Monotone Local Linear Estimation of Transducer Functions

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy

by

David Michael Hughes

September 2014

Extended Abstract

Local polynomial regression has received a great deal of attention in the past. It is a highly adaptable regression method when the true response model is not known. However, estimates obtained in this way are not guaranteed to be monotone. In some situations the response is known to depend monotonically upon some variables. Various methods have been suggested for constraining nonparametric local polynomial regression to be monotone. The earliest of these is known as the Pool Adjacent Violators algorithm (PAVA) and was first suggested by Brunk [4]. Kappenman [29] suggested that a non-parametric estimate could be made monotone by simply increasing the bandwidth used until the estimate was monotone. Dette et al. [9] have suggested a monotonicity constraint which they call the DNP method. Their method involves calculating a density estimate of the unconstrained regression estimate, and using this to calculate an estimate of the inverse of the regression function.

Fan, Heckman and Wand [17] generalized local polynomial regression to quasi-likelihood based settings. Obviously such estimates are not guaranteed to be monotone, whilst in many practical situations monotonicity of response is required. In this thesis I discuss how the above mentioned monotonicity constraint methods can be adapted to the quasi-likelihood setting. I am particularly interested in the estimation of monotone psychometric functions and, more generally, biological transducer functions, for which the response is often known to follow a distribution which belongs to the exponential family.

I consider some of the key theoretical properties of the monotonised local linear estimators in the quasi-likelihood setting. I establish asymptotic expressions for the bias and variance for my adaptation of the DNP method (called the LDNP method) and show that this estimate

is asymptotically normally distributed and first-order equivalent to competing methods. I demonstrate that this adaptation overcomes some of the problems with using the DNP method in likelihood based settings. I also investigate the choice of second bandwidth for use in the density estimation step.

I compare the LDNP method, the PAVA method and the bandwidth method by means of a simulation study. I investigate a variety of response models, including binary, Poisson and exponential. In each study I calculate monotone estimates of the response curve using each method and compare their bias, variance, MSE and MISE. I also apply these methods to analysis of data from various hearing and vision studies. I show some of the deficiencies of using local polynomial estimates, as opposed to local likelihood estimates.

Acknowledgements

I would like to thank my primary supervisor, Dr Kamila Żychaluk for her help in completing this thesis. I appreciate her help, advice and interest in my work over the course of this period of study, and also for increasing my interest in statistics. I am also grateful to Professor Damian Clancy, my secondary supervisor, who has helped me throughout my study, and particularly during my first year when Dr Żychaluk was on maternity leave.

I would especially like to thank my family, in particular my wife Hannah, whose daily support throughout these years of Ph.D. study has been a constant encouragement to me, and my daughter Grace, whose nappy changes and crying fits have been a happy distraction in these last few months of thesis writing.

I am grateful to the Department of Mathematics at the University of Liverpool and the EPSRC Doctoral Training Grant for their financial support during the course of my study.

I am grateful to Dr. Ian C. Smith of the Computing services department at the University of Liverpool for a lot of help in using the Condor system for my simulations. This allowed me to run a lot of the simulations presented in this thesis and I am grateful for his expert help.

Contents

Chapter 1

Introduction

This thesis considers the estimation of monotone Transducer functions and Psychometric functions using non-parametric regression methods. In this introduction I will give a brief description of non-parametric regression and issues involved with this. I will then describe transducer functions and psychometric functions and explain typical methods of estimation. After this I will explain how non-parametric methods can be used to estimate these functions. Finally I will discuss the issue of monotonicity, the methods that currently exists and the benefits of them.

1.1 Transducer functions

In this thesis I consider transducer functions. These are functions that measure the response of a subject to some stimulus. Perhaps the most common example of a transducer function is a psychometric function which is a transducer function with a binomial response model. For this reason, most of the following discussion is focused on psychometric functions. The response of a transducer function does not have to be binomial however. In this thesis I also consider Poisson responses and Exponential responses. In the following I will consider a psychometric function $P(x)$ although I could equally have considered a transducer function $T(x)$. Clearly the type of response considered will affect the range of values the transducer function can take. A psychometric function is a probability and hence takes values in $[0, 1]$. If the responses were

Poisson, the transducer function would take values in $(0, \infty)$.

Psychometric functions model the dependence of a subject's performance upon some physical stimulus. They are used in a wide variety of psychophysical experiments. A psychometrician/physchophysicist, is usually interested in measuring how an observer's performance improves with increasing stimulus level. Examples of such studies come from a wide range of fields including vision, hearing and brain function tests. For example, in a vision study a patient may have a light flashed in their eye repeatedly and they are asked if they can see it. The probability of successfully detecting the light signal as the intensity increases is recorded. In an auditory test, a noise of varying duration may be presented to the subject in one of two time intervals. The subject is required to state which of the two time intervals the noise occurred in and the probability of correctly identifying the noise as the length of signal changes is calculated.

A psychometric function, $P(x)$, measures the probability of a correct response at a given level of stimulus x. Psychophysical data is collected by testing an observer's response, r, at n different stimulus levels. In a typical psychometric study each of r individual observers will be tested in this way. Both here, and for the rest of this thesis, the stimulus level will be labelled x, the number of stimulus levels, n and the number of repeats r. Notice also that for any particular patient, responses at different stimulus levels will clearly be correlated. However, this is not considered in this thesis where I make the assumption that individual responses are independent.

The clinician is usually interested in two summary statistics. The **response threshold** is the stimulus intensity required to produce a given level of performance. For example, in the vision example mentioned above, a clinician may be interested in the light intensity required to allow 90% of observers to detect it. The **slope** of the psychometric function describes the rate at which the performance increases with an increasing stimulus intensity.

The most widely used method of estimating psychometric functions is a parametric one. The method is based on maximum likelihood and is described in detail in two papers by Wichmann and Hill ([48] and [49]). More recently Żychaluk and Foster have suggested a nonparametric

method of estimating the psychometric function and they describe some of the benefits of adopting this approach ([18] and [19]). A brief overview of these methods is provided in this introduction. I then explain the issue of monotonicity and present some of the existing methods described in the literature. Following this I give a brief overview of the chapters in this thesis.

1.2 Parametric estimation of a psychometric function

The general form of a psychometric function is often given as

$$P(x) = \gamma + (1 - \gamma - \lambda)F(x),$$

where $\gamma, \lambda \geq 0$, $P(x)$ is the proportion of correct responses at a stimulus level, x. The function, $F(x)$, is usually taken to be a smooth monotonic function with values in $[0, 1]$. Important features of the psychometric function are the asymptotes γ and λ. The lower asymptote, γ, is the base level of performance in the absence of any stimulus. In some situations this is described as the "guessing rate", i.e the probability that an observer guesses the correct response by chance. The upper asymptote, λ, is often called a "lapsing rate" and is interpreted as the rate at which an incorrect response is given regardless of the stimulus intensity. For example, in the vision study previously mentioned, an observer may blink and miss the light signal.

The function $F(x)$ takes the form of the inverse link function when the psychometric function is modelled as a generalized linear model (GLM). Most common choices of F are sigmoid functions such as the cumulative Gaussian, Weibull and logistic distributions. The definition of these distributions is given in Table 1.1.

A psychometric function is commonly estimated parametrically by fitting a specific GLM. A GLM consists of three parts. The first part is a random component. This component is a member of the exponential family. That is, it satisfies the form

$$f_Y(y; \theta, \phi) = \exp\left(y\theta - b(\theta)\right)/a(\phi) + c(y, \phi),$$

for some specific functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$. Here θ is the parameter to be estimated, often related to the mean of the distribution, whilst ϕ is a dispersion parameter (or sometimes called

Table 1.1: Some common $F(x)$ distributions

Distribution	$\boldsymbol{F(x)}$
Gaussian	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-z^2/2) dz$
Weibull	$0, \quad -\infty < x \leq 0$ $1 - \exp(-x^{\beta}), \quad 0 < x < \infty$ and $\beta > 0$
Logistic	$[1 + \exp(-x)]^{-1}, \quad -\infty < x < \infty$

nuisance parameter), and is often related to the variance of the distribution. More information about generalised linear models can be found in McCullagh and Nelder [33]. In the case of a psychometric function this is usually taken to be a binomial distribution with mean μ.

The second component, the systematic component, is a function η which is a linear predictor of the explanatory variables x. The final component is a link function, g. The link function describes the link between the systematic component and the random component. That is to say, it describes how the outcome, P, is influenced by the stimulus level x. The psychometric function, $P(x)$, is modelled as

$$\eta(x) = g[P(x)].$$

The choice of link function can ensure that the estimated response is within the correct bounds. In this case, that means constraining the estimates of the psychometric function to be in $[0, 1]$, which is, of course, an obvious requirement of a probability. The estimate of the psychometric function is then obtained by using the inverse of the link function, g^{-1}. The systematic component is almost always a first degree polynomial. That is,

$$\eta(x) = a_0 + a_1 x.$$

When psychometric data are collected, there are m_i trials at each of n stimulus levels x_i with $1 \leq i \leq n$. The responses y_i are typically assumed to have a binomial distribution with parameters $(m_i, p(x_i))$. This is the random component of the GLM. The log-likelihood is then given by

$$
\begin{aligned}
l(a_0, a_1) &= \sum_{i=1}^{n} \ln[P^{r_i}(x_i)(1 - P(x_i))^{m_i - r_i}] \\
&= \sum_{i=1}^{n} r_i \ln P(x_i) + (m_i - r_i) \ln[1 - P(x_i)] \\
&= \sum_{i=1}^{n} r_i \ln[g^{-1}(a_0 + a_1 x_i)] + (m_i - r_i) \ln[1 - g^{-1}(a_0 + a_1 x_i)].
\end{aligned}
$$

The coefficients a_0 and a_1 are then estimated using the iteratively reweighted least squares method outlined by McCullagh and Nelder [33] (page 40). This method is often referred to as the Fisher's scoring method and is a variant of the Newton-Raphson method. This is a numerical method that approximates the values of the parameters by successive weighted iterations of the likelihood function. Typically a few iterations are sufficient to obtain good estimates of the required parameters.

For this method the guessing and lapsing rates must be stated in the first place. Obviously, in many cases these will not be known and so must be estimated. It is possible to create an estimating procedure where the values for the guessing and lapsing rates, γ and λ, are estimated as well (see [51]).

Parametric modelling, as described above, is very commonly used in estimating psychometric functions but it is not without its problems. The correct form of the link function is rarely known and the use of an incorrect link function will lead to results that are not consistent. The specification of the guessing and lapsing rates can cause similar problems (see discussion in Foster and Żychaluk [18]).

1.3 Non-parametric regression

I first give a brief outline of some of the developments in the field of non-parametric regression. This gives the setting to the later discussion regarding psychometric and transducer functions. There are a number of methods of non-parametric regression including spline based methods (see for example the book by Green and Silverman [23]), kernel based methods (see the books by Fan and Gijbels [16] and Wand and Jones [44]) and wavelet transforms (see the overview

by Donoho, Johnstone, Kerkyacharian and Picard [12]). This thesis is only concerned with the kernel based methods. Kernel regression has been dealt with in many books, two of the most popular being those of Fan and Gijbels, [16] and Wand and Jones [44]. The very basic problem is to model a response Y_i and its dependence on some variable X_i. A realisation of the random variable X_i is denoted x_i. The response, Y_i is modelled as a smooth function, m depending on the variable , x,

$$Y_i = m(x_i) + v^{1/2}(x_i)\varepsilon_i, i = 1, \ldots, n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent random variables for which

$$E(\varepsilon_i) = 0 \text{ and } Var(\varepsilon_i) = 1.$$

The function, m is the regression function, whilst v is the variance function. If $v(x_i) = \sigma^2$ for all i then the model is said to be homoscedastic. If this is not the case the model is heteroscedastic.

Parametric methods of modelling are extensively described in the book by McCullagh and Nelder [33] on generalised linear models. In many cases a parametric regression estimate may be suitable for use in describing the data. However, parametric methods require the specification of a model and the parameters to be used in a model. In practice these are often not known. If a wrong model is chosen then any conclusions drawn from it may be spurious. For a generalised linear model, a link function must be specified. But there is often insufficient information to make a correct choice (See Foster and Żychaluk [18]). One way to overcome this problem associated with parametric fitting methods is to consider local smoothing. In practice in this method we fit a local regression at each point in a grid covering the range of x values. This local fit is used to provide an estimate of the true regression function in that locality.

1.3.1 Local constant estimates

Perhaps the simplest form of this is a local constant estimate. This is in some sense like a weighted average of the points in a locality. There are three common forms of a local constant estimator.

- **Nadaraya-Watson estimator** ([45] and [35])

  The Nadaraya-Watson estimator takes the form

  $$\hat{m}_h(x) = \frac{\sum_{i=1}^n K_h(X_i - x)Y_i}{\sum_{i=1}^n K_h(X_i - x)},$$

  where K is a real valued function that assigns weights, usually taken to be a symmetric probability density function. Here and for the remainder of this thesis, $K_h(\cdot) = K(\cdot/h)/h$ and h is a bandwidth which controls how much weight is applied to each point.

- **Gasser-Müller estimator** [21] Here the data have been ordered according the X variable

  $$\hat{m}_h(x) = \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} K_h(u - x)du\, Y_i,$$

  with $s_i = (X_{(i)} + X_{(i+1)})/2$, $X_0 = -\infty$ and $X_{n+1} = +\infty$.

- **Priestly-Chao estimator** [38]

  $$\hat{m}_h(x) = \sum_{i=1}^{n} \{X_{(i)} - X_{(i-1)}\} K_h(X_{(i)} - x)\, Y_i,$$

  where $X_{(i)}$ are the ordered X variables $X_1, \ldots X_n$.

The Nadaraya-Watson estimator has an increased asymptotic bias compared to the other two methods, although it has a smaller asymptotic variance. The disadvantage of the Gasser-Müller and Priestly-Chao estimators is that they have boundary bias problems. Some of these problems lead to the use of local linear regression methods described later. Discussions in Wand and Jones [44] (page 131) and Fan and Gijbels [16] (page 17) explain some of the properties of these estimators.

Two obvious questions arise from these estimators. The first is as to which kernel should be chosen, whilst the second is as to which bandwidth, h, should be chosen. The question of bandwidth choice is discussed in Section 1.4. Here the discussion focusses on the choice of kernel. More comprehensive detail and discussion, along with a list of some kernels is given in Wand and Jones [44] (page 31) and Silverman [42] (page 43). A list of some commonly used kernels and their properties is provided in this thesis in Chapter 3.

The first kernel to be suggested was the Epanechnikov kernel. This is the kernel that minimises the asymptotic mean integrated square error (AMISE) for kernel density estimation. The AMISE is a large sample approximation to the MISE. (See Wand [44] page 30). Other kernels are not much less efficient than the Epanechnikov kernel and may be easier to compute in practice (see Silverman [42]). Hence other kernels, such as the Gaussian kernel may be more commonly used in reality. The point is that there is not much loss in efficiency.

1.3.2 Local Polynomial estimates

One of the most commonly used forms of non-parametric regression is the local linear regression method. This is a type of local polynomial regression, and has many desirable properties. For example, since the modelling takes place locally, only a small degree of polynomial is required to achieve a small bias, whereas for parametric regression a much higher degree of polynomial would be required. This is obviously problematic as it can result in overparametisation. The properties of local polynomial regression can be found in Fan and Gijbels [16] and also articles by Fan [13][14]. The idea here is to estimate a locally linear regression within a local neighbourhood about a data point x_i. This linear regression estimate is then used to calculate an estimate of the value of the function at the point x. The form of the local linear estimate is given as

$$\hat{m}_h(x) = n^{-1} \sum_{i=1}^{n} \frac{\hat{s}_2(x;h) - \hat{s}_1(x;h)(x_i - x)K_h(x_i - x)Y_i}{\hat{s}_2(x;h)\hat{s}_0(x;h) - \hat{s}_1(x;h)^2} \tag{1.1}$$

where

$$\hat{s}_r(x;h) = n^{-1} \sum_{i=1}^{n} (x_i - x)^r K_h(x_i - x).$$

Fan and Gijbels [16] demonstrate that there is an advantage in terms of the asymptotic bias and variance of the estimates if the local linear method is used instead of either the Nadaraya-Watson or Gasser-Müller estimates. This is because the local linear method is design adaptive. The weighting scheme of the local linear method manages to maintain smooth weights in the same way as the Nadaraya-Watson estimator and to keep a smaller variance but also keeps

the smaller bias of the Gasser-Müller estimate. The local linear estimate adapts readily to boundary points whilst the other methods do not.

A more detailed discussion of these properties can be found on page 22 of Fan and Gijbels [16]. They show that there is an advantage to using odd degree polynomials in kernel polynomial regression. Clearly fitting higher order polynomials leads to a possible reduction of the bias. But another result of increasing the degree of polynomial used is that the variability increases. Fan and Gijbels [16] state that an odd degree polynomial of degree $2q + 1$ reduces the bias compared to a model of degree $2q$ but with no additional increase in variability. This reduction in bias, along with no increase in variability makes it desirable to use odd degree polynomials. Fan [14] also explains that even degree polynomial fits suffer from low efficiency and boundary effects, whereas odd degree polynomial fits do not have these problems. Fan and Gijbels [16] gives a comprehensive overview of local linear methods and explains the advantages of using them.

1.3.3 Non-Parametric density estimation

Before leaving this section on non-parametric methods I first give a brief overview of a related topic, that of density estimation. This is so as so give background to a later discussion of monotonicity constraint methods. Good overviews of kernel density estimation can be found in Silverman [42] and in Wand and Jones [44]. The challenge is to calculate an estimate of a density $f(x)$ without making distributional assumptions. The standard kernel density estimator is given by

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.2}$$

where K is a kernel function applying weights to data points depending on their closeness to x, and h is a bandwidth. Since a kernel is assumed to be a continuous function kernel methods have problems at the boundary of the range of x. This is a result of the fact that some of the neighbourhood created by the bandwidth h falls outside of the boundary. There will be no observations in this area, and hence bias is introduced in boundary regions. Methods have been suggested to compensate for this boundary bias. For and overview see Fan and Gijbels

(page 69, [16]) or Wand and Jones (46-48 [44]). In this thesis the boundary effects have been ignored and the results presented assume we are considering a region away from the boundary.

1.4 Bandwidth selection

Each of the regression estimates outlined so far have been dependent upon a choice of bandwidth, h. The bandwidth essentially determines how much weight is added to each point in the estimation procedure. Whilst the choice of kernel used is relatively unimportant, the choice of bandwidth can very significantly influence the estimate obtained. If the bandwidth chosen is too large then the estimate will be oversmoothed. This may mean that important features of the data are lost and the regression estimate will tend towards a simple linear regression. Conversely, if the bandwidth chosen is too small we may significantly under-smooth the data. We then risk drawing conclusion based upon spurious peaks and the regression estimate corresponds to little more than linear interpolation between the data points. The challenge, then, is to choose a bandwidth that is neither too large nor too small.

In an ideal situation we would choose a bandwidth, h that was optimal in terms in some metric such as the mean square error (MSE see (3.3)), or the mean integrated square error (MISE see (3.10)). However, the expression for the asymptotic MSE and MISE depends on unknown quantities. In practice these will be unknown since the precise form of the response function is not known. As a result approximation methods are needed in order to estimate the unknown quantities and calculate a suitable bandwidth. There are many ways in which to choose the bandwidth, with the two main families of method being cross-validation bandwidths and plug-in bandwidths. The plug-in methods provide ways of estimating the unknown terms in the MSE or MISE expressions. I will discuss only two methods in detail.

- **The plug in bandwidth method** (Ruppert, Sheather and Wand [40])

  We will assume that the errors are homoscedastic, with common variance σ^2, and also that the X_i's are from a compactly supported density on $[0, 1]$. The mean integrated square error (MISE) is the expected value of the integral of the squared distance between

a non-parametric regression estimate (of any of the common kernel methods) $\hat{m}(x;h)$ and the true function $m(x)$ (this is the L_2 risk funtion). That is to say,

$$MISE\{\hat{m}(\cdot;h)\} = E[\int\{\hat{m}(x;h) - m(x)\}^2 dx].$$

Ruppert, Sheather and Wand [40] propose weighting the MISE by a density $f(x)$ which puts more emphasis on regions where there are more data. They give the weighted MISE

$$MISE\{\hat{m}(\cdot;h)|X_1,\ldots,X_n\} = E[\int\{\hat{m}(x;h) - m(x)\}^2 f(x)dx|X_1,\ldots,X_n].$$

They show that the asymptotically optimal bandwidth, which minimises the asymptotic MISE is given by

$$h_{AMISE} = C_1(K)\left[\frac{\sigma^2}{\theta_{22}n}\right]^{1/5}$$

where $C_1(K) = \{R(K)/\mu_2(K)^2\}^{1/5}$ with $R(K) = \int K(u)^2 du$ and $\mu_2(K) = \int u^2 K(u)du$. The term θ_{22} is a special case of the notation

$$\theta_{rs} = \int m^{(r)}(x)m^{(s)}(x)f(x)dx$$

Wand and Jones [44] and Ruppert, Sheather and Wand [40] state that a natural estimator for θ_{22} is

$$\hat{\theta}_{22}(g) = n^{-1}\sum_{i=1}^{n}\widehat{m^{(2)}}(X_i;3,g)^2$$

where $\widehat{m^{(2)}}(X_i;3,g)$ is a third degree polynomial estimate of the second derivative of $m(x)$ with a bandwidth g, whilst a natural estimator for σ^2 is

$$\hat{\sigma}^2(\lambda) = \nu^{-1}\sum_{i=1}^{n}\{Y_i - \hat{m}(X_i;1,\lambda)\}^2$$

where

$$\nu = n - 2\sum_{i=1}^{n}w_{ii} + \sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}^2$$

and

$$w_{ij} = \boldsymbol{e}_1^T(\boldsymbol{X}_{X_i}\boldsymbol{W}_{X_i}\boldsymbol{X}_{X_i})^{-1}\boldsymbol{X}_{X_i}^T\boldsymbol{W}_{X_i}\boldsymbol{e}_j,$$

and λ is a bandwidth. In this notation $\boldsymbol{e}_j$ refers to a vector of length $(p+1)$ (where p is the degree of the local polynomial used) with 1 in the j'th entry and zeros everywhere else. The matrix $\boldsymbol{W}$ is a diagonal matrix with diagonal entries $K_h(x_1 - x), \ldots, K_h(x_n - x)$. The design matrix $\boldsymbol{X}$ is an $(n \times 2)$ matrix with the first column entries being 1, and the second column entries being $(x_1 - x), \ldots, (x_n - x)$. The direct plug in bandwidth is then calculated using

$$\hat{h}_{DPI} = C_1(K) \left[\frac{\hat{\sigma}_1^2(\lambda)}{\hat{\theta}_{22}(g)n} \right]^{1/5}$$

Wand and Jones [44] (page 139) give more details on this method of bandwidth estimation as well as details on selecting auxiliary bandwidths g and λ. Ruppert, Sheather and Wand [40] describe some of the properties of this direct plug-in bandwidth. In particular they show that this estimate has a good convergence rate to the MISE optimal bandwidth h_{MISE}.

- **Cross Validation methods** (See Fan and Gijbels [16])

  Recall that $\hat{m}_h(x)$ is an estimate of the regression function m at x using a particular bandwidth h. The idea for cross validation is to leave out the i^{th} observation and calculate an estimate of the regression function using the remaining data, $(X_j, Y_j), j \neq i$, and then validate (or check) this using the prediction error $Y_i - \hat{m}_{h,-i}(X_i)$. This can be repeated at each i. We want to find the bandwidth h_{CV} which minimises the weighted average of the squared errors,

$$\hat{h}_{CV} = \arg\min_h n^{-1} \sum_{i=1}^{n} \{Y_i - \hat{m}_{h,-i}(X_i)\}^2 w(X_i).$$

Wand and Jones [44] (page 85) discuss the relative performance of various bandwidth selection procedures. They conclude that direct plug in methods provide a good compromise between the bias and variance of the resulting estimate. However for some distributions the asymptotic results that this method relies upon may be less accurate. This is because the plug in methods attempt to approximate terms in the asymptotic mean integrated square error. This expression is only correct for large samples and may not be true for smaller samples, leading to poor estimates of a suitable bandwidth. The cross-validation methods do not rely on asymptotic

results since they approximate the integrated square error and so depend only upon the given sample. In this sense they can adapt easily to any sample and large samples are not required to provide a good estimate of a bandwidth. However cross validation methods can result in high sample variability and can sometimes be less useful then in practice.

As an aside, it should be noted that these methods all provide a constant bandwidth to be used across the range of x values. This need not necessarily be the case, although we confine our investigations to this situation in this thesis. A variable bandwidth could be used for example, which varies depending on how dense the stimulus values were in a region. More details and explanation of variable bandwidths may be found in Fan and Gijbels [15].

1.5 Local linear regression in exponential family models

As was described in Section 1.2, a psychometric function is usually estimated as a GLM, which allows the responses to follow any distribution in the exponential family. The use of a link function constrains the resulting estimates to lie within the correct range. Non-parametric modelling methods can also be adapted to a setting where the variance of the response is dependent upon its mean. This method has been proposed by Fan, Heckman and Wand, [17] and has been applied specifically to the psychometric function by Foster and Żychaluk [18].

1.5.1 Modelling the psychometric function

The procedure for obtaining a non-parametric estimate of a psychometric function is in many ways broadly similar to the parametric method. However, we now estimate the η function locally. We can accept any given link function and we estimate the η function, which in contrast to the parametric version we no longer assume to be a linear function. For any given point x we can estimate the value of the η function at any point close to x using a Taylor expansion,

$$\eta(u) \approx \eta(x) + (u - x)\eta'(x),$$

where $\eta'(x)$ is the derivative of η. This estimation is obviously better the closer u is to x.

As in the parametric case, the responses come from a binomial distribution with mean $P(x)$. The canonical link for a binomial distribution is the logit link,

$$g(P(x)) = \ln\left\{\frac{P(x)}{1-P(x)}\right\}.$$

We can use this information to express the local log-likelihood, which is very similar to the parametric log-likelihood function with the exception that we now have a weighting term.

$$\begin{aligned}
l(\alpha_0, \alpha_1; x) &= \sum_{i=1}^{n} \ln[P^{r_i}(x_i)(1-P(x_i))^{m_i-r_i}] K_h(x-x_i) \\
&= \sum_{i=1}^{n} \left[r_i \ln P(x_i) + (m_i - r_i)\ln[1-P(x_i)]\right] K_h(x-x_i) \\
&= \sum_{i=1}^{n} \left[r_i \ln[g^{-1}(\alpha_0+\alpha_1 x_i)] + (m_i-r_i)\ln[1-g^{-1}(\alpha_0+\alpha_1 x_i)]\right] K_h(x-x_i) \\
&= \sum_{i=1}^{n} \Big[r_i \ln\{(1+\exp(-(\alpha_0+\alpha_1 x_i)))^{-1}\} \\
&\quad + (m_i-r_i)\ln\{1-(1+\exp(-(\alpha_0+\alpha_1 x_i)))^{-1}\}\Big] K_h(x-x_i) \\
&= -\sum_{i=1}^{n} \left[r_i \ln\{1+\exp(-(\alpha_0+\alpha_1 x_i))\} + (m_i-r_i)\ln\{1+\exp(\alpha_0+\alpha_1 x_i)\}\right] K_h(x-x_i).
\end{aligned}$$

We can maximise this local log-likelihood using the Fisher scoring method (as with the parametric case and explained in [33] and [16]). This is a numerical approximation method since an exact solution is often not possible. This will give us the values of α_0 and α_1 that maximise the local log-likelihood. These values are then used to obtain the maximum likelihood estimate for $\eta(x)$ as $\hat{\eta}(x) = \hat{\alpha}_0$. We can use the link function to transform this in order to get an estimate for the psychometric function,

$$\hat{P}(x) = g^{-1}(\hat{\alpha}_0). \tag{1.3}$$

Fan and Gijbels [16], Wand and Jones [44] state that the influence of the weight function depends less on the kernel, K, chosen than on the bandwidth h.

This method has been applied to psychometric functions by Foster and Żychaluk [18] who discuss the properties of using this method for psychometric functions. They say that it is best if K has unbounded support which would be obtained if a Gaussian kernel was used for example.

They give various discussion of methods of bandwidth choice and say that cross-validation techniques performed no worse than other methods.

One of the advantages of the non-parametric method outlined above is that there is no need to specify guessing and lapsing rates of the psychometric function. This is desirable since often these values are not known and it can be costly to choose wrong values. By this, I mean that choosing wrong values can lead to inaccurate estimates of the psychometric function. The rest of their paper introduces seven data sets from psychometric studies in vision and hearing. They compare commonly used methods of parametric fitting and the non-parametric method they have outlined. Through various plots and discussions they demonstrate that the non-parametric estimate performs well in all of the examples. They conclude

> "As a matter of principle, a correct parametric model will always do better than a nonparametric one, simply because the parametric model assumes more about the data, but given an experimenter's ignorance of the correct model, the local linear method provides an impartial and consistent way of addressing this uncertainty."

One of the examples that they provide is taken from a study by Schofield *et al.* [41]. In their study they showed a subject a moving adaptation stimulus, followed by a test stimulus. The images moved either up or down. The subjects were asked whether the adapting stimulus and the test stimulus moved in the same direction. The number of correct responses was recorded at seven different relative modulation depths, with 10 trials conducted at each depth level. The results are shown in Figure 1.1. Various methods of estimating a psychometric function have been calculated and are also shown in the figure. The simplest regression method, that of simple linear regression is shown to demonstrate how unsatisfactory it is. Three different parametric fits with logit, probit and Weibull links are shown. These provide reasonable, though not exceptional, fit to the data. It is clear that the non-parametric fit of Żychaluk and Foster performs at least as well as the parametric methods.

To compare the methods more I calculated the deviance of each of these estimated models, following Foster and Żychaluk [18] . This is said to be a good way of comparing models

(McCullagh and Nelder [33]). The deviance for each model (with a binomial response) is given by

$$D = 2\sum_{i=1}^{n} r_i \ln \frac{r_i}{m_i \hat{P}(x_i)} + (m_i - r_i) \ln \frac{m_i - r_i}{m_i - m_i \hat{P}(x_i)} \tag{1.4}$$

The larger the value of D, the larger the discrepancy between the data and the fitted values. It is unclear as to exactly what form the distribution of D takes. Asymptotically it is $\chi^2(n-k)$ where k is the number of fitted parameters.

The assumptions required for this asymptotic results may not hold in the small sample size psychometric experiments we consider here. All this being said, the deviance is still a useful means of comparing the various fits. The degrees of freedom for the non-parametric fit is calculated by using the trace of the Hat matrix (see Hastie and Tibshirani [28] for more details). It will be non integer in most cases. For the parametric fits, the degrees of freedom is given by the number of points, n, minus the number of parameters, k.

The comparison of deviance statistics is given for each model in Table 1.2. Bearing in mind that this is merely an indication of model suitability and not an exact result, the indication is still that the linear model performs very poorly and models the data badly. The parametric fits all give unsatisfactory fits indicated by the low significance values, whilst the non-parametric estimate, performs considerably better than the alternatives with a much lower deviance and corresponding significance value.

Table 1.2: A comparison of the deviances of various estimates of the psychometric function.

	Deviance	DoF	p
Logit	14.46	5.00	0.01
Probit	17.33	5.00	0.004
Weibull	10.64	4.00	0.031
Linear	25.17	5.00	0.00
Non-Parametric	4.75	4.24	0.34

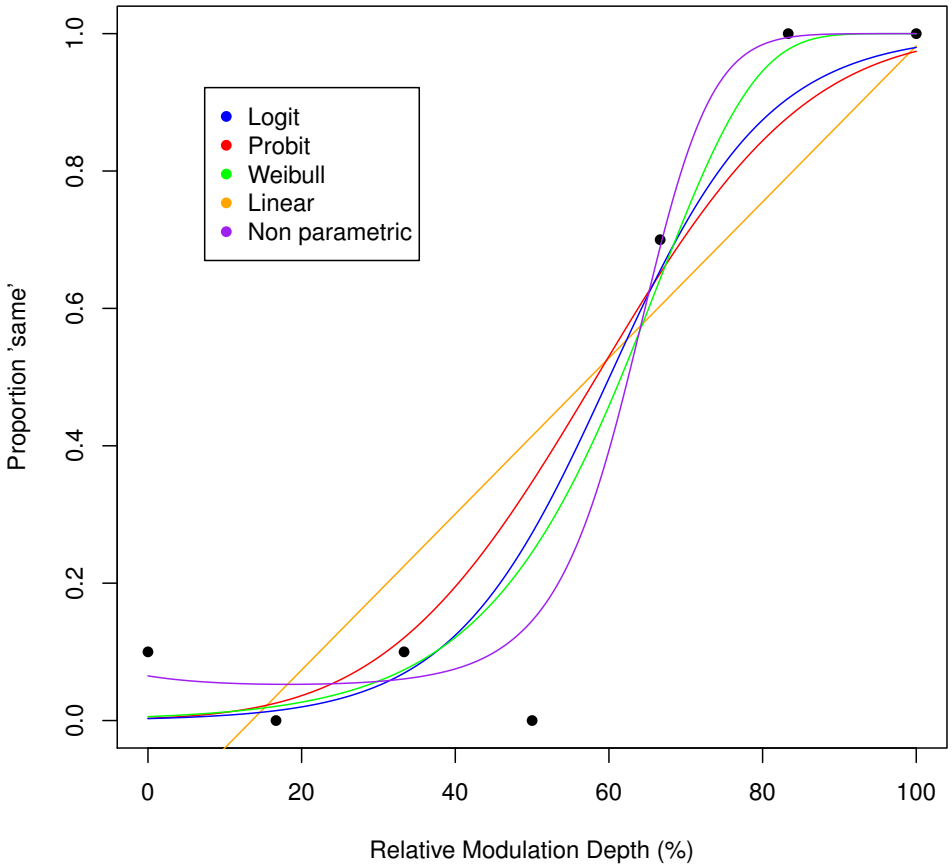

Figure 1.1: Psychometric function estimates obtained using various parametric methods and a non-parametric estimate.

1.5.2 Modelling Transducer functions

The method outlined by Foster and Żychaluk is not limited to psychometric functions. They have also discussed how it can be used for transducer functions [19]. A psychometric function is a particular kind of transducer function. Transducer functions measure the dependence of a response on the stimulus level. The difference in this case is that the response function is now not necessarily binary or binomial. This means that the link function used will be different. In addition the limits of the response function will be different. For binomial responses the response is constrained to lie in $[0, 1]$ since it is a probability and the logit link is used since

this is the canonical link for binomial responses. If the response follows a Poisson distribution then the log link would be the canonical link and the responses would lie in $[0, \infty)$. Any other response belonging to the exponential family could in theory be used with this method.

1.5.3 Wider Applications

There are many situations in clinical studies where dose-response curves are used. These test how a patient responds to a treatment as the dosage increases. A dose-response curve is used to determine the effective dose, that is, the dosage required to produce a positive response in $100\alpha\%$ of patients. In this sense it is comparable to the threshold that psychometricians are interested in. Dose-response curves are essentially similar to psychometric functions and so the method of Żychaluk and Foster could be used for them too. A survey of methods of estimating dose-response curves is given by Dette and Scheder [11]. This paper will be discussed in more detail later in this thesis. Closely related is the study of toxicity levels in environmental studies. Often scientists are interested in the level of chemicals required for a particular environment to become toxic and hazardous. This can be of interest with regards to chemical waste etc. and the method of Żychaluk and Foster could also be used in this setting.

Guidi *et al.* [24] have used the method of Żychaluk and Foster [18] in a study of visual illusions. They use the method to estimate psychometric functions for each subject from which they determine distance thresholds. Zhang *et al.* [52] use the model free method to estimate psychometric functions in a study of brain abnormality detection using magnetic resonance imaging (MRI). Lee and Noppeney [30] conduct a study of how long term music training (such as playing the piano) affects a person's perception. They use the model free method to estimate psychometric functions so as to avoid making distributional assumptions. In a study by Whitmal and DeRoy [46], investigating the role of importance functions in speech intelligibility, they say that the parametric method of estimating psychometric functions using Weibull or logistic functions (as in Wichmann and Hill [48]) provided poor fits to the data. To overcome this they use the model free method of Żychaluk and Foster. In a subsequent paper [47], the same authors discuss how the importance functions can develop the use of cochlear

implants.

1.6 Estimating monotone psychometric functions

In many situations scientific knowledge suggests that there should be a monotone relationship between the stimulus levels, x_i, and the psychometric function, $P(x)$. In the example considered in Figure 1.1, it makes sense that as the modulation depth increases so does the probability of correctly identifying the motion direction. It is difficult to think of reason why this would not be the case. Yet by observing Figure 1.1, you can see that there is a departure from monotonicity at areas with low stimulus levels in the nonparametric estimate. Granted, in this example the departure from monotonicity is hardly big enough to be considered significant, and appears unlikely to dramatically affect any inferences drawn. However, later in this thesis we will give examples of more serious departures from monotonicity.

The non-parametric method has no way of ensuring that the resulting estimate is monotone. The purpose of this thesis is to investigate methods of constraining an estimate of the psychometric function to be monotone. Various methods have been suggested for non-parametric regression estimators. I give a brief outline of these methods now and provide references to sources of more information on them. To the best of my knowledge, not all of these methods have been been applied to estimation of psychometric functions, or, more generally, likelihood based methods.

Throughout this thesis it is assumed that all mention of monotone implies monotone increasing. For ease of presentation and discussion I only consider the case when the response is monotone increasing. However, the methods described in this thesis are easily adapted to monotone decreasing settings and all the results still remain valid.

1.6.1 Methods of monotonicity constraint

Pool Adjacent Violators Algorithm (PAVA)

This method requires first ordering the responses Y_i according to the order statistic $X_{(1)} < \ldots < X_{(n)}$ of $\{X_i\}_{i=1}^n$. Brunk [4] suggested using

$$Y_i^* = \max_{s \leq i} \min_{t \geq i} \frac{1}{t-s+1} \sum_{j=s}^{t} Y_{[j]}$$

as an estimate of the regression function at the point $X_{(i)}$. Here $Y_{[j]}$ is the Y value corresponding to the ordered X variables. This process is known as the Pool-Adjacent-Violators-Algorithm, (PAVA) and is described by Härdle [27]. This procedure removes any dips in the data set and a monotone regression estimate can be obtained by linear interpolation between the Y_i^* points. Obviously this curve may well not be very smooth. So it is possible to apply one of the various non-parametric regression procedures outlined previously to the 'new' data (X_i, Y_i^*). We denote the estimate produced using this method as $\hat{m}_{IS}(x)$ since we first isotonize the data and then smooth it.

The main motivation for this method is to 'smooth' out any points within the set of responses that violate the monotonicity condition. The algorithm proceeds across the range of x values ensuring at each step that the adjusted 'data point' Y_i^* is greater than or equal to the previous data points. Hence the resulting responses essentially form a kind of step function. This is obviously not ideal for a realistic model of the response and hence these monotonised data points can be smoothed using any nonparametric regression method.

Alternatively we could first smooth the data and then isotonize the result as suggested in [20]. We denote the estimate obtained using that method as $\hat{m}_{SI}(x)$. Mammen [32] considers both of these estimates. He concludes that these estimates are asymptotically similar, but that their differences cannot be neglected in moderate sample sizes. Dette and Pilz [10] state that the MSE of an estimate obtained by smoothing a monotonised data set is usually smaller than the MSE of an estimate obtained by monotonising a smooth curve estimate and in this sense it is better to use in many cases. Mammen [32] concludes that the choice of which method to use depends mainly on the choice of kernel used, and specifically how smooth it is.

This method can obviously be easily adapted to likelihood based settings since the PAVA step is separate to the smoothing step. This has been done by Park and Park [37] for example. They use a likelihood based unconstrained regression estimate, calculated using a method similar to Fan, Heckman and Wand [17], and then apply the PAVA algorithm to this estimate to obtain a monotone regression estimate. Within this paper they also compare the estimate with the bandwidth monotone estimate described in 1.6.1. The paper does not contain any theoretical results and simply performs a simulation analysis comparing various methods of monotonicity constraint. They conclude that the estimate calculated by adjusting the bandwidth generally performs better than the PAVA estimate.

Bandwidth adjustment

A very simple way of obtaining a monotone regression estimate is to increase the bandwidth used. Obviously, the larger the bandwidth used, the more the data is smoothed and the closer to linear regression the estimate becomes. So a simple method of calculating a monotone regression estimate would be to select the smallest bandwidth h_0 such that for all $h \geq h_0$, $\hat{m}(x)$ is monotone.

Incidentally, the test for monotonicity suggested by Bowman *et al.* uses a similar idea to this. They select the first bandwidth h_c, for which $\hat{m}(x)$ is monotone and use it in their test. One issue is whether the regression estimate is then monotone for all $h \geq h_c$. They show that for some estimation methods this is true. For the local linear estimate, however, it is not true. They mention that through simulation studies only 2 out of 90,000 cases violated this condition. This suggests that simply selecting the first bandwidth for which the resulting estimate is monotone would be sufficient in most cases. More details are found in Section 1.6.2.

The first mention of this idea when using likelihood based regression comes from Kappenman [29] who is trying to calculate monotone estimates of dose-response curves. It is also discussed by Park and Park [37]. One advantage to this method is that the monotone estimate retains any desirable features from the unconstrained estimate, which is not always true when constraints are added. However, one obvious problem of this technique is that simply increasing

the bandwidth could often lead to oversmoothing which may mask important features of the response curve. This may lead to this method being inaccurate in practice.

In practice I calculate this estimate as follows. It is assumed that a first bandwidth, h_{init}, was selected and the regression estimate obtained using this bandwidth was not monotone. I then set up a grid of points from h_{init} to the maximum value of the range of the design points. I then proceed through this grid, fitting a regression estimate using each bandwidth. I continue until the regression estimate is monotone. In order to increase the speed of this calculation I test the first bandwidth in the grid first. If this is not monotone, I select the middle bandwidth in the grid and fit a regression estimate. If this estimate is not monotone I then select the grid value halfway in between the middle value and the final value. If the middle bandwidth did produce a monotone estimate then I choose the bandwidth halfway in between the first bandwidth and the middle bandwidth. I proceed in a similar manner until I find the first bandwidth for which monotonicity is achieved. This method of determining the appropriate bandwidth attempts to minimise the amount of calculations necessary and hence works fairly quickly.

A 'tilting' method

Hall and Huang [26] have suggested a weight modifying method for obtaining a monotone regression estimate. We require the unconstrained estimate to be such that it can be written in the form

$$\hat{m}(x) = \frac{1}{n}\sum_{i=1}^{n} A_i(x)Y_i,$$

where the weights $A_i(x)$ depend only on the explanatory variables X_i but not on the responses Y_i. (Most common estimates can be written in this form.) They suggest replacing this estimate with the estimate

$$\hat{m}(x|p) = \sum_{i=1}^{n} p_i A_i(x)Y_i,$$

and selecting the probability vector $\boldsymbol{p} = (p_1, \ldots, p_n)$ in such a way that it is as close as possible to the uniform vector $\boldsymbol{p}_u = ((1/n), \ldots, (1/n))$ whilst making the resulting estimate monotone. This requires some way of measuring the 'distance' between $\boldsymbol{p}$ and $\boldsymbol{p}_u$. Hall and Huang [26]

suggest various measures (as well as pointing to more options in Cressie and Read [7]) such as the Kullback-Leibler divergence

$$d_0(p,q) = -\sum_{i=1}^{n} q_i \log\left(\frac{p_i}{q_i}\right).$$

The value of p that minimises d_0 (where $q = p_u$) is used in the calculation of $\hat{m}(x|p)$. These measures are a way of comparing the monotone fit to the unconstrained fit. The resulting estimate will have the same smoothness properties as the unconstrained estimate and, in intervals where the regression function is strictly increasing, if the sample is large enough, $\hat{m}(x)$ and $\hat{m}(x|p)$ coincide with probability one.

