & \int_{-a/b}^{\infty} x^3 bc\left(1+\frac{1}{v-2}\left(\frac{bx+a}{1+\lambda}\right)^2\right)^{-(v+1)/2} dx.\end{aligned}$$

Let $y = \frac{bx+a}{1-\lambda}$ on the first integral and $y = \frac{bx+a}{1+\lambda}$ on the second integral. Then,

$$\begin{aligned}\alpha_3 \;=\; & \int_{-\infty}^{0}\left(\frac{1-\lambda}{b}y - a/b\right)^3 c\left(1+\frac{1}{v-2}y^2\right)^{-\frac{v+1}{2}}(1-\lambda)dy + \\ & \int_{0}^{\infty}\left(\frac{1+\lambda}{b}y - a/b\right)^3 c\left(1+\frac{1}{v-2}y^2\right)^{-\frac{v+1}{2}}(1+\lambda)dy.\end{aligned}$$

Simple manipulations yield, $\alpha_3 = \frac{8\kappa_3}{b^3}(\lambda^3+\lambda) - \frac{6\kappa_2 a}{b^3}(\lambda^3+3\lambda) + \frac{12a^2\kappa_1}{b^3}\lambda - \frac{a^3+3a(1-\lambda)^3}{b^3}$, where κ_i for $i = 1,2,3,4$ are as defined in **Proposition 1** . Using the same transformations for α_4, we have

$$\begin{aligned}\alpha_4 &= \frac{2\lambda^5+20\lambda^3+10\lambda}{b^4}(\kappa_4) - \frac{32a\kappa_3}{b^4}(\lambda^3+\lambda) + \frac{12a^2\kappa_2}{b^4}(\lambda^3+3\lambda) \\ &\quad -\frac{16a^3\lambda\kappa_1}{b^4} + \frac{1}{b^4}\left(a^4 + \frac{3(1-\lambda)^5(v-2)}{v-4} + 6a^2(1-\lambda)^3\right)\end{aligned}$$

Appendix C: Proof of Lemma 2 in Chapter 2

Proof of **Lemma2** . $E(X|X>z) = \int_z^\infty x f_{X>z}(x;v,\lambda)dx = (1-F(z))^{-1}\int_z^\infty x f(x;v,\lambda)dx$. Let $I = \int_z^\infty x f(x;v,\lambda)dx$ and put $y = \frac{bx+a}{1+\lambda}$, then

$$I = \frac{(1+\lambda)^2}{b}\int_\alpha^\infty x g(x;v)dx - \frac{(1+\lambda)a}{b}\int_\alpha^\infty g(x;v)dx \tag{22}$$

where $\alpha = \frac{bz+a}{1+\lambda}$ and $g(x;v) = c\left(1+\frac{1}{v-2}x^2\right)^{-(v+1)/2}$. From **Lemma 1** ,

$$\int_\alpha^\infty x g(x;v)dx = \frac{c(v-2)}{2}\int_0^{cos^2(arctan(\gamma))} u^{(v+1)/2-2}du \tag{23}$$

where $\gamma = (v-2)^{-0.5}\alpha$. Consequently, $\int x g(x;v)dx = \frac{c(v-2)}{v-1}\beta^{\frac{v-1}{2}}$. For the second integral note that from **Lemma 1** , it is easy to show that $\int_\alpha^\infty g(x;v)dx = 1 - \int_{-\infty}^\alpha g(x;v)dx = 1 - F_s\left(\frac{bz+a}{1+\lambda}\sqrt{\frac{v}{v-2}}\right)$, which combined with (23) and substituting back in (22) gives the desired expression.