As far as I know, this method has not been applied to likelihood based situations. In theory I believe that this method should be adaptable to likelihood based settings. This has not been considered in this thesis. One of the difficulties of adapting this method to exponential family settings is that there is no closed form for the term $A_i(x)$. However, it should be possible to extract a vector $\boldsymbol{p}_u$ from an unconstrained regression estimate. I think this would mean that these weights could then be altered so as to obtain a monotone regression estimate.

DNP method

Dette *et al.* [9] have proposed a method that combines density estimation and regression estimation techniques. Their method is motivated by considering N independent and identically distributed uniform random variables $U_1, \ldots, U_N$. The uniform random variable U has the following cumulative distribution function $F(u)$,

$$F(u) = \begin{cases} 0 & \text{if } u < 0; \\ u & \text{if } 0 \leq u \leq 1; \\ 1 & \text{if } u > 1. \end{cases} \tag{1.5}$$

and the corresponding probability density function is given by

$$f(u) = \begin{cases} 0 & \text{if } u < 0; \\ 1 & \text{if } 0 \leq u \leq 1; \\ 0 & \text{if } u > 1. \end{cases} \tag{1.6}$$

If we then consider a transformation of the uniform random variables $m(U_i)$, it is clear that this transformed random variable has the following cumulative distribution function and probability density function

$$F(m(u)) = \begin{cases} 0 & \text{if } u < m(0); \\ m^{-1}(u) & \text{if } m(0) \leq u \leq m(1); \\ 1 & \text{if } u > m(1). \end{cases} \tag{1.7}$$

$$f(m(u)) = \begin{cases} 0 & \text{if } u < m(0); \\ (m^{-1})'(u) & \text{if } m(0) \leq u \leq m(1); \\ 0 & \text{if } u > m(1). \end{cases} \tag{1.8}$$

Recalling (1.2) it is possible to construct a non-parametric kernel density estimate of $m(U_1)$. This would take the form,

$$\frac{1}{Nh_d} \sum_{i=1}^{N} K_d \left(\frac{m(U_i) - u}{h_d} \right). \tag{1.9}$$

It is clear then that this kernel density estimate is an estimate of the density $(m^{-1})'(u) I_{[m(0),m(1)]}(u)$. If we then integrate this density estimate we would have a consistent estimate of the function $m^{-1}(t)$, namely

$$\frac{1}{Nh_d} \int_{-\infty}^{t} \sum_{i=1}^{N} K_d \left(\frac{m(U_i) - u}{h_d} \right) du. \tag{1.10}$$

Finally, to obtain an estimate of the function m, we simply invert the estimate of $m^{-1}(t)$. The resulting estimate, $\hat{m}(t)$ is a strictly increasing function.

To develop this idea to a more general setting Dette *et al.* [9] propose taking an unconstrained regression estimate $\hat{m}(i/n)$ and then using a classical kernel density estimate

$$\hat{m}_I^{-1} := \frac{1}{N} h_d \sum_{i=1}^{N} \int_{-\infty}^{t} K_d \left(\frac{\hat{m}(i/N) - u}{h_d} \right) du$$

as an estimate of the inverse of the true regression function, $m^{-1}(t)$. I use the subscript 'I' here to denote that the estimate is isotonic. Technically isotonic would imply monotone increasing. However, for the remainder of this thesis it is assumed that monotone implies monotone increasing. The results would still be applicable if the required function was monotone decreasing. Here, h_d is a bandwidth to be used in the density estimation step. A bandwidth

h_r would be used in the initial regression step. This estimate can then be inverted to obtain a monotone estimate of m, which we denote $\hat{m}_{DNP}$.

The important difference between this case to the case described motivated by uniform random variables is the presence of the term (i/N) in place of U_i as the stimulus level in the regression estimate. Instead of using uniform random variables we now assume that the design points are equally spread across the region $(0, 1)$. The idea is that this will behave similarly to the uniform random variable case. It is also important to notice that the term N does not necessarily correspond to the number of stimulus levels (or sample size) n. In fact, if there is a small sample, or few stimulus levels are tested, then it is desirable that N is not equal to n. This would result in a very rough estimate, only calculated at a few points. It is desirable that N is reasonably large, or at least, assuming computational speed is not much of a problem, the larger N the better.

The motivation for this method is to consider the function

$$m_N^{-1}(t) = \frac{1}{Nh_d} \int_{-\infty}^{t} \sum_{i=1}^{N} K_d \left(\frac{m(i/N) - u}{h_d} \right) du = \int_0^1 I\{m(x) \leq t\} dx + o(1).$$

Dette *et al.* [9] show that $\hat{m}_I^{-1}(t)$ is a consistent estimate of $m_n^{-1}(t)$ which in turn approaches $m^{-1}(t)$ as $n \to \infty$. They give the precise order of this approximation. Hence, by estimating $\hat{m}_I^{-1}$ and inverting the result one obtains an estimate of $m(t)$.

Dette *et al.* [9] show that for sufficiently large N and small h_d, $\hat{m}_{DNP}$ and $\hat{m}$ coincide in regions where the unconstrained estimate is already monotone. They also show that if the Nadaraya-Watson or local linear methods are used to calculate the unconstrained estimate, $\hat{m}$, then $\hat{m}_{DNP}$ is asymptotically normally distributed. A comparison of this method for a local linear $\hat{m}$, the PAVA method and tilting method can be found in Dette and Pilz [10].

This method has been applied to dose-response curve estimation in Dette, Neumeyer and Pilz [8]. This approach uses the local linear method to calculate an estimate of the effective dose curve. I outline some problems with this approach in Chapter 2 of this thesis and suggest some modifications to overcome them.

In summary then, the DNP method proceeds as follows. First calculate a non-parametric regression estimate $\hat{m}(i/N)$ using a bandwidth for the regression h_r. Using this we then calcu-

late a classical kernel density estimate, $\hat{m}_I^{-1}(t)$ using a bandwidth for the density estimate h_d. Finally this estimate is inverted to obtain an estimate $\hat{m}_I(t)$. In this method, the subscript r refers to terms involved in the regression step of the procedure where as the subscript d refers to terms involved in the density estimation step of the procedure.

1.6.2 Testing for monotonicity

The question may arise as to how we justify applying a monotonicity constraint. How can we determine whether the model should be monotone or not? In many cases scientific knowledge determines this. However, various tests have been suggested to test for the presence of monotonicity in a data set. Details of three commonly used methods can be found in Bowman, Jones and Gijbels [2], Gijbels *et al.* [22] and Hall and Heckman [25].

It must be said that I have not come across many situations in which I have seen monotonicity tests used. In many cases the decision to enforce monotonicity is based on some scientific or clinical insight that suggests the response must be monotone. I have not come across many cases of a study that first tests for monotonicity and then enforces it. However, in some cases this may be a useful analysis and so in the proceeding sections I describe three possible tests to see if a response is monotone.

A "bandwidth" test

This method was suggested by Bowman, Jones and Gijbels [2]. Their method is very simple and leads to a conservative test of whether or not the regression function is monotone. They outline their method in four steps.

1. Find the critical bandwidth h_c which is the smallest bandwidth for which the regression function $\hat{m}(x; h_c)$ is monotone.

2. Calculate the error in the model at an initial bandwidth h_0. They give this to be the plug in bandwidth suggested by Ruppert, Sheather and Wand [40]. That is, calculate $\hat{\epsilon}_i = Y_i - \hat{m}(X_i; h_0)$ for $i = 1, \ldots, n$.

3. Generate a bootstrap sample, $\hat{\epsilon}_i^*$ from these errors and hence generate a bootstrap data set $Y_i^* = \hat{m}(X_i; h_c) + \hat{\epsilon}_i^*, i = 1, \ldots, n$.

4. Apply $\hat{m}$ using h_c to $\{(X_i, Y_i^*), i = 1, \ldots, n\}$ and observe whether or not the result is monotone.

5. Repeat steps 3 and 4 a large number of times and determine the proportion of estimates that are not monotonic.

The rationale behind this test is that we want to see how large (in some sense) the bandwidth must be for the data to appear monotone. If the data are truly monotone then we would expect this to be clear for a very small h, and for all h above this. However, if the data were not monotone then we would expect the data to appear monotone only for larger bandwidths when too much oversmoothing had occurred and any dips in the regression function had been masked. Hence, in the method proposed above, if we are correct in our assumption that the regression function is monotone then we would expect a large proportion of the bootstrap samples to appear non-monotone. This is because we have a relatively small critical bandwidth. As a result we would obtain a high p-value and hence accept the null hypothesis that the data were monotone.

A "gradient" test

This method has been proposed by Hall and Heckman [25]. They have suggested this method in an attempt to overcome a difficulty in the method suggested by Bowman *et al.* [2]. The problem is that the bandwidth method can fail to detect "flat points" or small dips within the regression function. If there is a dip in the data that is smaller in width than the bandwidth being used then when using the bootstrap samples this may be undetected and we may falsely believe that the regression function is monotone. The general idea of this method is to use "local"gradients to test for the presence of flat points or dips in the regression curve. They suggest using a test statistic as follows. Let $0 \leq r \leq s - 2 \leq n - 2$ be integers, let a, b be

constants and put

$$S(a, b|r, s) = \sum_{i=r+1}^{s} \{Y_i - (a + bx_i)\}^2.$$

For each choice of (r, s), define $\hat{a} = \hat{a}(r, s)$ and $\hat{b} = \hat{b}(r, s)$ by $(\hat{a}, \hat{b}) = \arg\min_{(a,b)} S(a, b|r, s)$, and let

$$Q(r, s)^2 = \sum_{i=r+1}^{s} \left\{ x_i - (s - r)^{-1} \sum_{j=r+1}^{s} x_j \right\}^2,$$

and use the test statistic

$$T_m = \max\{-\hat{b}(r, s)Q(r, s) : 0 \leq r \leq s - m \leq n - m\},$$

where m, satisfying $2 \leq m \leq n$, is an integer.

Effectively we are calculating a set of local gradients and standardising them using the variance within the locality. We select as a test statistic the most negative of these local gradients. This value is then compared with the value of a test of a flat function. This is stated by Hall and Heckman [25] to be the most difficult non-decreasing form of m for which to test. In practice we would calculate an unconstrained regression estimate from the data and then calculate the error between this estimate and the given data points. We then obtain a bootstrap sample of these errors, $\varepsilon_1^*, \ldots, \varepsilon_n^*$ and use

$$S(a, b|r, s) = \sum_{i=r+1}^{s} \{\varepsilon_i^* - (a + bx_i)\}^2$$

in the previously described method. The proportion of samples in which the T_m^* value is greater than the original calculated value, T_m is then calculated. If this proportion is greater than some predetermined value α then the hypothesis that the data is monotone is rejected.

A "count" method

Gijbels *et al.* [22] state that both the gradient method and the bandwidth method cannot guarantee the actual significance level of the test. To overcome this they have proposed a test for monotonicity with a guaranteed level. Their test is based on the signs of the differences $Y_{i+k} - Y_i$. They propose two test statistics. The first is simply a count of the number of + signs in a given sequence of differences. The second looks at strings of consecutive − signs. The idea

is that if the true regression really is increasing then we would expect mainly + signs whereas if there were dips in the curve then in some localities there would be a high proportion of − signs.

1. **Count Statistic**. Define $S_{i,k} \equiv sgn(Y_{i+k} - Y_i)$ for $1 \leq i \leq n-k$. Let $\ell_r = \ell_r(j,k)$, denote the sequence $S_{j+1,k}, \ldots, S_{j+r,k}$. Let $N(\ell_r)$ equal the number of + signs it contains and define

$$T_{\text{count}} = T_{\text{count}}(k, r_1, r_2) = \max_{r_1 \leq r \leq r_2} \max_{0 \leq j \leq n-r-k} r^{-1/2} \left[\frac{1}{2} r - N\{\ell_r(j,k)\} \right].$$

   Under the hypothesis that the data are monotone the signs have no particular tendency toward negativity. However if there are dips in the curve then we would expect more of the signs to be negative. This value is larger under the alternative hypothesis that there are dips in the curve.

2. **Run Statistic**. Define $R(j,k)$ to be the supremum of integers r such that $S_{j+1,k}, \ldots, S_{j+r,k}$ is a sequence solely of -1's and we define a test statistic,

$$T_{\text{run}} = T_{\text{run}}(k) = \max_{0 \leq j \leq n-k} R(j,k).$$

   Again we would expect a larger run of negative signs, and hence a larger T_{run} if the data truly did contain a dip than if the data were monotone.

Irrespective of whether we use T_{run} or T_{count}, the algorithm suggested by Gijbels *et al.* [22] is as follows. Generate a large number of samples, $Y^* = (Y_1^*, \ldots, Y_n^*)$ and calculate T^* for each sample. We could, for example generate these samples using bootstrap resampling of the signs $S_{i,k}$. Count the proportion of samples where T^* is greater than or equal to T. If this proportion is less than or equal to the significance level α then reject the hypothesis that the regression function is non decreasing.

1.7 Outline of thesis

In this thesis, I discuss the estimation of monotone psychometric functions and transducer functions. In Chapter 2 I explore the DNP method when applied to estimation of transducer

functions. I show that it has various problems. For example it does not constrain the estimate to lie within the correct range. I propose a change to this method which involves using the likelihood based local linear regression and estimating a monotone version of the η function which can then be transformed by the inverse link function to obtain a monotone estimate of a transducer function. I call this method the LDNP method since it is a likelihood based adaptation of the DNP method. I develop the asymptotic theory of the LDNP method. In particular, I show that an estimate produced using these methods is asymptotically normally distributed with a bias that is of order $O(h_r^2)$ where h_r is the bandwidth used in the unconstrained regression step (See Chapter 2). In this sense I show that my estimate is asymptotically first order equivalent to other competing methods. The LDNP estimate has a variance that is of order $O(1/nh_r)$ which is again equivalent to other competing methods.

In Chapter 3 I explore the choice of the bandwidth h_d to be used in the density estimation step of the LDNP procedure. In particular, I explore the relationship between the choice of bandwidth h_d and the choice of bandwidth h_r and show that in most cases it is best to choose a bandwidth $h_d = o(h_r)$. I investigate this through a study of asymptotic properties and also through a simulation study.

Chapter 4 and Chapter 5 compare the performance of various competing monotonicity constraints described in Section 1.6.1. Chapter 4 considers the estimation of monotone psychometric functions whilst Chapter 5 considers estimating monotone transducer functions, where the responses are either Poisson or Exponential response functions. I compare a variety of different factors that may affect the estimation such as the number of stimulus levels used, the number of repeats used and the choice of bandwidth h_r. I compare the methods by using metrics such as the bias, variance, MSE and MISE.

Finally, in Chapter 6 I give a summary of my research and suggest possible future work.

Chapter 2

The LDNP method

2.1 Introduction

In the introduction I have outlined some of the commonly used methods of monotonising nonparametric estimators of a regression function. Dette and Scheder [11] compared the performance of some of these methods when applied to the estimation of the effective dose (threshold/lethal concentration). They conclude that their method is probably the most effective of the methods considered. They come to this conclusion as a result of many simulations of response functions, after which they calculate the square bias, variance, MSE and MISE for the estimates obtained using the various methods.

Despite the fact that this paper is only concerned with estimating the effective dose, it is still reasonable to assume that this method would be useful in estimating monotone psychometric functions. Nevertheless, we have tried using this method on the data sets found in Żychaluk and Foster [18], and have found that in some cases, the DNP method is not good for estimating psychometric functions which are effectively the inverse of the effective dose, as they result in estimates taking values outside the interval $[0, 1]$. This is explained in Section 2.2. In Section 2.3, I propose a modification to the method proposed by Dette, Neumeyer and Pilz [8] which overcomes the problems we encountered, by applying a link function. The new approach also takes into account the changing variance of the responses. We demonstrate, with the use

of the examples in Żychaluk and Foster [18], that this new method performs well. We also offer theoretical justifications for the new method.

2.2 Initial investigations from psychometric studies

Żychaluk and Foster [18] propose using non-parametric methods to estimate psychometric functions. This method has been described in the introduction, along with some justification for its use. They also introduce six data sets which they use in comparing the non-parametric and parametric methods of estimating psychometric functions. These data sets come from studies of vision and hearing. More details about the sets can be found in Żychaluk and Foster [18], and also in the papers concerning the original studies.

Levi and Tripathy [31] study path deviation from a moving visual stimulus. A patient was presented with the image of a dot that moved on a linear path to the right until it reached the middle of the display. It then moved either up or down and the subject had to indicate which way the dot had moved. The proportion of correct responses from 30 trials is shown in Figure 2.1(a).

Figure 2.1(b) shows data from an unpublished study by S. Carcagno of the University of Lancaster in which the discrimination of pitch was studied. A patient was presented with three intervals during which a tone was played. In one of the three intervals the tone was of a different frequency to the other two and the patient was asked to identify this interval.

Xie and Griffin [50] studied the visual perception of fragmented images. A display was split into two parts, one containing a pair of patches from the same image and the other with patches from two different images. The patient had to identify which pair of patches came from the same image. This data is shown in Figure 2.1(c).

Schofield *et al.* [41] have studied visual motion aftereffect.The subject was presented with a moving adaptation stimulus, followed by a test stimulus. The patient was required to indicate whether the adapting stimulus moved in the same direction as the test stimulus. The proportion of correct responses is shown in Figure 2.1(d).

Nasciemento *et al.* [36] studied visual discrimination. A patient was shown an image of a natural scene and then an approximation of this scene based on a principal component analysis. The patient was required to distinguish between the two responses and the proportion of correct responses is shown in Figure 2.1(e).

Finally, Baker *et al.* [1] have studied the detection of a gap in noise. A 300-msec noise burst containing a gap of 2-8 msec duration or no gap was presented to the subject. The subject was required to say whether they noticed a gap or not. The proportion of correct responses is shown in Figure 2.1(f).

When I estimated a monotone psychometric function for each data set, using the DNP method, along with an direct plug in bandwidth of Ruppert, Sheather and Wand [40] I found that in some cases the estimates contained values that were greater than one.

In addition, I examined 485 data sets from the studies of Baker *et al.* [1], Levi and Tripathy [31], Nascimento *et al.* [36] and Schofield *et al.* [41] and found this to be the case in many sets, whilst some values were also less than zero. As a simple demonstration of this, I looked at 485 psychometric data sets from the studies cited above. For each of these data sets I calculated a cross validation bandwidth and then an estimate using the DNP method. Of the 485 data sets, 370 had points which were below 0 or above 1. That is approximately 76% of the data sets. This highlights the fact that the DNP method regularly gives estimates of psychometric functions taking impossible values. This obviously makes no sense when considering a psychometric function (which is a probability) and hence for these situations the DNP method provides an estimate that does not really accurately reflect scientific knowledge about the model. This problem has been highlighted previously by Chu and Cheng [5]. They note that the local linear regression method is not really appropriate in binary response models since it often leads to estimates that are less than 0 or greater than 1.

Figure 2.1 shows these six data sets with a monotone estimate plotted using the DNP method. The data sets and fitted psychometric functions, (a) - (f) are presented in the order in which they occur in [18]. For functions (b), (c) and (e) this method produces a very reasonable estimate of the psychometric function. However, function (d) in particular, and also function

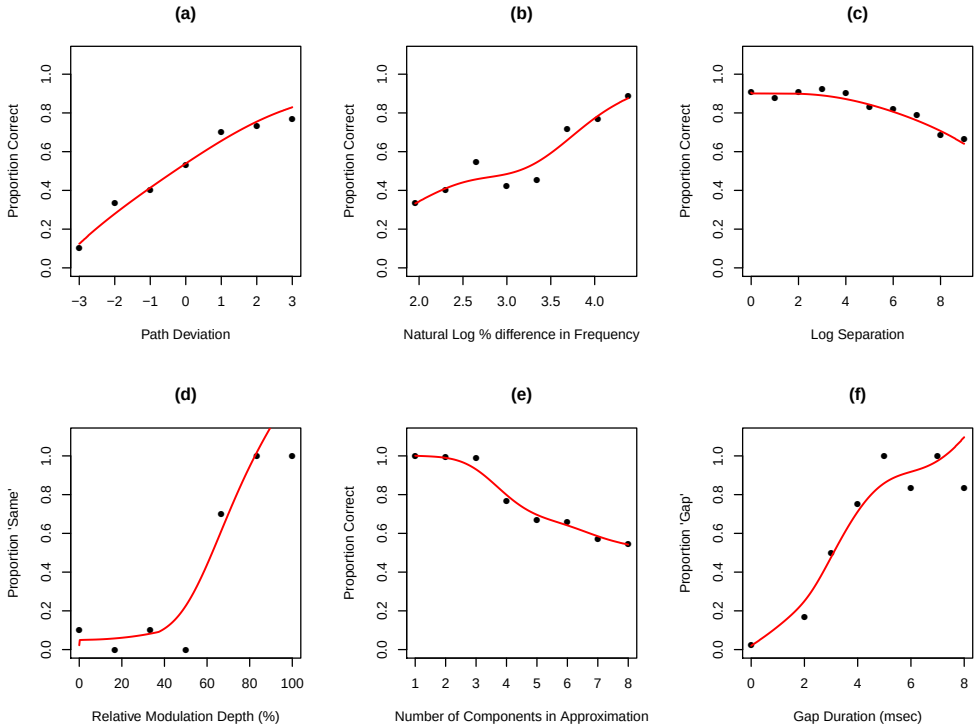

Figure 2.1: Psychometric function estimates obtained using DNP method. Data taken from Żychaluk and Foster [18]

(f) show that the DNP method fails to recognise the asymptotes at 0 or 1. The reason that this happens is because the local linear method does not make use of a link function. Using an appropriate link function ensures that the resulting estimates lie within the correct bounds. In the case demonstrated here using a logit link function constrains the responses to be in (0,1). The local linear estimating procedure uses no link function whereas maximum likelihood procedures can use this extra information. This obviously contravenes the fact that $P(x)$ is supposed to be a probability. Function (a) provides a reasonable fit but it does look as if the estimate is not recognising what seems to be a lapsing rate of about $P = 0.8$.

It must be noted that in many cases the DNP method will produce a good estimate of the psychometric function that will be suitable for practical purposes. But the examples presented above show that in some cases the DNP method produces an estimate of the psychometric function that is not correct and may produce misleading conclusions. It should also be noted that

not all of the above examples required a monotonicity constraint. Some of the unconstrained estimates were already monotone. This emphasises that the problem is with the unconstrained estimation procedure used in the initial regression step of the DNP method, rather than the monotonicity constraint in itself. I have therefore proposed a modification to the method to enable monotone psychometric functions to be estimated easily and in a way that does not violate any range constraints. This method is outlined in the next section.

2.3 A proposed modification of the DNP method

Dette, Neumeyer and Pilz [8] propose using a local polynomial regression to estimate the effective dose in quantal bioassay. They use the local polynomial regression estimate as the unconstrained regression step of the DNP method. I propose that the psychometric function (or indeed any function for which the response is a member of the exponential family) would be better modelled using local log-likelihood as explained in Section 1.5 (based on the work of Fan, Heckman and Wand [17] and Foster and Żychaluk [18]) and then applying the monotonising step to the η function. This method ensures that the psychometric function lies within the correct bounds of $[0, 1]$ through the use of the canonical link, the logit function. It also takes into account the fact that the variance of the response varies with its mean, a fact that is not considered in the local polynomial method.

The concept of local likelihood smoothing was first introduced by Tibshirani and Hastie [43]. They proposed smoothing the local likelihood by using a running lines smoother. This meant that essentially they only use a set number of points to calculate estimates at each value of x. Further, they use no method of weighting observations depending on their distance from x. In contrast Fan Heckman and Wand [17] and consequently Foster and Żychaluk [18] weight observations using a kernel function. As I have already discussed, a kernel function gives most weight to observations closest to the value x. In addition then allow for situations where the variance is only expressed in terms of the mean and a quasi-likelihood would need to be used.

Denote by $\hat{p}_{LL}$ the local linear unconstrained estimate used in the first step of the DNP

method. At this point in the thesis I change notation. Instead of calling a regression function $m(x)$, I now call it $p(x)$ This is to emphasise that I am estimating a psychometric function. The subscript LL serves to remind the reader that this estimate is simply a local linear regression estimate. At this point no account has been taken of the likelihood information. Recall that the DNP method estimates the inverse of the response curve p^{-1},

$$\hat{p}^{-1}(\alpha) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{\alpha} K_d \left(\frac{\hat{p}_{LL}(i/N) - u}{h_d} \right) du.$$

In Foster and Żychaluk [18]), an estimate of $\eta(x)$ is obtained by maximizing the local log-likelihood. This method is described in Section 1.5 and involves maximising the local log-likelihood,

$$\begin{aligned}
l(\alpha_0, \alpha_1; x) &= \sum_{i=1}^{n} \ln[P^{r_i}(x_i)(1 - P(x_i))^{m_i - r_i}] K_h(x - x_i) \\
&= \sum_{i=1}^{n} \left[r_i \ln P(x_i) + (m_i - r_i) \ln[1 - P(x_i)] \right] K_h(x - x_i) \\
&= \sum_{i=1}^{n} \left[r_i \ln[g^{-1}(\alpha_0 + \alpha_1 x_i)] + (m_i - r_i) \ln[1 - g^{-1}(\alpha_0 + \alpha_1 x_i)] \right] K_h(x - x_i) \\
&= \sum_{i=1}^{n} \Big[r_i \ln\{(1 + \exp(-(\alpha_0 + \alpha_1 x_i)))^{-1}\} \\
&\quad + (m_i - r_i) \ln\{1 - (1 + \exp(-(\alpha_0 + \alpha_1 x_i)))^{-1}\} \Big] K_h(x - x_i) \\
&= -\sum_{i=1}^{n} [r_i \ln\{1 + \exp(-(\alpha_0 + \alpha_1 x_i))\} \\
&\quad + (m_i - r_i) \ln\{1 + \exp(\alpha_0 + \alpha_1 x_i)\}] K_h(x - x_i).
\end{aligned}$$

This log-likelihood expresses the fact that in a local area, the size of which depends on a bandwidth h, the function η is locally linear, and an estimate of the value at x is given by the value $\hat{\alpha}_0$. This value will change at each different x value. This procedure is also explained in Section 1.5. I approximate values around x by a linear function and then interpolate from this linear function the value of $\eta(x)$, which is given by α_0. This is repeated for each value of x, where a new local neighbourhood is created each time.

This is then maximised using the Fisher scoring method (as with the parametric case and explained in [33] and [16]). This is a numerical approximation method since an exact solution

is often not possible. This will give us the values of α_0 and α_1 that maximise the local log-likelihood. These values are then used to obtain the maximum likelihood estimate for $\eta(x)$ as $\hat{\eta}(x) = \hat{\alpha}_0$. This estimate, $\hat{\eta}(x)$ is then transformed using the inverse link function to obtain an estimate of the psychometric function,

$$\hat{P}(x) = g^{-1}(\hat{\eta}(x)).$$

We propose that, prior to this final transformation we apply the monotonicity constraint of Dette *et al.* [9] to the systematic component. That is to say we obtain a monotone estimate of the inverse of the systematic component

$$\hat{\eta}_M^{-1}(x) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{x} K_d \left(\frac{\hat{\eta}_L(i/N) - u}{h_d} \right) du, \tag{2.1}$$

and then invert this estimate. In the above equation the subscript M simply indicates that this estimate is monotone. The subscript L in $\hat{\eta}_L$ is used to denote that this estimate comes from a local likelihood based estimate as opposed to the local linear estimate that had was used by Dette *et al.* [8] Two bandwidths are used, h_r, which is used in the unconstrained regression step, and h_d, which is used in the density estimation section of the monotonicity constraint. Finally the inverse link transformation is applied and a monotone estimate of the psychometric function is obtained.

$$\hat{P}_M(x) = g^{-1}(\hat{\eta}_M(x)).$$

This method of estimating a psychometric function has the advantage of constraining the estimates to lie within the range $[0, 1]$ and hence avoid some of the irregularities discussed previously in Section 2.2. This method also will guarantee monotonicity of the estimate so long as the link function is monotonicity preserving. We shall refer to this method as the LDNP method since we are incorporating likelihood information into the DNP method.

In this adaptation I have monotonised the η function before the final transformation to obtain $\hat{P}(x)$. This is so as to ensure that the estimate $\hat{P}(x)$ is definitely monotone. This is a result of the fact that the link function is monotonicity preserving. When I investigated the possibility of simply apply the monotonicity step after the inverse link transformation I

found that in some cases the numerical methods of the DNP method did not always keep the final estimates less than 1. The breach of this condition was never severe and to all practical purposes the final estimates could be considered to be 1 in cases where they exceed 1. However, for the sake of accuracy it seemed better to apply the monotonicity step to the η function and then apply the inverse link, thus avoiding any problems.

Initial investigations into the suitability of this method involved estimating psychometric functions using the data in [18]. We repeat the plots shown in Figure 2.1 and this time also add an estimate of the psychometric function obtained by using the LDNP method. This is shown in Figure 2.2. A visual inspection of the plots shows that in the cases where the DNP fit already provided a reasonable fit (functions (b), (c) and (e)) the LDNP fit is at least as good and certainly no worse a fit. In the case of function (a) the LDNP method seems to capture a lapsing rate more effectively. In functions (d) and (f), for which the DNP method had some problems, by returning estimates outside the correct bounds, the LDNP method has corrected these problems and ensured that the estimates lie within the correct range.

In the remainder of this chapter we offer some theoretical justifications of the use of the LDNP method, by examining its asymptotic behaviour.

2.4 Aysmptotic Behaviour

For the remainder of this chapter I will be considering transducer functions in general. The motivation for this chapter comes from the study of psychometric functions. I have already discussed in this chapter the problems associated with the DNP method when using it to estimate psychometric functions. The key problem is that the resulting estimates are not necessarily within the required bounds for a psychometric function. I have already explained that a psychometric function is a type of transducer function, specifically, a transducer function for which the responses are binomially distributed. In the rest of this chapter, I generalise the adaptation to the DNP method, which I have called the LDNP method to allow its use in estimating transducer functions with any transducer function where the responses come from

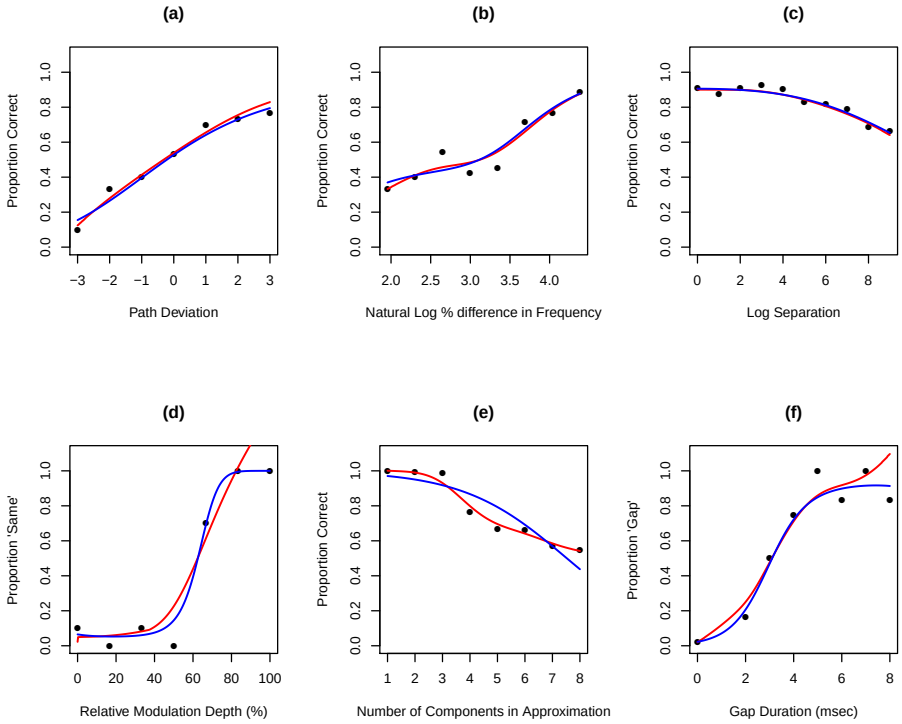

Figure 2.2: Psychometric function estimates obtained using DNP method (red lines) and LDNP method (Blue lines). Data taken from Żychaluk and Foster [18]

the exponential family.

Dette *et al.* [9] demonstrate that the DNP estimate is asymptotically normal if kernel type or local polynomial estimators are used for the preliminary estimation of the regression function. They also state that their estimator is asymptotically first-order equivalent to the PAVA estimates and the tilting estimate of Hall and Huang [26]. In this section I show that the LDNP method also performs well asymptotically. I show that the LDNP estimate is asymptotically normal and, like the DNP method, first order asymptotically equivalent to competing methods of monotonicity constraint.

In this section we consider a pair of n i.i.d observations $\{(X_i, Y_i)\}_{i=1}^n$ and a mean function $\mu(x)$ and variance $V\{\mu(x)\}$, for a given function $V(\cdot)$. These are the mean and variance of Y given $X = x$. We are interested in estimating $\eta(x) = g(\mu(x))$ for some link function g. In this section I move from using $P(x)$ to denote a psychometric function to using $\mu(x)$ to denote any

response from the exponential family model. Even though the setting of psychometric functions is the motivation for this thesis the asymptotic results I present hold true for any exponential family responses and so the LDNP method is more widely applicable. So then, $\mu(x)$ in these asymptotic results may represent a binomial response such as the psychometric function, $P(x)$, or a Poisson response such as the Transducer function $T(x)$ considered in Chapter 5.

In the following asymptotic results I assume that the number of design points N is equal to the sample size n. This need not be the case but is assumed from here on simply to simplify the notation and make the following results more easy to follow. Define $q_l(x, y) = (\partial^l/\partial x^l)Q\{g^{-1}(x), y\}$ following Fan and Gijbels [16] (page 211), Here, Q is the quasi-likelihood function that satisfies

$$\frac{\partial}{\partial \mu} = \frac{y - \mu}{V(\mu)} \tag{2.2}$$

Notice that q_l is linear in y for fixed x. It can also be seen that

$$q_1\{\eta(x), \mu(x)\} = 0 \tag{2.3}$$

$$q_2\{\eta(x), \mu(x)\} = -\rho(x) \tag{2.4}$$

where

$$\rho(x) = ([g'\{P(x)\}]^2 V\{P(x)\})^{-1}. \tag{2.5}$$

In the remainder of this chapter I need to make various assumptions. These assumptions fall largely into three categories.

The first set of assumptions, (A1)-(A6), are requirements for the unconstrained regression step. These are largely taken from Fan and Gijbels [16] (page 211) and ensure that the asymptotic distribution of $\hat{\eta}(x)$, is normal (see Section 2.5). They are fairly standard regularity conditions and are not especially restrictive.

A1 The functions $\eta''(\cdot)$, $(\eta^{-1})'''(\cdot)$, $Var(Y|X = x)$, $V'(\cdot)$ and $g'''(\cdot)$ are continuous.

A2 The function $q_2(x, y) < 0$ for $x \in \mathbb{R}$ and y in the range of the response variable.

A3 $\rho(x) \neq 0$, $Var(Y|X = x) \neq 0$, $g'\{\mu(x) \neq 0\}$.

A4 $E(Y^4|X = \cdot)$ is bounded in a neighborhood of x.

A5 The kernel, K is a symmetric probability density function with compact support.

A6 Y_i is a member of the exponential family and has density f.

The second set of assumptions, (B1)-(B6) are conditions required for the monotonicity constraint and are similar to those in Dette, Neumeyer and Pilz [9]. They control the rate of convergence of the bandwidths h_d and h_r relative to the sample size n as well as imposing regularity conditions on the density of X_i and the regression function $\mu(x)$.

B1 X_i has a positive, twice continuously differentiable density f_X with compact support, say $[0, 1]$.

B2 $f_X(x) \geq 0$, and has three continuous derivatives.

B3 The mean function $\mu(x)$ is assumed to be twice continuously differentiable.

B4 $h_r \to 0$, $h_d \to 0$, $nh_r \to \infty$, $nh_d \to \infty$.

B5 $nh_r^5 = O(1)$.

B6 $\int_0^1 |K_d''(t)|dt$ is finite.

The final set of assumptions, (C1)-(C6) are additional requirements due to the fact that I am incorporating likelihood information into the monotonicity constraint procedure. These assumptions are required in the proof of Theorem 2.

C1 $\frac{\log n}{h_r\sqrt{nh_d}} = o(1)$

C2 $\frac{\log n}{nh_r h_d^2} = o(1)$

C3 $\frac{h_r^2}{h_d} = o(1)$

C4 $h_r^4 \log(n) = o(1)$

C5 $\frac{h_d^3 \log(n)}{h_r} = o(1)$

C6 $h_d^2 \log(n) = o(1)$

Firstly, I demonstrate that the monotone estimate, or $\hat{\eta}_M^{-1}(x)$ provides a good approximation for the function

$$\eta_n^{-1}(t) = \frac{1}{nh_d} \int_{-\infty}^{t} \sum_{i=1}^{n} K_d \left(\frac{\eta(i/n) - u}{h_d} \right) du, \tag{2.6}$$

which in turn approaches the true value $\eta^{-1}(x)$. Theorem 1 shows that the estimate, or $\hat{\eta}_M^{-1}(x) - \eta_n^{-1}(x)$, is asymptotically normally distibuted, with a known bias that is of order $O(h_r^2)$ and a known variance that is of order $O((nh_r)^{-1})$. Theorem 1 is proved at the end of this chapter in section 2.5.

Theorem 1. *If the assumptions (A1)-(A6) and (B1)-(B5) are satisfied,* $\lim_{n\to\infty} h_r/h_d = c \in [0, \infty)$ *exists and* η *is strictly increasing, then it follows that for all* $t \in (\eta(0), \eta(1))$ *with* $\eta'(\eta^{-1}(t)) > 0$,

$$\sqrt{nh_d} \left(\hat{\eta}_M^{-1}(t) - \eta_n^{-1}(t) + \kappa_2(K_r) h_r^2 \frac{\eta''(\eta^{-1}(t))}{\eta'(\eta^{-1}(t))} \right) \xrightarrow{D} N(0, r^2(t)),$$

where $\kappa_2(K) = \frac{1}{2} \int_{-1}^{1} \nu^2 K(\nu) d\nu$, *and the asymptotic variance is given by*

$$\begin{aligned} r^2(t) = & \frac{[g'(g^{-1}(t)]^2 Var[Y|X = \eta^{-1}(t)]}{\eta'(\eta^{-1}(t)) f(\eta^{-1}(t))} \\ & \times \int \int \int K_d \left(u + \frac{h_r}{h_d} \eta'(\eta^{-1}(t))(\nu - w) \right) K_d(u) K_r(w) K_r(\nu) du dw d\nu \end{aligned} \tag{2.7}$$

If $\lim_{n\to\infty} h_r/h_d = \infty$, *then we have, for all* $t \in (\eta(0), \eta(1))$ *with* $\eta'(\eta^{-1}(t)) > 0$, *and the assumptions (A1)-(A4) and (B1)-(B5) are satisfied*

$$\sqrt{nh_r} \left(\hat{\eta}_M^{-1}(t) - \eta_n^{-1} + \kappa_2(K_r) h_r^2 \frac{\eta''(\eta^{-1}(t))}{\eta'(\eta^{-1}(t))} \right) \xrightarrow{D} N(0, \tilde{r}^2(t)),$$

where the asymptotic variance is given by

$$\begin{aligned} \tilde{r}^2(t) = & \frac{[g'(g^{-1}(t)]^2 Var[Y|X = \eta^{-1}(t)]}{(\eta'(\eta^{-1}(t)))^2 f(\eta^{-1}(t))} \\ & \times \int K_r(u) du \end{aligned} \tag{2.8}$$

In order to establish the relationship the relationship between $\eta_M(t) - \eta_n(t)$ three preliminary lemmas are required. Lemma 1 considers the precise order of the difference between $\eta_n^{-1}(t)$ and

the true value $\eta^{-1}(t)$. It is clear that if $\hat{\eta}_M^{-1}(t)$ is an approximation of $\eta^{-1}(t)$ then the inverse $\hat{\eta}_M(t)$ is an approximation of $\eta(t)$. Lemma 2 gives the precise relationship and is adapted from Dette, Neumeyer and Pilz [9]. The proofs of these two lemmas are provided at the end of this chapter in section 2.5, for the sake of completeness although they are largely adapted from the proofs in Dette, Neumeyer and Pilz [9].

Lemma 1. *If the regression function, $\eta(t)$ is strictly increasing and the assumptions(A1)-(A6) and (B1)-(B4) are satisfied, then we have for any $t\epsilon(\eta(0),\eta(1))$ with $\eta'(\eta^{-1}(t))>0$,*

$$\eta_n^{-1}(t)=\eta^{-1}(t)+\kappa_2(K_d)h_d^2(\eta^{-1})''(t)+o(h_d^2)+O\left(\frac{1}{nh_d}\right)$$

where the constant, $\kappa_2(K)=\frac{1}{2}\int_{-1}^{1}\nu^2K(\nu)d\nu$.

In Lemma 1 I have stated a result about $\eta_n^{-1}(t)$. In the following lemma I consider the inverse of this function, $\eta_n(t)$. This is again analogous to Lemma 2.2 of Dette *et al.* [9]

Lemma 2. *If the assumptions (A1)-(A6), (B1)-(B5) hold and η is a strictly increasing function, then for any $t\in(0,1)$ with $\eta'(t)>0$,*

$$\eta_n(t)=\eta(t)+\kappa_2(K_d)h_d^2\frac{\eta''(t)}{(\eta'(t))^2}+o(h_d^2)+O\left(\frac{1}{nh_d}\right). \tag{2.9}$$

I am finally in a position to state a theorem regarding the asymptotic behavior of $\hat{\eta}_M(t)$. I will show that it is asymptotically normal, in a similar way to that of Theorem 3.1 in Dette *et al.* [9]. I have shown that the bias of $\hat{\eta}_M(x)$ is of order $O(h_r^2)$ and known variance of order $O((nh_r)^{-1})$. The precise statement of the theorem is as follows.

Theorem 2. *Assume (A1)-(A6), (B1)-(B6) and (C1)-(C3) and (C5) stated at the beginning of section 2.4 hold. Define $\hat{\eta}_M$ as the inverse of $\hat{\eta}_M^{-1}$ and η_n as the inverse of η_n^{-1}. If $\lim_{n\to\infty}h_r/h_d=c\in[0,\infty)$ exists, then for every $t\in(0,1)$ with $\eta(t)>0$,*

$$\sqrt{nh_d}\left(\hat{\eta}_M(t)-\eta_n(t)-\kappa_2(K_r)h_r^2\eta''(t)\right)\xrightarrow{D}N(0,s^2(t)) \tag{2.10}$$

where the asymptotic variance is given by

$$\begin{aligned}s^2(t)=&\frac{\eta'(t)[g'(g^{-1}(\eta(t)))]^2Var[Y|X=t]}{f(t)}\\&\times\int\int\int K_d\left(u+\frac{h_r}{h_d}\eta'(t)(v-w)\right)K_d(u)K_r(w)K_r(v)dudwdv\end{aligned} \tag{2.11}$$

If $\lim_{n\to\infty} h_r/h_d = \infty$*, then for every* $t \in (0,1)$ *with* $\eta(t) > 0$*, and assumptions (A1)-(A4), (B1)-(B5) and (C4)-(C6) hold,*

$$\sqrt{nh_d}\left(\hat{\eta}_M(t) - \eta_n(t) - \kappa_2(K_r)h_r^2\eta''(t)\right) \xrightarrow{D} N(0, \tilde{s}^2(t)) \tag{2.12}$$

where the asymptotic variance is given by

$$\tilde{s}^2(t) = \frac{[g'(g^{-1}(\eta(t)))]^2 Var[Y|X=t]}{f(t)} \int K_d^2(u)\,du \tag{2.13}$$

Up until this point I have been focusing on $\eta(x)$. Finally I will present a theorem that is a corollary of the two theorems presented so far. Fan and Gijbels [16] (Theorem 5.3, page 197) state a theorem regarding the behavior of $\hat{\mu(x)} - \mu(x)$. They state that the asymptotic behaviour is the same as for $\hat{\eta}(x) - \eta(x)$ except that the asymptotic bias is divided by $g'\{\mu(x)\}$ and the asymptotic variance is divided by $[g'\{\mu(x)\}]^2$. So far, I have shown in Theorem 2 that $\hat{\eta}(x) - \eta_n(x)$ is asymptotically normally distributed with a known mean and variance. In Lemma 2, I have also shown a relationship between $\eta_n(x)$ and $\eta(x)$. The expression in Lemma 2 is deterministic and so incorporating this into Theorem 2, will affect the mean, but not the variance. Therefore, as a result of Theorem 5.3 of Fan and Gijbels [16], Theorem 2 and Lemma 2, I have a corollary regarding the behavior of $\hat{\mu}(x) - \mu(x)$.

Corollary 1. *Assume that the conditions for Theorems 2 and 2 hold. Define* $\hat{\eta}_M$ *as the inverse of* $\hat{\eta}_M^{-1}$ *and* η_n *as the inverse of* η_n^{-1}*. If* $\lim_{n\to\infty} h_r/h_d = c \in [0,\infty)$ *exists, then for every* $t \in (0,1)$ *with* $\eta(t) > 0$*,*

$$\sqrt{nh_d}\left(\hat{\eta}_M(t) - \eta(t) + \eta''(t)\left(\frac{\kappa_2(K_d)h_d^2}{(\eta'(t))^2} - \kappa_2(K_r)h_r^2\right)\right) \xrightarrow{D} N(0, s^2(t)) \tag{2.14}$$

where the asymptotic variance is given by

$$\begin{aligned} s_1^2(t) = {} & \frac{\eta'(t)[g'(g^{-1}(\eta(t)))]^2 Var[Y|X=t]}{f(t)} \\ & \times \int\int\int K_d\left(u + \frac{h_r}{h_d}\eta'(t)(v-w)\right) K_d(u)K_r(w)K_r(v)dudwdv \end{aligned} \tag{2.15}$$

If $\lim_{n\to\infty} h_r/h_d = \infty$*, then for every* $t \in (0,1)$ *with* $\eta(t) > 0$*, and assumptions (A1)-(A4), (B1)-(B5) and (C4)-(C5) hold,*

$$\sqrt{nh_r}\left(\hat{\eta}_M(t) - \eta(t) + \eta''(t)\left(\frac{\kappa_2(K_d)h_d^2}{(\eta'(t))^2} - \kappa_2(K_r)h_r^2\right)\right) \xrightarrow{D} N(0, \tilde{s}^2(t)) \tag{2.16}$$

where the asymptotic variance is given by

$$\tilde{s}_1^2(t) = \frac{[g'(g^{-1}(\eta(t)))]^2 Var[Y|X=t]}{f(t)} \int K_d^2(u)\, du. \tag{2.17}$$

Corollary 2. *Assume that the conditions for Theorems 1 and 2 hold. Noting that* $\hat{\mu}(\cdot) = g^{-1}(\hat{\eta}(\cdot))$*, then we have that if* $\lim_{n\to\infty} h_r/h_d = c \in [0,\infty)$ *exists, then for every* $t \in (0,1)$ *with* $\eta(t) > 0$*,*

$$\sqrt{nh_d}\left(\hat{\mu}_M(t) - \mu(t) - \frac{\eta''(t)}{g'\{\mu(x)\}}\left[\kappa_2(K_r)h_r^2 + \frac{h_d^2\kappa_2(K_d)}{(\eta'(t))^2}\right]\right) \xrightarrow{D} N(0, c^2(t)) \tag{2.18}$$

where

$$c^2(t) = \frac{\eta'(t)Var[Y|X=t]}{f(t)} \times \int\int\int K_d\left(u + \frac{h_r}{h_d}\eta'(t)(v-w)\right) K_d(u)K_r(w)K_r(v)dudwdv \tag{2.19}$$

If $\lim_{n\to\infty} h_r/h_d = \infty$*, then for every* $t \in (0,1)$ *with* $\eta(t) > 0$*,*

$$\sqrt{nh_r}\left(\hat{\mu}_M(t) - \mu(t) - \frac{\eta''(t)}{g'\{\mu(x)\}}\left[\kappa_2(K_r)h_r^2 + \frac{h_d^2\kappa_2(K_d)}{(\eta'(t))^2}\right]\right) \xrightarrow{D} N(0, \tilde{c}^2(t)) \tag{2.20}$$

where the asymptotic variance is given by

$$\tilde{c}^2(t) = \frac{Var[Y|X=t]}{f(t)} \int K_d^2(u)\, du. \tag{2.21}$$

2.5 Proof of Theorems and Lemmas

2.5.1 Proof of Theorem 1

The proof of Theorem 1 requires a theorem given by Fan and Gijbels [16]. I state it here so as not to interrupt the flow of the proof later on. The reader is referred to the above mentioned book for details of the proof of the theorem. Fan and Gijbels [16] (Theorem 5.2, pages 196,197) state that

Theorem 3. *If x is an interior point of the design density, we have*

$$\{\hat{\eta}(x) - \eta(x)\} \xrightarrow{D} N(b(x)h_r^2, \sigma^2(x)\frac{1}{nh_r}) \tag{2.22}$$

provided that $h_r \to 0$ and $nh_r \to \infty$, where

$$\varphi(x) = \kappa_2(K_r)\eta''(x)$$

and

$$\sigma^2(x) = \int K^2(\nu)d\nu \frac{[g'\{\mu(x)\}]^2 Var[Y|X=x]}{f(x)}.$$

Proof. The majority of this proof follows closely the proof in Dette *et al.* [9]. Adaptations have been made where necessary to allow for the use of likelihood. I have also used results from Fan and Gijbels [16].

First, recall the definition,

$$\eta_n^{-1}(t) = \frac{1}{nh_d}\int_{-\infty}^{t}\sum_{i=1}^{n} K_d\left(\frac{\eta(i/n) - u}{h_d}\right)du.$$

and notice that

$$\hat{\eta}_M^{-1} = \frac{1}{nh_d}\int_{-\infty}^{t}\sum_{i=1}^{n} K_d\left(\frac{\hat{\eta}(i/n) - u}{h_d}\right)du = \eta_n^{-1}(t) + \Delta_n(t),$$

where

$$\begin{aligned}\Delta_n(t) &= \hat{\eta}_M^{-1}(t) - \eta_n^{-1}(t) \\ &= \frac{1}{nh_d}\sum_{i=1}^{n}\int_{-\infty}^{t}\left\{K_d\left(\frac{\hat{\eta}(i/n) - u}{h_d}\right) - K_d\left(\frac{\eta(i/n) - u}{h_d}\right)\right\}du \end{aligned} \tag{2.23}$$

Then use a Taylor expansion,

$$K_d\left(\frac{\hat{\eta}(i/n)-u}{h_d}\right) = K_d\left(\frac{\eta(i/n)-u}{h_d}\right) + K_d'\left(\frac{\eta(i/n)-u}{h_d}\right)\left\{\frac{\hat{\eta}(i/n)-\eta(i/n)}{h_d}\right\} + \frac{1}{2}K_d''\left(\frac{\xi_i-u}{h_d}\right)\left\{\frac{\hat{\eta}(i/n)-\eta(i/n)}{h_d}\right\}^2$$

with $|\xi_i - \eta(i/n)| < |\hat{\eta}(i/n) - \eta(i/n)|$ and $\Delta_n(t)$ can be expressed as

$$\Delta_n(t) = \Delta_n^{(1)}(t) + \frac{1}{2}\Delta_n^{(2)}(t), \tag{2.24}$$

where

$$\Delta_n^{(1)}(t) = \frac{1}{nh_d^2}\sum_{i=1}^{n}\int_{-\infty}^{t} K_d'\left(\frac{\eta(i/n)-u}{h_d}\right)\{\hat{\eta}(i/n)-\eta(i/n)\}\,du$$

$$\Delta_n^{(2)}(t) = \frac{1}{nh_d^3}\sum_{i=1}^{n}\int_{-\infty}^{t} K_d''\left(\frac{\xi_i-u}{h_d}\right)\{\hat{\eta}(i/n)-\eta(i/n)\}^2\,du.$$

First I focus attention on the second term $\Delta_n^{(2)}(t)$ and show that it is negligible compared to $\Delta_n^{(1)}(t)$. It is clear that by integrating out the derivative and replacing the sum by a Riemann integral one obtains

$$\Delta_n^{(2)}(t) = \frac{1}{h_d^2}\int_0^1 K_d'\left(\frac{\eta(x)-t}{h_d}\right)\{\hat{\eta}(x)-\eta(x)\}^2 dx(1+o(1)). \tag{2.25}$$

Making the substitution $u = (\eta(x)-t)/h_d$ and using a Taylor expansion for the resulting terms involving $t + uh_d$ gives

$$\begin{aligned}\Delta_n^{(2)}(t) &= \frac{1}{h_d}\{\hat{\eta}(\eta^{-1}(t)) - \eta(\eta^{-1}(t))\}^2(\eta^{-1})'(t)\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d} K_d'(u)du(1+o(1)).\\ &= \frac{C(t)}{h_d}\{\hat{\eta}(\eta^{-1}(t)) - \eta(\eta^{-1}(t))\}^2(1+o(1))\end{aligned}$$

where $C(t) = (\eta^{-1})'(t)\int K_d'(u)du$. For the theorem there are two cases to consider. The first is the case when $\lim_{n\to\infty} h_r/h_d = c \in [0,\infty)$. In this case one has

$$\begin{aligned}\sqrt{nh_d}\Delta_n^{(2)}(t) &= \frac{C(t)\sqrt{nh_d}}{h_d}\{\hat{\eta}(\eta^{-1}(t)) - \eta(\eta^{-1}(t))\}^2\\ &= C(t)\sqrt{\frac{n}{h_d}}\{\hat{\eta}(\eta^{-1}(t)) - \eta(\eta^{-1}(t))\}^2\\ &= \{\sqrt[4]{\frac{n}{h_d}}\sqrt{C(t)}(\hat{\eta}(\eta^{-1}(t)) - \eta(\eta^{-1}(t)))\}^2.\end{aligned} \tag{2.26}$$

Theorem 3 implies that

$$\left\{\sqrt[4]{\frac{n}{h_d}}\sqrt{C(t)}(\hat{\eta}(\eta^{-1}(t))-\eta(\eta^{-1}(t)))\right\} \xrightarrow{D} N(\varphi(\eta^{-1}(t))h_r^2\sqrt[4]{\frac{n}{h_d}}\sqrt{C(t)}, \sigma^2(\eta^{-1}(t))\frac{C(t)}{h_r\sqrt{nh_d}}) \tag{2.27}$$

If a normal random variable has mean υ and variance τ^2, then the sum of squares of the random variable has a non central χ^2 distribution with two parameters, k the degrees of freedom, and $\lambda = \upsilon^2/\tau^2$. The mean of such a χ^2 distribution is $k+\lambda$ whilst the variance is equal to $2(k+2\lambda)$. Applying this here, I have one random variable and hence $k = 1$. Combining (2.26) and (2.27) I have in this case that

$$\frac{nh_rh_d\Delta_n^{(2)}(t)}{\sigma^2(\eta^{-1}(t))C(t)} \xrightarrow{D} \chi_1^2(\lambda) \tag{2.28}$$

with

$$\lambda = \frac{\varphi^2(\eta^{-1}(t))nh_r^5}{\sigma^2(\eta^{-1}(t))}. \tag{2.29}$$

The expected value and variance of this chi-squared distribution are

$$E\left[\frac{nh_rh_d\Delta_n^{(2)}(t)}{\sigma^2(\eta^{-1}(t))C(t)}\right] = 1 + \frac{\varphi^2(\eta^{-1}(t))nh_r^5}{\sigma^2(\eta^{-1}(t))} \tag{2.30}$$

$$Var\left[\frac{nh_rh_d\Delta_n^{(2)}(t)}{\sigma^2(\eta^{-1}(t))C(t)}\right] = 2 + \frac{2\varphi^2(\eta^{-1}(t))nh_r^5}{\sigma^2(\eta^{-1}(t))}. \tag{2.31}$$

Recalling assumption (B5), it is clear that the expected value is $1 + O(1)$, whilst the variance is $2 + O(1)$. The variance τ^2 has order

$$\frac{\sigma^2(\eta^{-1}(t))C(t)}{h_r\sqrt{nh_d}} = O\left(\frac{1}{\sqrt{nh_dh_r^2}}\right) = O\left(\sqrt{\frac{h_r^3}{h_d}}\right) = O(h_r) = o(1), \tag{2.32}$$

we can now see that

$$\sqrt{nh_d}\Delta_n^{(2)}(t) = o(1) \tag{2.33}$$

in distribution.

In the case where $\lim_{n\to\infty} h_r/h_d = \infty$, I need to calculate

$$\begin{aligned}\sqrt{nh_r}\Delta_n^{(2)}(t) &= \frac{C(t)\sqrt{nh_r}}{h_d}\{\hat{\eta}(\eta^{-1}(t))-\eta(\eta^{-1}(t))\}^2 \\ &= \left\{\sqrt[4]{\frac{nh_r}{h_d^2}}\sqrt{C(t)}(\hat{\eta}(\eta^{-1}(t))-\eta(\eta^{-1}(t)))\right\}^2.\end{aligned} \tag{2.34}$$

Proceeding as before I now have that

$$\frac{nh_rh_d\Delta_n^{(2)}(t)}{\sigma^2(\eta^{-1}(t))C(t)} \xrightarrow{D} \chi_1^2(\lambda). \tag{2.35}$$

The term λ is the same as in (2.29), although the asymptotic mean and variance of (2.34) are different than in (2.27). This gives the expected value of this distribution being $1+O(1)$ and the variance being $2+O(1)$. In this case the variance τ^2 is given by

$$\frac{\sigma^2(\eta^{-1}(t))C(t)}{\sqrt{nh_d^2h_r}} = O\left(\frac{1}{\sqrt{nh_d^2h_r}}\right) = O\left(\sqrt{\frac{h_r^4}{h_d^2}}\right) = O\left(\frac{h_r^2}{h_d}\right). \tag{2.36}$$

Using the assumption (C3), I can say that $\sqrt{nh_r}\Delta_n^{(2)}(t) = o(1)$, as required.

Using (2.33) we now can prove the first assertion of Theorem 1 if it can be shown that

$$\sqrt{nh_d}\left(\Delta_n^{(1)} + \frac{\mu_2(K)h_r^2}{2}\eta''(\eta^{-1}(t))(\eta^{-1})'(t)\right) \Longrightarrow^D N(0, r^2(t)). \tag{2.37}$$

I firstly consider the asymptotic expected value and variance of $\Delta_n^{(1)}(t)$. After this I will show the asymptotic normality step. Consider $\Delta_n^{(1)}(t)$,

$$\begin{aligned}
\Delta_n^{(1)}(t) &= \frac{1}{nh_d^2}\sum_{i=1}^{n}\int_{-\infty}^{t} K_d'\left(\frac{\eta(i/n)-u}{h_d}\right)\{\hat{\eta}(i/n)-\eta(i/n)\}\,du \\
&= \frac{1}{nh_d}\sum_{i=1}^{n} K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\{\hat{\eta}(i/n)-\eta(i/n)\} \\
&= \frac{1}{nh_d}\sum_{i=1}^{n} K_d\left(\frac{\eta(i/n)-t}{h_d}\right) \\
&\quad\times\left\{\frac{-1}{nh_r\rho(i/n)f(i/n)}\sum_{j=1}^{n} q_1\{\bar{\eta}(i/n,X_j),Y_j\}K_r\left(\frac{X_j-i/n}{h_r}\right)\right\}(1+o_p((nh_r)^{-1/2}))
\end{aligned} \tag{2.38}$$

where $\bar{\eta}(x,z) = \eta(x)+\eta'(x)(z-x)$, and $q_1(x,y)$ and $\rho(x)$ are as defined in (2.3) and (2.5) respectively. The last line above comes from the results in Fan and Gijbels ([16] page 212). Throughout the proof I will make use of the identities $\int_{-\infty}^{\infty}K(z)=1$, and $\int_{-\infty}^{\infty}zK(z)=0$, which follow from the assumption that K is a symmetric bounded probability density function with finite fourth moment. Now consider

$$\Delta_n^{(1)}(t) = \frac{1}{nh_d}\sum_{i=1}^{n} K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{-1}{nh_r\rho(i/n)f(i/n)}\sum_{j=1}^{n} q_1\{\bar{\eta}(i/n,X_j),Y_j\}K_r\left(\frac{X_j-i/n}{h_r}\right)\right\}. \tag{2.39}$$

The next task is to calculate the asymptotic expectation of $\Delta_n^{(1)}(t)$.

$$\begin{aligned}
E[\Delta_n^{(1)}(t)] &= E\left[\sum_{j=1}^{n}\frac{-1}{n^2h_rh_d}\sum_{i=1}^{n}K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(i/n,X_j),Y_j\}}{\rho(i/n)f(i/n)}K_r\left(\frac{X_j-i/n}{h_r}\right)\right\}\right] \\
&= \frac{-1}{n^2h_rh_d}\sum_{j=1}^{n}E\left[\sum_{i=1}^{n}K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(i/n,X_j),Y_j\}}{\rho(i/n)f(i/n)}K_r\left(\frac{X_j-i/n}{h_r}\right)\right\}\right] \\
&= \frac{-1}{nh_rh_d}\sum_{j=1}^{n}E\left[\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(x,X_j),Y_j\}}{\rho(x)f(x)}K_r\left(\frac{X_j-x}{h_r}\right)\right\}dx\right](1+o(1)) \\
&= \frac{-1}{h_rh_d}E\left[\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(x,X_1),\mu(X_1)\}}{\rho(x)f(x)}K_r\left(\frac{X_1-x}{h_r}\right)\right\}dx\right](1+o(1)) \\
&= \frac{-1}{h_rh_d}\int_0^1\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(x,z),\mu(z)\}}{\rho(x)f(x)}K_r\left(\frac{z-x}{h_r}\right)\right\}f(z)dzdx(1+o(1)) \\
&= \frac{-1}{h_d}\int_0^1\int_{-x/h_r}^{(1-x)/h_r}K_d\left(\frac{\eta(x)-t}{h_d}\right) \\
&\quad\times\left\{\frac{q_1\{\bar{\eta}(x,x+uh_r),\mu(x+uh_r)\}}{\rho(x)f(x)}K_r(u)\right\}f(x+uh_r)dudx(1+o(1)) \qquad (2.40)
\end{aligned}$$

where we have used the substitution $u=(z-x)/h_r$, and the fact that q_1 is linear in y for fixed x. Fan and Gijbels ([16] page 213, section 5.6) state that

$$q_1\{\bar{\eta}(x,x+uh_r),\mu(x+uh_r)\} = \rho(x+uh_r)(uh_r)^2\frac{\eta''(x)}{2}+o(h_r^2). \qquad (2.41)$$

To show (2.41), recall (2.3) and notice that $\bar{\eta}(x,x+uh_r)=\eta(x)+\eta'(x)uh_r$, and consider the Taylor expansion,

$$\eta(x+uh_r)=\eta(x)+\eta'(x)uh_r+\eta''(x)\frac{(uh_r)^2}{2}+\eta^{(3)}(\tilde{x})\frac{(uh_r)^3}{6}=\bar{\eta}(x,x+uh_r)+\Delta\eta,$$

where $x\le\tilde{x}\le x+uh_r$. Then, using (2.4)

$$\begin{aligned}
q_1\{\bar{\eta}(x,x+uh_r),\mu(x+uh_r)\} &= q_1\{(\eta(x+uh_r)-\Delta\eta),\mu(x+uh_r)\} \\
&= q_1\{\eta(x+uh_r),\mu(x+uh_r)\}-\Delta\eta q_2\{\eta(x+uh_r),\mu(x+uh_r)\}+O(h_r^4) \\
&= \left(\eta''(x)\frac{(uh_r)^2}{2}+\eta^{(3)}(\tilde{x})\frac{(uh_r)^3}{6}\right)\rho(x+uh_r)+O(h_r^4) \\
&= \eta''(x)\frac{(uh_r)^2}{2}\rho(x+uh_r)+o(h_r^2)
\end{aligned}$$

as required. Here, the $O(h_r^4)$ term stems from the fact that all further terms of the Taylor expansion will have a term involving $(\Delta\eta)^2$, or a greater power.

The leading term of this expression is $(\eta''(x))^2(uh_r)^4/4 = O(h_r^4)$. Before returning to (2.40) consider the following Taylor expansions,

$$\eta''(\eta^{-1}(t+uh_d)) = \eta''(\eta^{-1}(t)) + \eta^{(3)}(\eta^{-1}(t))(\eta^{-1})'(t)uh_d + \frac{u^2h_d^2}{2}\left\{\eta^{(4)}(\eta^{-1}(\tilde{t}))(\eta^{-1})'^2(\tilde{t}) + \eta^{(3)}(\eta^{-1}(t))(\eta^{-1})''(\tilde{t})\right\} \tag{2.42}$$

for $t < \tilde{t} < t + uh_d$, and

$$(\eta^{-1})'(t+uh_d) = (\eta^{-1})'(t) + (\eta^{-1})''(t)uh_d + (\eta^{-1})^{(3)}(t^*)\frac{u^2h_d^2}{2} \tag{2.43}$$

for $t < t^* < t+uh_d$, are also required. Define $b(x) = \rho(x)f(x)$, and then using Taylor expansions it can be seen that,

$$b(x+uh_r) = b(x) + b'(x)uh_r + b''(\tilde{x})\frac{u^2h_r^2}{2} \tag{2.44}$$

for $x < \tilde{x} < x + uh_r$. It is then clear that $b(x+uh_r)/b(x) = 1 + O(h_r)$. Using equation (2.41) and the additional substitution $v = (\eta(x) - t)/h_d$, followed by expansions (2.42)-(2.44), equation (2.40) becomes

$$\begin{aligned}
E[\Delta_n^{(1)}(t)] &= -\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}\int_{(-\eta^{-1}(t+vh_d))/h_r}^{(1-\eta^{-1}(t+vh_d))/h_r} K_d(v)(uh_r)^2\frac{\eta''(\eta^{-1}(t+vh_d))}{2} \\
&\quad \times K_r(u)\eta^{-1}(t+vh_d)dudv(1+O(h_r)) \\
&= -\frac{h_r^2}{2}\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}\int_{(-\eta^{-1}(t+vh_d))/h_r}^{(1-\eta^{-1}(t+vh_d))/h_r} u^2K_r(u)K_d(v)(\eta''(\eta^{-1}(t)) + O(h_d)) \\
&\quad \times((\eta^{-1})'(t) + O(h_d))dudv(1+O(h_r)) \\
&= \eta''(\eta^{-1}(t))(\eta^{-1})'(t)\frac{-h_r^2}{2}\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d} K_d(v)\int_{(-\eta^{-1}(t+vh_d))/h_r}^{(1-\eta^{-1}(t+vh_d))/h_r} u^2K_r(u)dudv\{1+o(1)\} \\
&= \eta''(\eta^{-1}(t))(\eta^{-1})'(t)\kappa_2(K_r)\frac{-h_r^2}{2}\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d} K_d(v)dv\{1+o(1)\} \\
&= -h_r^2\eta''(\eta^{-1}(t))(\eta^{-1})'(t)\kappa_2(K_r)\{1+o(1)\}
\end{aligned}$$

Notice that the expectation is of order $O(h_r^2)$. The next task is to calculate the asymptotic

variance of $\Delta_n^{(1)}(t)$.

$$\begin{aligned}
Var[\Delta_n^{(1)}(t)] &= Var\left[\sum_{j=1}^{n}\frac{-1}{n^2h_rh_d}\sum_{i=1}^{n}K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(i/n,X_j),Y_j\}}{b(i/n)}K_r\left(\frac{X_j-i/n}{h_r}\right)\right\}\right] \\
&= \sum_{j=1}^{n}\frac{1}{n^4h_r^2h_d^2}Var\left[\mathcal{Z}_j\right] \\
&= \frac{1}{n^3h_r^2h_d^2}Var\left[\mathcal{Z}_j\right] \\
&= \frac{1}{n^3h_r^2h_d^2}\left\{E\left[\mathcal{Z}_j^2\right]-E\left[\mathcal{Z}_j\right]^2\right\}. \qquad (2.45)
\end{aligned}$$

I will first consider the first term of this expression.

$$\begin{aligned}
E[\mathcal{Z}_j^2] &= n^2E\left[\left(\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(x,X_1),Y_1\}}{b(x)}K_r\left(\frac{X_1-x}{h_r}\right)\right\}dx\right)^2\right]\{1+o(1)\} \\
&= n^2\int_0^1\int_0^1\left(\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(x,z),y\}}{b(x)}K_r\left(\frac{z-x}{h_r}\right)dx\right\}\right)^2 \\
&\quad \times f(z,y)dzdy\{1+o(1)\} \\
&= n^2\left[\int_0^1\int_0^1\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(x,z),y\}}{b(x)}K_r\left(\frac{z-x}{h_r}\right)dx\right\}\right. \\
&\quad \left.\times\int_0^1 K_d\left(\frac{\eta(a)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(a,z),y\}}{b(a)}K_r\left(\frac{z-a}{h_r}\right)da\right\}f(z,y)dzdy\right]\{1+o(1)\} \\
&= n^2\left[\int_0^1 K_d\left(\frac{\eta(a)-t}{h_d}\right)\frac{1}{b(a)}\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\frac{1}{b(x)}\int_0^1\int_0^1 q_1\{\bar{\eta}(a,z),y\}\right. \\
&\quad \left.\times q_1\{\bar{\eta}(x,z),y\}K_r\left(\frac{z-a}{h_r}\right)K_r\left(\frac{z-x}{h_r}\right)f(z,y)dydzdxda\right]\{1+o(1)\}. \qquad (2.46)
\end{aligned}$$

Now it is necessary to recall the definitions of $q_1(x,y) = (\partial/\partial x)Q\{g^{-1}(x),y\}$, and $(\partial/\partial\nu)Q(\nu,y) = (y-\nu)/V(\nu)$. Using these definitions and the chain rule it can be seen that

$$\begin{aligned}
q_1(x,y) &= \frac{\partial}{\partial g^{-1}(x)}Q(g^{-1}(x),y)\frac{\partial g^{-1}(x)}{\partial x} \\
&= \frac{y-g^{-1}(x)}{V(g^{-1}(x))}\frac{\partial g^{-1}(x)}{\partial x} \\
&= \frac{y-g^{-1}(x)}{V(g^{-1}(x))}\frac{1}{g'(g^{-1}(x))} \\
&= \frac{y-g^{-1}(x)}{G(x)}
\end{aligned}$$

where $G(x) = V(g^{-1}(x))g'(g^{-1}(x))$, and hence

$$q_1(\bar{\eta}(x,z),y) = \frac{y-g^{-1}(\bar{\eta}(x,z))}{G(\bar{\eta}(x,z)))}. \qquad (2.47)$$

To proceed, recall that $\bar{\eta}(x,z) = \eta(x) + \eta'(x)(z-x)$, and also make use of the following Taylor expansions.

$$\begin{aligned} G(\bar{\eta}(x,z)) &= G(\eta(x) + \eta'(x)(z-x)) \\ &= G(\eta(x)) + G'(\eta(\tilde{x}))\eta'(x)(z-x) \end{aligned} \tag{2.48}$$

where $\eta(x) < \eta(\tilde{x}) < \eta(x) + \eta'(x)(z-x)$, and

$$g^{-1}(\bar{\eta}(x,z)) = g^{-1}(\eta(x)) + (g^{-1})'(\eta(\check{x}))\eta'(x)(z-x) \tag{2.49}$$

with $\eta(x) < \eta(\check{x}) < \eta(x) + \eta'(x)(z-x)$.

I'm now in a position to return to (2.46) and will use the substitutions, $u = (\eta(a) - t)/h_d$, $v = (\eta(x) - t)/h_d$ and $w = (z - \eta^{-1}(t + vh_d))/h_r$.

$$\begin{aligned} E[\mathcal{Z}_j^2] = n^2 \Bigg[& \int_0^1 K_d\left(\frac{\eta(a)-t}{h_d}\right)\frac{1}{b(a)}\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\frac{1}{b(x)}\int_0^1\int_0^1 \\ & \times \frac{\{y - g^{-1}(\eta(x)) + (g^{-1})'(\eta(\check{x}))(z-x)\eta'(x)\}}{G(\eta(x)) + G'(\eta(\tilde{x}))\eta'(x)(z-x)} \frac{\{y - g^{-1}(\eta(a)) + (g^{-1})'(\eta(\check{a}))(z-a)\eta'(a)\}}{G(\eta(a)) + G'(\eta(\tilde{a}))\eta'(a)(z-a)} \\ & \times K_r\left(\frac{z-a}{h_r}\right) K_r\left(\frac{z-x}{h_r}\right) f(z,y)dydzdxda \Bigg] \{1 + o(1)\} \\ = n^2 h_d^2 \Bigg[& \int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d} \frac{K_d(u)}{b(\eta^{-1}(t+uh_d))} \int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d} \frac{K_d(v)}{b(\eta^{-1}(t+vh_d))} \\ & \times \int_0^1\int_0^1 \frac{\{y - g^{-1}(\eta(\eta^{-1}(t+vh_d))) + (g^{-1})'(\eta(\check{x}))(z-\eta^{-1}(t+vh_d))\eta'(\eta^{-1}(t+vh_d))\}}{G(\eta(\eta^{-1}(t+vh_d))) + G'(\eta(\tilde{x}))\eta'(\eta^{-1}(t+vh_d))(z-\eta^{-1}(t+vh_d))} \\ & \times \frac{\{y - g^{-1}(\eta(\eta^{-1}(t+uh_d))) + (g^{-1})'(\eta(\check{a}))(z-\eta^{-1}(t+uh_d))\eta'(\eta^{-1}(t+uh_d))\}}{G(\eta(\eta^{-1}(t+uh_d))) + G'(\eta(\tilde{a}))\eta'(\eta^{-1}(t+uh_d))(z-\eta^{-1}(t+uh_d))} \\ & \times K_r\left(\frac{z-\eta^{-1}(t+uh_d)}{h_r}\right) K_r\left(\frac{z-\eta^{-1}(t+vh_d)}{h_r}\right) f(z,y)dydz \\ & \times (\eta^{-1})'(t+vh_d)dv(\eta^{-1})'(t+uh_d)du \Bigg] \{1 + o(1)\} \end{aligned}$$

$$
\begin{aligned}
E[\mathcal{Z}_j^2] = n^2 h_d^2 h_r \Bigg[& \int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d} \frac{K_d(u)}{b(\eta^{-1}(t+uh_d))} \int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d} \frac{K_d(v)}{b(\eta^{-1}(t+vh_d))} \\
& \times \int_{(-\eta^{-1}(t+vh_d))/h_r}^{(1-\eta^{-1}(t+vh_d))/h_r} \int_0^1 \frac{\{y - g^{-1}(t+vh_d) + wh_r (g^{-1})'(\eta(\check{x}))\eta'(\eta^{-1}(t+vh_d))\}}{G(t+vh_d) + wh_r G'(\eta(\tilde{x}))\eta'(\eta^{-1}(t+vh_d))} \\
& \times \frac{\{y - g^{-1}(t+uh_d) + (g^{-1})'(\eta(\check{a}))(wh_r + \eta^{-1}(t+vh_d) - \eta^{-1}(t+uh_d))\eta'(\eta^{-1}(t+uh_d))\}}{G(t+uh_d) + G'(\eta(\tilde{a}))\eta'(\eta^{-1}(t+uh_d))(wh_r + \eta^{-1}(t+vh_d) - \eta^{-1}(t+uh_d))} \\
& \times K_r(w) K_r\left(w + \frac{\eta^{-1}(t+vh_d) - \eta^{-1}(t+uh_d)}{h_r}\right) f(wh_r + \eta^{-1}(t+vh_d), y) dy dw \\
& \times (\eta^{-1})'(t+vh_d) dv (\eta^{-1})'(t+uh_d) du \Bigg] \{1 + o(1)\}
\end{aligned} \tag{2.50}
$$

To finish calculating this asymptotic variance I will make use of the Taylor expansions (2.51) to (2.55).

$$
\eta^{-1}(t+uh_d) = \eta^{-1}(t) + uh_d(\eta^{-1})'(\tilde{t}_u), \text{ for } t \leq \tilde{t}_u \leq t + uh_d \tag{2.51}
$$

$$
g^{-1}(t+uh_d) = g^{-1}(t) + uh_d(g^{-1})'(t_u^*), \text{ for } t \leq t_u^* \leq t + uh_d \tag{2.52}
$$

$$
(\eta^{-1})'(t+uh_d) = (\eta^{-1})'(t) + uh_d(\eta^{-1})''(\check{t}_u), \text{ for } t \leq \check{t}_u \leq t + uh_d \tag{2.53}
$$

$$
\begin{aligned}
b(\eta^{-1}(t+uh_d)) &= b(\eta^{-1}(t) + uh_d(\eta^{-1})'(\tilde{t})) \\
&= b(\eta^{-1}(t)) + b'(\eta^{-1}(\check{t}_u)) uh_d(\eta^{-1})'(\tilde{t}), \text{ for } \eta^{-1}(t) \leq \eta^{-1}(\check{t}) \leq \eta^{-1}(t) + uh_d(\eta^{-1})'(\tilde{t})
\end{aligned} \tag{2.54}
$$

$$
G(t+uh_d) = G(t) + uh_d G'(\acute{t}), \text{ for } t \leq \acute{t}_u \leq t + uh_d. \tag{2.55}
$$

Returning to (2.50) and noticing that $g^{-1}(t) = g^{-1}(\eta(\eta^{-1}(t))) = \mu(\eta^{-1}(t))$, I show the expres-

sion for the asymptotic variance.

$$
\begin{aligned}
E[\mathcal{Z}_j^2] &= n^2h_d^2h_r\left[\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}\frac{K_d(u)}{b(\eta^{-1}(t))}\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}\frac{K_d(v)}{b(\eta^{-1}(t))}\int_{(-\eta^{-1}(t+vh_d))/h_r}^{(1-\eta^{-1}(t+vh_d))/h_r}\right.\\
&\quad\times\int_0^1\left(\frac{y-g^{-1}(t)}{G(t)}\right)^2K_r(w)K_r\left(w+\frac{\eta^{-1}(t+vh_d)-\eta^{-1}(t+uh_d)}{h_r}\right)\\
&\quad\left.\times f(\eta^{-1}(t),y)dydw(\eta^{-1})'^2(t)dvdu\right]\{1+o(1)\}\\
&= \frac{n^2h_d^2h_r(\eta^{-1})'^2(t)}{b^2(\eta^{-1}(t))G^2(t)}\left[\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}K_d(u)\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}K_d(v)\right.\\
&\quad\times\int_{(-\eta^{-1}(t+vh_d))/h_r}^{(1-\eta^{-1}(t+vh_d))/h_r}\int_0^1(y-g^{-1}(t))^2f_{Y|X}(y|x=\eta^{-1}(t))f_X(\eta^{-1}(t))dy\\
&\quad\left.\times K_r(w)K_r\left(w+\frac{\eta^{-1}(t+vh_d)-\eta^{-1}(t+uh_d)}{h_r}\right)dwdvdu\right]\{1+o(1)\}\\
&= \frac{n^2h_d^2h_r(\eta^{-1})'^2(t)[g'(g^{-1}(t)]^2Var[Y|X=\eta^{-1}(t)]}{f(\eta^{-1}(t))}\\
&\quad\times\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}\int_{(-\eta^{-1}(t+vh_d))/h_r}^{(1-\eta^{-1}(t+vh_d))/h_r}K_d(u)K_d(v)K_r(w)\\
&\quad\times K_r\left(w+\frac{\eta^{-1}(t+vh_d)-\eta^{-1}(t+uh_d)}{h_r}\right)dwdvdu\{1+o(1)\}. \qquad (2.56)
\end{aligned}
$$

Recalling (2.45) it is now clear then that the first term of this is

$$
\begin{aligned}
\frac{1}{n^3h_r^2h_d^2}E[\mathcal{Z}_j^2] &= \frac{(\eta^{-1})'^2(t)[g'(g^{-1}(t)]^2Var[Y|X=\eta^{-1}(t)]}{nh_rf(\eta^{-1}(t))}\\
&\quad\times\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}\int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d}\int_{(-\eta^{-1}(t+vh_d))/h_r}^{(1-\eta^{-1}(t+vh_d))/h_r}K_d(u)K_d(v)K_r(w)\\
&\quad\times K_r\left(w+\frac{\eta^{-1}(t+vh_d)-\eta^{-1}(t+uh_d)}{h_r}\right)dwdvdu\{1+o(1)\}. \qquad (2.57)
\end{aligned}
$$

I now need to consider the second term of (2.45).

$$
\begin{aligned}
\frac{-1}{n^3h_r^2h_d^2}E\left[\mathcal{Z}_j\right]^2 &= E\left[\frac{1}{n^3h_r^2h_d^2}(nh_rh_d\Delta_n^{(1)}(t))^2\right]\\
&= E\left[\frac{(\Delta_n^{(1)}(t))^2}{n}\right]. \qquad (2.58)
\end{aligned}
$$

It is clear that, since the expectation of $\Delta_n^{(1)}(t))$ has $O(h_r^2)$, the term (2.58) is $O(h_r^4/n)$ and the order of (2.59) is $O(1/(nh_r))$. I can say then that (2.58) is $o((nh_r)^{-1})$. The dominant part of

the variance is given by (2.59). The variance of $\Delta_n^{(1)}(t))$ is given by

$$Var[\Delta_n^{(1)}(t))] = \frac{(\eta^{-1})'^2(t)[g'(g^{-1}(t)]^2 Var[Y|X=\eta^{-1}(t)]}{nh_r f(\eta^{-1}(t))} \times \int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d} \int_{(\eta(0)-t)/h_d}^{(\eta(1)-t)/h_d} \int_{(-\eta^{-1}(t+vh_d))/h_r}^{(1-\eta^{-1}(t+vh_d))/h_r} K_d(u)K_d(v)K_r(w) \times K_r\left(w + \frac{\eta^{-1}(t+vh_d) - \eta^{-1}(t+uh_d)}{h_r}\right) dwdvdu\{1+o(1)\}. \quad (2.59)$$

I now consider the case where $\lim_{n\to\infty} hr/h_d \to c \in [0,\infty)$. Dette, Neumeyer and Pilz [9] state the following relation

$$\lim_{h_r\to 0, h_d\to 0, h_r/h_d\to c} K_d\left(\frac{\eta(\eta^{-1}(t+h_d w) + h_r(v-u)) - t}{h_d}\right) = K_d(w + c\eta'(\eta^{-1}(t))(v-u)). \quad (2.60)$$

At this point, making the substitution $\nu = \{\eta(\eta^{-1}(t+h_d w) + h_r(v-u)) - t\}/h_d$ in (2.56), and using the identity (2.60) yields

$$Var[\Delta_n^{(1)}(t)] = \frac{[g'(g^{-1}(t)]^2 Var[Y|X=\eta^{-1}(t)]}{\eta'(\eta^{-1}(t) nh_d f(\eta^{-1}(t))} \times \int\int\int K_d\left(u + \frac{h_r}{h_d}\eta'(\eta^{-1}(t))(\nu - w)\right) K_d(u)K_r(w)K_r(\nu) dudwd\nu$$

since $(\eta^{-1})'(t) = 1/\eta'(\eta^{-1}(t))$. Hence $\sqrt{nh_d}Var[\Delta_n^{(1)}(t)] = r(t)$ as required (see (2.7)).

Now consider the second case. For the second part of the theorem, (2.8) is obtained when the factor of $\sqrt{nh_r}$ is considered. This factor cancels with the nh_r in the denominator of (2.56) to give the required variance.

Now considering the second case, when $\lim_{n\to\infty} hr/h_d \to \infty$

$$K_r\left(w + \frac{\eta^{-1}(t+vh_d) - \eta^{-1}(t+uh_d)}{h_r}\right) \to K_r(w).$$

Remembering that $\int K(\nu)d\nu = 1$, gives the required results. Hence $\sqrt{nh_r}Var[\Delta_n^{(1)}(t)] = \tilde{r}(t)$ as required (see (2.8)).

The final task is to show the asymptotic normality of $\Delta_n^{(1)}(t)$. Consider the first case of

Theorem 1.

$$
\begin{aligned}
&\sqrt{nh_d}\left(\Delta_n^{(1)}(t)+\kappa_2(K_r)h_r^2\eta''(\eta^{-1}(t))(\eta^{-1})'(t)\right)\\
&=\sqrt{nh_d}\left(\sum_{j=1}^{n}\frac{-1}{n^2h_rh_d}\sum_{i=1}^{n}K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(i/n,X_j),Y_j\}}{b(i/n)}K_r\left(\frac{X_j-i/n}{h_r}\right)\right\}\right.\\
&\quad\left.+\kappa_2(K)h_r^2\eta''(\eta^{-1}(t))(\eta^{-1})'(t)\right)\\
&=\sum_{j=1}^{n}\left\{\frac{-\sqrt{nh_d}}{n^2h_rh_d}\sum_{i=1}^{n}K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(i/n,X_j),Y_j\}}{b(i/n)}K_r\left(\frac{X_j-i/n}{h_r}\right)\right\}\right\}\\
&\quad+\sqrt{nh_d}\kappa_2(K)h_r^2\eta''(\eta^{-1}(t))(\eta^{-1})'(t)\\
&=\sum_{j=1}^{n}\left\{\frac{-\sqrt{nh_d}}{n^2h_rh_d}\sum_{i=1}^{n}K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(i/n,X_j),Y_j\}}{b(i/n)}K_r\left(\frac{X_j-i/n}{h_r}\right)\right\}\right.\\
&\quad\left.+\frac{h_r^2\sqrt{nh_d}}{n}\kappa_2(K)\eta''(\eta^{-1}(t))(\eta^{-1})'(t)\right\}\\
&=\sum_{j=1}^{n}(Z_j-\mu_j).
\end{aligned}
\tag{2.61}
$$

Here,

$$
Z_j=\frac{-\sqrt{nh_d}}{n^2h_rh_d}\sum_{i=1}^{n}K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(i/n,X_j),Y_j\}}{b(i/n)}K_r\left(\frac{X_j-i/n}{h_r}\right)\right\}
\tag{2.62}
$$

is a random variable with mean,

$$
\begin{aligned}
\mu_j&=\frac{h_r^2\sqrt{nh_d}}{n}\kappa_2(K)\eta''(\eta^{-1}(t))(\eta^{-1})'(t)\\
&=\sqrt{\frac{h_d}{n}}E[\Delta_n^{(1)}(t)]
\end{aligned}
\tag{2.63}
$$

and variance

$$
\sigma_j^2=Var[Z_j]=h_dVar[\Delta_n^{(1)}(t)].
\tag{2.64}
$$

The Lyapunov central limit theorem states that for a series of independent identically distributed random variables $X_1,\dots X_n$, with finite expected value μ_i and variance σ_i^2, if for some $\delta>0$

$$
\lim_{n\to\infty}\frac{1}{s_n^{2+\delta}}\sum_{j=1}^{n}E[|X_j-\mu_j|^{\delta}]=0
\tag{2.65}
$$

is satisfied, where $s_n^2=\sum_{i=1}^{n}\sigma_i^2$, then

$$
\sum_{j=1}^{n}(Z_j-\mu_j)\xrightarrow{D}N(0,r^2(t)).
$$

I use the Lyapunov central limit theorem to prove asymptotic normality of $\Delta_n^{(1)}(t)$, setting $\delta = 2$. First notice that since $Var[\Delta_n^{(1)}(t)] = O(1/(nh_d))$, $\sigma_i^2 = O(1/n)$. Hence $s_n^2 = \sum_{i=1}^n \sigma_i^2 = O(1)$ and $s_n^4 = O(1)$.