Appendix D: Proof of Lemma 3 in Chapter 3

Lemma 3 Let $\{Z_i\}_{i=1}^n$ be a sequence of i.i.d. random variables and $\psi_n(Z_1,\cdots,Z_k)$ be a symmetric function with $k \leq n$. Let $u_n = \begin{pmatrix} n \\ k \end{pmatrix}^{-1} \sum_{(n,k)} \psi_n(Z_{i_1},\cdots,Z_{i_k})$ and $\hat{u}_n = \frac{k}{n}\sum_{i_1=1}^n (\psi_{1n}(Z_i) - \theta_n) + \theta_n$, where $\sum_{(n,k)}$ denotes a sum over all subsets $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ of $\{1,2,\cdots,n\}$, $\psi_{1n}(Z_i) = E(\psi_n(Z_1,\cdots,Z_k)|Z_i)$, $\theta_n = E(\psi_n(Z_1,\cdots,Z_k))$. If $E(\psi_n^2(Z_1,\cdots,Z_k)) = o(n)$ then $\sqrt{n}(u_n - \hat{u}_n) = o_p(1)$.

Proof [Lemma 3] Using Hoeffding's(1961) decomposition for U-statistics we write, $u_n = \theta_n + \sum_{j=1}^k \begin{pmatrix} k \\ j \end{pmatrix} H_n^{(j)}$ where $H_n^{(j)} = \begin{pmatrix} n \\ j \end{pmatrix}^{-1} \sum_{(n,j)} h_n^{(j)}(Z_{v_1},\cdots,Z_{v_j})$, $h_n^{(1)}(Z_{v_1}) = \psi_{1n}(Z_{v_1}) - \theta_n$, $h_n^{(c)}(Z_{v_1},\cdots,Z_{v_c}) = \psi_{cn}(Z_{v_1},\cdots,Z_{v_c}) - \sum_{j=1}^c \sum_{(c,j)} h_n^{(j)}(Z_{i_1},...,Z_{i_j}) - \theta_n$ where
$\psi_{cn}(Z_{v_1},\cdots,Z_{v_c}) = E(\psi_n(Z_1,\cdots,Z_k)|Z_1,\cdots,Z_c)$ and $c = 2,...,k$. Then,
$u_n - \hat{u}_n = \sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix} H_n^{(j)}$ and it is straightforward to show that $E(u_n - \hat{u}_n) = 0$. Also,

$$\begin{aligned} V(n^{1/2}(u_n - \hat{u}_n)) &= nE\left(\left(\sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix} H_n^{(j)}\right)^2\right) = nE\left(\sum_{j'=2}^k \sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix} \begin{pmatrix} k \\ j' \end{pmatrix} H_n^{(j)} H_n^{(j')}\right) \\ &= n\sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix}^2 \begin{pmatrix} n \\ j \end{pmatrix}^{-1} E\left(h_n^{(j)}(Z_1,...,Z_j)^2\right) \end{aligned}$$

where the last equality follows from theorem 3 in Lee(1990, p.30). By Chebyshev's inequality, for all $\epsilon > 0$, $P\left(|n^{1/2}(u_n - \hat{u}_n)| \geq \epsilon\right) \leq nE((u_n - \hat{u}_n)^2)/\epsilon^2$. Therefore, it suffices to show that

$$n\sum_{j=2}^k \begin{pmatrix} k \\ j \end{pmatrix}^2 \begin{pmatrix} n \\ j \end{pmatrix}^{-1} E\left(h_n^{(j)}(Z_1,...,Z_j)^2\right) = o(1).$$

If for all $j = 2,...,k$

$$E\left((h_n^{(j)}(Z_1,...,Z_k))^2\right) = O\left(E\left(\psi_n^2(Z_1,...,Z_k)\right)\right) \tag{24}$$

then for some $\Delta > 0$,

$$nE((u_n - \hat{u}_n)^2) \leq n \sum_{j=2}^{k} \binom{k}{j}^2 \binom{n}{j}^{-1} \Delta E(\psi_n^2(Z_1, ..., Z_k)) =$$

$$n^2 \sum_{j=2}^{k} \binom{k}{j}^2 \frac{(n-j)!j!}{n!} n^{-1} \Delta E(\psi_n^2(Z_1, ..., Z_k)).$$