I next want to calculate the value of $E[Z_1^k]$ for $k = 1, 2, 3, 4$.

$$
\begin{aligned}
E[Z_1^k] &= E\left[\left(\frac{-\sqrt{nh_d}}{n^2 h_r h_d}\sum_{i=1}^n K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(i/n,X_1),Y_1\}}{b(i/n)}K_r\left(\frac{X_1-i/n}{h_r}\right)\right\}\right)^k\right] \\
&= \left(\frac{-\sqrt{nh_d}}{nh_r h_d}\right)^k E\left[\left(\frac{1}{n}\sum_{i=1}^n K_d\left(\frac{\eta(i/n)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(i/n,X_1),Y_1\}}{b(i/n)}K_r\left(\frac{X_1-i/n}{h_r}\right)\right\}\right)^k\right] \\
&= \left(\frac{-\sqrt{nh_d}}{nh_r h_d}\right)^k E\left[\left(\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(x,X_1),Y_1\}}{b(x)}K_r\left(\frac{X_1-x}{h_r}\right)\right\}dx(1+o(1))\right)^k\right] \\
&= \left(\frac{-\sqrt{nh_d}}{nh_r h_d}\right)^k \int_0^1\int_0^1\left(\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(x,z),y\}}{b(x)}K_r\left(\frac{z-x}{h_r}\right)\right\}dx\right)^k \\
&\quad \times f(z,y)dzdy(1+o(1)) \\
&= \left(\frac{-\sqrt{nh_d}}{nh_r h_d}\right)^k \int_0^1\int_0^1\left\{\prod_{i=1}^k\int_0^1 K_d\left(\frac{\eta(x)-t}{h_d}\right)\left\{\frac{q_1\{\bar{\eta}(\mathfrak{x}_i,z),y\}}{b(\mathfrak{x}_i)}K_r\left(\frac{z-\mathfrak{x}_i}{h_r}\right)\right\}d\mathfrak{x}_i\right\} \\
&\quad \times f(z,y)dzdy(1+o(1)). \qquad (2.66)
\end{aligned}
$$

At this point I need to use two substitutions, $\mathfrak{y}_i = (z-\mathfrak{x}_i)/h_r$, and $u = (\eta(z-h_r\mathfrak{y}_1)-t)/h_d$ in turn. In addition, I use the Taylor expansions (2.47)-(2.49) and (2.51)-(2.55). The following calculations proceed very similarly to the calculations for the variance case which I have already considered here and so only the main steps are shown. To continue then from (2.66),

$$
\begin{aligned}
E[Z_1^k] &= \left(\frac{1}{\sqrt{nh_d}}\right)^k \int_0^1\int_0^1\left\{\prod_{i=1}^k\int_{z/h_r}^{(z-1)/h_r} K_d\left(\frac{\eta(z-h_r\mathfrak{y}_i)-t}{h_d}\right)\frac{q_1\{\bar{\eta}(z-h_r\mathfrak{y}_i,z),y\}}{b(z-h_r\mathfrak{y}_i)}K_r(\mathfrak{y}_i)d\mathfrak{y}_i\right\} \\
&\quad \times f(z,y)dzdy(1+o(1))
\end{aligned}
$$

$$
\begin{aligned}
&= h_d \left(\frac{1}{\sqrt{nh_d}}\right)^k \int_0^1 \int_{(\eta(-h_r\mathfrak{y}_1)-t)/h_d}^{(\eta(1-h_r\mathfrak{y}_1)-t)/h_d} \\
&\quad \times \left\{ \prod_{i=2}^{k} \int_0^1 K_d \left(\frac{\eta^{-1}(t+uh_d) + h_r\mathfrak{y}_1 - h_r\mathfrak{y}_i}{h_d} \right) \right. \\
&\quad \left. \times \frac{q_1\{\bar{\eta}(\eta^{-1}(t+uh_d) + h_r\mathfrak{y}_1 - h_r\mathfrak{y}_i, \eta^{-1}(t+uh_d) + h_r\mathfrak{y}_1), y\}}{b(\eta^{-1}(t+uh_d) + h_r\mathfrak{y}_1 - h_r\mathfrak{y}_i)} K_r(\mathfrak{y}_i) d\mathfrak{y}_i \right\} \\
&\quad \times \left\{ \int_{(\eta^{-1}(t+uh_d)+h_r\mathfrak{y}_1)/h_r}^{(\eta^{-1}(t+uh_d)+h_r\mathfrak{y}_1-1)/h_r} K_d(u) \frac{q_1\{\bar{\eta}(\eta^{-1}(t+uh_d), \eta^{-1}(t+uh_d) + h_r\mathfrak{y}_1), y\}}{b(\eta^{-1}(t+uh_d))} K_r(\mathfrak{y}_1) d\mathfrak{y}_1 \right\} \\
&\quad \times f(\eta^{-1}(t+uh_d), y)(\eta^{-1})'(t+uh_d) du dy (1+o(1))
\end{aligned}
$$

$$
\begin{aligned}
E[Z_1^k] &= h_d \left(\frac{1}{\sqrt{nh_d}}\right)^k \frac{(\eta^{-1})'(t) f(\eta^{-1}(t)}{b^k(\eta^{-1}(t)) G^k(t)} \int\int (y - g^{-1}(t))^k f_{Y|X}(y|x = \eta^{-1}(t)) dy \\
&\quad \times \left\{ \prod_{i=1}^{k} \int K_d \left(u + \frac{h_r}{h_d} (\eta'(\eta^{-1}(t))(\eta^{-1})'(t)(\mathfrak{y}_1 - \mathfrak{y}_i)) \right) K_r(\mathfrak{y}_i) d\mathfrak{y}_i \right\} \\
&\quad \times \int K_d(u) K_r(\mathfrak{y}_1) d\mathfrak{y}_1 du (1+o(1)) \\
&= O\left(\frac{h_d^{(2-k)/2}}{n^{k/2}} \right).
\end{aligned}
$$

It is clear from (2.63) that $\mu_j = O(h_r^2 h_d^{1/2} n^{-1/2})$. The next step is to consider the expression $(Z_j - \mu_j)^4$. Obviously,

$$
(Z_j - \mu_j)^4 = Z_j^4 - 4Z_j^3\mu_j + 6Z_j^2\mu_j^2 - 4Z_j\mu_j^3 + \mu_j^4 \tag{2.67}
$$

and then considering each term in turn I have

$$
E[Z_j^4] = O\left(\frac{1}{n^2 h_d}\right) = o(1) \tag{2.68}
$$

$$
E[Z_j^3\mu_j] = O\left(\frac{h_d^{1/2}}{n^{3/2}}\right) \times O\left(h_r^2 h_d^{1/2} n^{-1/2}\right) = O\left(\frac{h_r^2}{n^2}\right) = o(1) \tag{2.69}
$$

$$
E[Z_j^2\mu_j^2] = O\left(n^{-1}\right) \times O\left(\frac{h_r^4 h_d}{n}\right) = O\left(\frac{h_r^4 h_d}{n^2}\right) = o(1) \tag{2.70}
$$

$$
E[Z_j\mu_j^3] = O\left(\frac{h_d^{1/2}}{n^{1/2}}\right) \times O\left(\frac{h_r^6 h_d^{3/2}}{n^{3/2}}\right) = \left(\frac{h_r^6 h_d^2}{n^2}\right) = o(1) \tag{2.71}
$$

$$
E[\mu_j^4] = O\left(\frac{h_r^8 h_d^2}{n^2}\right) = o(1). \tag{2.72}
$$

From (2.68)-(2.72) it is clear that $E[(Z_j - \mu_j)^4] = o(1)$. Taking the sum over j results in obtaining $\sum_{j=1}^{n} E[(Z_j - \mu_j)^4] = o(1)$. Recalling that $s_n^4 = O(1)$, the Lyapunov condition is satisfied and so I can say that asymptotic normality is proved by the Lyapunov Central Limit theorem. □

2.5.2 Proof of Lemma 1

Proof. This follows the proof given in Dette *et al.* [9]. Recall the definition,

$$\eta_n^{-1}(t) = \frac{1}{nh_d}\int_{-\infty}^{t}\sum_{i=1}^{n} K_d\left(\frac{\eta(i/n)-u}{h_d}\right)du.$$

We can approximate this as a Riemann integral,

$$\eta_n^{-1}(t) = \int_0^1\int_{-\infty}^{t} K_d\left(\frac{\eta(x)-u}{h_d}\right)\frac{1}{h_d}dudx\left(1+O\left(\frac{1}{nh_d}\right)\right).$$

Recall that $t \in (\eta(0),\eta(1))$ is fixed, and observing that the kernel, K_d has compact support, $[-1,1]$ we know that $\eta(x)-h_d \le u \le \eta(x)+h_d$ and so we integrate with respect to u from $\eta(x)-h_d$ to t. Also $\eta^{-1}(u-h_d) \le x \le \eta^{-1}(u+h_d)$ and so we integrate with respect to x from 0 to $\eta^{-1}(t+h_d)$. So the leading term of the right hand side becomes

$$\begin{aligned}
A(h_d) &= \int_0^1\int_{-\infty}^{t} K_d\left(\frac{\eta(x)-u}{h_d}\right)\frac{1}{h_d}dudx \\
&= \int_0^{\eta^{-1}(t+h_d)}\int_{\eta(x)-h_d}^{t} K_d\left(\frac{\eta(x)-u}{h_d}\right)\frac{du}{h_d}dx \\
&= \int_0^{\eta^{-1}(t-h_d)}\int_{\eta(x)-h_d}^{\eta(x)+h_d} K_d\left(\frac{\eta(x)-u}{h_d}\right)\frac{du}{h_d}dx + \int_{\eta^{-1}(t-h_d)}^{\eta^{-1}(t+h_d)}\int_{\eta(x)-h_d}^{t} K_d\left(\frac{\eta(x)-u}{h_d}\right)\frac{du}{h_d}dx \\
&= \eta^{-1}(t-h_d) + \int_{\eta^{-1}(t-h_d)}^{\eta^{-1}(t+h_d)}\int_{\eta(x)-h_d}^{t} K_d\left(\frac{\eta(x)-u}{h_d}\right)\frac{du}{h_d}dx.
\end{aligned}$$

Now we make the substitutions $z = (\eta(x)-t)/h_d$ and $v = (zh_d+t-u)/h_d$ and obtain,

$$A(h_d) = \eta^{-1}(t-h_d) + h_d\int_{-1}^{1}(\eta^{-1})'(t+zh_d)\int_z^1 K_d(v)dvdz.$$

Using the Taylor expansions

$$\eta^{-1}(t-h_d) = \eta^{-1}(t) - h_d(\eta^{-1})'(t) + \frac{h_d^2(\eta^{-1})''(t)}{2} - \frac{h_d^3(\eta^{-1})^{(3)}(\tilde{t})}{6}, \tag{2.73}$$

for $t-h_d < \tilde{t} < t$, and

$$(\eta^{-1})'(t+zh_d) = (\eta^{-1})'(t) + (\eta^{-1})''(t)zh_d + \frac{(\eta^{-1})^{(3)}(t^*)z^2h_d^2}{2} \tag{2.74}$$

for $t < t^* < t + zh_d$, it can be seen that

$$\begin{aligned}
A(h_d) &= \eta^{-1}(t) - h_d(\eta^{-1})'(t) + \frac{h_d^2(\eta^{-1})''(t)}{2} - \frac{h_d^3(\eta^{-1})^{(3)}(\tilde{t})}{6} \\
&\quad + h_d \int_{-1}^{1} \left((\eta^{-1})'(t) + (\eta^{-1})''(t) z h_d + \frac{(\eta^{-1})^{(3)}(t^*) z^2 h_d^2}{2} \right) \int_z^1 K_d(v) dv dz \\
&= \eta^{-1}(t) - h_d(\eta^{-1})'(t) + \frac{h_d^2(\eta^{-1})''(t)}{2} - \frac{h_d^3(\eta^{-1})^{(3)}(\tilde{t})}{6} \\
&\quad + h_d(\eta^{-1})'(t) \int_{-1}^{1} \int_z^1 K_d(v) dv dz + + (\eta^{-1})''(t) h_d^2 \int_{-1}^{1} z \int_z^1 K_d(v) dv dz \\
&\quad + (\eta^{-1})^{(3)}(\tilde{t}) h_d^3 \int_{-1}^{1} z^2 \int_z^1 K_d(v) dv dz \\
&= \eta^{-1}(t) + h_d^2(\eta^{-1})''(t) \kappa_2(K_d) + o(h_d^2)
\end{aligned}$$

where we have used the two identities,

$$\int_{-1}^{1} \int_z^1 K_d(v) dv dz = 1$$

$$\int_{-1}^{1} z \int_z^1 K_d(v) dv dz = \frac{1}{2} \int_{-1}^{1} v^2 K_d(v) dv - \frac{1}{2}.$$

Hence,

$$\eta_n^{-1}(t) = \eta^{-1}(t) + \kappa_2(K_d) h_d^2 (\eta^{-1})''(t) + o(h_d^2) + O\left(\frac{1}{n h_d} \right),$$

as required. □

Lemma 3 gives details of how an operator maps a non-decreasing function η to its 'quantile' $\eta^{-1}(t)$, is taken from Dette, Neumeyer and Pilz [9] and is stated here, without proof. The interested reader is referred to [9] for details. I now need to consider the operator which maps a non-decreasing function η to its quantile $\eta^{-1}(t)$. Lemma 3 and the surrounding discussion are taken from Dette *et al.* [9] although I have replaced their m with η to emphasise that I consider this in a likelihood setting. First I consider a fixed t, and a set $\mathcal{M}$ which is the set of all twice continuously differentiable functions $\mathcal{H} \in C^2[0,1]$, with positive derivative on the interval $[0,1]$, for which t is an interior point of their image. Then consider the functional,

$$\Phi : \begin{cases} \mathcal{M} & \mapsto [0,1] \\ \mathcal{H} & \mapsto \mathcal{H}^{-1}(t) \end{cases}$$

and define $\mathcal{H}_1$, $\mathcal{H}_2 \in \mathcal{M}$ the function

$$Q : \begin{cases} [0,1] & \mapsto \mathbb{R} \\ \lambda & \mapsto \Phi(\mathcal{H}_1 + \lambda(\mathcal{H}_2 - \mathcal{H}_1)). \end{cases} \tag{2.75}$$

Lemma 3 demonstrates that the derivative of Q exists.

Lemma 3. *The mapping $Q : [0,1] \to \mathbb{R}$ defined by (2.75) is twice continuously differentiable with*

$$Q'(\lambda) = -\frac{\mathcal{H}_2 - \mathcal{H}_1}{h_1 + \lambda(h_2 - h_1)} \circ (\mathcal{H}_1 + \lambda(\mathcal{H}_2 - \mathcal{H}_1))^{-1}(t), \tag{2.76}$$

$$Q''(\lambda) = Q'(\lambda) \left\{ \frac{-2(h_2 - h_1)}{h_1 + \lambda(h_2 - h_1)} + \frac{(\mathcal{H}_2 - \mathcal{H}_1)(h_1' + \lambda(h_2' - h_1'))}{\{h_1 + \lambda(h_2 - h_1)\}^2} \right\} \circ Q(\lambda), \tag{2.77}$$

where h_1, h_2 denote the derivatives of $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively.

2.5.3 Proof of Lemma 2

Proof. To prove Lemma 2, I follow the proof of Dette *et al.* [9], and use Lemma 3, with $\mathcal{H}_1 = \eta^{-1}$ and $\mathcal{H}_2 = \eta_n^{-1}$. Then, with the use of a Taylor expansion for Q, it is clear to see that $\eta_n(t) - \eta(t) = \Phi(\eta_n^{-1}) - \Phi(\eta^{-1}) = Q(1) - Q(0) = Q'(\lambda^*)$ for some $\lambda^* \in [0,1]$ and

$$Q'(\lambda^*) = -\frac{\eta_n^{-1} - \eta^{-1}}{(\eta^{-1} + \lambda^*(\eta_n^{-1} - \eta^{-1}))'} \circ (\eta^{-1} + \lambda^*(\eta_n^{-1} - \eta^{-1}))^{-1}(t). \tag{2.78}$$

Using Lemma 1 it can be seen that $(\eta^{-1} + \lambda^*(\eta_n^{-1} - \eta^{-1})) \to \eta^{-1}$ as $n \to \infty$. In addition it is clear that setting $t_n = (\eta^{-1} + \lambda^*(\eta_n^{-1} - \eta^{-1}))^{-1}(t)$, then $t_n \to \eta(t)$. Next, by the use of a Taylor expansion, it can be shown that

$$(\eta_n^{-1} - \eta^{-1})(t_n) - (\eta_n^{-1} - \eta^{-1})(\eta(t)) = (\eta_n^{-1} - \eta^{-1})'(\xi_n)(t_n - \eta(t)) \tag{2.79}$$

where $t_n \leq \xi_n \leq \eta(t)$. Taking the first factor on the right hand side of (2.79) gives,

$$\begin{aligned}
(\eta_n^{-1} - \eta^{-1})'(\xi_n) &= \left(\frac{1}{nh_d}\int_\infty^{\xi_n} \sum_{i=1}^n K_d\left(\frac{\eta(i/n) - u}{h_d}\right) du\right)' - (\eta^{-1})'(\xi_n) \\
&= \frac{1}{nh_d}\sum_{i=1}^n K_d\left(\frac{\eta(i/n) - \xi_n}{h_d}\right) - (\eta^{-1})'(\xi_n) \\
&= \int_0^1 K_d\left(\frac{\eta(x) - \xi_n}{h_d}\right)\frac{dx}{h_d} + O\left(\frac{1}{nh_d}\right) - (\eta^{-1})'(\xi_n) \\
&= \int_{\frac{\eta(0)-\xi_n}{h_d}}^{\frac{\eta(1)-\xi_n}{h_d}} K_d(v)(\eta^{-1})'(\xi_n + vh_d)dv - (\eta^{-1})'(\xi_n) + O\left(\frac{1}{nh_d}\right) \\
&= \int_{\frac{\eta(0)-\xi_n}{h_d}}^{\frac{\eta(1)-\xi_n}{h_d}} K_d(v)(\eta^{-1})'(\xi_n)dv + \int_{\frac{\eta(0)-\xi_n}{h_d}}^{\frac{\eta(1)-\xi_n}{h_d}} K_d(v)(\eta^{-1})''(\xi_n)vh_d dv \\
&\quad + \int_{\frac{\eta(0)-\xi_n}{h_d}}^{\frac{\eta(1)-\xi_n}{h_d}} K_d(v)(\eta^{-1})^{(3)}(\tilde{\xi}_n)\frac{v^2h_d^2}{2}dv - (\eta^{-1})'(\xi_n) + O\left(\frac{1}{nh_d}\right) \\
&= \frac{(\eta^{-1})^{(3)}(\tilde{\xi}_n)h_d^2}{2}\int v^2 K_d(v)dv + O\left(\frac{1}{nh_d}\right) \\
&= O(h_d^2) + O\left(\frac{1}{nh_d}\right).
\end{aligned} \tag{2.80}$$

To obtain (2.80) I have used the substitution $v = (\eta(x) - \xi_n)/h_d$ and also the Taylor expansion,

$$(\eta^{-1})'(\xi_n + vh_d) = (\eta^{-1})'(\xi_n) + (\eta^{-1})''(\xi_n)vh_d + (\eta^{-1})^{(3)}(\tilde{\xi}_n)\frac{v^2h_d^2}{2} \tag{2.81}$$

for $\xi_n \leq \tilde{\xi}_n \leq \xi_n + vh_d$. From (2.80) and (2.79) it's clear that

$$(\eta_n^{-1} - \eta^{-1})(t_n) = (\eta_n^{-1} - \eta^{-1})(\eta(t)) + o(h_d^2) + o\left(\frac{1}{nh_d}\right), \tag{2.82}$$

and hence it is obvious that

$$\begin{aligned} Q'(\lambda^*) &= -\frac{(\eta_n^{-1} - \eta^{-1}) \circ \eta(t)}{(\eta^{-1})'(\eta(t))} + o(h_d^2) + o\left(\frac{1}{nh_d}\right) \\ &= -\frac{\kappa_2(K_d) h_d^2 (\eta^{-1})''(\eta(t)) + o(h_d^2) + O\left(\frac{1}{nh_d}\right)}{(\eta^{-1})'(\eta(t))} + o(h_d^2) + o\left(\frac{1}{nh_d}\right) \end{aligned} \tag{2.83}$$

where we have used Lemma 1. Lemma 2 follows from a rearrangement of (2.83), using the identity $(\eta^{-1})''(\eta(t)) = -\eta''(t)/\{\eta'(t)\}^3$. □

2.5.4 Proof of Theorem 2

Proof. The proof of this theorem follows similar lines to that of Lemma 2, and is also similar to the proof for Theorem 3.2 in Dette *et al.* [9]. The key differences for my method will be noted.

I begin by using Lemma 3, and setting $\mathcal{H}_2 = \hat{\eta}_M^{-1}$ and $\mathcal{H}_1 = \eta_n^{-1}$. Using a Taylor expansion it can be seen that

$$\mathcal{H}_2^{-1}(t) - \mathcal{H}_1^{-1}(t) = Q(1) - Q(0) = Q'(0) + \frac{1}{2}Q''(\lambda^*)$$

for some $\lambda^* \in [0, 1]$. This means that

$$\hat{\eta}_M(t) - \eta_n(t) = A_n + \frac{1}{2}B_n \tag{2.84}$$

where

$$A_n = -\frac{\hat{\eta}_M^{-1} - \eta_n^{-1}}{(\eta_n^{-1})'} \circ \eta_n(t) \tag{2.85}$$

$$\begin{aligned} B_n = &\frac{2(\hat{\eta}_M^{-1} - \eta_n^{-1})(\hat{\eta}_M^{-1} - \eta_n^{-1})'}{\{(\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1}))'\}^2} \circ (\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1}))^{-1}(t) \\ &- \frac{(\hat{\eta}_M^{-1} - \eta_n^{-1})^2(\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1}))''}{\{(\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1}))'\}^3} \circ (\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1}))^{-1}(t). \end{aligned} \tag{2.86}$$

First consider (2.85) and consider the Taylor expansion,

$$(\hat{\eta}_M^{-1} - \eta_n^{-1}) \circ \eta_n(t) - (\hat{\eta}_M^{-1} - \eta_n^{-1}) \circ \eta(t) = (\hat{\eta}_M^{-1} - \eta_n^{-1})'(\xi_n)(\eta_n(t) - \eta(t)), \tag{2.87}$$

where $|\xi_n - \eta(t)| \leq |\eta_n(t) - \eta(t)|$. Now consider (2.23) and using a Taylor expansion,

$$\begin{aligned} (\hat{\eta}_M^{-1} - \eta_n^{-1})'(\xi_n) &= \frac{1}{nh_d}\sum_{i=1}^{n}\left\{K_d\left(\frac{\hat{\eta}(i/n) - \xi_n}{h_d}\right) - K_d\left(\frac{\eta(i/n) - \xi_n}{h_d}\right)\right\} \\ &= \frac{1}{nh_d}\sum_{i=1}^{n}\left[K_d'\left(\frac{\eta(i/n) - \xi_n}{h_d}\right)\left\{\frac{\hat{\eta}\left(\frac{i}{n}\right) - \eta\left(\frac{i}{n}\right)}{h_d}\right\}\right. \\ &\quad \left. + K_d''\left(\frac{\zeta_{i,n} - \xi_n}{h_d}\right)\frac{\left\{\hat{\eta}\left(\frac{i}{n}\right) - \eta\left(\frac{i}{n}\right)\right\}^2}{2h_d^2}\right] \end{aligned} \tag{2.88}$$

where $|\zeta_{i,n} - \eta(i/n)| \leq |\hat{\eta}(i/n) - \eta(i/n)| = O(R_n)$, almost surely, where

$$R_n = \left(\frac{\log n}{nh_r}\right)^{1/2} + h_r^2. \tag{2.89}$$

This result comes from Claeskens and Van Keilegom [6] (Theorem 2.1). This is a uniform bound on the difference $|\hat{\eta}(x) - \eta(x)|$ for all values of x. It is clear that this is different to the bound, for $\hat{m}(t) - m(t)$ given in Dette *et al.* [9]

$$R_n = \left(\frac{\log h_r^{-1}}{n h_r} \right)^{1/2}. \tag{2.90}$$

The difference in our case is due to the use of likelihood information. It is clear from considering the bound (2.89) that the first term dominates the second due to assumption (B5). This means that the rate of convergence is mainly dependent on the first term. Comparing the first term of (2.89) with the bound of Dette *et al.* [9], given in (2.90, using assumption (B5) it can be seen that the only difference is that the bound in (2.89) depends on a term $\log n$ whilst the bound in (2.90) depends on a term $logn^{1/5}$. Notice that this second term can also be written as $1/5 \log n$ as so the rate of convergence is roughly the same. This goes to show that the two bounds convergence at the same rate and the main result is not affected.

Returning to (2.88) with the above Taylor expansion gives

$$\begin{aligned}(\hat{\eta}_M^{-1} - \eta_n^{-1})'(\xi_n) &= \frac{1}{nh_d} \sum_{i=1}^{n} \left[K_d' \left(\frac{\eta(i/n) - \xi_n}{h_d} \right) \left\{ \frac{\hat{\eta}\left(\frac{i}{n}\right) - \eta\left(\frac{i}{n}\right)}{h_d} \right\} + K_d'' \left(\frac{\zeta_{i,n} - \xi_n}{h_d} \right) \frac{\left\{ \hat{\eta}\left(\frac{i}{n}\right) - \eta\left(\frac{i}{n}\right) \right\}^2}{2h_d^2} \right] \\ &= \frac{1}{nh_d^2} \sum_{i=1}^{n} K_d' \left(\frac{\eta(i/n) - \xi_n}{h_d} \right) \left\{ \hat{\eta}\left(\frac{i}{n}\right) - \eta\left(\frac{i}{n}\right) \right\} + O\left(\frac{R_n^2}{h_d^3} \right) a.s. \end{aligned} \tag{2.91}$$

This can be seen by considering the term

$$\begin{aligned} & \left| \frac{1}{2nh_d^3} \sum_{i=1}^{n} K_d'' \left(\frac{\zeta_{i,n} - \xi_n}{h_d} \right) \left\{ \hat{\eta}\left(\frac{i}{n}\right) - \eta\left(\frac{i}{n}\right) \right\}^2 \right| \\ \leq & \frac{1}{2h_d^3} R_n^2 \frac{1}{n} \sum_{i=1}^{n} \left| K_d'' \left(\frac{\zeta_{i,n} - \xi_n}{h_d} \right) \right| \\ = & O\left(\frac{R_n^2}{h_d^2} \right). \end{aligned}$$

To proceed I state a further Taylor expansion,

$$K_d' \left(\frac{\eta(x) - \eta_n(t)}{h_d} \right) = K_d' \left(\frac{\eta(x) - \eta(t)}{h_d} \right) + K_d'' \left(\frac{\eta(x) - \psi_t}{h_d} \right) \frac{\eta(t) - \eta_n(t)}{h_d} \tag{2.92}$$

for some $\eta(t) < \psi_t < \eta_n(t)$, and then returning to (2.91) and using a Riemann Integral it can

be seen that

$$\begin{aligned}
\left|(\hat{\eta}_M^{-1}-\eta_n^{-1})'(\xi_n)\right| &\leq \frac{1}{h_d^2}\int\left|K_d'\left(\frac{\eta(x)-\eta_n(t)}{h_d}\right)\{\hat{\eta}(x)-\eta(x)\}\right|dx(1+O(n^{-1}))+O\left(\frac{R_n^2}{h_d^2}\right)\\
&= \frac{1}{h_d^2}\int\left|\left\{K_d'\left(\frac{\eta(x)-\eta(t)}{h_d}\right)+K_d''\left(\frac{\eta(x)-\psi_t}{h_d}\right)\frac{\eta(t)-\eta_n(t)}{h_d}\right\}\right|\\
&\quad\times|\{\hat{\eta}(x)-\eta(x)\}|\,dx(1+O(n^{-1}))+O\left(\frac{R_n^2}{h_d^2}\right)\\
&= \frac{1}{h_d^2}\int\left|K_d'\left(\frac{\eta(x)-\eta(t)}{h_d}\right)\{\hat{\eta}(x)-\eta(x)\}\right|dx(1+O(n^{-1}))\\
&\quad+O\left(R_n+\frac{R_n^2}{h_d^2}\right)a.s
\end{aligned}\tag{2.93}$$

since

$$\begin{aligned}
&\frac{1}{h_d^2}\int\left|K_d''\left(\frac{\eta(x)-\psi_t}{h_d}\right)\frac{\eta(t)-\eta_n(t)}{h_d}\{\hat{\eta}(x)-\eta(x)\}\right|dx\\
&=\left|\frac{\eta(t)-\eta_n(t)}{h_d^3}\right|\int\left|K_d''\left(\frac{\eta(x)-\psi_t}{h_d}\right)\{\hat{\eta}(x)-\eta(x)\}\right|dx\\
&=\left|\frac{\eta(t)-\eta_n(t)}{h_d^2}\right|\int\left|K_d''(u)\{\hat{\eta}(\eta^{-1}(\psi_t+uh_d))-\eta(\eta^{-1}(\psi_t+uh_d))\}(\eta^{-1})'(\psi_t+uh_d)\right|du\\
&=\left|\frac{\eta(t)-\eta_n(t)}{h_d^2}\right|\int\left|K_d''(u)\{\hat{\eta}(\eta^{-1}(\psi_t))-\eta(\eta^{-1}(\psi_t))\}(\eta^{-1})'(\psi_t)\right|du\{1+o(1)\}\\
&=\left|\frac{(\eta(t)-\eta_n(t))\{\hat{\eta}(\eta^{-1}(\psi_t))-\eta(\eta^{-1}(\psi_t))\}(\eta^{-1})'(\psi_t)}{h_d^2}\right|\int|K_d''(u)|\,du\\
&\leq\frac{R_n(\eta(t)-\eta_n(t))(\eta^{-1})'(\psi_t)}{h_d^2}\int|K_d''(u)|\,du\\
&=O(R_n)
\end{aligned}$$

due to the fact that $\eta(t)-\eta_n(t)=O(h_d^2)$ by Lemma 2. In the above I have used the substitution $u=(\eta(x)-\psi_t)/h_d$. The next step is to return to (2.93) and consider a substitution $v=$

$(\eta(x) - \eta(t))/h_d$.

$$
\begin{aligned}
|(\hat{\eta}_M^{-1} - \eta_n^{-1})'(\xi_n)| &= \frac{1}{h_d} \int \Big| K_d'(v)\{\hat{\eta}(\eta^{-1}(\eta(t) + vh_d)) - \eta(\eta^{-1}(\eta(t) + vh_d))\} \\
&\quad (\eta^{-1})'(\eta(t) + vh_d) \Big| \, dv\{1 + O(n^{-1})\} \; + O\left(R_n + \frac{R_n^2}{h_d^2}\right) \\
&= \frac{1}{h_d} \int \Big| K_d'(v)\{\hat{\eta}(\eta^{-1}(\eta(t))) - \eta(\eta^{-1}(\eta(t)))\}(\eta^{-1})'(\eta(t)) \Big| \, dv\{1 + o(1)\} \\
&\quad + O\left(R_n + \frac{R_n^2}{h_d^2} + \frac{1}{nh_d}\right) \\
&\leq \frac{R_n(\eta^{-1})'(\eta(t))}{h_d} \int |K_d'(v)| \, dv\{1 + o(1)\} + O\left(R_n + \frac{R_n^2}{h_d^3} + \frac{1}{nh_d}\right) \\
&= O\left(\frac{R_n}{h_d} + \frac{R_n^2}{h_d^2} + \frac{1}{nh_d}\right) a.s. \qquad (2.94)
\end{aligned}
$$

The $O(1/nh_r)$ term in the second line comes from the $h_d^{-1}O(n^{-1})$ in the previous line. Now recalling Lemma 2 and returning to (2.87) we have

$$
(\hat{\eta}_M^{-1} - \eta_n^{-1}) \circ \eta_n(t) - (\hat{\eta}_M^{-1} - \eta_n^{-1}) \circ \eta(t) = O\left(R_n h_d + R_n^2 + \frac{h_d}{n}\right). \qquad (2.95)
$$

In the case, $\lim_{n\to\infty} h_r/h_d = c \in [0, \infty)$, it can be shown that this term is $o(1/\sqrt{nh_d})$ almost surely as follows, using assumption (B4), (B5), and then assumption (C5) for (2.97), and (C2) for (2.98).

$$
\frac{\sqrt{nh_d}h_d}{n} = \sqrt{\frac{h_d^3}{n}} = o(1) \qquad (2.96)
$$

$$
\begin{aligned}
\sqrt{nh_d}R_n h_d &= \sqrt{nh_d}\left\{\left(\frac{\log n}{nh_r}\right)^{1/2} + h_r^2\right\} h_d \\
&= \sqrt{h_d^3 \frac{\log n}{h_r}} + \sqrt{nh_d^3 h_r^4} \qquad = o(1) \qquad (2.97)
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{nh_d}R_n^2 &= \sqrt{nh_d}\left\{\left(\frac{\log n}{nh_r}\right)^{1/2} + h_r^2\right\}^2 \\
&= \sqrt{nh_d}\left(\frac{\log n}{nh_r} + 2h_r^2\left(\frac{\log n}{nh_r}\right)^{1/2} + h_r^4\right) \\
&= \sqrt{nh_d}\left(\frac{\log n}{nh_r}\right) + \sqrt{h_r^3 h_d \log n} + \sqrt{nh_d h_r^8} \\
&= o(1). \qquad (2.98)
\end{aligned}
$$

Using lemma 1, (2.85) and (2.95), it is clear to see that

$$A_n = -\frac{\hat{\eta}_M^{-1} - \eta_n^{-1}}{(\eta^{-1})'} \circ \eta(t) + o_p\left(\frac{1}{\sqrt{nh_d}}\right). \tag{2.99}$$

In the case where $\lim_{n\to\infty} h_r/h_d = \infty$, (2.95) is of order $o(1/\sqrt{nh_r})$ by calculations which are analagous to those presented above, and also use assumptions (B4) and (B5). Assumption (C4) is required for (2.102) and (C6) is required for (2.101). Here it is the case that

$$\frac{\sqrt{nh_r}h_d}{n} = \sqrt{\frac{h_r h_d^2}{n}} = o(1) \tag{2.100}$$

$$\begin{aligned}
\sqrt{nh_d}R_n h_d &= \sqrt{nh_r}\left(\left(\frac{\log n}{nh_r}\right)^{1/2} + h_r^2\right) h_d \\
&= \sqrt{h_d^2 \log n} + \sqrt{nh_r^5 h_d^2} \\
&= o(1)
\end{aligned} \tag{2.101}$$

$$\begin{aligned}
\sqrt{nh_r}R_n^2 &= \sqrt{nh_r}\left(\frac{\log n}{nh_r} + 2h_r^2\left(\frac{\log n}{nh_r}\right)^{1/2} + h_r^4\right) \\
&= \frac{\log n}{\sqrt{nh_r}} + \sqrt{h_r^4 \log n} + \sqrt{nh_r^9} \\
&= o(1).
\end{aligned}$$

Combining these results gives

$$A_n = -\frac{\hat{\eta}_M^{-1} - \eta_n^{-1}}{(\eta^{-1})'} \circ \eta(t) + o_p\left(\frac{1}{\sqrt{nh_r}}\right). \tag{2.102}$$

I now need to consider the estimate B_n and, following Dette *et al.* want to show that it is $o_p(1/\sqrt{nh_d})$, or $o_p\left(\frac{1}{\sqrt{nh_r}}\right)$, in the case where $\lim_{n\to\infty} h_r/h_d = \infty$. The calculations for this run in a similar way to those I have just shown for A_n and so I just show the main steps. First notice that $B_n = 2B_{n1} - B_{n2}$, where

$$B_{n1} = \frac{(\hat{\eta}_M^{-1} - \eta_n^{-1})(\hat{\eta}_M^{-1} - \eta_n^{-1})'(t_n)}{\{(\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1}))'\}^2(t_n)} \tag{2.103}$$

$$B_{n2} = \frac{(\hat{\eta}_M^{-1} - \eta_n^{-1})^2(\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1}))''(t_n)}{\{(\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1}))'\}^3(t_n)} \tag{2.104}$$

where $t_n = (\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1}))^{-1}(t)$. Notice that $t_n \to \eta(t)$, $(\eta_n^{-1} + \lambda^*(\hat{\eta}_M^{-1} - \eta_n^{-1})) \xrightarrow{P} \eta^{-1}$. Recall that

$$\begin{aligned}
\Delta_n'(t_n) &= (\hat{\eta}_M^{-1} - \eta_n^{-1})'(t_n) \\
&= \frac{1}{nh_d} \sum_{i=1}^{n} \left\{ K_d \left(\frac{\hat{\eta}(i/n) - t_n}{h_d} \right) - K_d \left(\frac{\eta(i/n) - t_n}{h_d} \right) \right\} \\
&= \frac{1}{nh_d} \sum_{i=1}^{n} \left\{ K_d' \left(\frac{\eta(i/n) - t_n}{h_d} \right) \frac{\{\hat{\eta}(i/n) - \eta(i/n)\}}{h_d} \right. \\
&\quad \left. + K_d'' \left(\frac{\Lambda_{i,n} - t_n}{h_d} \right) \frac{\{\hat{\eta}(i/n) - \eta(i/n)\}^2}{2h_d^2} \right\} \qquad (2.105)
\end{aligned}$$

for some $|\Lambda_{i,n} - \eta(i/n)| \leq |\hat{\eta}(i/n) - \eta(i/n)| = O(R_n)$. Considering first the second term it is clear that

$$\begin{aligned}
\left| \frac{1}{nh_d^3} \sum_{i=1}^{n} K_d'' \left(\frac{\Psi_{i,n} - t_n}{h_d} \right) \{\hat{\eta}(i/n) - \eta(i/n)\}^2 \right| &\leq \frac{1}{nh_d^3} R_n^2 \sum_{i=1}^{n} \left| K_d'' \left(\frac{\Psi_{i,n} - t_n}{h_d} \right) \right| \\
&= O\left(\frac{R_n^2}{h_d^2} \right). \qquad (2.106)
\end{aligned}$$

Returning to (2.105) gives

$$\begin{aligned}
|\Delta_n'(t_n)| &= \left| \frac{1}{nh_d^2} \sum_{i=1}^{n} K_d' \left(\frac{\eta(i/n) - t_n}{h_d} \right) \{\hat{\eta}(i/n) - \eta(i/n)\} + O\left(\frac{R_n^2}{h_d^2} \right) \right| \\
&= \left| \frac{1}{h_d^2} \int K_d' \left(\frac{\eta(x) - \eta(t)}{h_d} \right) \{\hat{\eta}(x) - \eta(x)\} dx (1 + O(n^{-1})) \right| + O\left(\frac{R_n^2}{h_d^2} \right) \\
&= \left| \frac{1}{h_d} \int K_d'(u) \{\hat{\eta}(\eta^{-1}(\eta(t) + uh_d)) - \eta(\eta^{-1}(\eta(t) + uh_d))\} (\eta^{-1})'(\eta(t) + uh_d) du (1 + o(1)) \right| \\
&\quad + O\left(\frac{R_n^2}{h_d^2} \right) \\
&= \left| \frac{\{\hat{\eta}(t) - \eta(t)\}(\eta^{-1})'(\eta(t))}{h_d} \int K_d'(u) du (1 + o(1)) \right| + O\left(\frac{R_n^2}{h_d^2} \right) \\
&\leq \left| \frac{R_n (\eta^{-1})'(\eta(t))}{h_d} \int K_d'(u) du (1 + o(1)) \right| + O\left(\frac{R_n^2}{h_d^2} \right) \\
&= O\left(\frac{R_n}{h_d} \right) + O\left(\frac{R_n^2}{h_d^2} \right) \\
&= O\left(\frac{R_n}{h_d} + \frac{R_n^2}{h_d^2} \right). \qquad (2.107)
\end{aligned}$$

Now, consider Theorem 1. Since the bias of $\hat{\eta}(\cdot) - \eta(\cdot)$ is of order $O(h_r^2) = o(1)$ and its variance is of order $O(1)$ it is clear that $(\hat{\eta}_M^{-1} - \eta_n^{-1})(t_n) = O(\sqrt{nh_d})$. Thus we have

$$B_{n1} = O\left(\frac{1}{\sqrt{nh_d}} \left(\frac{R_n}{h_d} + \frac{R_n^2}{h_d^2} \right) \right) = o\left(\frac{1}{\sqrt{nh_d}} \right). \qquad (2.108)$$

which can be shown using assumptions (C2) and (C3) as follows

$$\begin{aligned}\frac{R_n}{h_d} &= \frac{1}{h_d}\left\{\left(\frac{\log n}{nh_r}\right)^{1/2} + h_r^2\right\} \\ &= \sqrt{\frac{\log n}{nh_r h_d^2}} + \frac{h_r^2}{h_d} \\ &= o(1).\end{aligned}$$

In the case where $\lim_{n\to\infty} h_r/h_d = \infty$, it can be shown that $B_{n1} = o(1/\sqrt{nh_r})$ in exactly the same way.

The next step is to consider the term B_{n2}. From the above discussion it can be seen that

$$\begin{aligned}B_{n2} &\to (\hat{\eta}_M^{-1} - \eta_n^{-1})^2(t_n)\frac{(\eta^{-1})''(\eta(t))}{(\eta^{-1})'^3(\eta(t))} \\ &= O\left(\frac{1}{nh_d}\right).\end{aligned}$$

I am now in a position to finish the proof of the theorem. Recall Theorem 1 and (2.84). Theorem 1 implies that when $\lim_{n\to\infty} h_r/h_d = c \in [0,\infty)$,

$$\sqrt{nh_d}\left(\hat{\eta}_M(\eta(t)) - \eta_n^{-1}(\eta(t)) + \kappa_2(K_r)h_r^2\frac{\eta''(t)}{\eta'(t)}\right) \xrightarrow{D} N(0, r^2(\eta(t))) \tag{2.109}$$

From here it is clear that

$$\begin{aligned}-\eta'(t)\sqrt{nh_d}\left(\hat{\eta}_M^{-1}(\eta(t)) - \eta_n^{-1}(\eta(t)) + \kappa_2(K_r)h_r^2\frac{\eta''(t)}{\eta'(t)}\right) &\xrightarrow{D} N(0,(\eta'(t))^2 r^2(\eta(t))) \\ -\sqrt{nh_d}\left(\frac{\hat{\eta}_M^{-1}(\eta(t)) - \eta_n^{-1}(\eta(t)) + \kappa_2(K_r)h_r^2\frac{\eta''(t)}{\eta'(t)}}{(\eta^{-1})'(\eta(t))}\right) &\xrightarrow{D} N(0,(\eta'(t))^2 r^2(\eta(t))) \\ \sqrt{nh_d}\left(-\frac{\hat{\eta}_M^{-1} - \eta_n^{-1}}{(\eta^{-1})'} \circ (\eta(t)) - \kappa_2(K_r)h_r^2\eta''(t)\right) &\xrightarrow{D} N(0,(\eta'(t))^2 r^2(\eta(t))) \\ \sqrt{nh_d}\left(\hat{\eta}_M(t) - \eta_n(t) - \kappa_2(K_r)h_r^2\eta''(t)\right) &\xrightarrow{D} N(0,(\eta'(t))^2 r^2(\eta(t)))\end{aligned} \tag{2.110}$$

which gives Theorem 2 as required. In the alternative case, where $\lim_{n\to\infty} h_r/h_d = \infty$, the proof is analogous. □

Chapter 3

Choosing the bandwidths for the LDNP method

3.1 Introduction

The bandwidths chosen for nonparametric regression estimates have a significant effect on the performance of an estimate. In an ideal situation we would choose a bandwidth, h_r that was optimal in terms of some metric such as the MSE (see (3.3)), or the MISE (see (3.10)). However, the expression for the asymptotic MSE and MISE, for estimating $\eta(x)$, depends on quantities which are unknown in practice since the precise form of the response function is not known. As a result, approximation methods are needed in order to estimate the unknown quantities and calculate a suitable bandwidth. There are many ways in which to choose the bandwidth, with the two main families of methods being cross-validation bandwidths and plug-in bandwidths. These methods are briefly described in Chapter 1.

For the LDNP method, clearly two bandwidths must be chosen, one for the unconstrained regression step, h_r and a second for the density estimation step h_d. Dette *et al.* [9] state that any standard method may be used to choose the bandwidth for the initial regression and then explore the effect of the choice of h_d. The choice of bandwidth for the unconstrained regression, h_r can affect the resulting estimate. This choice is explored in detail in a simulation study

in Chapter 4. In this chapter, I concentrate on studying the effect of the choice of a second bandwidth, h_d, to be used in the density estimation procedure in the monotonisation step.

3.2 MSE optimal bandwidth

In this section, I will concentrate on MSE optimal bandwidth and assume that it varies with location t in a similar way as Dette *et al.* [9]. For the regression estimate I will use the bandwidth that minimises the Mean Square Error (MSE). For a quasi-likelihood method as described in Fan and Gijbels ([16] pages 196-198) the bias and the variance of the estimator $\hat{\eta}$ are given by (3.1) and (3.2) respectively.

$$b(x) = \kappa_2(K_r)\eta''(x)h_r^2 \tag{3.1}$$

where $\kappa_2(K_r) = (1/2)\int z^2 K_r(z)dz$.

$$\sigma^2(x) = \int K_r^2(z)dz \frac{[g'\{\mu(x)\}]^2 Var[Y|X=x]}{f(x)nh_r}. \tag{3.2}$$

The MSE is the sum of the square bias and the variance, and hence a bandwidth h_r that minimises the MSE can be calculated resulting in the following asymptotically MSE optimal bandwidth

$$h_{r,OPT} = \left[\frac{\int K_r^2(z)dz[g'\{\mu(x)\}]^2 Var[Y|X=x]}{4nf(x)\kappa_2^2(K_r)(\eta''(x))^2}\right]^{1/5}. \tag{3.3}$$

From Chapter 2 it is clear that for the LDNP method the leading terms of the bias and the variance can be obtained from Theorem 2 and Corollary 1. The leading term of the bias is

$$Bias(t) = \kappa_2(K_d)h_d^2\frac{\eta''(t)}{(\eta'(t))^2} + \kappa_2(K_r)\eta''(t)h_r^2 \tag{3.4}$$

and the variance is given by

$$\begin{aligned} Var(t) &= \frac{\eta'(t)[g'\{\mu(t)\}]^2 Var[Y|X=t]}{nh_d f(t)} \int\int\int K_d\left(u + \frac{h_r}{h_d}\eta'(t)(v-w)\right) \\ &= \times K_d(u)K_r(w)K_r(v)dudwdv. \end{aligned}$$

Equations (3.4) and (3.5) are true in the case where $\lim_{n\to\infty} h_r/h_d = c \in [0,\infty)$. In this section I will explore the relationship of h_d to h_r. Following Dette *et al.* [9], I will set

$h_d = \gamma\eta'(t)h_r$ for some constant $\gamma > 0$. Using this substitution it is evident that the leading term of the bias now becomes

$$Bias(t) = \eta''(t)h_r^2 \left[\kappa_2(K_r) + \gamma^2\kappa_2(K_d)\right]. \tag{3.5}$$

The variance term becomes

$$Var(t) = \frac{[g'\{\mu(t)\}]^2 Var[Y|X=t]}{n\gamma h_r f(t)} \int\int\int K_d\left(u + \frac{v-w}{\gamma}\right) K_d(u)K_r(w)K_r(v)dudwdv. \tag{3.6}$$

Making the substitution $s = (v-w)/\gamma$ gives

$$\begin{aligned} Var(t) &= \frac{[g'\{\mu(t)\}]^2 Var[Y|X=t]}{nh_r f(t)} \int \left(\int K_d(u+s)K_d(u)du\right)\left(\int K_r(w)K_r(w+s\gamma)dw\right)ds \\ &= \frac{[g'\{\mu(t)\}]^2 Var[Y|X=t]}{nh_r f(t)} D_K^2(\gamma). \end{aligned} \tag{3.7}$$

The second part of Theorem 2 considers the case where $\lim_{n\to\infty} h_r/h_d = \infty$. We can include this by assuming this is what is meant by allowing $\gamma = 0$. In this case it can be seen that when $\gamma = 0$, the integrals in $D_K^2(\gamma)$ simplify and $D_K^2(0) = \int K_r^2(w)dw$.

It is clear that the bias term, (3.5) is increasing with γ. It is less clear how the variance behaves with increasing γ. To investigate this I have considered some commonly used kernels which are listed, along with several key properties in Table 3.1. A plot of these kernels is shown in Figure 3.1.

For each of these six kernels I have calculated the function $D_K^2(\gamma)$ for a range of γ values. I have here assumed that the two kernel used are the same, i.e. $K_d = K_r$. I then plotted these functions in Figure 3.2. It is clear to see that the function D_K^2 is decreasing with γ for these kernels. This implies that the variance of $\hat{\eta}$ decreases with γ, for the kernels listed here.

These investigations have shown that the bias increases with increasing γ whilst the variance decreases with increasing γ. The question that now needs to be considered is how much the choice of γ affects the performance of the estimate. I can consider this by thinking about the value of the asymptotic MSE function when an MSE optimal bandwidth, (3.3), is chosen for the unconstrained regression. In this case a first order approximation of the asymptotic MSE

Table 3.1: **Some commonly used kernels**

Kernel	$K(z)$	$\int z^2K(z)dz$	$\int K^2(z)dz$
Uniform	$\frac{1}{2}\mathbf{1}_{\lvert z\rvert\leq 1}$	$\frac{1}{3}$	$\frac{1}{2}$
Epanechnikov	$\frac{3}{4}(1-z^2)\mathbf{1}_{\lvert z\rvert\leq 1}$	$\frac{1}{5}$	$\frac{3}{5}$
Biweight	$\frac{15}{16}(1-u^2)^2\mathbf{1}_{\lvert z\rvert\leq 1}$	$\frac{1}{7}$	$\frac{5}{7}$
Triangular	$(1-\lvert z\rvert)\mathbf{1}_{\lvert z\rvert\leq 1}$	$\frac{1}{6}$	$\frac{2}{3}$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-z^2/2)$	1	$\frac{1}{2\sqrt{\pi}}$
Tricube	$\frac{70}{81}(1-\lvert z\rvert^3)^3\mathbf{1}_{\lvert z\rvert\leq 1}$	$\frac{35}{243}$	$\frac{175}{247}$

is

$$\begin{aligned}
MSE(\gamma) &= \left(\eta''(t)h_r^2\left[\kappa_2(K_r)+\gamma^2\kappa_2(K_d)\right]\right)^2 + \frac{[g'\{\mu(t)\}]^2Var[Y|X=t]}{nh_rf(t)}D_K^2(\gamma)\\
&= (\eta''(t))^2\left[\kappa_2(K_r)+\gamma^2\kappa_2(K_d)\right]^2\left\{\frac{D_K^2(0)[g'\{\mu(x)\}]^2Var[Y|X=x]}{4nf(x)\kappa_2^2(K_r)(\eta''(x))^2}\right\}^{4/5}\\
&\quad\times\frac{[g'\{\mu(t)\}]^2Var[Y|X=t]}{nf(t)}D_K^2(\gamma)\left\{\frac{D_K^2(0)[g'\{\mu(x)\}]^2Var[Y|X=x]}{4nf(x)\kappa_2^2(K_r)(\eta''(x))^2}\right\}^{-1/5}\\
&= \left[\frac{D_K^2(0)[g'\{\mu(x)\}]^2Var[Y|X=x]}{4nf(t)}\right]^{4/5}\left\{\eta''(t)\kappa_2(K_r)\right\}^{2/5}\\
&\quad\times\left[\left(\frac{\kappa_2(K_d)}{\kappa_2(K_r)}\gamma^2+1\right)^2+\frac{4D_K^2(\gamma)}{D_K^2(0)}\right].
\end{aligned}\tag{3.8}$$

Following Dette *et al.* [9] I consider the ratio of the MSE for the monotone estimate and the MSE of the unconstrained estimate. The purpose of this is to see how well the monotone estimate performs in comparison to the unconstrained version. Setting $\gamma = 0$ would imply that $h_d = 0$. This is equivalent to not doing the monotonisation step and so $MSE(0)$ can be considered as the MSE of the unconstrained estimate. Consider the function

$$e(\gamma) = \frac{MSE(\gamma)}{MSE(0)} = \frac{(1+\gamma^2\kappa_2(K_d)/\kappa_2(K_r))^2 + (4/D_K^2(0))D_K^2(\gamma)}{5}.\tag{3.9}$$

In the case where $K_d = K_r$, the effect of γ on $e(\gamma)$ is shown in Figure 3.3 for the kernels considered in Table 3.1. It is not easy to distinguish all of the six kernels in this plot due

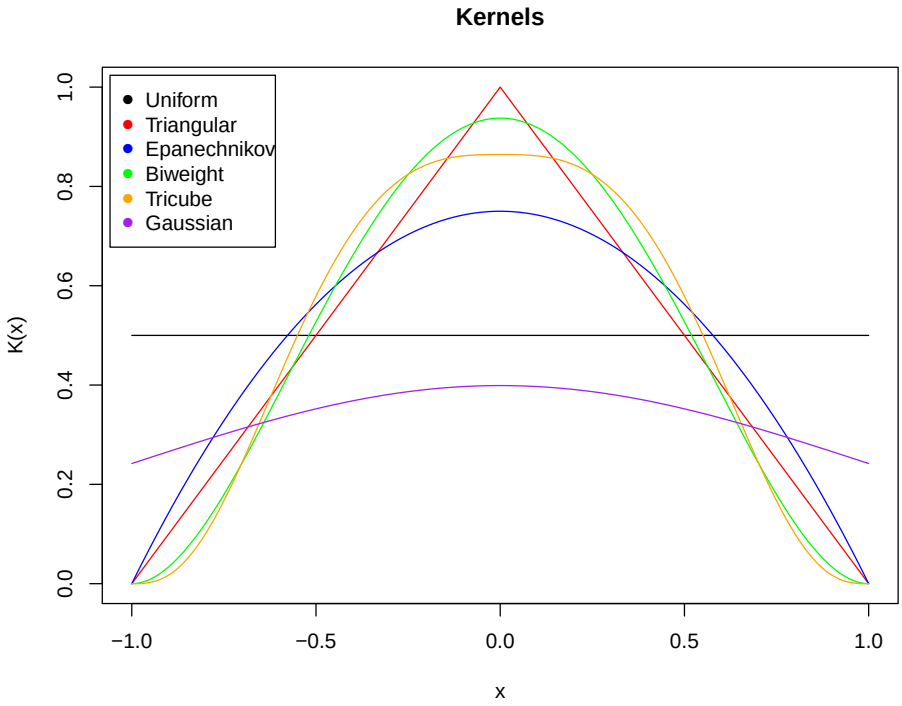

Figure 3.1: Some commonly used kernels

to similarities in the results, but the overall pattern is plain to see. It is clear that for all kernels except the uniform kernel, the value of $e(\gamma)$ is minimal when $\gamma = 0$. This suggests that an optimal choice of bandwidth, h_d, would be one for which $\lim_{n\to\infty} h_r/h_d = \infty$. From an asymptotic point of view then, a bandwidth which satisfies $h_d = o(h_r)$ is desirable for the monotone estimate. In the case of the uniform kernel the minimum occurs at approximately $\gamma = 0.306$. For this kernel then, it is asymptotically optimal, in terms of the MSE, to choose a bandwidth h_d of the same order as that of the regression bandwidth h_r. These results are in agreement with those in Dette *et al.* [9] for the DNP method.

3.3 MISE optimal bandwidth

In the previous section I have followed the analysis of Dette *et al.* [9] and considered a choice of bandwidth for the monotonisation step of the form $h_d = \gamma\eta'(t)h_r$. In this case the MSE optimal

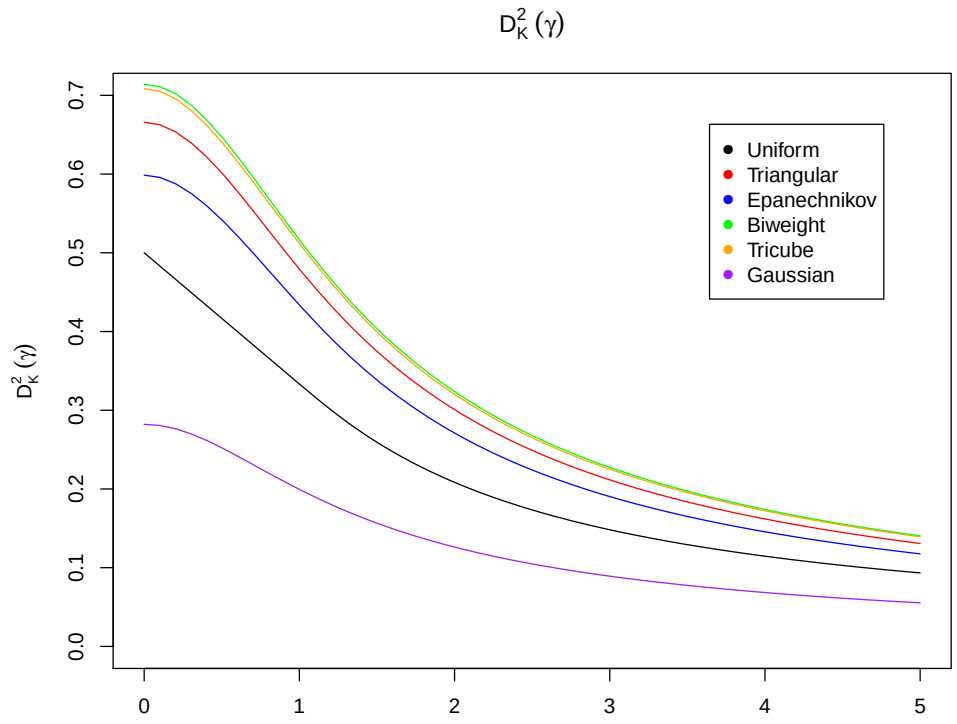

Figure 3.2: The effect of γ on $D_K^2(\gamma)$

bandwdith was used for my analysis. It should be noticed that in this case the bandwidth changes with t. In practice, a global bandwidth is often used for the monotonisation step. I want to explore in this section how γ should be chosen in the case where a global bandwidth is used instead of a variable bandwidth. In this case it may be more appropriate to use a mean integrated square error (MISE) optimal bandwidth for the unconstrained regression h_r, since this is a global measure of performance.

Now I use a bandwidth of the form $h_d = \gamma h_r$, with a similar understanding of the case when $\gamma = 0$ to that of section 3.2. The bandwidth that minimises the asypmtotic MISE is given by Fan and Gijbels ([16] page 198)

$$h_{r,opt} = \left[\frac{D_{K_1}^2(0) \int [g'(g^{-1}(\eta(x)))]^2 Var[Y|X=x] w(x) dx}{4n\kappa_2^2(K_r) \int (\eta''(x))^2 f(x) w(x) dx} \right]^{1/5}. \tag{3.10}$$

In this equation there are obviously a number of unknown terms. A link function g must be specified, as must a variance term $Var[Y|X=x]$. The function f describes the design density

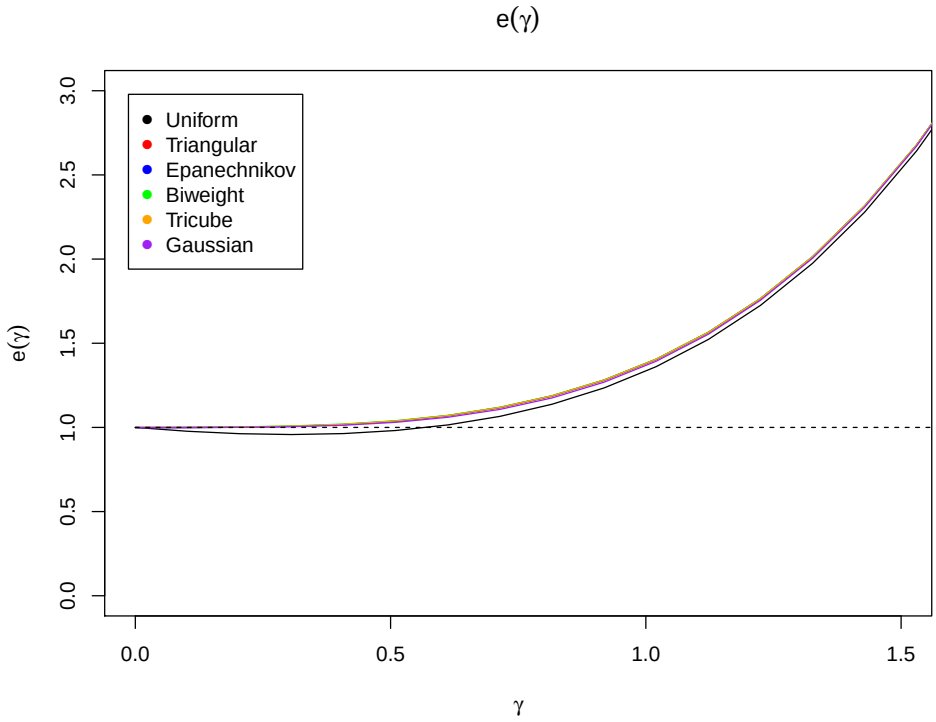

Figure 3.3: The function $e(\gamma)$ for the six kernels in Table 3.1 $K_d = K_r$.

of the stimulus levels, w is a weighting function. In practice, plug in values of the $\eta''(x)$ would need to be provided in order to estimate the asymptotically optimal bandwidth. This optimal bandwidth is for the regression step is then used to determine the value of γ that would give the best MISE performance. As I show in what follows the optimum value would be situation dependent and will require plug in values of a number of unknown values. This limits the usefulness of the approach. It is included here with a number of known functions so as to indicate the relationship between h_d and h_r. Here $D^2_{K_1}$ is similar, though not identical to the value of D^2_K in the previous section. This is due to the different relationship between h_d and h_r. In this case this constant depends upon t.

$$D^2_{K_1}(\gamma, t) = \int\int\int K_d(u+s)K_d(u)K_r(w)K_r\left(w + \frac{s\gamma}{\eta'(t)}\right) dudwds. \tag{3.11}$$

I then use the bias and variance expressions (3.4) and (3.5) and calculate an expression for the asymptotic MISE when using the optimal bandwidth (3.10).

$$
\begin{aligned}
MISE(\gamma) &= \int (Bias^2 + Var) w(t) f(t) dt \\
&= h_{r,opt}^4 \int (\eta''(t))^2 \left[\frac{\kappa_2(K_d)\gamma^2}{\eta'(t)} + \kappa_2(K_r) \right]^2 f(t) w(t) dt \\
&\quad + \frac{1}{n h_{r,opt}} \int [g'(g^{-1}(\eta(t)))]^2 Var[Y|X=t] D_{K_1}^2(\gamma, t) w(t) dt. \qquad (3.12)
\end{aligned}
$$

As in the previous section I will consider the ratio of the MISE for the monotonised version and the unconstrained version. I denote by $MISE(0)$ the MISE for the unconstrained regression, and then, following simplification of the expression I consider the function

$$
\begin{aligned}
e_1(\gamma) &= \frac{MISE(\gamma)}{MISE(0)} \\
&= \frac{4}{5} \left[\frac{\int (\eta''(t))^2 \left(\frac{\gamma^2}{\eta'(t)} + 1 \right)^2 w(t) f(t) dt}{4 \int (\eta''(t))^2 f(t) w(t) dt} + \frac{\int [g'(g^{-1}(\eta(t)))]^2 Var[Y|X=t] D_{K_1}^2(\gamma, t) w(t) dt}{D_{K_1}^2(0) \int [g'(g^{-1}(\eta(t)))]^2 Var[Y|X=t] w(t) dt} \right] \\
&= \frac{4}{5} \left[E_1(\gamma) + E_2(\gamma) \right] \qquad (3.13)
\end{aligned}
$$

Clearly now the situation differs from that of the previous case. When I considered the function $e(\gamma)$ there were no unknown terms in the expression, whereas now, for $e_1(\gamma)$ there are. The main consequence of this is that $e_1(\gamma)$ depends on the (unknown) function η, and potentially, the efficiency $e_1(\gamma)$ varies with γ in different ways for different functions η. For example for some functions it might be decreasing, but for others it might increase or have a minimum/maximum.

The term of $E_1(\gamma)$ in 3.13 is an increasing function of γ since it is a polynomial of order 4 with positive coefficients. For a fixed value of t, $D_{K_1}^2(\gamma, t) = D_K^2(\tilde{\gamma})$, where $\tilde{\gamma} = \gamma/\eta'(t)$. In Figure 3.2 I showed that for commonly used kernels, $D_K^2(\gamma)$ was a decreasing function, and hence, for fixed t, $D_{K_1}^2(\gamma, t)$ is a decreasing function of γ. However, this does not tell us how the function $E_2(\gamma)$ behaves with changing γ, and its behaviour might be different for different functions η. Therefore, in the following section I will present some simulation studies to investigate the effect of the choice of bandwidth h_d on the performance of the LDNP estimator.

I will focus the investigations on psychometric functions only in this section. I will use six functions as a basis for my investigations. These functions will be considered again in Chapter 4

as part of a more extensive simulation study of the monotone estimation of psychometric functions. The functions are discussed in more detail in Section 5.2. The functions are

$$
\begin{aligned}
p_1(x) &= (1+\exp(5-15x))^{-1} \\
p_2(x) &= (1+\exp(-(20x^2+1)))^{-1} \\
p_3(x) &= \Phi\left(\frac{x-\mu}{\sigma}\right), && \mu = .5, \sigma = .5 \\
p_4(x) &= \Phi\left(\frac{x-\mu}{\sigma}\right), && \mu = .5, \sigma = .1 \\
p_5(x) &= 1-\exp(-x^{\gamma}), && \gamma = .52876 \\
p_6(x) &= \eta\Phi\left(\frac{x-\mu_1}{\tau}\right) + (1-\eta)\Phi\left(\frac{x-\mu_2}{\tau}\right) \\
&\qquad \mu_1 = 0.4, \mu_2 = 1, \eta = .64946, \tau = .13546
\end{aligned}
$$

I will not consider the function $p_1(x)$ in this analysis since for this function $\eta''(x) = 0$ and hence the expression for bias in 3.12 is zero. Notice that this doesn't necessary mean that the estimator is unbiased.

For a psychometric function the canonical link, g is the logistic link and thus η is defined as

$$\eta(x) = log\left(\frac{p(x)}{1-p(x)}\right).$$

For the purpose of this study I assume that Y is Bernoulli distributed and hence $Var[Y|X = x] = p(x)(1-p(x))$, and now have all the information I need to be able to calculate the function $e_1(\gamma)$. I will define $f = w \equiv 1$. If a canonical link is used then $[g'(g^{-1}(\eta(t)))] = 1/Var[Y|X = t]$ which means that $E_2(\gamma)$ can be simplified to

$$\frac{\int[g'(g^{-1}(\eta(t)))]D^2_{K_1}(\gamma,t)dt}{D^2_{K_1}(0)\int[g'(g^{-1}(\eta(t)))]dt}. \tag{3.14}$$

Since I define the functions $p_2(x)$-$p_6(x)$ I am able to calculate the value of $e_1(\gamma)$ for each model. By considering the function $e_1(\gamma)$ for each of these functions I will give an indication of how the two parts of the integral, ($E_1(\gamma)$ and $E_2(\gamma)$), affect the value of $e_1(\gamma)$.

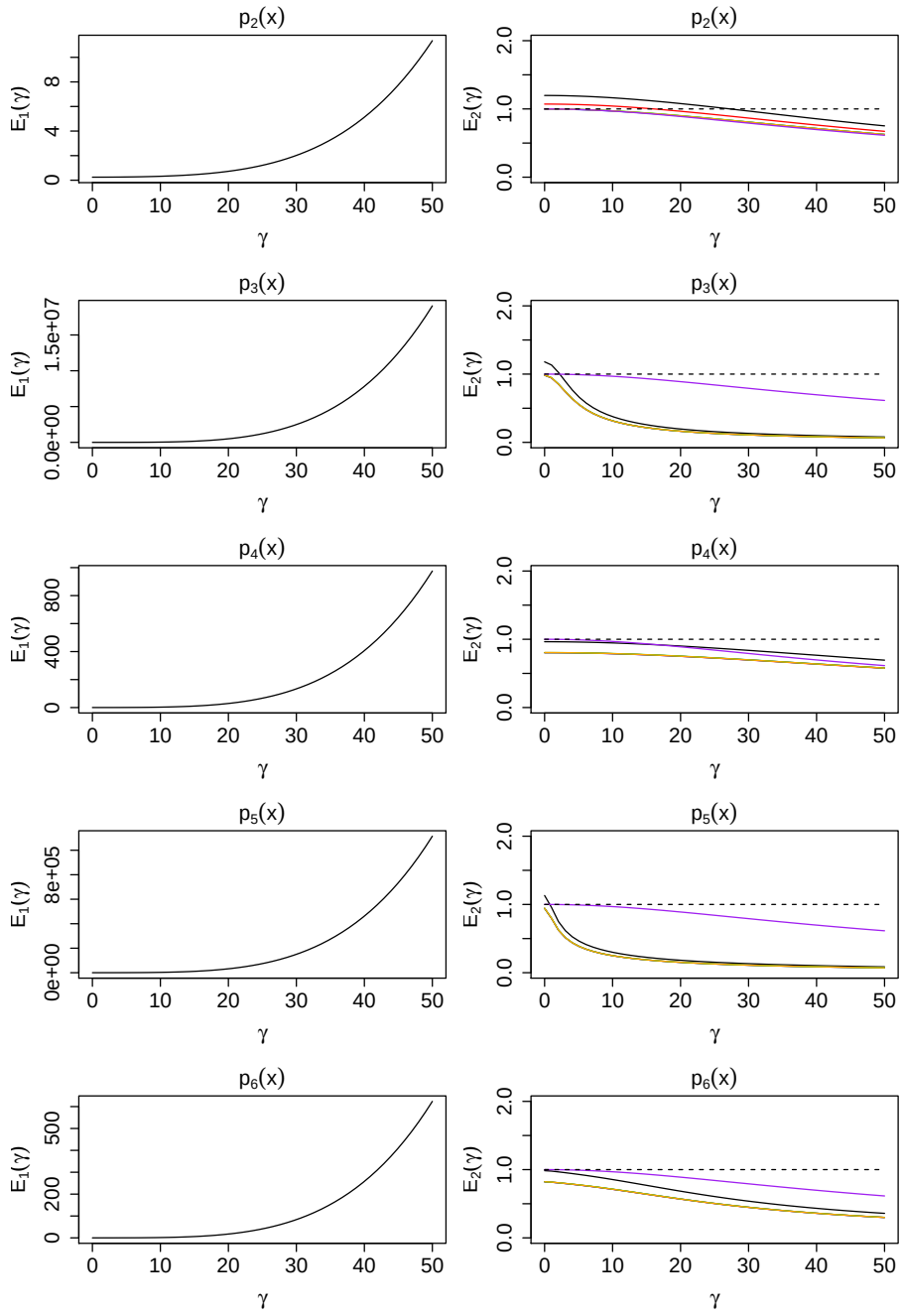

Figure 3.4: The functions $E_1(\gamma)$ and $E_2(\gamma)$ for the six kernels in Table 3.1 $K_d = K_r$ and functions $p_2(x)$-$p_6(x)$. The Uniform Kernel is shown by the black curves, the triangular by the red, the tricube by orange, Epanechnikov by blue, Gaussian by purple and biweight by green.

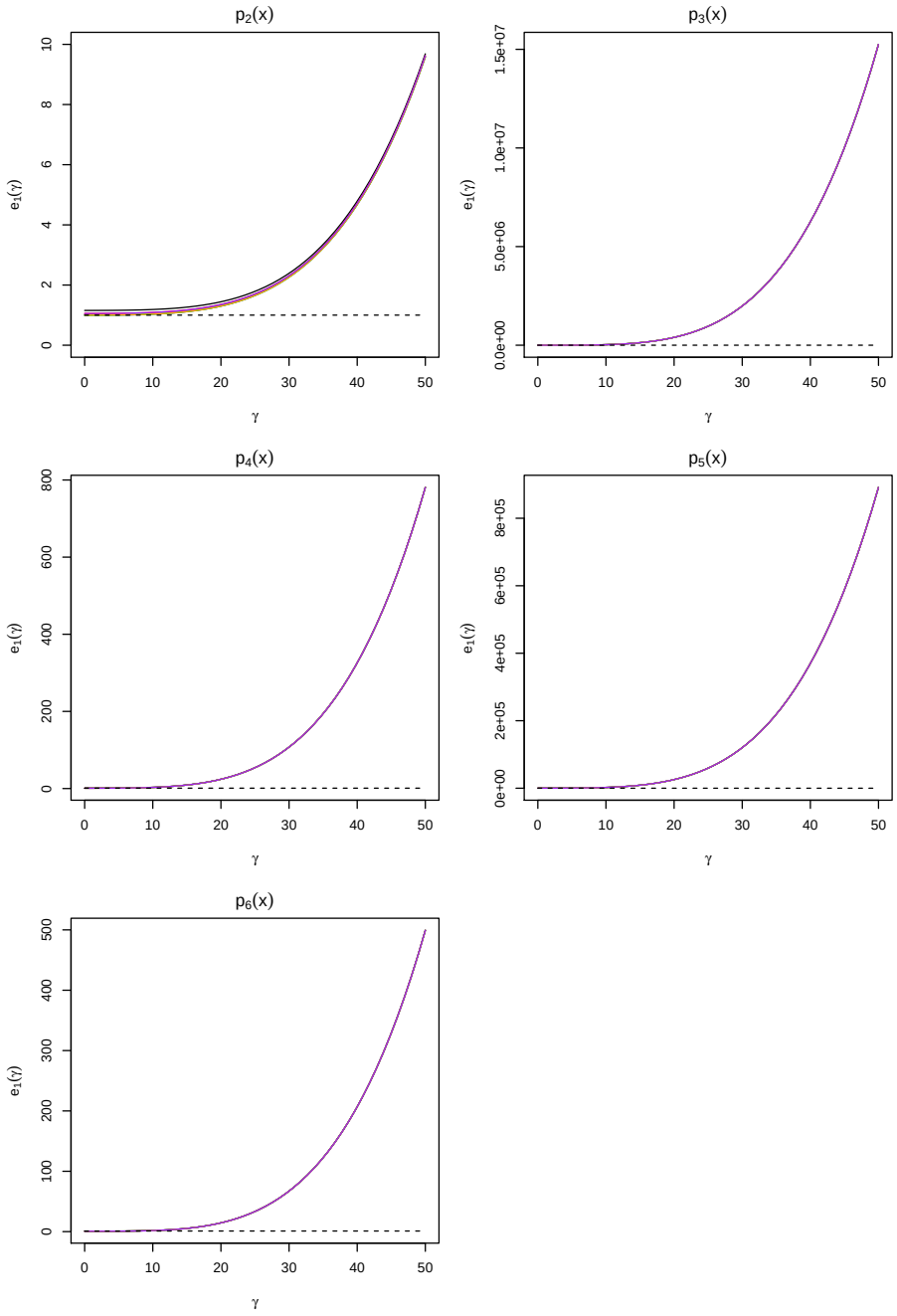

Figure 3.5: The function $e_1(\gamma)$ for the six kernels in Table 3.1 $K_d = K_r$ and functions $p_2(x)$-$p_6(x)$. The Uniform Kernel is shown by the black curves, the triangular by the red, the tricube by orange, Epanechnikov by blue, Gaussian by purple and biweight by green.

I have plotted the two functions $E_1(\gamma)$ and $E_2(\gamma)$ for each of the models in Figure 3.4. The function $E_1(\gamma)$ doesn't depend on a kernel, and for all psychometric functions $p_2(x)-p_6(x)$ is an increasing function with γ since the terms are all squared and hence positive. The function $E_2(\gamma)$ is a decreasing function for each of the commonly used kernels. However the magnitude of $E_1(\gamma)$ is much larger and so this dominates $E_2(\gamma)$. This is made even more clear when the function $e_1(\gamma)$ is plotted in Figure 3.5. Clearly for this function the value $e_1(\gamma)$ is minimal when $\gamma = 0$ and so we should choose a bandwidth which satisfies $h_d = o(h_r)$. This is in agreement with the results for MSE optimal bandwidth choices in Section 3.2.

An analytical solution for the value of $e_1(\gamma)$ is not easily obtained. It is not clear how the value of $e_1(\gamma)$ changes from function to function. Furthermore, in practical situations the true function would not be known and so this method would not be very useful in determining the relationship of h_d to h_r.

3.4 Simulations

3.4.1 Design of Simulations

In this section I present a small simulation study in order to investigate the effect of the choice of h_d on the performance of the LDNP estimator. For all of the simulations in this section I have used the Gaussian kernel for K_r and the Epanechnikov kernel for K_d since this is the built in kernel for the monoproc procedure in the R package monoProc that calulates the DNP estimate. I use the four functions that I introduced in section 3.3, $p_2(x)$-$p_5(x)$. I do not use the function $p_6(x)$ for reasons described in Chapter 4. Essentially the MISE optimal bandwidth for this function was so small that it caused computational problems when trying to calculate estimates of the psychometric function. I use these four functions to generate binomial samples assuming a binomial distribution with mean $p_i(x)$ and r repeats at each of n equally spaced stimulus levels in $(0, 1)$, $x_1, \ldots x_n$.

I considered the effect of varying values of both r and n. I considered $n = 5, 10, 20, 50$ stimulus levels and $r = 1, 5, 10, 20$ repeats at each stimulus level. In total I generated 10000

samples for each model, and for each combination of r and n. A more detailed description of the setup of these simulations can be found in Chapter 4.

For each sample I fitted an unconstrained regression estimate described in Section 1.5 and a bandwidth that minimised the asymptotic MISE of each psychometric function $p_i(x)$. If the estimate was not monotone then I used the LDNP monotonicity constraint described in this thesis to fit monotone regression estimates. If the unconstrained regression estimate was already monotone then this sample was discarded and excluded from any further analysis.

I used four different choices of the second bandwidth $h_d = \sqrt{h_r}, h_r, h_r^2, h_r^3$. In total then for each non monotone sample I created four monotone estimates. I then compared these estimates to see which choice of h_d produced the best results. I compared the square bias, the variance, the MSE and the MISE for each of the methods.

3.4.2 Results of Simulations

The numbers of samples that required monotonisation are recorded in Tables 4.1-4.6 in Chapter 4 (See the discussion there for details). For each model, and each combination of r and n I have compared the square bias, variance and MSE for the LDNP estimates with each of the four possible choices for the value of h_d.

Figures (3.6)-(3.9) show the results for the models $p_2(x)$-$p_5(x)$ when comparing the effect of changing the value of n. In each of these plots the value of r is kept constant at $r = 5$. For the function $p_2(x)$ it is clear that the four choices of bandwidth h_d have almost no effect on the performance of the LNDP estimate except for very small values of x. All four estimates result in comparable estimates of the psychometric function. This is also true for functions $p_4(x)$ and $p_5(x)$. In all three cases the choice of bandwidth h_d makes little difference to the final estimate. This suggests that for these models the sample sizes are not large enough for the asymptotic theory of section 3.3 to come into effect.

Function $p_3(x)$ is a little more interesting. When $n = 5$ the performance of the bandwidth $h_d = \sqrt{h_r}$ seems to perform better than the competing choices in terms of the square bias, variance and MSE. This is true as the value of n increases but it is obvious that by the time

$n = 50$ the four estimates are largely the same. This agrees with the asymptotic theory that as $n \to \infty$ it is desirable to choose a bandwidth such that $h_d = o(h_r)$. However, these simulation results suggest that the asymptotic theory does not hold for small sample sizes and for small samples the choice of h_d is less crucial.

Figures (3.10)-(3.13) show the results for the models $p_2(x)$-$p_5(x)$ when comparing the effect of changing the value of r. In these plots the sample size is kept constant at $n = 10$. Functions $p_2(x)$, $p_4(x)$ and $p_5(x)$ all seem to show that for small sample sizes the choice of h_d has no noticeable affect on the MSE performance of the LDNP method. Any of the values of h_d studied would provide similar estimates of the psychometric function. For the function $p_3(x)$ there again seems to be a slight advantage in choosing $h_d = \sqrt{h_r}$ although this advantage was not huge compared to the other options.

The main conclusion from these simulations was that the asymptotic results stated in this chapter do not hold for small sample sizes. We have a hint from the function $p_3(x)$ that as the sample size increases the best choice of bandwidth would be one such that $h_d = o(h_r)$, but the asymptotic results stated in this chapter seem to need sample sizes larger than $n = 50$ to be of much use. This is not a common case in psychometric studies. However, the simulation study has shown that there is no disadvantage in choosing a bandwidth h_d such that $h_d = o(h_r)$. Bandwidths of this sort performed just as well as other choices.

I would conclude, then, that since the asymptotic theory shows that for large sample sizes it is best to choose h_d such that $h_d = o(h_r)$, and the simulation study shows that this choice of bandwidth is not a bad choice in small samples sizes it makes sense to choose $h_d = h_r^3$ or $h_d = h_r^2$ for the denstity estimation step of the LDNP procedure.

3.5 Figures and Plots

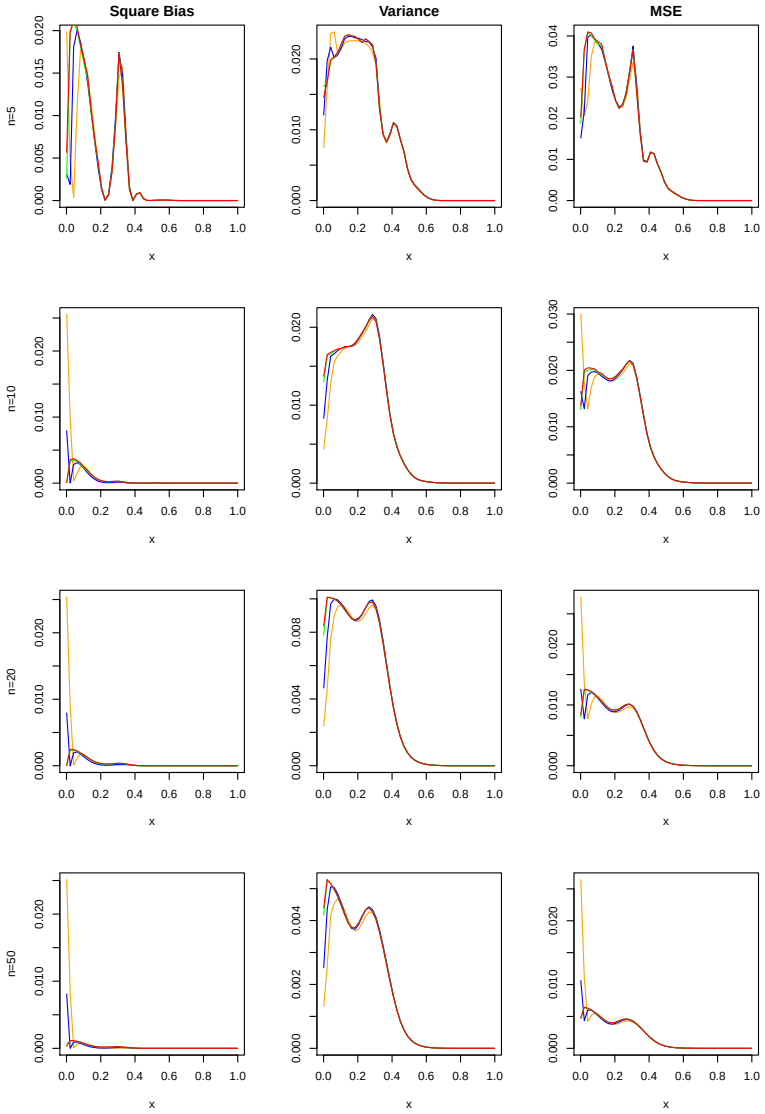

Figure 3.6: A comparison of the square bias, variance and MSE of the LDNP estimate of $p_2(x)$, with four choices of bandwith h_d with $r = 5$ and h_r set as $h_{P,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The orange curve represents the bandwidth $h_d = \sqrt{h_r}$, the blue curve represents the bandwidth $h_d = h_r$, the green curve, $h_d = h_r^2$ and the red curve $h_d = h_r^3$.