Since $E(\psi_n^2(Z_1, ..., Z_k)) = o(n)$ by assumption, for fixed k, there are a finite number of terms in $\sum_{j=2}^{k}$, the magnitude determined by $j = 2$. For some $\Delta' > 0$,
$nE((u_n - \hat{u}_n)^2) \leq \Delta' n^2 \binom{k}{2}^2 \frac{(n-2)!2!}{n!} \frac{E(\psi_n^2(Z_1,...,Z_k))}{n} \leq O(1)o(1)$. We now use induction to prove that $E\left((h_n^{(j)}(Z_1, ..., Z_k))^2\right) = O\left(\psi_n^2(Z_1, ..., Z_k)\right)$. Note that for $j = 2, ..., m$,

$$h_n^{(j)}(Z_1, ..., Z_j) = \psi_{jn}(Z_1, ..., Z_j) + \sum_{d=1}^{j-1} (-1)^d \sum_{(j,j-d)} \psi_{(j-d)n}(Z_{i_1}, ..., Z_{i_{j-d}}) + (-1)^j \theta_n$$

We first establish the result for $j = 2$.

$$(h_n^{(2)}(Z_1, Z_2))^2 = \psi_{2n}^2(Z_1, Z_2) - \psi_{1n}^2(Z_1) - \psi_{1n}^2(Z_2) + \theta_n^2 - 2\psi_{2n}(Z_1, Z_2)\psi_{1n}(Z_1)$$

$$-2\psi_{2n}(Z_1, Z_2)\psi_{1n}(Z_2) + 2\psi_{2n}(Z_1, Z_2)\theta_n + 2\psi_{1n}(Z_1)\psi_{1n}(Z_2) - 2\psi_{1n}(Z_1)\theta_n - 2\psi_{1n}(Z_2)\theta_n$$

By Cauchy-Schwarz's inequality, the expected value of each term on the righthand side can be shown to be less than $E\left(\psi_n^2(Z_1, Z_2)\right)$. Since there are a finite number of terms $E\left((h_n^{(2)}(Z_1, Z_2))^2\right) = O\left(E\left(\psi_n^2(Z_1, ..., Z_k)\right)\right)$. Now suppose that the statement is true for all $2 \leq j \leq k-1$. For $j = k$

$$\begin{aligned} E(h_n^{(k)}(Z_1, ..., Z_k)^2) &= E\left(\psi_n(Z_1, ..., Z_k)^2\right) + E\left(\left(\sum_{j=1}^{k-1} \sum_{(k,j)} h_n^{(j)}(Z_{i_1}, ..., Z_{i_j})\right)^2\right) + \theta_n^2 \\ &\quad -2 \sum_{j=1}^{k-1} \sum_{(k,j)} E\left(h_n^{(j)}(Z_{i_1}, ..., Z_{i_j})\psi_n(Z_1, ..., Z_k)\right) \\ &\quad -2E\left(\psi_n(Z_1, ..., Z_k)\theta_n\right) + 2\theta_n \sum_{j=1}^{k-1} \sum_{(k,j)} E\left(h^{(j)}(Z_{i_1}, ..., Z_{i_j})\right) \end{aligned}$$

and by Theorem 3 in Lee(1990)

$$
\begin{aligned}
& E\left(\left(\sum_{j=1}^{k-1}\sum_{(k,j)} h_n^{(j)}(Z_{i_1},...,Z_{i_j})\right)^2\right) \\
= & \sum_{j=1}^{k-1}\sum_{(k,j)}\sum_{j'=1}^{k-1}\sum_{(k,j')} E\left(h_n^{(j)}(Z_{i_1},...,Z_{i_j})h_n^{(j')}(Z_{i_1},...,Z_{i_{j'}})\right) \\
= & \sum_{j=1}^{k-1}\sum_{(k,j)} E\left((h_n^{(j)}(Z_{i_1},...,Z_{i_j}))^2\right)
\end{aligned}
$$