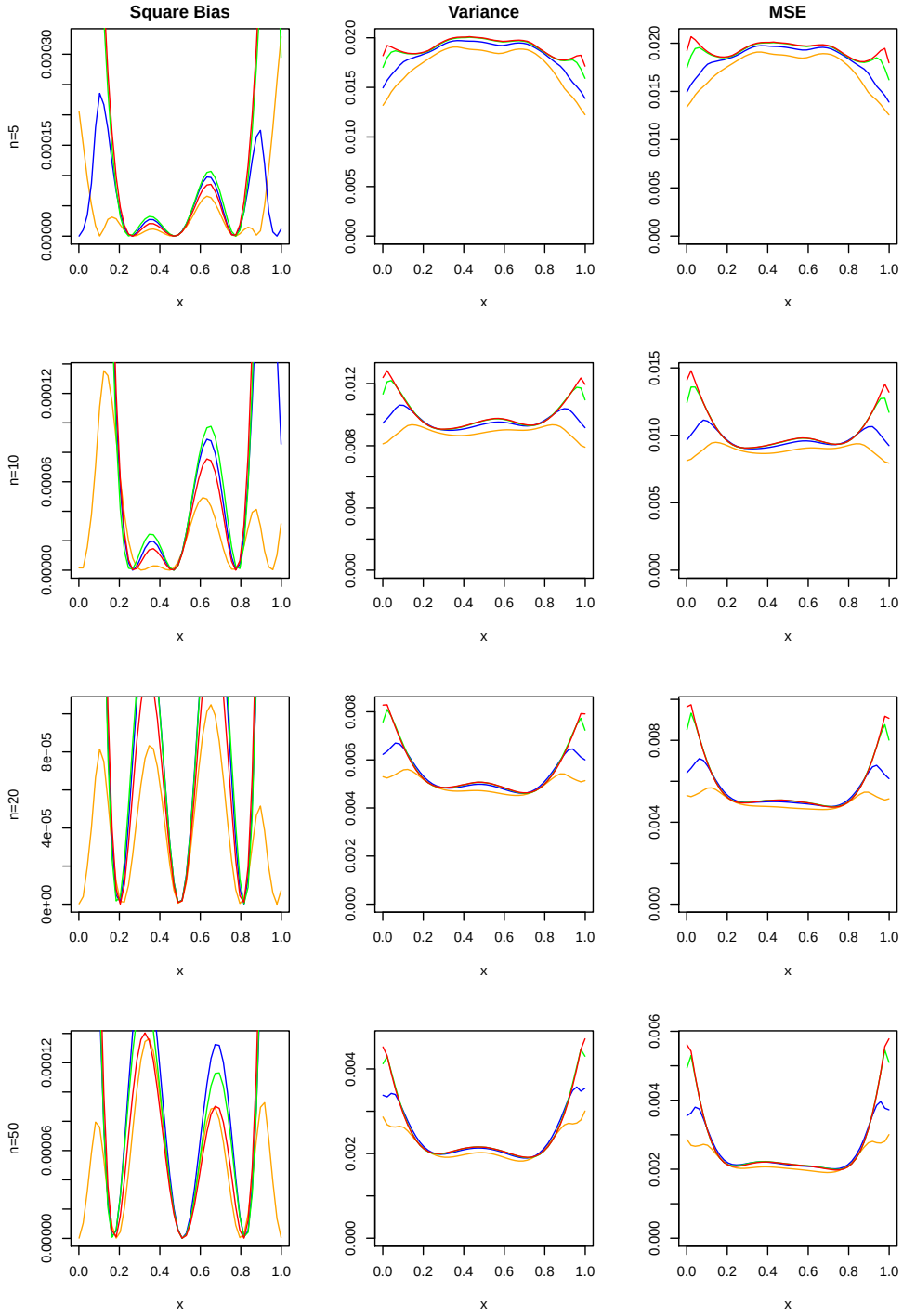

Figure 3.7: A comparison of the square bias, variance and MSE of the LDNP estimate of $p_3(x)$, with four choices of bandwidth h_d with $r = 5$ and h_r set as $h_{P,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The orange curve represents the bandwidth $h_d = \sqrt{h_r}$, the blue curve represents the bandwidth $h_d = h_r$, the green curve, $h_d = h_r^2$ and the red curve $h_d = h_r^3$.

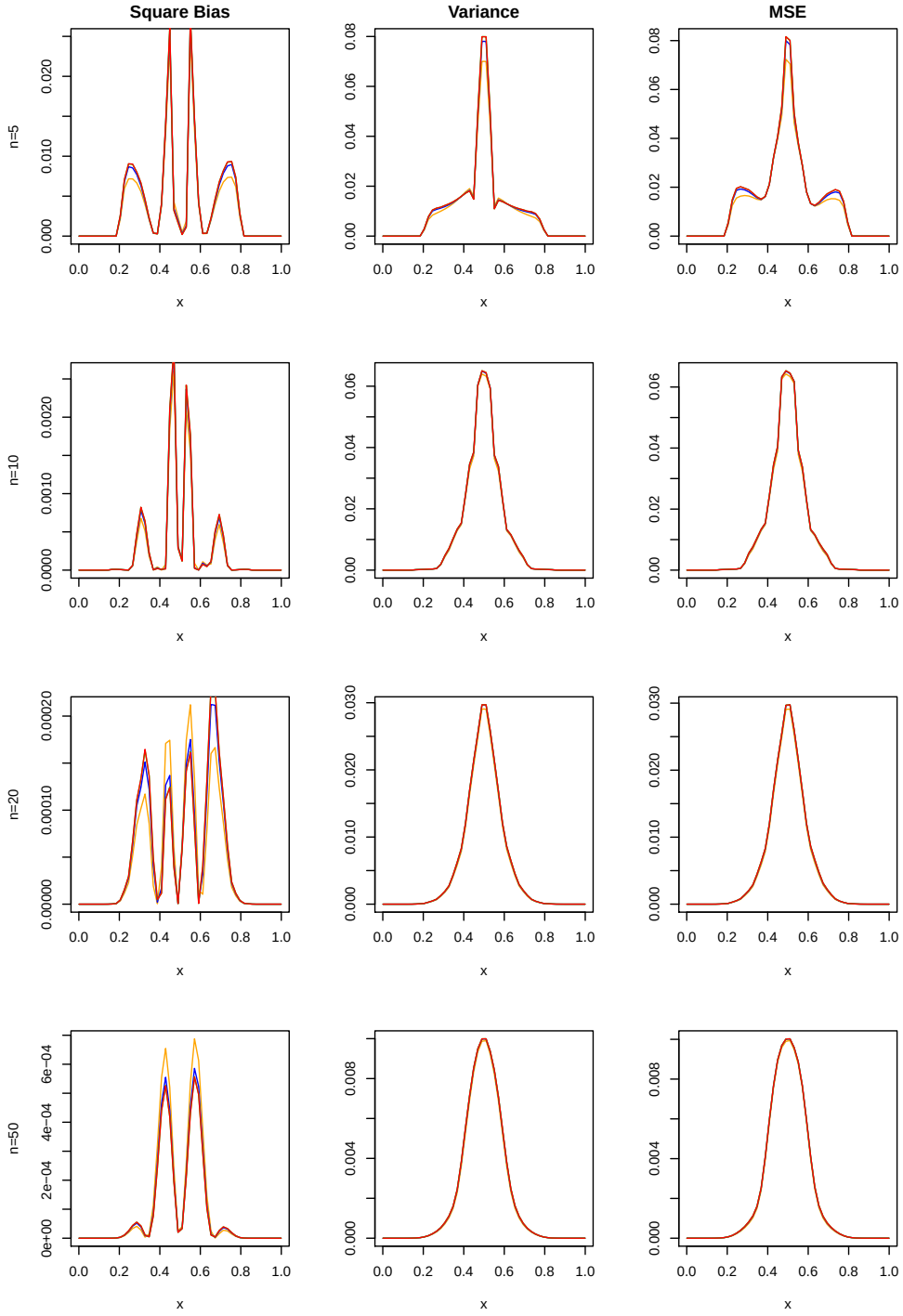

Figure 3.8: A comparison of the square bias, variance and MSE of the LDNP estimate of $p_4(x)$, with four choices of bandwidth h_d with $r = 5$ and h_r set as $h_{P,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The orange curve represents the bandwidth $h_d = \sqrt{h_r}$, the blue curve represents the bandwidth $h_d = h_r$, the green curve, $h_d = h_r^2$ and the red curve $h_d = h_r^3$.

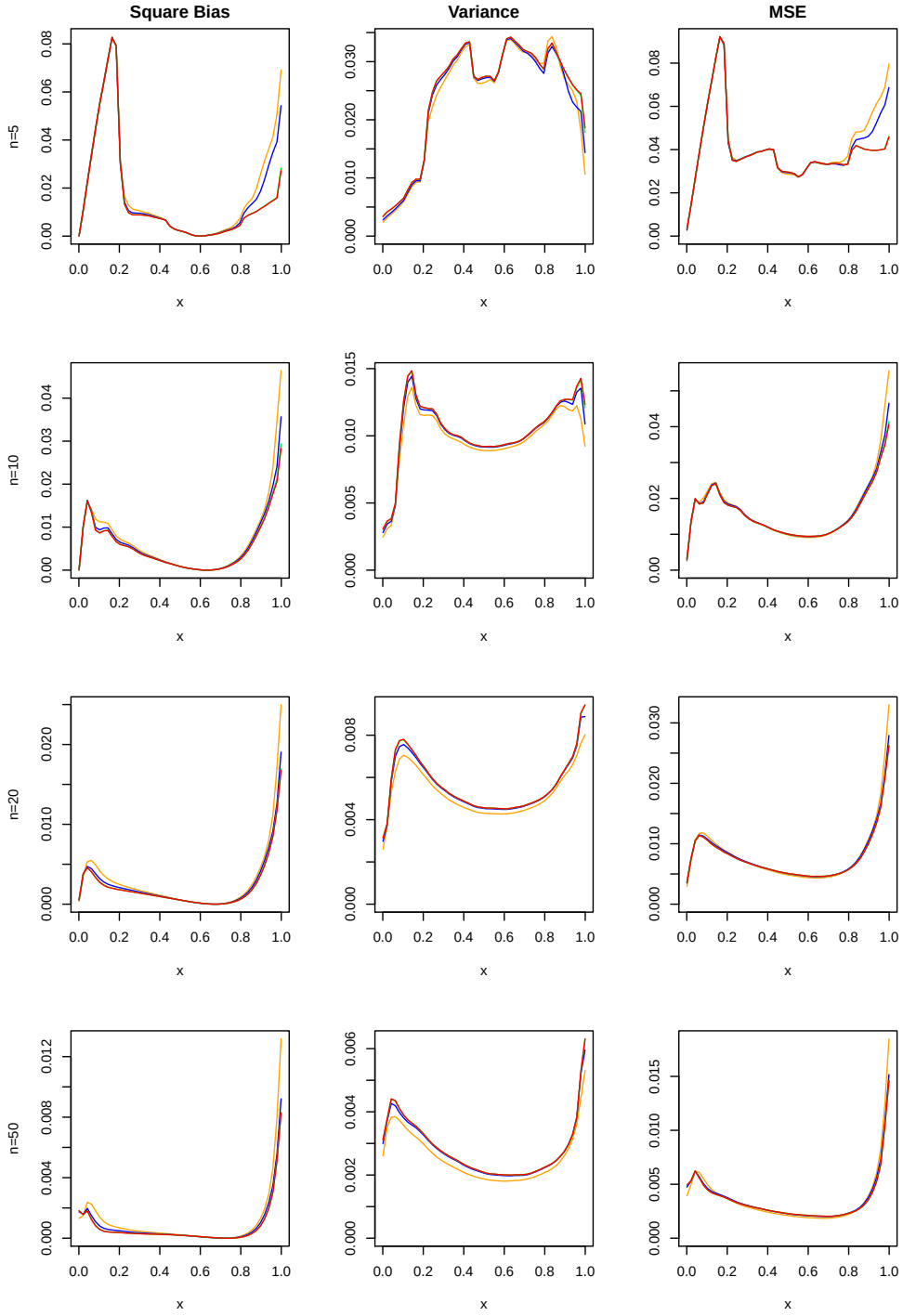

Figure 3.9: A comparison of the square bias, variance and MSE of the LDNP estimate of $p_5(x)$, with four choices of bandwidth h_d with $r = 5$ and h_r set as $h_{P,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The orange curve represents the bandwidth $h_d = \sqrt{h_r}$, the blue curve represents the bandwidth $h_d = h_r$, the green curve, $h_d = h_r^2$ and the red curve $h_d = h_r^3$.

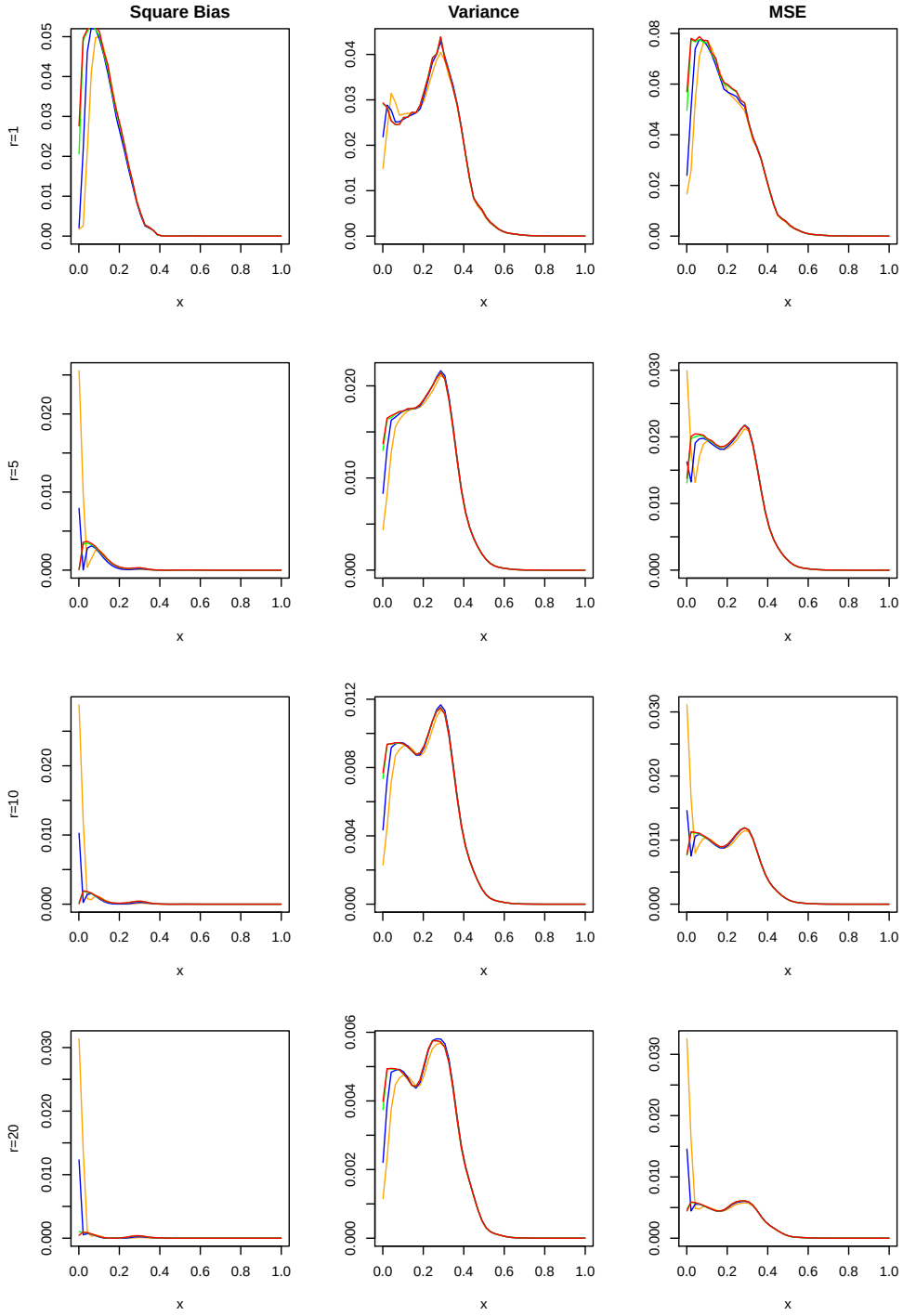

Figure 3.10: A comparison of the square bias, variance and MSE of the LDNP estimate of $p_2(x)$, with four choices of bandwidth h_d with $n = 10$ and h_r set as $h_{P,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The orange curve represents the bandwidth $h_d = \sqrt{h_r}$, the blue curve represents the bandwidth $h_d = h_r$, the green curve, $h_d = h_r^2$ and the red curve $h_d = h_r^3$.

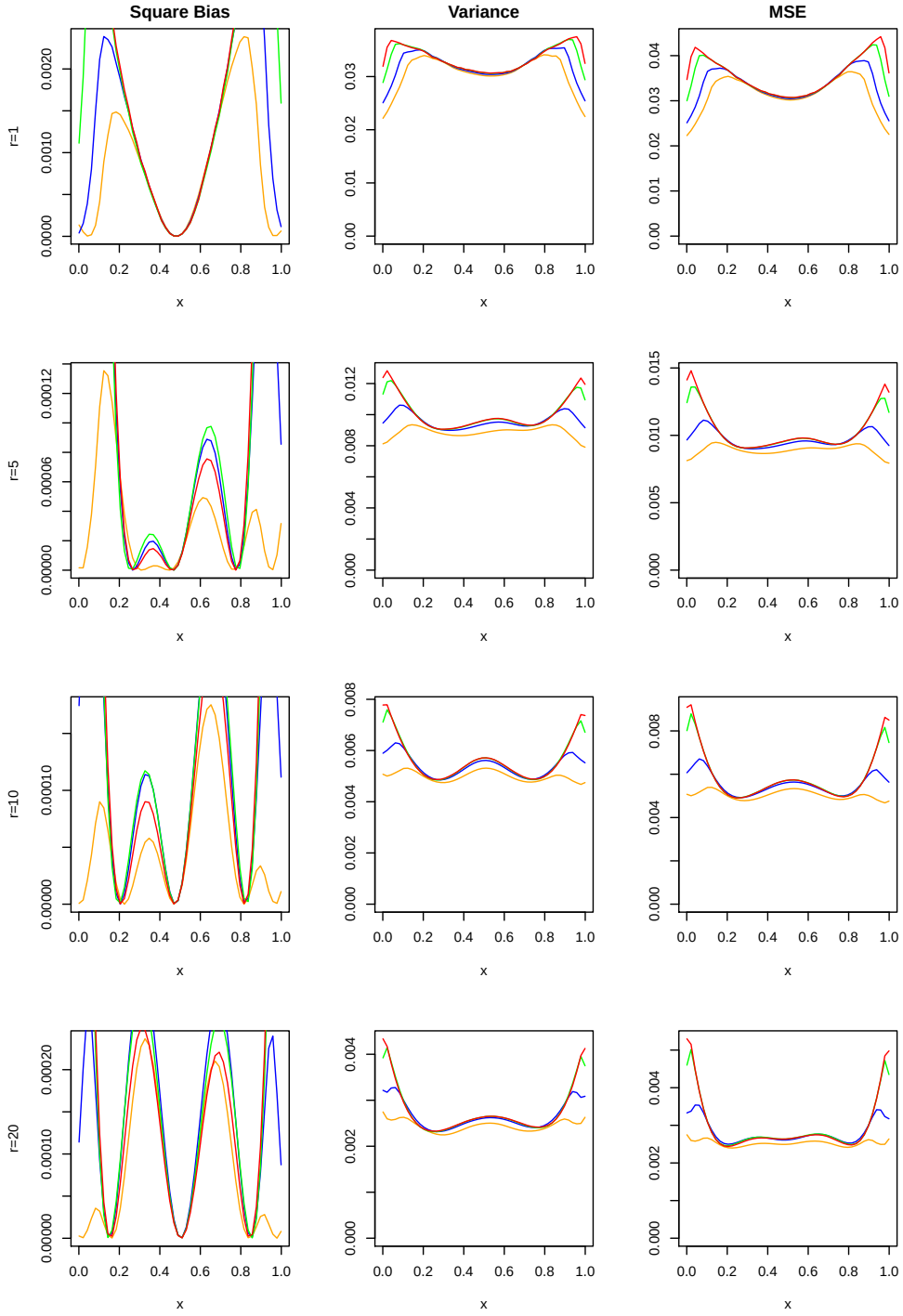

Figure 3.11: A comparison of the square bias, variance and MSE of the LDNP estimate of $p_3(x)$, with four choices of bandwidth h_d with $n = 10$ and h_r set as $h_{P,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The orange curve represents the bandwidth $h_d = \sqrt{h_r}$, the blue curve represents the bandwidth $h_d = h_r$, the green curve, $h_d = h_r^2$ and the red curve $h_d = h_r^3$.

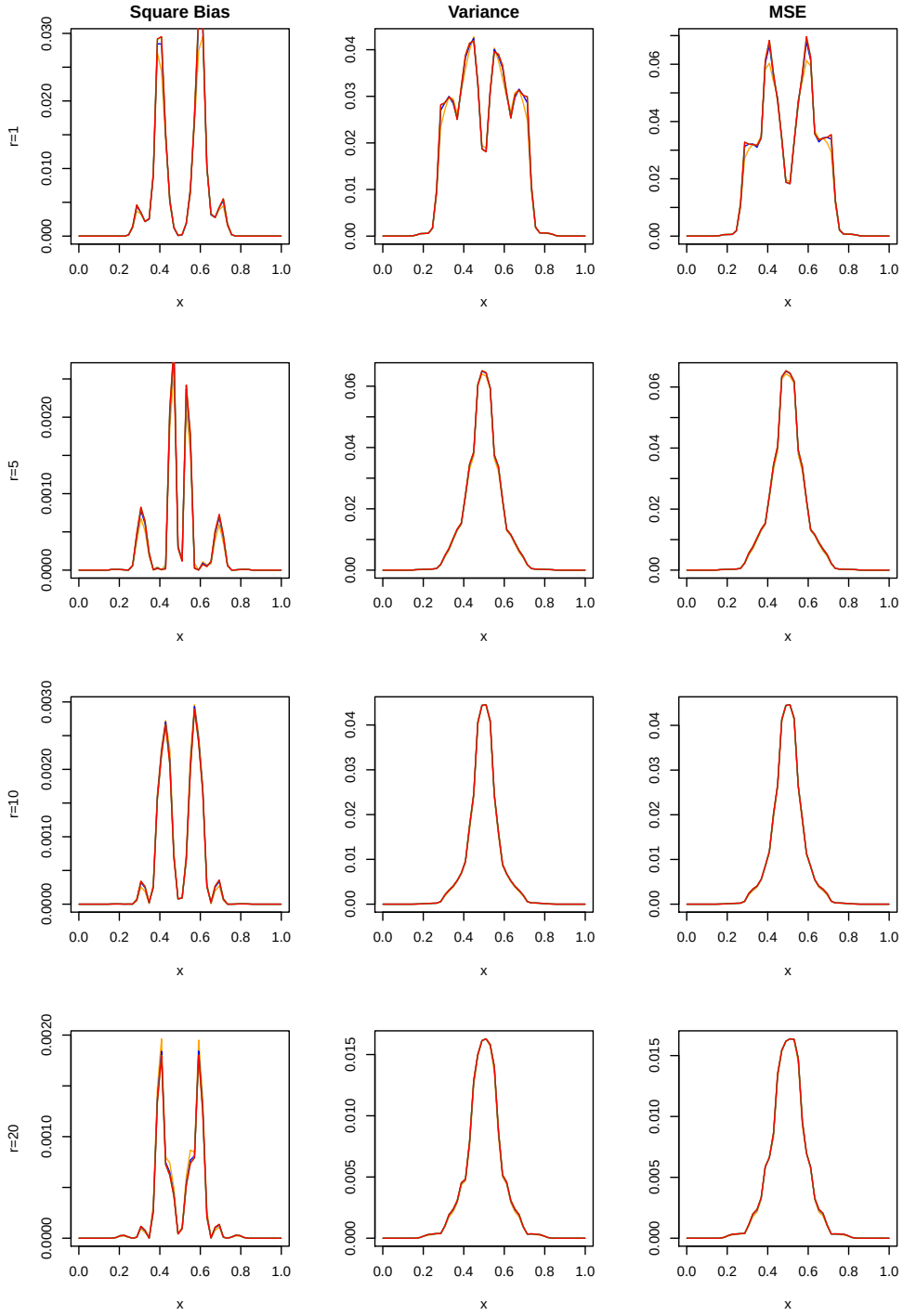

Figure 3.12: A comparison of the square bias, variance and MSE of the LDNP estimate of $p_4(x)$, with four choices of bandwidth h_d with $n = 10$ and h_r set as $h_{P,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The orange curve represents the bandwidth $h_d = \sqrt{h_r}$, the blue curve represents the bandwidth $h_d = h_r$, the green curve, $h_d = h_r^2$ and the red curve $h_d = h_r^3$.

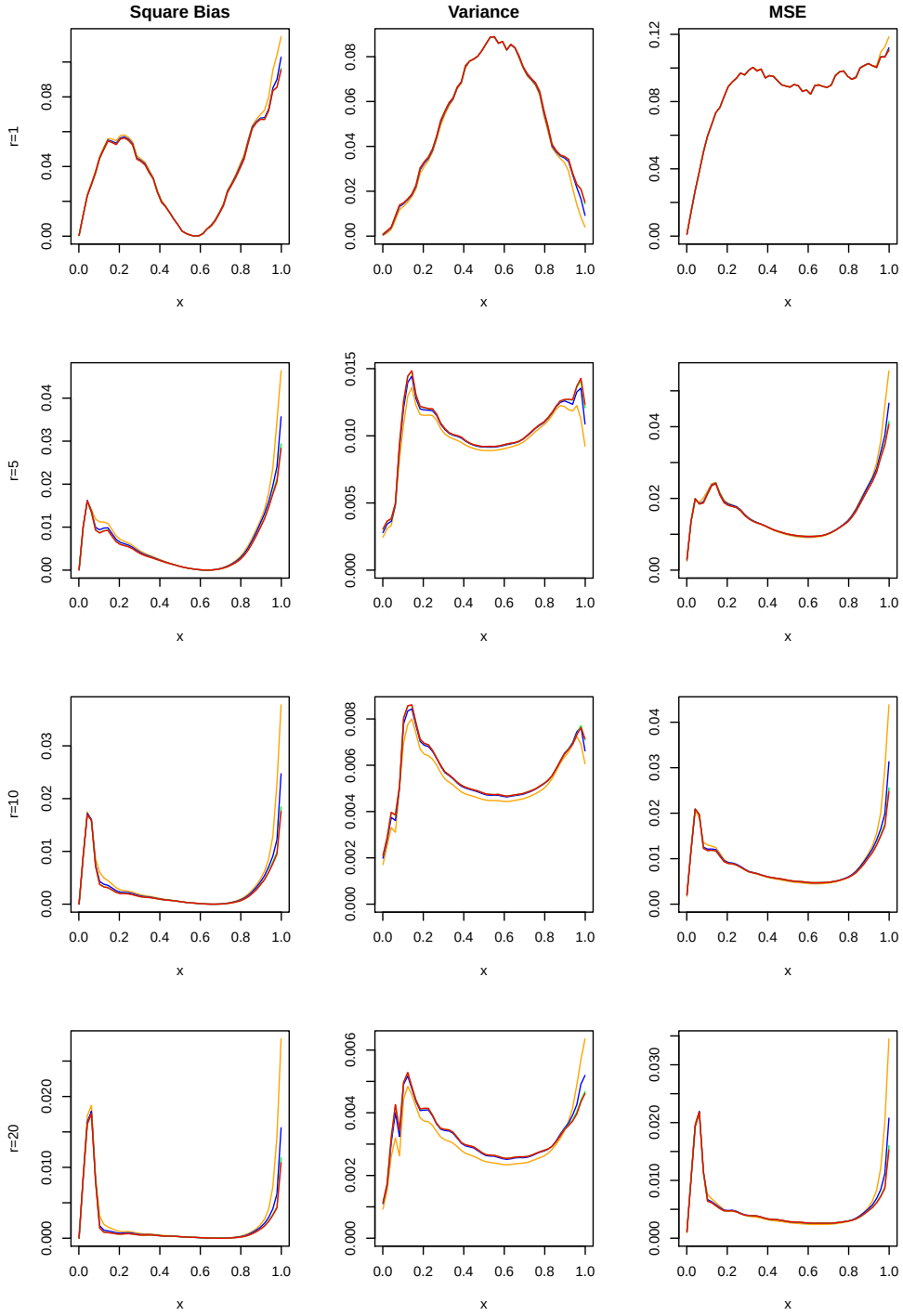

Figure 3.13: A comparison of the square bias, variance and MSE of the LDNP estimate of $p_5(x)$, with four choices of bandwidth h_d with $n = 10$ and h_r set as $h_{P,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The orange curve represents the bandwidth $h_d = \sqrt{h_r}$, the blue curve represents the bandwidth $h_d = h_r$, the green curve, $h_d = h_r^2$ and the red curve $h_d = h_r^3$.

Chapter 4

A comparison of monotone estimates of psychometric functions

4.1 Introduction

In Chapter 1 of this thesis I have outlined various methods of estimating monotone psychometric functions. I have considered the bandwidth method of Kappenman [29] which consists of gradually increasing the bandwidth used in the regression until the resulting estimate is monotone, the PAVA method (Brunk [4] and [3], Mammen [32] and Mukerjee [34]), which 'averages' out any points causing bumps in the response estimate, and the DNP method (see Dette Neumeyer and Pilz [9] and [8]) which is a combination of non-parametric regression and non-parametric density estimation. All three methods are widely used to obtain monotone non-parametric regression estimates. All three methods are explained in more detail in Chapters 1 and 2.

In this thesis I have developed an alteration to the DNP method which I have labelled the LDNP method and takes into account the likelihood. Without this adaptation, the DNP method has problems in estimating monotone psychometric functions. In particular, the main problem was that the DNP method does not constrain the resulting monotone estimate to lie in the region $[0, 1]$, which is a requirement of a psychometric function, since it is a probability. This is largely due to the fact that a link function is not used and information about the changing

variance is not used in the DNP estimation procedure. This problem has been overcome in the LDNP method. I have developed some of the asymptotic properties of the LDNP estimator in Chapter 2 and have shown that it is first order asymptotically equivalent to the other methods I have outlined above.

However, in psychometric studies there is rarely a large sample size. This can be due to a number of factors, such as the cost of obtaining data, ethical issues and time constraints. Typically psychometric studies are carried out with small sample sizes. In this chapter, I compare the performance of monotone estimates of finite sample psychometric functions.

It should be noted that various authors have compared performance of monotonicity constraints before. In a general setting, for the Nadaraya-Watson estimator with normal responses, Dette and Pilz [10] compare the performance of the tilting method, the DNP method and the PAVA method. They use six functions and simulate data samples from each of these functions. They calculate monotone regression estimates for each sample using the DNP method, the PAVA method and the 'tilting method' and then compare these estimates with the true function. They then calculate the square bias, variance and Mean Square Error (MSE) for each sample and use these to compare the performance of each of the three methods under consideration. They conclude that the PAVA method is the worst performing method. The DNP method and 'tilting method' perform comparably but the DNP method has the advantage of being computationally more simple.

In a setting closer to that of this thesis, Dette and Scheder [11] compare methods of calculating a monotone estimate of the effective dose in quantal bioassay. The effective dose is equivalent to the study of the threshold in psychometric studies. They perform a similar analysis to that described above in Dette and Pilz [10] although this time using eight model functions. The 'tilting method' is no longer considered in this setting, whilst a number of additional methods are also considered, including the bandwidth method of Kappenman [29] described in Chapter 1. They consider the case where there are no repeats at each stimulus level and also the case where there are five repeated measurements at each stimulus level. They fix a bandwidth choice of $h_r = 0.1$ for all of their calculations.

Interestingly, in Dette and Scheder [11], they use quasi-likelihood versions of the PAVA method and the bandwidth method in their comparisons but not for the DNP method. As I have mentioned, this can cause problems in the estimation of psychometric functions (or any other bounded functions). Park and Park [37] performed similar simulations focusing on the small sample properties of various monotonicity constraints. They concluded that the DNP method performs at least as well as other estimates in terms of MSE. The purpose of this chapter is to analyse the performance of the LDNP method in comparison to other methods of monotonicity constraint.

There are a number of factors that affect the final estimate of the psychometric function. In a typical psychometric study, a clinician will examine n stimulus levels. Clearly as the number of stimulus levels increases we have more information about the response and so should be able to estimate the true function more accurately. At each stimulus level there may be r trials, either by testing r people once or repeated observations on fewer individuals. As the value of r is increased we again would have more information about the response and would hope to estimate the true function more accurately.

In addition, when estimating the psychometric function, a bandwidth, h_r, must be chosen to be used in the unconstrained step. For the LDNP estimate there is also a second bandwidth, h_d, used in the density estimation step. Changing the value of h_r will clearly affect the ability of the estimating procedure to correctly estimate the true function. Too small a choice of bandwidth would give too much significance to individual points which may pull the estimate away from the true value. Too large a choice of h_r would result in oversmoothing and we may miss important features of the true response since the resulting estimate had been flattened too much. The choice of h_d has been considered in a previous chapter and is not dealt with in this chapter. We will alter the other three factors in turn.

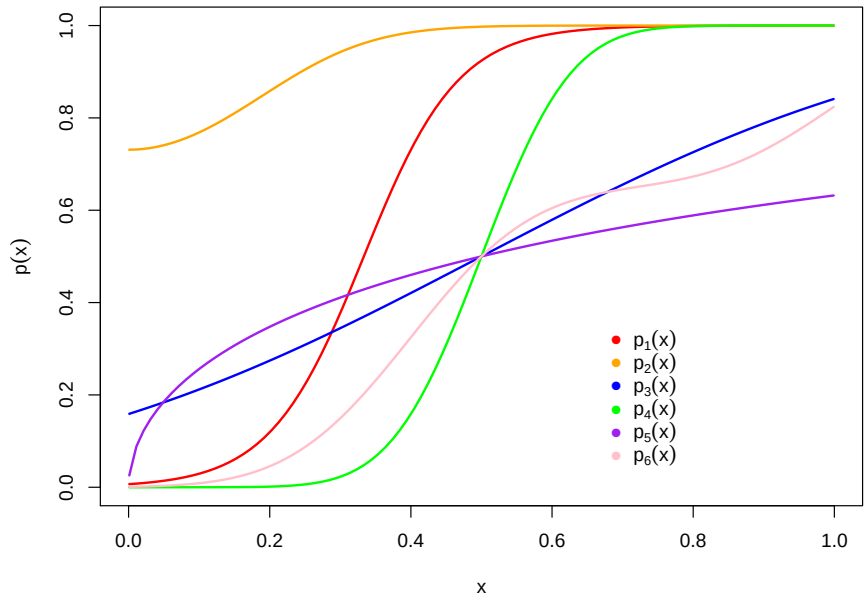

Figure 4.1: Psychometric functions, $p_1(x)$-$p_6(x)$ used in simulations

4.2 Design of Simulation Study

In this chapter I will use five of the functions which were used by Dette and Scheder [11] in their simulations. I used these functions in simulations in Chapter 3. A plot of these functions is shown in Figure 4.1, whilst a plot of the the corresponding η functions (obtained with logit link) is shown in Figure 4.2.

$$
\begin{aligned}
p_1(x) &= (1+\exp(5-15x))^{-1} \\
p_2(x) &= (1+\exp(-(20x^2+1)))^{-1} \\
p_3(x) &= \Phi\left(\frac{x-\mu}{\sigma}\right), && \mu = .5, \sigma = .5 \\
p_4(x) &= \Phi\left(\frac{x-\mu}{\sigma}\right), && \mu = .5, \sigma = .1 \\
p_5(x) &= 1-\exp(-x^{\gamma}), && \gamma = .52876 \\
p_6(x) &= \eta\Phi\left(\frac{x-\mu_1}{\tau}\right) + (1-\eta)\Phi\left(\frac{x-\mu_2}{\tau}\right) && \mu_1 = 0.4, \mu_2 = 1, \eta = .64946, \tau = .13546
\end{aligned}
$$

In all of the following simulations I will generate n equally spaced stimulus levels, $x_1 \dots x_n$,

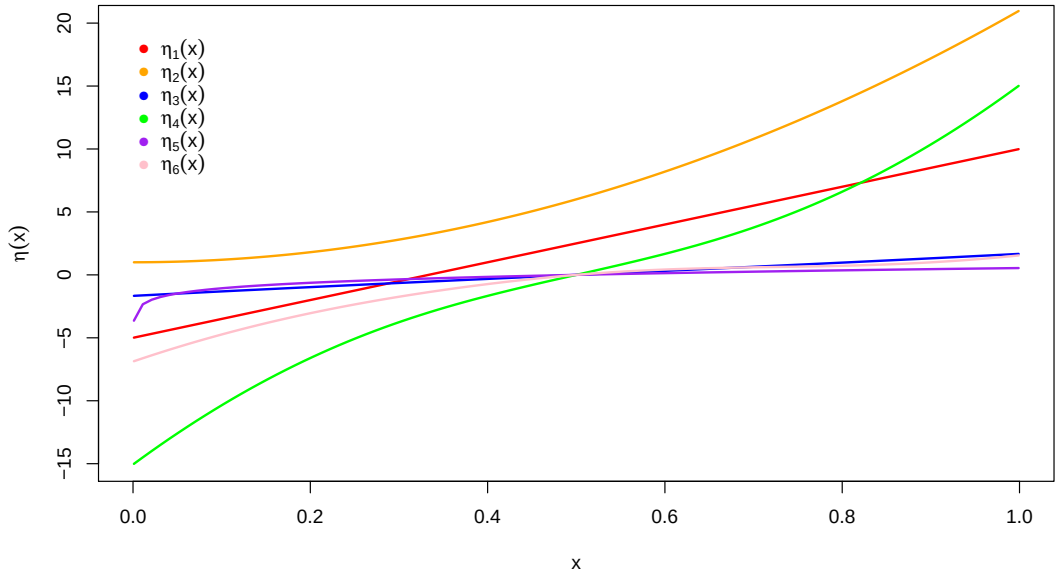

Figure 4.2: The functions $\eta_1(x)$-$\eta_6(x)$ corresponding to the functions $p_1(x)$-$p_6(x)$ (see Fig 4.1) after transformation by the logit link.

in the interval $[0, 1]$. For each model in turn, at each stimulus level I will simulate results from a binomial distribution with r repeats and based on the probability of success $p_j(x_i)$. For each of these data sets I will calculate a regression estimate and then, if not monotone, monotinise using the PAVA method, the bandwidth method and the LDNP method. I will calculate the LDNP estimate using four different options for the choice of h_d, as described in Section 3.4. I calculate the MISE for each of these options and use the option that produces the lowest MISE for comparison with the PAVA and bandwidth methods. In all simulations, the Gaussian kernel defined in Table 3.1 is used for K_r and the Epanechnikov kernel for K_d. In theory any kernel could be used and the choice of kernel does not really affect the final results and so the choice is not crucial here.

For the unconstrained estimate I will use three different types of bandwidth: MISE-optimal, ISE-optimal and cross-validation. The MISE optimal bandwidth will be dependent upon the psychometric function $p_j(x)$, the number of stimulus levels n and the number of repeats r,

but not upon the individual samples. So for a fixed value of r and n this bandwidth will be kept constant for all samples from a psychometric function $p_j(x)$. I will consider a bandwidth $h_{P,MISE}$, given in (4.1), which optimises the MISE in terms of the psychometric function $P(x)$, and also a bandwidth $h_{\eta,MISE}$, given in (4.2), which is the optimal bandwidth for the MISE in terms of $\eta(x)$.

$$h_{P,MISE} = \left[\frac{1}{2mn\sqrt{\pi}} \frac{\int_0^1 P(x)(1-P(x))dx}{\int_0^1 (\eta''(x))^2 P^2(x)(1-P(x))^2 dx}\right]^{1/5} \tag{4.1}$$

$$h_{\eta,MISE} = \left[\frac{1}{2mn\sqrt{\pi}} \frac{\int_0^1 \frac{1}{P(x)(1-P(x))}dx}{\int_0^1 (\eta''(x))^2 dx}\right]^{1/5}. \tag{4.2}$$

At this point I must note two exceptions. When considering the function $p_1(x)$ it is clear that for this function $\eta''(x) = 0$. Clearly the formulas (4.1) and (4.2) cannot be used here. The two formulas would suggest though that the MISE would be minimised by selecting a very large bandwidth h_r for use in the unconstrained regression. This does make sense as the function $\eta(x)$ is in this case a linear function and a local linear regression, (1.3), with a large bandwidth becomes close to logistic regression. A logistic regression would be most suitable in this case to estimate a monotone $\hat{\eta}(x)$ as it is the correct model. For this reason I have not considered the MISE optimal bandwidth for the function $p_1(x)$. In my simulations I did try using a large bandwidth of $h_r = 10$ just to check what happened and, as expected, in each case the unconstrained estimate was monotone. It should be noted that this was also the case for $h_r = 1$.

For the function $p_6(x)$ the MISE optimal bandwidths calculated using (4.1) and (4.2) were very small. These very small bandwidths caused computational problems when trying to calculate an estimate of the psychometric function. For this reason I have not included the MISE optimal bandwidths for the model $p_6(x)$ and just use the remaining four bandwidth choices for this model.

The second type of bandwidth I will use is an ISE optimal bandwidth. This bandwidth is the best possible bandwidth that could be used for each individual sample. I will use a P optimal ISE bandwidth, $h_{P,ISE}$ and also an η optimal bandwidth, $h_{\eta,ISE}$. These bandwidths

are obtained by minimising the integrated square errors (4.3) and (4.4) respectively.

$$ISE_P = \int_0^1 (P(x) - \hat{P}(x))^2 dx \tag{4.3}$$

$$ISE_\eta = \int_0^1 (\eta(x) - \hat{\eta}(x))^2 dx. \tag{4.4}$$

Finally, I will use cross-validation bandwidths, denoted $h_{P,cv}$ and $h_{\eta,cv}$ that are calculated using formulas (4.5) and (4.6) given in Fan and Gijbels ([16] page 150)

$$CV_P(h) = \int \hat{P}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{P}_{-i}(X_i) \tag{4.5}$$

$$CV_\eta(h) = \int \hat{\eta}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{\eta}_{-i}(X_i). \tag{4.6}$$

where $\hat{P}_{-i}(X_i)$ and $\hat{\eta}_{-i}(X_i)$ are the estimates obtained by leaving out the i^{th} observation and calculating a regression estimate using the remaining observations. The effect of the bandwidth choice is discussed in section 4.3.3. MISE and ISE optimal bandwidths can be calculated here because the true psychometric functions are known. In practice, one needs to estimate the optimal bandwidth, as discussed in Section 1.4. One of the methods is the cross-validation method which is recommended by Żychaluk and Foster [18]. I use it here since it is a commonly used bandwidth selection method.

I will compare the methods for a range of different values of n. I will use the values $n = 5, 10, 20, 50$. It should be said that 50 stimulus levels would be larger than usually experienced in psychometric studies with the other three values much more common. Each of the methods I am comparing are first order asymptotically equivalent. The point of this investigation is to investigate their finite sample performance.

I will also compare the effect of changing the number of repeats at each level. As the value of r increases then we have much more information and would expect better performance for the estimates. But many psychometric studies only consider a small number of repeats, if any. So I want to investigate which method performs the best when there are only small numbers of repeats at each stimulus level. The values that I have considered are $r = 1, 5, 10, 20$. I did not consider the case of $r = 1$ with only 5 stimulus levels as this is a very small amount of data

and non-monotone samples are very unlikely. It is clear that the situation of $r = 1$ corresponds to non repeated measurements. In this case the responses are Bernoulli.

For each model and values of r and n I generated 10,000 samples. In the following analysis of these simulations, I will report how many of the simulated samples required monotonicity constraints to be used. Only unconstrained estimates that required monotonisation were kept and analysed. This is so as to avoid confusing the analysis. I am interested in comparing how good each of the three methods are at estimating monotone psychometric functions. If the unconstrained regression is already monotone then no constraint is required and so these samples are not considered in my analysis. To the best of my knowledge other studies (see Dette and Pilz [10] and Dette and Scheder [11]) do not remove monotone samples from their comparisons, but I feel this is a more clear and fairer comparison.

Once I have obtained the simulated samples I will follow the example of Dette and Scheder [11], and estimate the square bias, variance and mean square error (MSE) for each of the models where the MISE optimal bandwidth was used. The other bandwidth choices vary between samples and so I can't take averages over samples in this case and hence this kind of analysis would be inappropriate. In the cases where a P optimal bandwidth is used, I will compare the MSE in terms of estimating $P(x)$ and, in the case where I use an η optimal bandwidth, I will compare the MSE in terms of the estimation of $\eta(x)$.

$$MSE_P = E\left[(\hat{P}(x) - P(x))^2\right]$$

$$MSE_\eta = E\left[(\hat{\eta}(x) - \eta(x))^2\right]$$

These are standard ways of comparing the performance of the estimates and will give a good idea of whether any of the estimates are consistently better than others. Obviously, the closer to 0 each of these values is, the better the model. The MSE assesses the performance of an estimator at individual points within the range. I will also compare the mean integrated square error (MISE), to give an idea of the performance of the estimator over the whole range. The square bias, variance and MSE were calculated on a grid of 50 equally spaced points between 0 and 1. They were averaged across all the samples. The MISE was calculated by numerical integration of the MSE in R [39].

When a different bandwidth has been used for each sample this kind of analysis is no longer appropriate as results cannot be averaged over samples in this case. In these cases, where ISE optimal bandwidths, or cross validation bandwidths have been used, then I will compare the methods in a different way. I will calculate the ISE for both the unconstrained fit and the monotone fit for each sample and then calculate the ratio,

$$R = \frac{ISE_{monotone}}{ISE_{unconstrained}}. \tag{4.7}$$

I will plot density estimates of the ratios for each of the methods of monotonisation. and show these plots as a means of comparing the performance of each of the three estimators. Obviously the method with the smallest R values will be the best method. In general, I expect $R > 1$ and that the density with most of its mass close to 1 indicates better performance. However, in some cases $R < 1$, indicating that the monotonised estimate has lower ISE than the unconstrained fit. In these cases, density concentrated near 0 indicates better performance.

In all the simulations in this chapter and in Chapter 5 I have used R version 2.12.2 [39]. I particular I have used the code for the DNP method found in the R package 'monoProc' produced by Dette *et al.* [9] which I have adapted for use in the LDNP method. I have also adapted the code local linear regression estimates of exponential family models found in the package 'modelfree', produced by Żychaluk and Foster [18]. I have used the University of Liverpool Condor system to run the simulations.

4.3 Results

4.3.1 Number of samples requiring monotonisation

The numbers of samples out of the original 10,000 that required monotonisation for each of the six possible bandwidth choices are shown in Tables 4.1-4.6. It should be mentioned that not all of the 10,000 samples will be unique. Inevitably some samples will be repeated. This is because there is a finite number of possible samples. I record the number of unique samples for the sake of interest and completeness. This figure is recorded in brackets in Tables 4.1-4.6. Recall that

only samples which required monotonisation were considered in the following analysis. Clearly as the values of r and n increase, the possibility of repeated samples decreases and more of the samples are unique.

I can see no definitive pattern in the number of samples that required monotonisation. However it would seem to be generally true that as the number of repeats r and the number of stimulus levels, n, increases the number of samples requiring monotonisation increases, although clearly this is not always true (see functions $p_2(x)$ and $p_4(x)$ for example). In many cases, for a fixed n, as r increases the number of samples requiring monotonisation decreases (see functions $p_1(x)$ and $p_3(x)$ with $n = 10$, for example). For a lot of the samples most of the P optimal bandwidth methods produced a lot more samples that required monotonisation than the η optimal bandwidths (see functions $p_1(x)$ and $p_2(x)$ for example).

4.3.2 Mean Integrated Square Error

For each of the sets of investigations that involved MISE optimal bandwidths, I calculated the MISE values. I calculated these using numerical integration of the Mean Square Error. These are given in Tables 4.7-4.10. In each table the best performing method has been highlighted in red. An entry of NA records the fact that for this model there were no samples that required monotonisation. LDNP represents the LDNP estimate with $h_d = \sqrt{h_r}$, LDNP1 represents the LDNP estimate with $h_d = h_r$, LDNP2 represents the LDNP estimate with $h_d = h_r^2$ and LDNP3 represents the LDNP estimate with $h_d = h_r^3$. See Chapter 3 for more detailed discussion on the choice of the bandwidth h_d.

In Table 4.7 a number of the cells also show the standard error of the MISE estimate. This has been included to show that the values of these standard errors are very small and also to give a more complete picture of the MISE results. I omit the standard errors from all of the other calculations. They could just as easily be calculated but have been omitted to avoid cluttering the results.

It is clear from these tables that a general pattern, as the number of repeats increases, the MISE values decreases, indicating that the estimate is closer to the true values when more

repeats are used. This is to be expected. A similar pattern is observed as the number of stimulus levels is increased.

In general, the values of MISE are of similar order for all of the methods, although for $p_5(x)$ with small sample sizes, the bandwidth adjustment performs considerably better. For the function $p_2(x)$, in general the bandwidth adjustment method performs best in terms of the MISE values. But as the sample size increases, then usually at least one of the LDNP estimates is better. Most often the choice of $h_d = h_r^2$ or $h_d = h_r^3$ is the best choice for the LDNP method which supports the asymptotic theory of Chapter 3.

For function $p_3(x)$, the bandwidth adjustment method consistently performs better than the rest of the methods. For the LDNP method the choice of bandwidth $h_d = \sqrt{h_r}$ seems to be the best here. The bandwidth adjustment method and the LDNP method both do well for function $p_4(x)$ although in general the bandwidth method is better. Finally for the function $p_5(x)$, the bandwidth adjustment method does considerably better at small sample sizes but as the sample size increases at least one of the LDNP estimates performs better than it.

4.3.3 Choice of bandwidth h_r

In these comparisons I have used six different bandwidths, as described section 5.2 which were used for the unconstrained regression step.

In my initial investigations, I had fixed the bandwidth h_r as 0.1, 0.2 and 0.3. The simulations ran analogously to those described in this chapter. The results indicated that for a bandwidth choice of 0.1, as used in Dette and Scheder [11], the bandwidth method performed much better than both the LDNP method and the PAVA method, whose performances were broadly comparable.

For $h_r = 0.2$ the bandwidth method was still better than the other two, although less clearly than before, whilst for $h_r = 0.3$ the bandwidth method no longer performed as well as the other two methods. This was a surprising result initially. But in reality the bandwidth method was performing well in these simulations precisely because it was the method that was able to increase the bandwidth when required. The LDNP method and the PAVA method were stuck

with a bandwidth that was much smaller than the optimal one and the bandwidth method was able to increase it to be closer to the optimal bandwidth. This was an unfair comparison. The good performance of the bandwidth method was not due to the fact that it was better than the other methods but simply due to the fact that it could change bandwidth.

The results of these simulation caused me to consider only the optimal bandwidths to be used in the unconstrained regression so as to provide a fair comparison. The results to these initial investigations are not shown here for the sake of space and are only mentioned so as to highlight the effect bandwidth choice can make.

In the general pattern seems to be that for MISE optimal bandwidths and for cross-validations bandwidths as the bandwidth for the unconstrained regression, the bandwidth adjustment method performs the best. If the ISE optimal bandwidth is used for the unconstrained regression then the LNDP method consistantly performs well compared to the other methods.

4.3.4 Changing the number of stimulus levels

The next factor that I considered was the effect the number of stimulus levels had on the monotone estimate. I considered the cases when $n = 5, 10, 20, 50$. In each of the simulations that I report in this section I fix the value of $r = 5$. I also only consider the versions of the MISE, ISE and cross-validation bandwidths that were optimal for $P(x)$ although mention here that results for the other bandwidths followed a similar pattern. It is clear that the magnitude of the square bias, variance and MSE decreases as n increases.

Figures 4.3-4.6 show the comparison of square bias variance and MSE for each of the functions $p_2(x)$-$p_5(x)$ respectively. Recall that I do not consider the MISE optimal bandwidth for functions $p_1(x)$ and $p_6(x)$. A general pattern in all of these plots was that as the number of stimulus levels increases the three methods became more similar.

For function $p_2(x)$ the square bias values are generally higher for lower values of x with much higher bias at the boundary. There is little to distinguish between the three models although it is clear that as the number of stimulus levels increases the bandwidth method performs less well compared to the other methods. As the value of n increases the square bias of the bandwidth

method remains far higher than the other two methods. This results in higher MSE values for this method.

The bandwidth adjustment method is clearly performs the best in terms of MSE performance for estimating $p_3(x)$ as can be seen from Figure 4.4. The square bias of the PAVA estimate and the bandwidth estimate are very high at both boundaries compared to the LDNP estimate but are broadly similar over the rest of the values of x. The variance of the PAVA estimate is always higher than that of the other two estimates suggesting that this method is not very good for estimating $p_3(x)$.

The bandwidth adjustment method also performs the best in terms of MSE performance for estimating $p_4(x)$ as can be seen from Figure 4.5. This seems to be largely due to the fact that the variance for this estimate is significantly lower than for the other methods with the bias values being reasonably similar.

Figure 4.6, shows that the bandwidth adjustment method again performs well for estimating $p_5(x)$. The LDNP method seems to perform better than the PAVA method. For this model it is clear that the methods become more similar as the number of stimulus levels increases. As the value of n increases the bias performance of the LDNP method and PAVA method becomes much better in comparison to the bandwidth method but the variance remains higher. In general then, as the value of n increases there is less of a difference between the three methods.

A comparison of the ISE ratios for the ISE optimal and cross validation bandwidths can be seen in Figures 4.7-4.12 for functions $p_1(x)$-$p_6(x)$ respectively. By comparing the first column we can see that as the number of stimulus levels increases the PAVA method and the LDNP method become more and more similar. They also become more tightly bound around 1. This indicates that as the value of n increases the LDNP estimate and the PAVA estimate approach the unconstrained estimate. In general the LDNP method performs the best when the ISE optimal bandwidth is used for the unconstrained regression. The bandwidth method performs poorly in this case. In addition the density estimate for this method becomes much more flat as the number of stimulus levels increases. This mean that this method is becoming progressively worse at obtianing a monotone estimate and in many cases produces very bad estimates. A

flatter denstiy estimate indicates a lot of variation in ISE ratio.

By considering the second column of Figures 4.7-4.12 it can be seen what affect increasing the sample size has on the estimates when a cross-validation estimate has been used for the unconstrained regression. At low values of n the bandwidth adjustment method performs well in comparison to the others. However as the number of stimulus levels increases the density estimates for this method become more flat suggesting more variability in this method and hence poorer performance. The PAVA estimates and the LDNP methods perform comparably to each other although the LDNP method is usually a little bit better. This suggests that for small values of n the cross-validation bandwidth estimate is too small and hence the best way to calculate a monotone estimate is to simply increase the bandwidth. As the value of n increases the cross-validation bandwidth is closer to the optimal one and the bandwidth method performs less well.

In summary, then, it seems that when a good bandwidth is used (ISE optimal) then the bandwidth adjustment method does not work well in terms of ISE. The other two methods are similar but the LDNP is usually better. As n increases the performance of the methods is more similar but there are still noticeable differences for $n = 50$. If a poor bandwidth is chosen then the best method is to simply increase the bandwidth.

4.3.5 Changing the number of trials

In this section I compare the effect of changing the number of trials at each stimulus level. I keep the number of stimulus levels constant at $n = 10$. As when comparing the change in n in section 4.3.4 I compare only the P optimal bandwidths. I consider the values $r = 1, 5, 10, 20$.

Figures 4.13-4.16 show a comparison of the square bias, variance and MSE for functions $p_2(x)$-$p_5(x)$ when an MISE optimal bandwidth is used for the unconstrained regression. Figure 4.13 shows that for $p_2(x)$ with the MISE optimal bandwidth the LDNP method performs well at all values of r. The bandwidth method gets progressively worse whilst the PAVA method becomes more similar to the LDNP method with and increase in r. The bandwidth method has a much larger bias than the other two methods, although generally smaller variance.

Figure 4.14 shows that for $p_3(x)$ with the MISE optimal bandwidth, for small values of r the LDNP method performs well. As n increases the bandwidth adjustment method performs well, largely due to its lower variance. The PAVA estimate has high bias and variance at the boundaries causing poor performance in this case. For $p_4(x)$ Figure 4.15 shows that the bandwidth method is the best to choose from when the MISE optimal bandwidth is used in the unconstrained regressions step. Both the PAVA and LDNP methods have large peaks in the bias plots at approximately $x = 0.4$ and $x = 0.6$ and a peak in the variance at approximately $x = 0.5$. This is largely the reason for their poor performance for this function. For $p_5(x)$ Figure 4.16 shows that the bandwidth method is the best to choose from when the MISE optimal bandwidth is used in the unconstrained regressions step although this becomes less clear as r increases as the methods become more similar.

The ISE ratios when using an ISE optimal bandwidth for the unconstrained regression can be compared in the first column of Figures 4.17-4.22. These show that the bandwidth adjustment method performs poorly when a good bandwidth is chosen. The density estimates for the bandwidth method get flatter as the number of trials increases indicating increased variability and poorer performance. When the number of trials is low then the LDNP estimate is often clearly the best method. As the number of trials increases the PAVA estimate and the LDNP method become much more similar and tightly bound about 1.

The second column of Figures 4.17-4.22 compare the ISE ratios when using a cross-validation bandwidth for the unconstrained regression. Although the bandwidth method generally does well for this choice of unconstrained bandwidth it is noticeable that as the number of trials increases the density estimate often gets very flat indicating poor performance. When the number of trials is low the cross-validation bandwidth is too small and so the best monotonicity method is to increase the bandwidth. This is as was seen in comparing the effect of changing n in section 4.3.4. The LDNP method and PAVA method are broadly similar although the LDNP method is often slightly better.

In this chapter I have investigated various methods of enforcing monotonicity on a nonparametric regression estimate. In general it has been seen that if a poor choice of initial bandwidth

has been chosen then the simplest thing to do is simply to increase the bandwidth. However, if a good initial bandwidth has been chosen then this is no longer a good option. In this case I recommend the use of the LDNP bandwidth since it has been shown in the simulations in this chapter to work well in a variety of conditions. The LDNP method produces a smooth estimate (as opposed to a stepwise function in the case of the PAVA method), and is easy to compute. These reasons make the LDNP method attractive for use in practical situations.

4.4 Figures and Tables

Table 4.1: Number of samples requiring monotonisation for $p_1(x)$ for each of four different bandwidths. The number of unique samples is given in brackets. In this table, and also in Tables 4.2-4.6, a P in the subscript of the bandwidth refers to the fact that the bandwidth is chosen for the estimation of $P(x)$, whilst η in the subscript refers to the fact that the bandwidth is chosen for estimating $\eta(x)$.

Model	$h_{P,cv}$	$h_{\eta,cv}$	$h_{P,ISEP}$	$h_{\eta,ISE}$
$p_1(x)$ **{n=5, r=5}**	28 (4)	1 (1)	4 (4)	0
$p_1(x)$ **{n=5, r=10}**	39 (5)	0	63 (9)	0
$p_1(x)$ **{n=5, r=20}**	47 (15)	0	40 (14)	0
$p_1(x)$ **{n=10, r=1}**	522 (23)	5 (5)	41 (6)	1 (1)
$p_1(x)$ **{n=10, r=5}**	331 (82)	0	91 (47)	0
$p_1(x)$ **{n=10, r=10}**	191 (111)	0	133 (77)	0
$p_1(x)$ **{n=10, r=20}**	118 (115)	2 (2)	75 (58)	0
$p_1(x)$ **{n=20, r=1}**	1946 (98)	0	66 (36)	0
$p_1(x)$ **{n=20, r=5}**	627 (533)	0	131 (127)	0
$p_1(x)$ **{n=20, r=10}**	480 (480)	7 (7)	124 (124)	0
$p_1(x)$ **{n=20, r=20}**	448 (448)	20 (20)	99 (99)	0
$p_1(x)$ **{n=50, r=1}**	636 (467)	3 (3)	132 (124)	0
$p_1(x)$ **{n=50, r=5}**	503 (503)	9 (9)	75 (75)	0
$p_1(x)$ **{n=50, r=10}**	546 (546)	16 (16)	73 (73)	0
$p_1(x)$ **{n=50, r=20}**	482 (482)	42 (42)	70 (70)	0

Table 4.2: Number of samples requiring monotonisation for $p_2(x)$ for each of six different bandwidths. The number of unique samples is given in brackets.

Model	$h_{P,MISE}$	$h_{\eta MISE}$	$h_{P,cv}$	$h_{\eta,cv}$	$h_{P,ISE}$	$h_{\eta,ISE}$
$p_2(x)$ **{n=5 r=5}**	878 (24)	0	69 (11)	0	326 (2)	0
$p_2(x)$ **{n=5 r=10}**	556 (62)	0	84 (19)	0	181 (11)	0
$p_2(x)$ **{n=5 r=20}**	362 (69)	0	54 (23)	0	66 (19)	0
$p_2(x)$ **{n=10 r=1}**	2656 (35)	5 (2)	1487 (19)	28 (3)	703 (5)	28 (3)
$p_2(x)$ **{n=10 r=5}**	3410 (656)	0	1805 (258)	0	613 (66)	0
$p_2(x)$ **{n=10 r=10}**	3179 (1428)	0	1478 (586)	0	1136 (393)	0
$p_2(x)$ **{n=10 r=20}**	2817 (2224)	0	1347 (1023)	0	1395 (1051)	0
$p_2(x)$ **{n=20 r=1}**	3343 (392)	1 (1)	2386 (190)	1 (1)	279 (12)	1 (1)
$p_2(x)$ **{n=20 r=5}**	3791 (3630)	0	2118 (1991)	0	1259 (1177)	0
$p_2(x)$ **{n=20 r=10}**	3734 (3729)	0	2225 (2221)	0	1905 (1901)	0
$p_2(x)$ **{n=20 r=20}**	3681 (3681)	0	2486 (2486)	0	2550 (2550)	0
$p_2(x)$ **{n=50 r=1}**	3737 (3635)	0	1640 (1596)	0	986 (938)	0
$p_2(x)$ **{n=50 r=5}**	3848 (3848)	0	2393 (2393)	0	2089 (2089)	0
$p_2(x)$ **{n=50 r=10}**	3943 (3943)	0	2819 (2819)	0	2771 (2771)	0
$p_2(x)$ **{n=50 r=20}**	4012 (4012)	0	3315 (3315)	0	3236 (3236)	0

Table 4.3: Number of samples requiring monotonisation for $p_3(x)$ for each of six different bandwidths. The number of unique samples is given in brackets.

Model	$h_{P,MISE}$	$h_{\eta MISE}$	$h_{P,cv}$	$h_{\eta,cv}$	$h_{P,ISE}$	$h_{\eta,ISE}$
$p_3(x)$ **{n=5 r=5}**	3497 (841)	3528 (856)	1137 (290)	614 (205)	286 (143)	69 (58)
$p_3(x)$ **{n=5 r=10}**	2602 (1570)	2651 (1608)	578 (345)	302 (199)	136 (103)	6 (6)
$p_3(x)$ **{n=5 r=20}**	1574 (1413)	1595 (1431)	256 (220)	129 (107)	60 (53)	0
$p_3(x)$ **{n=10 r=1}**	3892 (642)	3924 (646)	3702 (491)	1583 (426)	1438 (457)	946 (389)
$p_3(x)$ **{n=10 r=5}**	3523 (3498)	3699 (3673)	1390 (1379)	678 (675)	87 (87)	10 (10)
$p_3(x)$ **{n=10 r=10}**	2756 (2756)	2933 (2933)	959 (959)	506 (506)	15 (15)	0
$p_3(x)$ **{n=10 r=20}**	1937 (1937)	2077 (2077)	503 (503)	260 (260)	11 (11)	0
$p_3(x)$ **{n=20 r=1}**	4440 (4259)	4556 (4370)	2110 (1942)	712 (707)	798 (796)	345 (345)
$p_3(x)$ **{n=20 r=5}**	3071 (3071)	3249 (3249)	1020 (1020)	468 (468)	11 (11)	0
$p_3(x)$ **{n=20 r=10}**	2457 (2457)	2621 (2621)	788 (788)	469 (469)	4 (4)	0
$p_3(x)$ **{n=20 r=20}**	1854 (1854)	1989 (1989)	557 (557)	361 (361)	2 (2)	0
$p_3(x)$ **{n=50 r=1}**	3794 (3794)	3972 (3972)	1343 (1343)	235 (235)	177 (177)	20 (20)
$p_3(x)$ **{n=50 r=5}**	2484 (2484)	2623 (2623)	709 (709)	396 (396)	6 (6)	0
$p_3(x)$ **{n=50 r=10}**	1862 (1862)	2009 (2009)	513 (513)	348 (348)	0	0
$p_3(x)$ **{n=50 r=20}**	1256 (1256)	1396 (1396)	329 (329)	216 (216)	0	0

Table 4.4: Number of samples requiring monotonisation for $p_4(x)$ for each of six different bandwidths. The number of unique samples is given in brackets.

Model	$h_{P,MISE}$	$h_{\eta MISE}$	$h_{P,cv}$	$h_{\eta,cv}$	$h_{P,ISE}$	$h_{\eta,ISE}$
$p_4(x)$ **{n=5 r=5}**	114 (5)	0	0	0	0	0
$p_4(x)$ **{n=5 r=10}**	9 (2)	0	0	0	0	0
$p_4(x)$ **{n=5 r=20}**	1 (1)	1 (1)	0	0	0	0
$p_4(x)$ **{n=10 r=1}**	1517 (21)	1 (1)	297 (12)	0	0	0
$p_4(x)$ **{n=10 r=5}**	1387 (185)	1 (1)	181 (30)	0	43 (10)	0
$p_4(x)$ **{n=10 r=10}**	1506 (408)	4 (4)	42 (29)	0	19 (8)	0
$p_4(x)$ **{n=10 r=20}**	1126 (755)	3 (3)	14 (13)	0	3 (3)	0
$p_4(x)$ **{n=20 r=1}**	1645 (183)	0	1452 (67)	0	22 (9)	0
$p_4(x)$ **{n=20 r=5}**	2279 (1813)	0	558 (368)	0	19 (18)	0
$p_4(x)$ **{n=20 r=10}**	2547 (2489)	7 (7)	282 (275)	0	19 (18)	0
$p_4(x)$ **{n=20 r=20}**	2820 (2812)	15 (15)	254 (251)	0	16 (16)	0
$p_4(x)$ **{n=50 r=1}**	2374 (1949)	10 (10)	728 (397)	0	70 (67)	0
$p_4(x)$ **{n=50 r=5}**	2440 (2440)	14 (14)	636 (636)	0	18 (18)	0
$p_4(x)$ **{n=50 r=10}**	2681 (2681)	17 (17)	562 (562)	0	14 (14)	0
$p_4(x)$ **{n=50 r=20}**	3075 (3075)	25 (25)	574 (574)	0	64 (64)	0

Table 4.5: Number of samples requiring monotonisation for $p_5(x)$ for each of six different bandwidths. The number of unique samples is given in brackets.

Model	$h_{P,MISE}$	$h_{\eta MISE}$	$h_{P,cv}$	$h_{\eta,cv}$	$h_{P,ISE}$	$h_{\eta,ISE}$
$p_5(x)$ **{n=5 r=5}**	9499 (1162)	9499 (1162)	2453 (364)	2041 (301)	2014 (357)	529 (170)
$p_5(x)$ **{n=5 r=10}**	8789 (2924)	8789 (2924)	2119 (754)	1990 (682)	2046 (779)	229 (136)
$p_5(x)$ **{n=5 r=20}**	7770 (5249)	7770 (5249)	1677 (1196)	1714 (1230)	1851 (1363)	72 (65)
$p_5(x)$ **{n=10 r=1}**	9484 (673)	9484 (673)	4450 (411)	2735 (353)	3115 (391)	2105 (324)
$p_5(x)$ **{n=10 r=5}**	9896 (9749)	9896 (9749)	3029 (2990)	2446 (2408)	3160 (3110)	450 (448)
$p_5(x)$ **{n=10 r=10}**	9871 (9866)	9917 (9912)	3004 (3004)	2849 (2849)	3590 (3587)	153 (153)
$p_5(x)$ **{n=10 r=20}**	9682 (9682)	9743 (9743)	2714 (2714)	3045 (3045)	3803 (3803)	35 (35)
$p_5(x)$ **{n=20 r=1}**	9916 (9661)	9967 (9706)	3430 (3357)	2140 (2116)	3202 (3166)	1554 (1546)
$p_5(x)$ **{n=20 r=5}**	9990 (9990)	9992 (9992)	3042 (3042)	2383 (2383)	3610 (3610)	206 (206)
$p_5(x)$ **{n=20 r=10}**	9984 (9984)	9984 (9984)	3318 (3318)	3277 (3277)	3979 (3979)	55 (55)
$p_5(x)$ **{n=20 r=20}**	9937 (9937)	9937 (9937)	3480 (3480)	4221 (4221)	4307 (4307)	3 (3)
$p_5(x)$ **{n=50 r=1}**	9963 (9963)	9969 (9969)	2910 (2910)	1252 (1252)	3414 (3414)	706 (706)
$p_5(x)$ **{n=50 r=5}**	9971 (9971)	9971 (9971)	3195 (3195)	2485 (2485)	4191 (4191)	32 (32)
$p_5(x)$ **{n=50 r=10}**	9947 (9947)	9947 (9947)	3808 (3808)	4034 (4034)	4618 (4618)	3 (3)
$p_5(x)$ **{n=50 r=20}**	9950 (9950)	9950 (9950)	4349 (4349)	5604 (5604)	5012 (5012)	0

Table 4.6: Number of samples requiring monotonisation for $p_6(x)$ for each of four different bandwidths. The number of unique samples is given in brackets.

Model	$h_{P,cv}$	$h_{\eta,cv}$	$h_{P,ISE}$	$h_{\eta,ISE}$
$p_6(x)$ **{n=5 r=5}**	1330 (97)	260 (56)	1029 (103)	42 (20)
$p_6(x)$ **{n=5 r=10}**	890 (233)	241 (98)	678 (213)	4 (4)
$p_6(x)$ **{n=5 r=20}**	583 (328)	190 (119)	457 (297)	0
$p_6(x)$ **{n=10 r=1}**	3312 (126)	709 (82)	1229 (117)	295 (65)
$p_6(x)$ **{n=10 r=5}**	2047 (1677)	147 (146)	2102 (1751)	1 (1)
$p_6(x)$ **{n=10 r=10}**	1944 (1916)	88 (88)	2302 (2273)	0
$p_6(x)$ **{n=10 r=20}**	2022 (2021)	84 (84)	1990 (1989)	0
$p_6(x)$ **{n=20 r=1}**	2620 (1453)	135 (129)	1717 (1296)	51 (47)
$p_6(x)$ **{n=20 r=5}**	2849 (2849)	30 (30)	2899 (2899)	0
$p_6(x)$ **{n=20 r=10}**	3096 (3096)	21 (21)	2755 (2755)	0
$p_6(x)$ **{n=20 r=20}**	3121 (3121)	12 (12)	2387 (2387)	0
$p_6(x)$ **{n=50 r=1}**	2425 (2425)	3 (3)	2414 (2414)	3 (3)
$p_6(x)$ **{n=50 r=5}**	3424 (3424)	2 (2)	3074 (3074)	0
$p_6(x)$ **{n=50 r=10}**	3529 (3529)	0	2760 (2760)	0
$p_6(x)$ **{n=50 r=20}**	3412 (3412)	1 (1)	2325 (2325)	0

Table 4.7: MISE values for model $p_2(x)$ for various monotonicity constraints with the two MISE optimal unconstrained bandwidths. Here addthe column MISEP represents the MISE value when the MISE optimal bandwidth for $P(x)$ was chosen, whilst the column MISEETA represents the MISE value when the MISE optimal bandwidth for $\eta(x)$ was chosen. The number of stimulus levels is denoted n and the number of trials was denoted r.

	$p_2(x)$ {n=5 r=5}		$p_2(x)$ {n=5 r=10}		$p_2(x)$ {n=5 r=20}		$p_2(x)$ {n=10 r=1}		$p_2(x)$ {n=10 r=5}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.016097 (0.0149)	NA	0.012485	NA	0.006574	NA	0.034801	0.085037	0.008438	NA
Band	0.013638 (0.0115)	NA	0.009452	NA	0.004573	NA	0.026906	0.097307	0.007013	NA
LDNP	0.011711 (0.0131)	NA	0.010275	NA	0.007326	NA	0.022016	0.05541	0.007581	NA
LDNP1	0.012162 (0.0136)	NA	0.010358	NA	0.00708	NA	0.02318	0.041466	0.0076	NA
LDNP2	0.012477 (0.0139)	NA	0.010491	NA	0.006243	NA	0.024318	0.003469	0.007751	NA
LDNP3	0.01253 (0.0139)	NA	0.010519	NA	0.00584	NA	0.024541	0.042055	0.00779	NA
	$p_2(x)$ {n=10 r=10}		$p_2(x)$ {n=10 r=20}		$p_2(x)$ {n=20 r=1}		$p_2(x)$ {n=20 r=5}		$p_2(x)$ {n=20 r=10}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.004361 (0.0039)	NA	0.00227	NA	0.020621	0.097103	0.004415	NA	0.002359	NA
Band	0.00412 (0.0040)	NA	0.002583	NA	0.016132	0.100265	0.003925	NA	0.00247	NA
LDNP	0.004296 (0.0034)	NA	0.00258	NA	0.015443	0.08179	0.004105	NA	0.00246	NA
LDNP1	0.004074 (0.0036)	NA	0.002219	NA	0.016193	0.073788	0.003971	NA	0.002206	NA
LDNP2	0.004106 (0.0038)	NA	0.002167	NA	0.01691	0.039872	0.004065	NA	0.002216	NA
LDNP3	0.00412 (0.0038)	NA	0.00217	NA	0.017049	0.001515	0.004087	NA	0.002224	NA
	$p_2(x)$ {n=20 r=20}		$p_2(x)$ {n=50 r=1}		$p_2(x)$ {n=50 r=5}		$p_2(x)$ {n=50 r=10}		$p_2(x)$ {n=50 r=20}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.001242 (0.00102)	NA	0.008682	NA	0.002021	NA	0.001095	NA	0.000603	NA
Band	0.001695 (0.00115)	NA	0.007224	NA	0.002139	NA	0.001555	NA	0.001213	NA
LDNP	0.001549 (0.00089)	NA	0.007324	NA	0.002104	NA	0.001338	NA	0.000933	NA
LDNP1	0.001222 (0.00092)	NA	0.007445	NA	0.001869	NA	0.001057	NA	0.000624	NA
LDNP2	0.001188 (0.00098)	NA	0.007679	NA	0.001887	NA	0.00104	NA	0.000587	NA
LDNP3	0.00119 (0.00099)	NA	0.007725	NA	0.001895	NA	0.001042	NA	0.000587	NA

Table 4.8: MISE values for model $p_3(x)$ for various monotonicity constraints with the two MISE optimal unconstrained bandwidths. Notation is as in 4.7.

	$p_3(x)$ {n=5 r=5}		$p_3(x)$ {n=5 r=10}		$p_3(x)$ {n=5 r=20}		$p_3(x)$ {n=10 r=1}		$p_3(x)$ {n=10 r=5}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.020617	0.020825	0.010666	0.010818	0.005918	0.006029	0.044744	0.045426	0.011018	0.011047
Band	0.016405	0.01644	0.00821	0.008246	0.004145	0.004151	0.042492	0.042654	0.008461	0.008475
LDNP	0.017308	0.017475	0.008998	0.009107	0.005059	0.005131	0.031615	0.032063	0.008902	0.008958
LDNP1	0.018323	0.01849	0.00968	0.009807	0.005481	0.00557	0.033382	0.033824	0.009688	0.009734
LDNP2	0.019098	0.019237	0.010138	0.010261	0.005735	0.005828	0.035107	0.035508	0.01023	0.010248
LDNP3	0.019324	0.019454	0.010232	0.010351	0.005758	0.005852	0.035773	0.036167	0.010336	0.010346
	$p_3(x)$ {n=10 r=10}		$p_3(x)$ {n=10 r=20}		$p_3(x)$ {n=20 r=1}		$p_3(x)$ {n=20 r=5}		$p_3(x)$ {n=20 r=10}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.006094	0.006094	0.003023	0.003042	0.024369	0.024433	0.00607	0.006069	0.00322	0.003227
Band	0.004455	0.004446	0.002033	0.00204	0.021022	0.020944	0.00438	0.00436	0.002133	0.002134
LDNP	0.00503	0.005033	0.002515	0.002541	0.01867	0.018708	0.004988	0.004998	0.002634	0.002648
LDNP1	0.005514	0.005519	0.00279	0.002815	0.019949	0.019972	0.005466	0.005473	0.002939	0.002952
LDNP2	0.005809	0.005801	0.002924	0.00294	0.020975	0.020968	0.005747	0.005737	0.00309	0.003093
LDNP3	0.005847	0.005838	0.00293	0.002946	0.021296	0.021278	0.00578	0.005768	0.003103	0.003105
	$p_3(x)$ {n=20 r=20}		$p_3(x)$ {n=50 r=1}		$p_3(x)$ {n=50 r=5}		$p_3(x)$ {n=50 r=10}		$p_3(x)$ {n=50 r=20}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.001784	0.00179	0.011351	0.011344	0.002769	0.002777	0.001527	0.001526	0.000828	0.000834
Band	0.001077	0.00108	0.008643	0.008617	0.001764	0.001762	0.000889	0.000885	0.000447	0.000451
LDNP	0.001462	0.001472	0.009013	0.009023	0.002238	0.00225	0.001251	0.001254	0.000688	0.000696
LDNP1	0.001654	0.001664	0.009805	0.009805	0.002518	0.002529	0.001423	0.001424	0.000791	0.000799
LDNP2	0.001726	0.001731	0.010338	0.010309	0.002649	0.002651	0.001477	0.001474	0.000807	0.000813
LDNP3	0.001731	0.001735	0.010437	0.0104	0.002661	0.002663	0.001481	0.001478	0.000809	0.000815

Table 4.9: MISE values for model $p_4(x)$ for various monotonicity constraints with the two MISE optimal unconstrained bandwidths. Notation is as in 4.7.

	$p_4(x)$ {n=5 r=5}		$p_4(x)$ {n=5 r=10}		$p_4(x)$ {n=5 r=20}		$p_4(x)$ {n=10 r=1}		$p_4(x)$ {n=10 r=5}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.015518	NA	0.014365	NA	0.014334	0.006209	0.027572	0.03887	0.013162	0.066862
Band	0.009858	NA	0.016049	NA	0.001762	0.001879	0.015859	0.039088	0.004178	0.045614
LDNP	0.014019	NA	0.013658	NA	0.013748	0.006262	0.018374	0.030068	0.011015	0.063876
LDNP1	0.015069	NA	0.014141	NA	0.013725	0.006231	0.018945	0.030848	0.011312	0.065624
LDNP2	0.0154	NA	0.014287	NA	0.013718	0.006215	0.019155	0.03517	0.011374	0.066542
LDNP3	0.015415	NA	0.014293	NA	0.013717	0.006216	0.019176	0.03747	0.011378	0.06668
	$p_4(x)$ {n=10 r=10}		$p_4(x)$ {n=10 r=20}		$p_4(x)$ {n=20 r=1}		$p_4(x)$ {n=20 r=5}		$p_4(x)$ {n=20 r=10}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.007612	0.01031	0.003399	0.001477	0.021494	NA	0.005801	NA	0.002906	0.002096
Band	0.002438	0.005746	0.001513	0.000895	0.009113	NA	0.002464	NA	0.001349	0.001457
LDNP	0.006753	0.010103	0.003133	0.001486	0.015968	NA	0.005248	NA	0.002739	0.002127
LDNP1	0.006877	0.010223	0.003192	0.001464	0.016516	NA	0.005432	NA	0.00281	0.002103
LDNP2	0.006896	0.010264	0.003203	0.001473	0.016668	NA	0.005466	NA	0.002822	0.002094
LDNP3	0.006896	0.010269	0.003204	0.001474	0.016679	NA	0.005467	NA	0.002822	0.002093
	$p_4(x)$ {n=20 r=20}		$p_4(x)$ {n=50 r=1}		$p_4(x)$ {n=50 r=5}		$p_4(x)$ {n=50 r=10}		$p_4(x)$ {n=50 r=20}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.001498	0.001597	0.010982	0.012817	0.002251	0.001518	0.001176	0.001052	0.00061	0.000295
Band	0.000725	0.001186	0.004971	0.006806	0.001108	0.001077	0.000573	0.000865	0.000306	0.000285
LDNP	0.001446	0.001643	0.009424	0.012441	0.002141	0.001536	0.001138	0.001081	0.0006	0.000325
LDNP1	0.001471	0.001611	0.009752	0.012649	0.002187	0.001522	0.001156	0.00106	0.000605	0.000303
LDNP2	0.001474	0.001597	0.009821	0.012753	0.002194	0.001517	0.001158	0.001052	0.000605	0.000295
LDNP3	0.001474	0.001595	0.009825	0.012779	0.002194	0.001517	0.001158	0.001052	0.000605	0.000295

Table 4.10: MISE values for model $p_5(x)$ for various monotonicity constraints with the two MISE optimal unconstrained bandwidths. Notation is as in 4.7.