Given that this sum has a finite number of terms and the induction hypothesis we have that the left-hand side of the last equality is $O\left(E\left(\psi_n^2(Z_1,...,Z_k)\right)\right)$. Second, again by Theorem 3 in Lee(1990)

$$
E\left(h_n^{(j)}(Z_{i_1},...,Z_{i_j})\psi_n(Z_1,...,Z_k)\right) = E\left((h_n^{(j)}(Z_{i_1},...,Z_{i_j}))^2\right),
$$

therefore by the induction hypothesis

$\sum_{j=1}^{k-1}\sum_{(k,j)} E\left((h_n^{(j)}(Z_{i_1},...,Z_{i_j}))^2\right) = O\left(E\left(\psi_n^2(Z_1,...,Z_k)\right)\right)$.

Finally, $E\left(\psi_n(Z_1,...,Z_k)\theta_n\right) = \theta_n^2 \leq E(\psi_n^2(Z_1,...,Z_k))$ and the last term is zero. Hence, $E(h_n^{(k)}(Z_1,...,Z_k)^2) = O\left(E\left(\psi_n^2(Z_1,...,Z_k)\right)\right)$ for all $j = 2,...,k$.

Appendix E: Proof of Lemma 4 in Chapter 3

Lemma 4 Assume that A1, A2, A3 and A4. If $h_n \to 0$, $\frac{nh_n^3}{ln(h_n)} \to \infty$, and $X_t \in G$ a compact subset of $\Re$, then $\hat{b} - b = o_p(1)$

Proof [Lemma 4] We write $\hat{b} - b = \theta_1 - \theta_2 + \theta_3 + \theta_4 - \theta_5$, where

$\theta_1 = \frac{b}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)}\left(\frac{1}{n}\sum_{t=1}^n \sigma(X_t)(\hat{\sigma}(X_t) - \sigma(X_t))\right)$,

$\theta_2 = \frac{b}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)}\left(\frac{1}{n}\sum_{t=1}^n (\hat{\sigma}^2(X_t) - \sigma^2(X_t))\right)$,

$\theta_3 = \frac{\frac{1}{n}\sum_{t=1}^n \sigma^2(X_t)\epsilon_t}{\frac{1}{n}\sum_{t=1}^n \sigma^2(X_t)}$, $\theta_4 = \frac{\frac{1}{n}\sum_{t=1}^n \sigma(X_t)(\hat{\sigma}(X_t)-\sigma(X_t))\epsilon_t}{\frac{1}{n}\sum_{t=1}^n \hat{\sigma}^2(X_t)}$ and,

$\theta_5 = \frac{\frac{1}{n}\sum_{t=1}^{n}(\hat{\sigma}^2(X_t)-\sigma^2(X_t))}{\frac{1}{n}\sum_{t=1}^{n}\hat{\sigma}^2(X_t)} \frac{\frac{1}{n}\sum_{t=1}^{n}\sigma^2(X_t)\epsilon_t}{\frac{1}{n}\sum_{t=1}^{n}\sigma^2(X_t)}$.

Under assumptions A1-A4 a routine application of Kolmogorov's law of large numbers gives $\theta_3 = o_p(1)$. Now,

$$\begin{aligned}
\theta_1 + \theta_4 &= \frac{1}{\frac{1}{n}\sum_{t=1}^{n}\hat{\sigma}^2(X_t)} n^{-1}\sum_{t=1}^{n}(\hat{\sigma}(X_t) - \sigma(X_t))Y_t \\
&= \frac{1}{\frac{1}{n}\sum_{t=1}^{n}\hat{\sigma}^2(X_t)} n^{-1}\sum_{t=1}^{n}\left(\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}}\right) \\
&\times \quad (\hat{\sigma}^2(X_t) - \sigma^2(X_t))Y_t + \frac{1}{\frac{1}{n}\sum_{t=1}^{n}\hat{\sigma}^2(X_t)} n^{-1}\sum_{t=1}^{n}\frac{1}{2\sigma(X_t)}(\hat{\sigma}^2(X_t) - \sigma^2(X_t))Y_t \\
&= \frac{1}{\frac{1}{n}\sum_{t=1}^{n}\hat{\sigma}^2(X_t)}(D_{1n} + D_{2n}),
\end{aligned}$$

where $\sigma_b^2(X_t) = \theta\sigma^2(X_t) + (1-\theta)\hat{\sigma}^2(X_t)$ for some $0 \leq \theta \leq 1$ and for all $X_t \in G$. Since $\frac{1}{\frac{1}{n}\sum_{t=1}^{n}\hat{\sigma}^2(X_t)} = O_p(1)$ from Theorem 1, it suffices to consider D_{1n} and D_{2n}. We first consider D_{1n}. It is easy to see that if $\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}} = o_p(h_n)$ uniformly in G and $n^{-1}\sum_{t=1}^{n}|\hat{\sigma}^2(X_t) - \sigma^2(X_t)||Y_t| = o_p(h_n)$, then $D_{1n} = o_p(h_n^2)$. Now, $\left|\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}}\right| \leq \frac{1}{2}\underline{B}_\sigma^{-1}\frac{1}{\sqrt{\sigma_b^2(X_t)}}|\sigma(X_t) - \sigma_b(X_t)|$ and since $\sigma^2(X_t) - \sigma_b^2(X_t) = (1-\theta)(\sigma^2(X_t) - \hat{\sigma}^2(X_t))$ we have by Theorem 1 that $\sigma^2(X_t) - \sigma_b^2(X_t) = o_p(h_n)$ uniformly for $X_t \in G$. From Corollary 1 it follows that $\sigma(X_t) - \sigma_b(X_t) = o_p(h_n)$ and $\frac{1}{\sqrt{\sigma_b^2(X_t)}} = O_p(1)$ uniformly in G. Hence,

$$sup_{X_t \in G}\left|\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}}\right| = o_p(h_n)$$

and

$$\begin{aligned}
|D_{1n}| &\leq n^{-1}\sum_{t=1}^{n}|Y_t| sup_{X_t \in G}\left|\frac{1}{2\sqrt{\sigma_b^2(X_t)}} - \frac{1}{2\sqrt{\sigma^2(X_t)}}\right| sup_{X_t \in G}|\hat{\sigma}^2(X_t) - \sigma^2(X_t)| \\
&\leq o_p(h_n^2)n^{-1}\sum_{t=1}^{n}|Y_t| = o_p(h_n^2)
\end{aligned}$$

where the last equality follows from the fact that $n^{-1}\sum_{t=1}^{n}|Y_t| = O_p(1)$ by Chebyshev's inequality. Now $D_{2n} \leq n^{-1}\sum_{t=1}^{n}\frac{|Y_t|}{2\sigma(X_t)} sup_{X_t \in G}|\hat{\sigma}^2(X_t) - \sigma^2(X_t)| = o_p(h_n)n^{-1}\sum_{t=1}^{n}\frac{|Y_t|}{2\sigma(X_t)} =$

$o_p(h_n)n^{-1}\sum_{t=1}^{n}\frac{1}{2}|b+\epsilon_t| = o_p(h_n)$, where the last equality follows from $n^{-1}\sum_{t=1}^{n}\frac{1}{2}|b+\epsilon_t| = O_p(1)$ by Chebyshev's inequality. Hence, $\theta_1 + \theta_4 = o_p(h_n)$. Now,

$$|\theta_2| \leq \left|\frac{b}{\sum_{t=1}^{n}\hat{\sigma}^2(X_t)}\right| \left|n^{-1}\sum_{t=1}^{n}(\hat{\sigma}^2(X_t) - \sigma^2(X_t))\right| = O_p(1)o_p(h_n) = o_p(h_n)$$

by Theorem 1. Finally, $|\theta_5| = o_p(h_n)$ by the results from the analysis of θ_2 and θ_3. Combining all the convergence results $\hat{b} - b = o_p(1)$.