	$p_5(x)$ {n=5 r=5}		$p_5(x)$ {n=5 r=10}		$p_5(x)$ {n=5 r=20}		$p_5(x)$ {n=10 r=1}		$p_5(x)$ {n=10 r=5}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.046493	0.049657	0.021652	0.023426	0.014785	0.016301	0.118549	0.152174	0.016991	0.024408
Band	0.019054	0.019354	0.011356	0.011492	0.007528	0.007614	0.041165	0.041675	0.01018	0.010335
LDNP	0.042705	0.046251	0.023661	0.025148	0.016409	0.01752	0.086457	0.114652	0.01648	0.023643
LDNP1	0.041377	0.044907	0.022619	0.024287	0.01544	0.016739	0.086171	0.114493	0.016085	0.023125
LDNP2	0.039634	0.043219	0.021293	0.022776	0.014554	0.015752	0.086139	0.114429	0.015809	0.022598
LDNP3	0.03963	0.043217	0.021272	0.022773	0.014533	0.015728	0.086134	0.114427	0.015762	0.022507
	$p_5(x)$ {n=10 r=10}		$p_5(x)$ {n=10 r=20}		$p_5(x)$ {n=20 r=1}		$p_5(x)$ {n=20 r=5}		$p_5(x)$ {n=20 r=10}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.00897	0.010818	0.005288	0.00589	0.055141	0.088327	0.008459	0.009982	0.004727	0.005305
Band	0.005872	0.005948	0.003686	0.00371	0.023524	0.024059	0.005781	0.005879	0.003377	0.003409
LDNP	0.009044	0.011283	0.00554	0.006236	0.041538	0.060328	0.007882	0.009829	0.004293	0.005055
LDNP1	0.008611	0.010918	0.005148	0.005949	0.041138	0.059952	0.007752	0.009609	0.004234	0.004936
LDNP2	0.008364	0.010476	0.004965	0.005736	0.04094	0.059798	0.007625	0.009445	0.004156	0.004836
LDNP3	0.008326	0.010369	0.004942	0.005732	0.040916	0.059787	0.007611	0.009433	0.004149	0.00483
	$p_5(x)$ {n=20 r=20}		$p_5(x)$ {n=50 r=1}		$p_5(x)$ {n=50 r=5}		$p_5(x)$ {n=50 r=10}		$p_5(x)$ {n=50 r=20}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	0.002756	0.003052	0.016957	0.023975	0.00396	0.004391	0.002301	0.002527	0.001356	0.001485
Band	0.002128	0.002124	0.010849	0.011027	0.002903	0.002934	0.001816	0.001808	0.001219	0.001192
LDNP	0.002482	0.002814	0.015456	0.020654	0.003506	0.004053	0.002011	0.00224	0.001192	0.001303
LDNP1	0.002466	0.002758	0.015206	0.020327	0.003487	0.003975	0.002017	0.002209	0.001208	0.001299
LDNP2	0.002421	0.002702	0.015031	0.020129	0.003431	0.003899	0.001984	0.002167	0.001191	0.001278
LDNP3	0.002417	0.002699	0.01501	0.020114	0.003426	0.003895	0.001982	0.002165	0.00119	0.001277

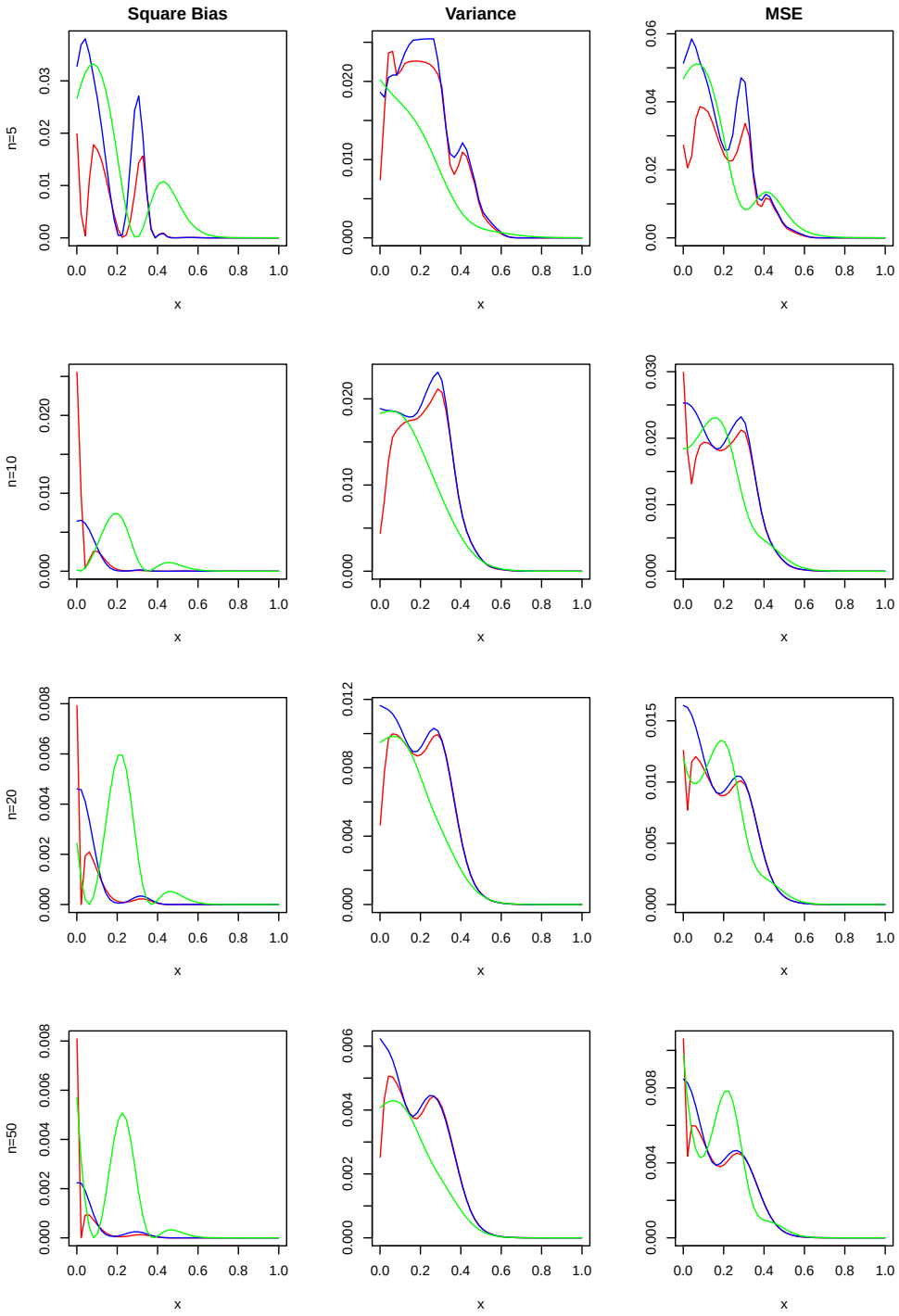

Figure 4.3: A comparison of the square bias, variance and MSE of the three monotone estimates of $p_2(x)$ with $r = 5$ and h_r set as $h_{P,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

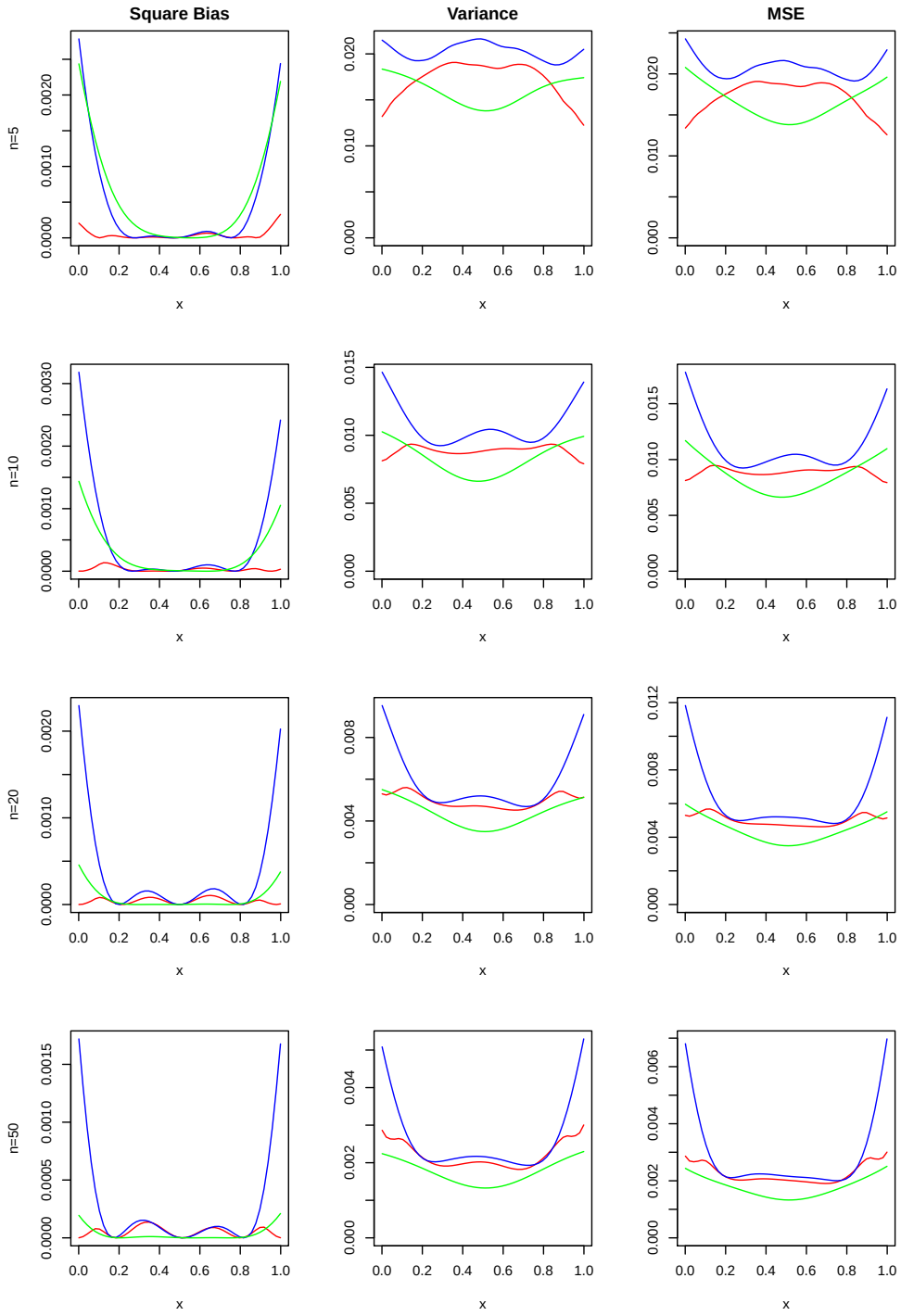

Figure 4.4: A comparison of the square bias, variance and MSE of the three monotone estimates of $p_3(x)$ with $r = 5$ and h_r set as $h_{P,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

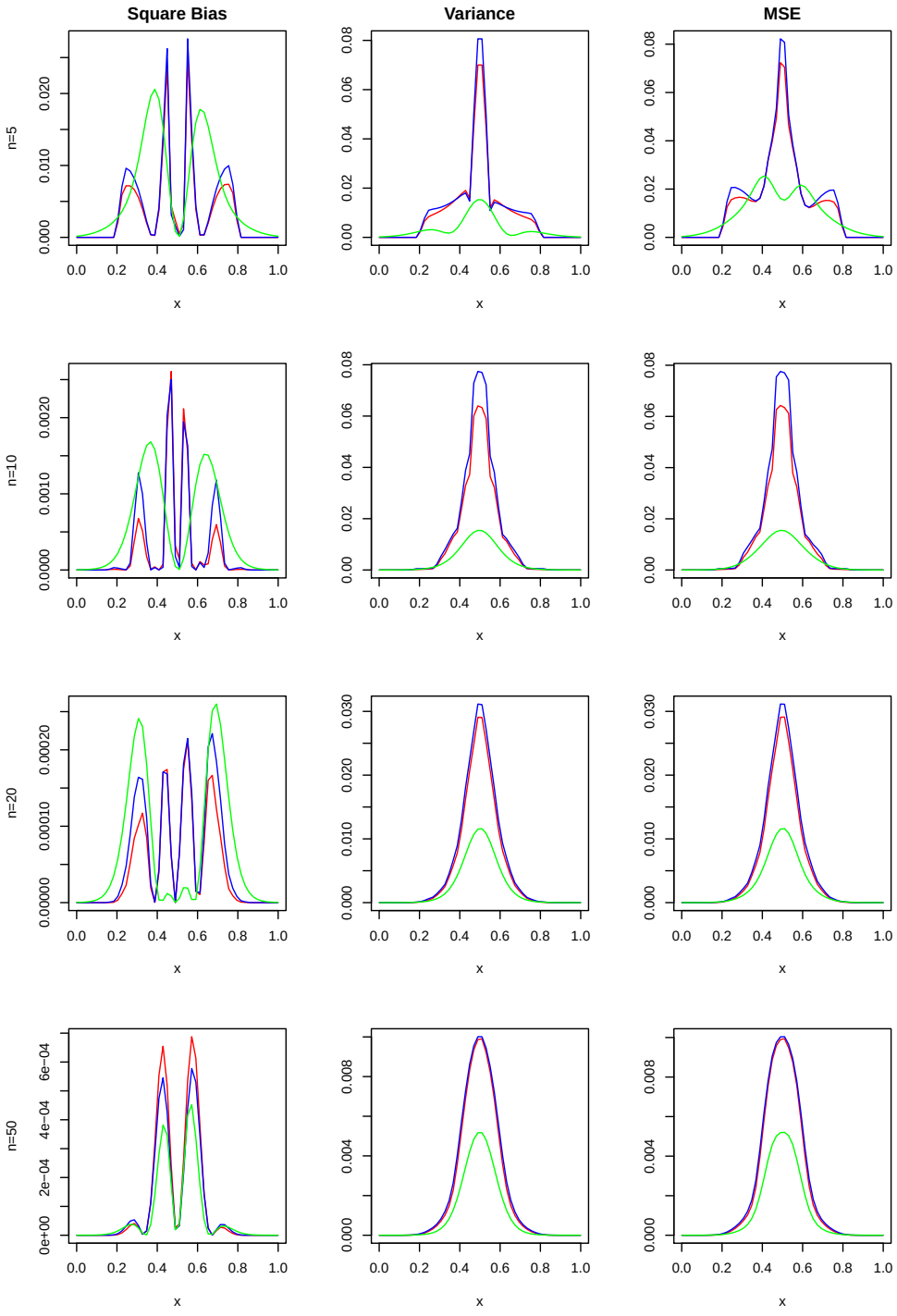

Figure 4.5: A comparison of the square bias, variance and MSE of the three monotone estimates of $p_4(x)$ with $r = 5$ and h_r set as $h_{P,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

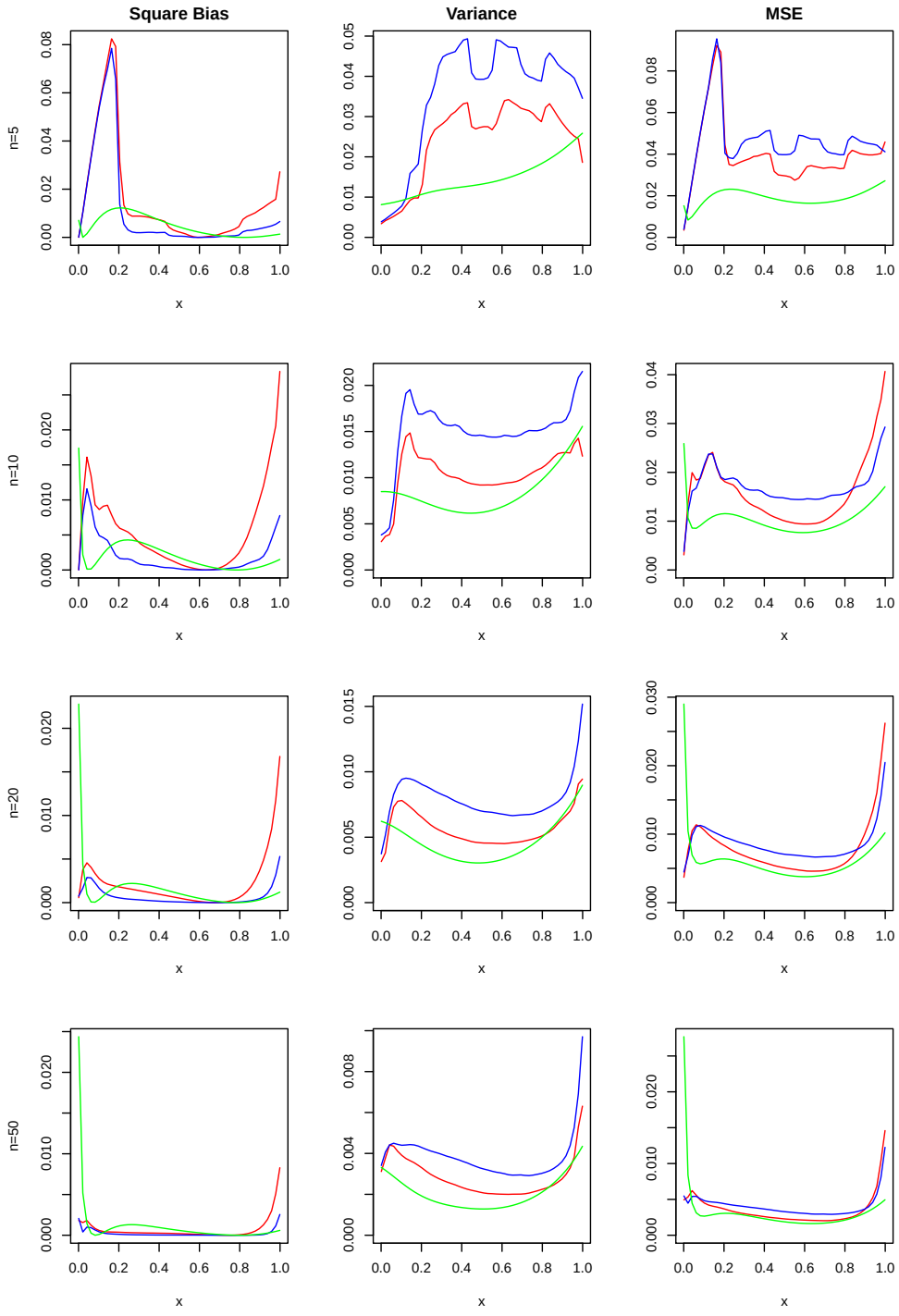

Figure 4.6: A comparison of the square bias, variance and MSE of the three monotone estimates of $p_5(x)$ with $r = 5$ and h_r set as $h_{P,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

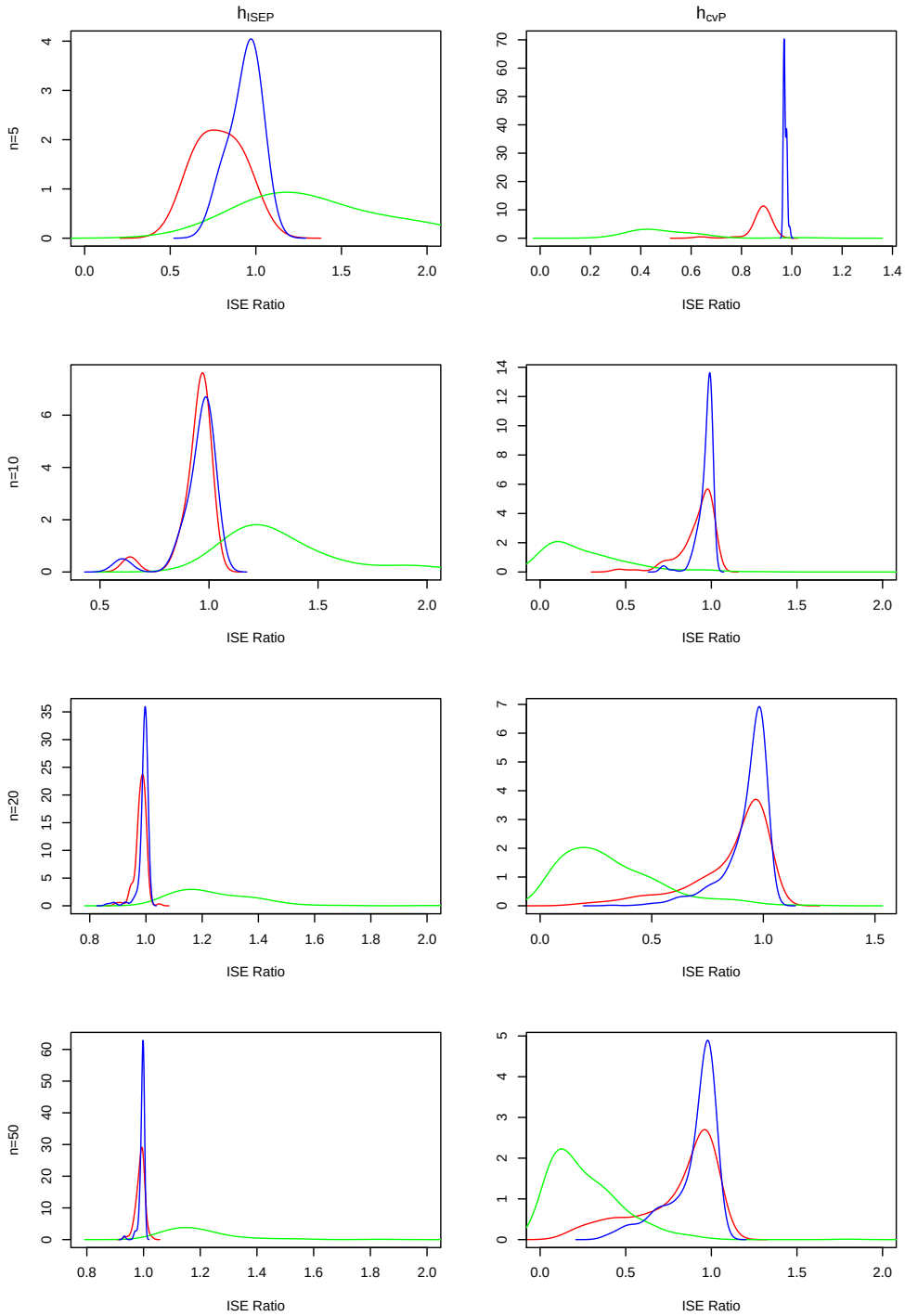

Figure 4.7: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_1(x)$ with $r = 5$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

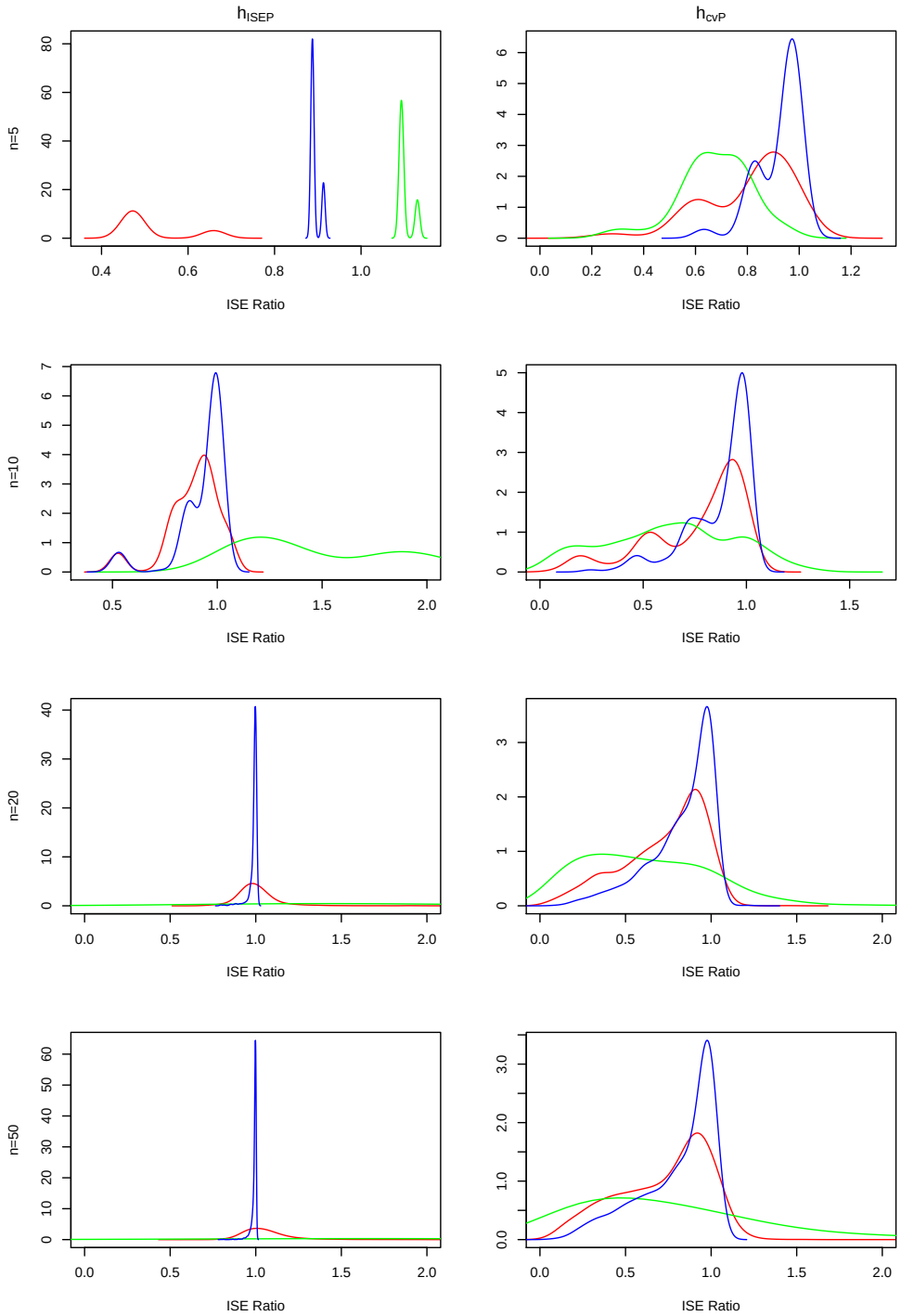

Figure 4.8: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_2(x)$ with $r = 5$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

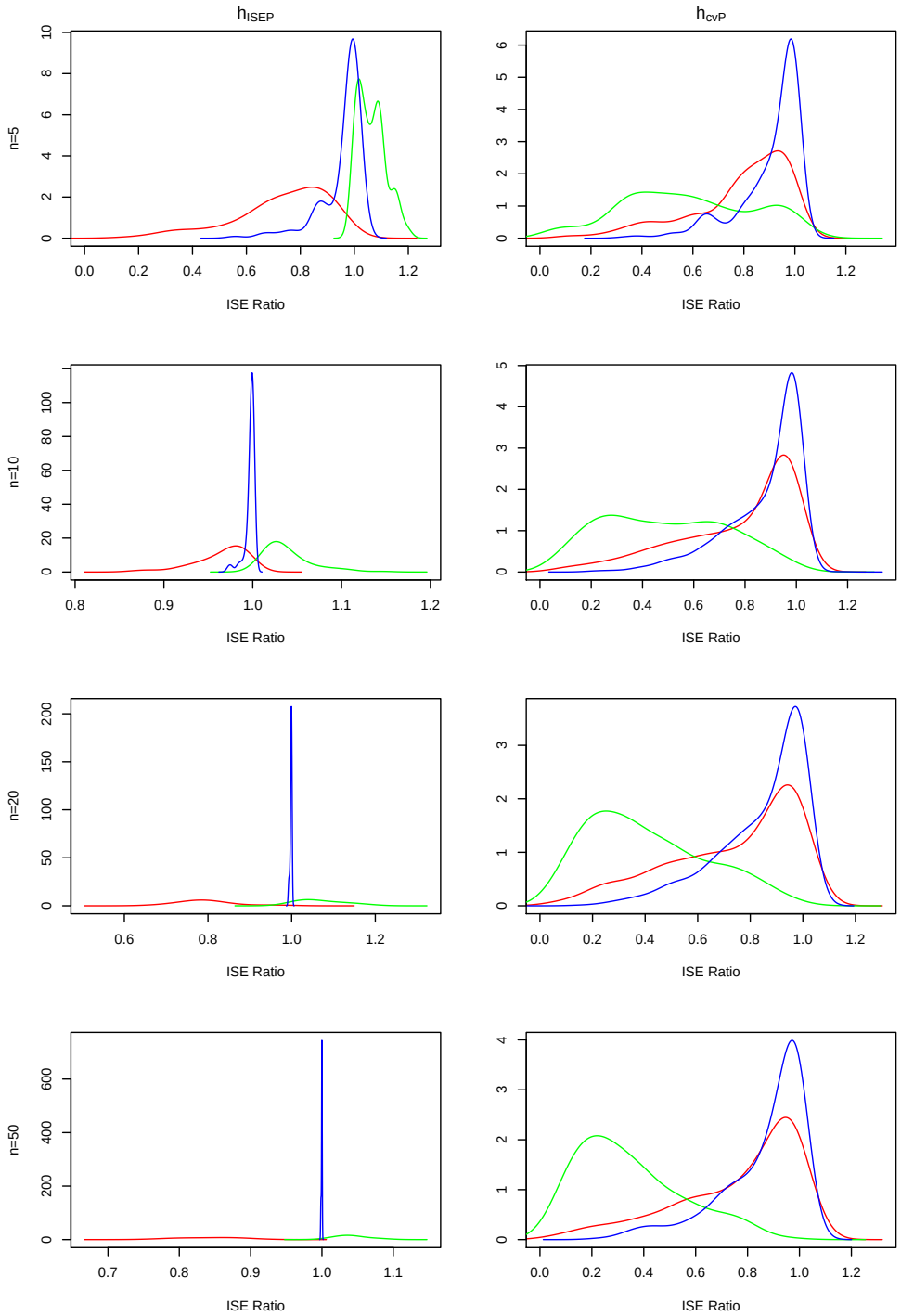

Figure 4.9: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_3(x)$ with $r = 5$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

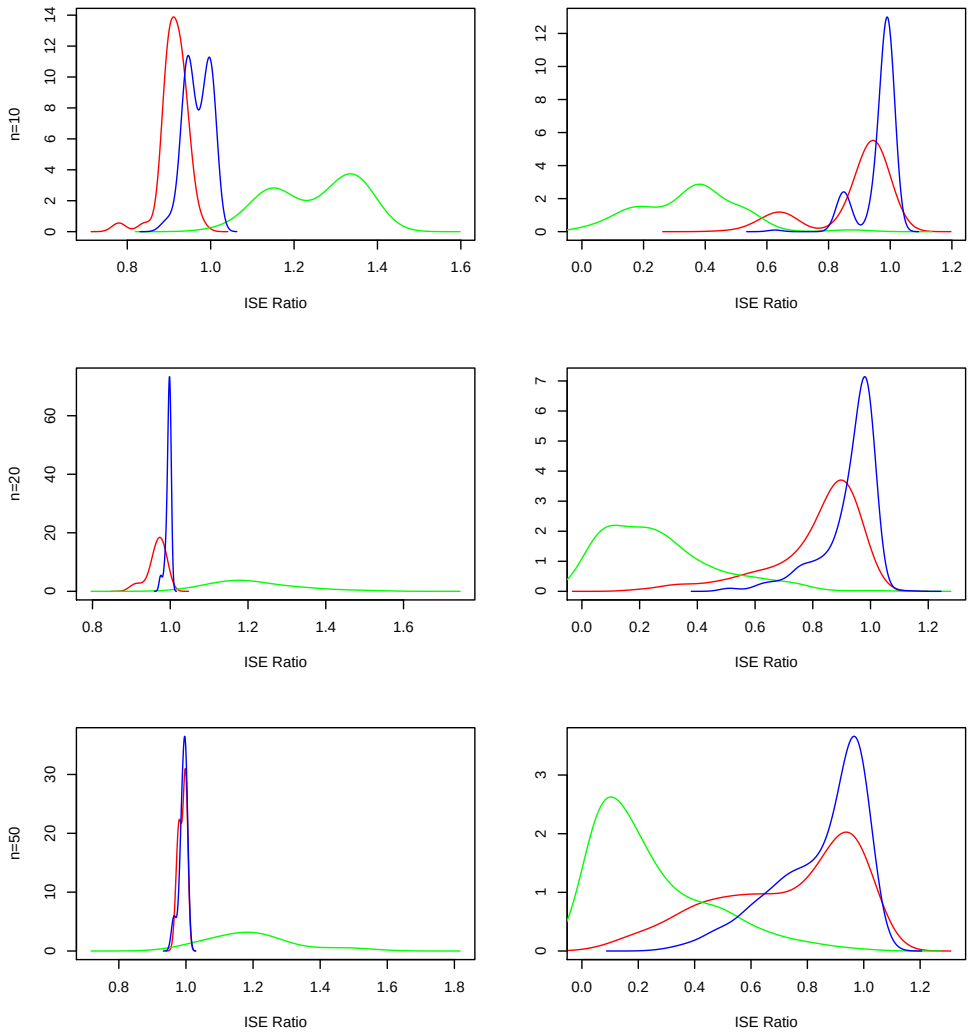

Figure 4.10: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_4(x)$ with $r = 5$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $n = 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

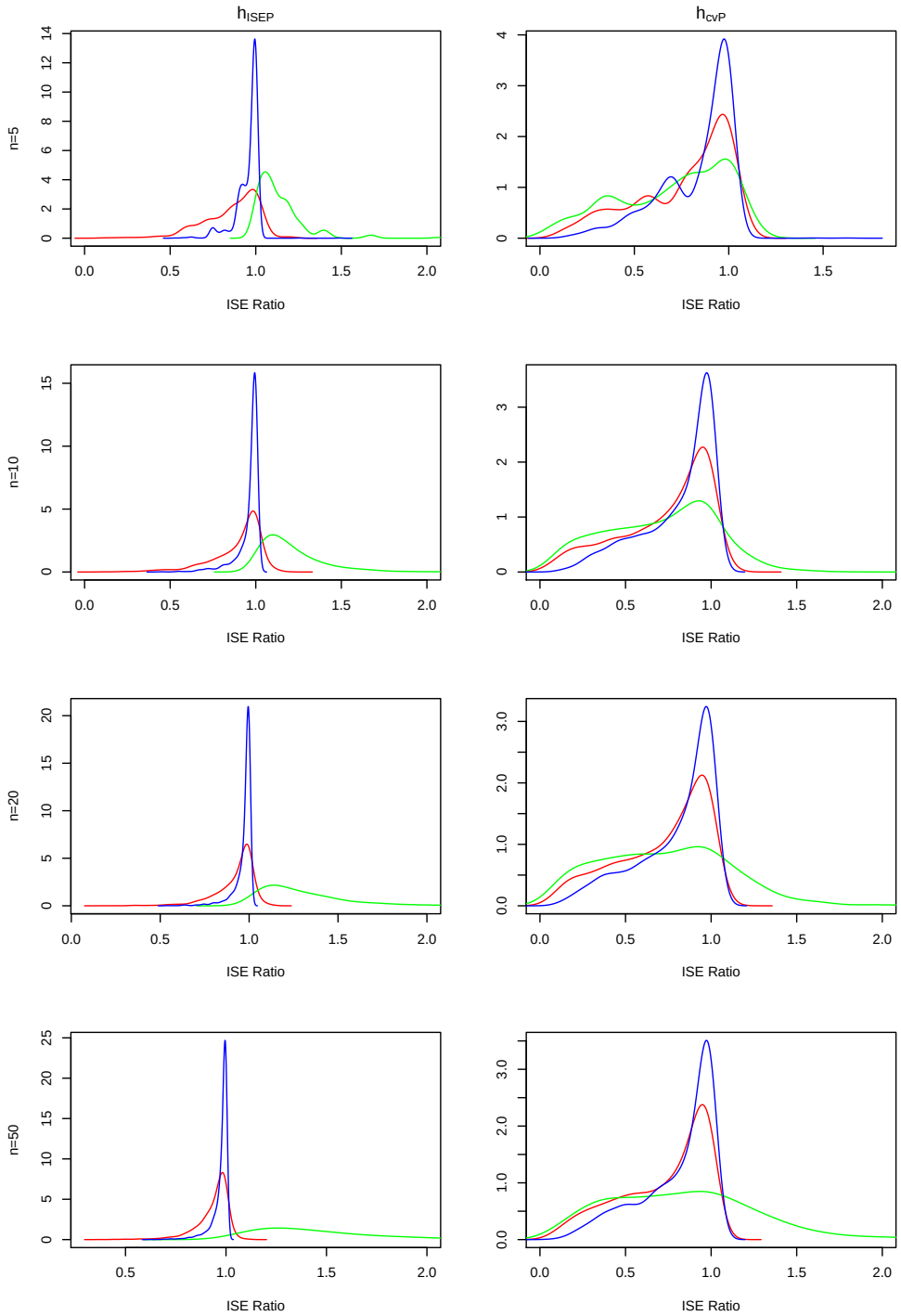

Figure 4.11: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_5(x)$ with $r = 5$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

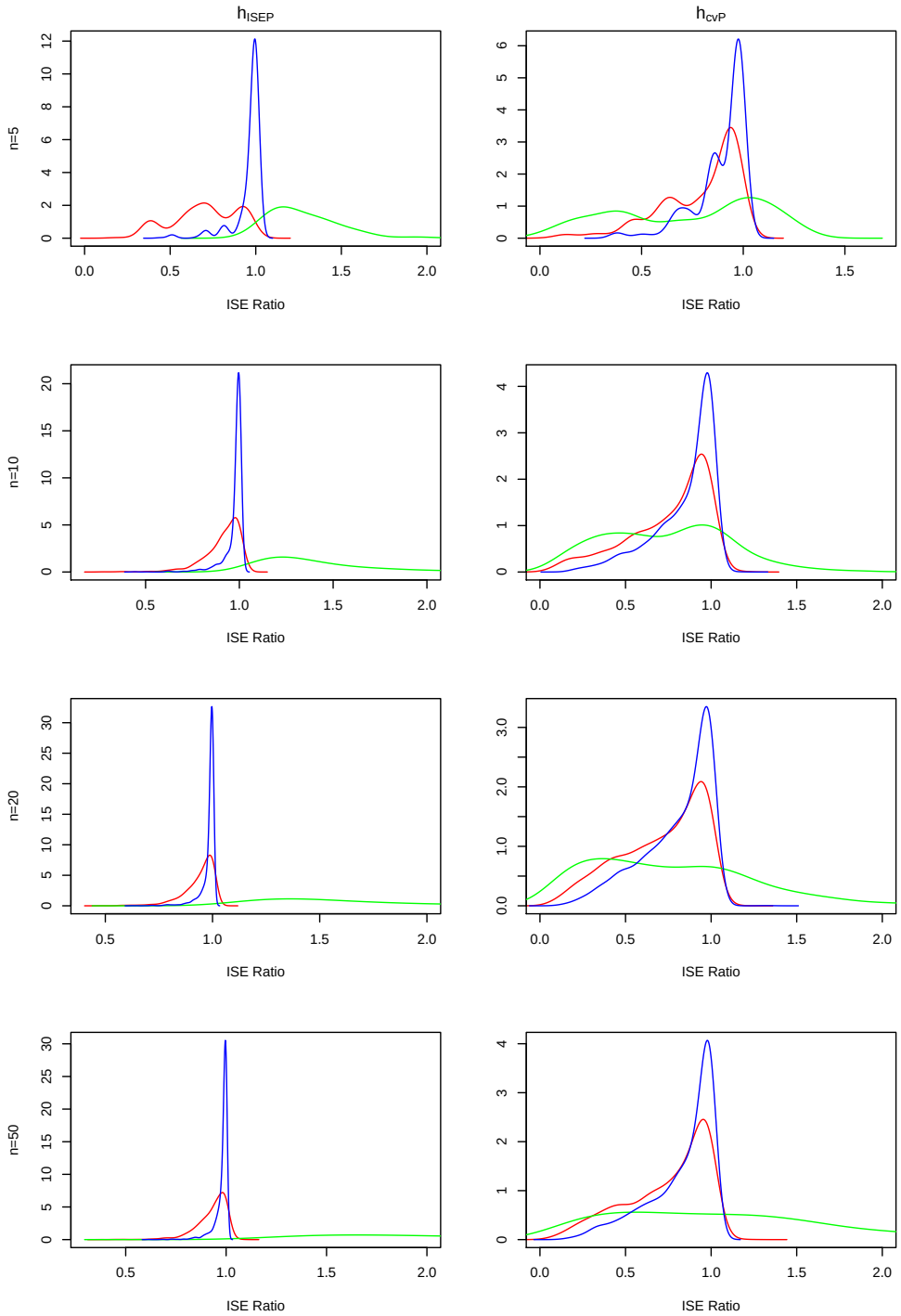

Figure 4.12: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_6(x)$ with $r = 5$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

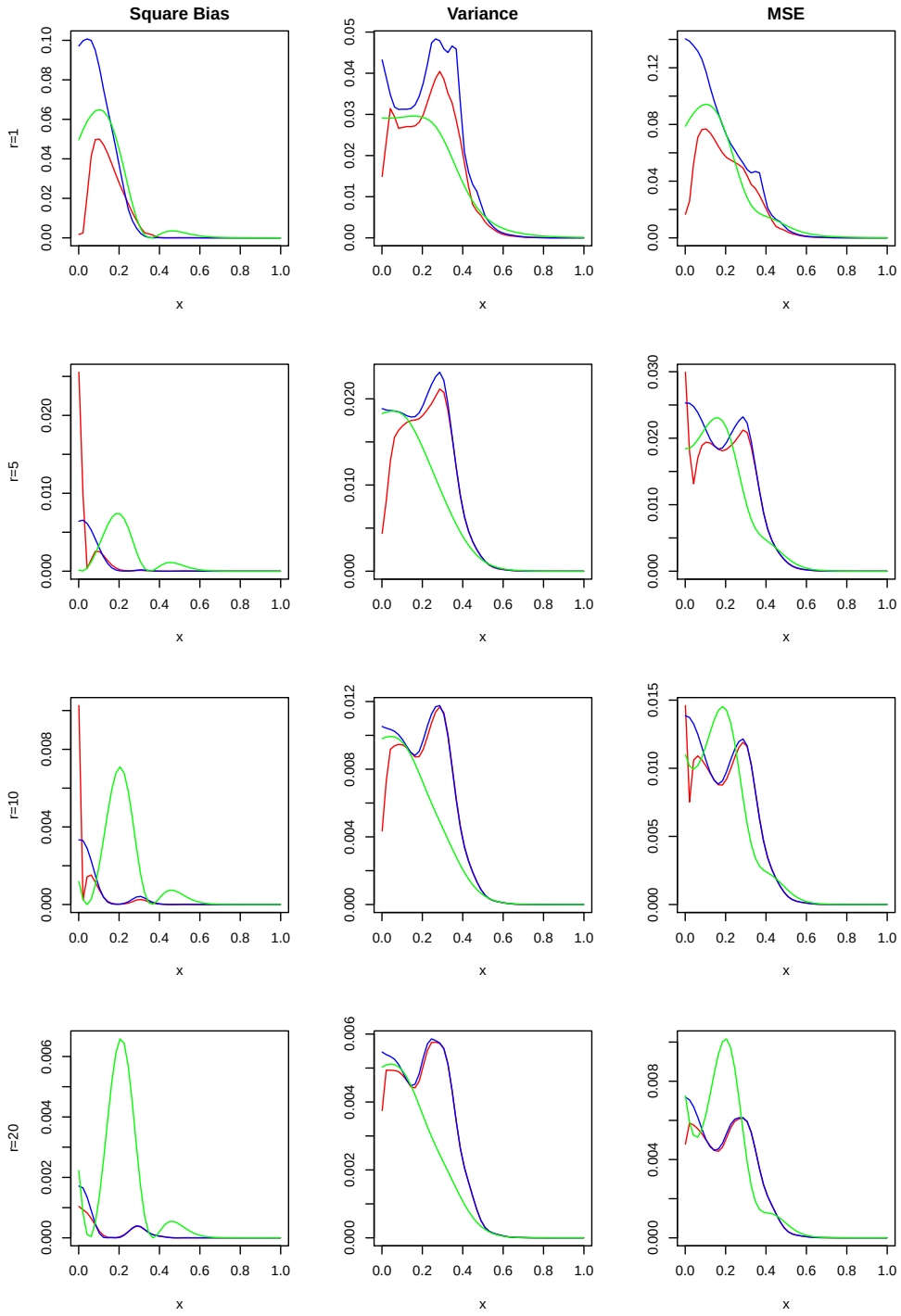

Figure 4.13: A comparison of the square bias, variance and MSE of the three monotone estimates of $p_2(x)$ with $n = 10$ and h_r set as $h_{P,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

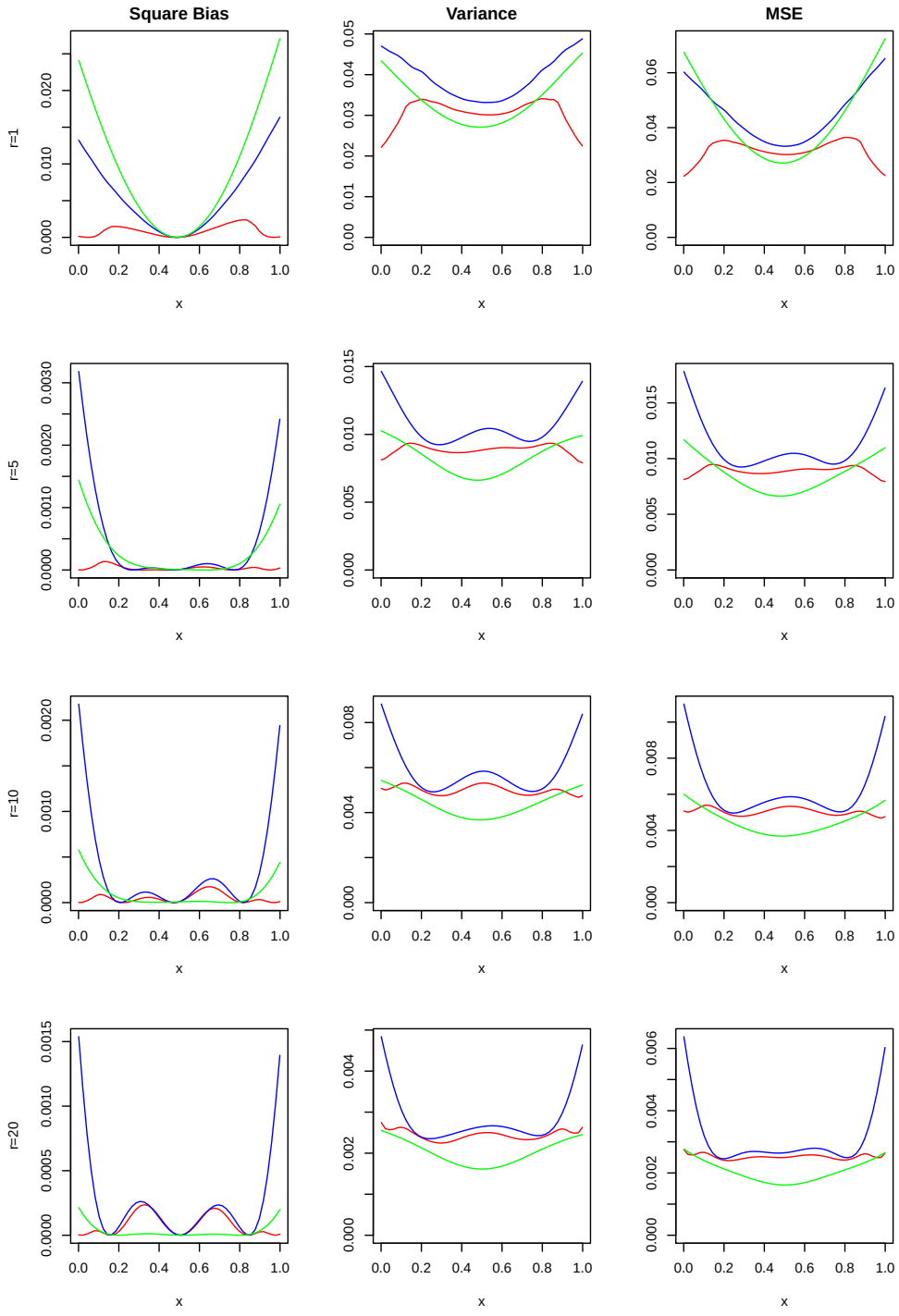

Figure 4.14: A comparison of the square bias, variance and MSE of the three monotone estimates of $p_3(x)$ with $n = 10$ and h_r set as $h_{P,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

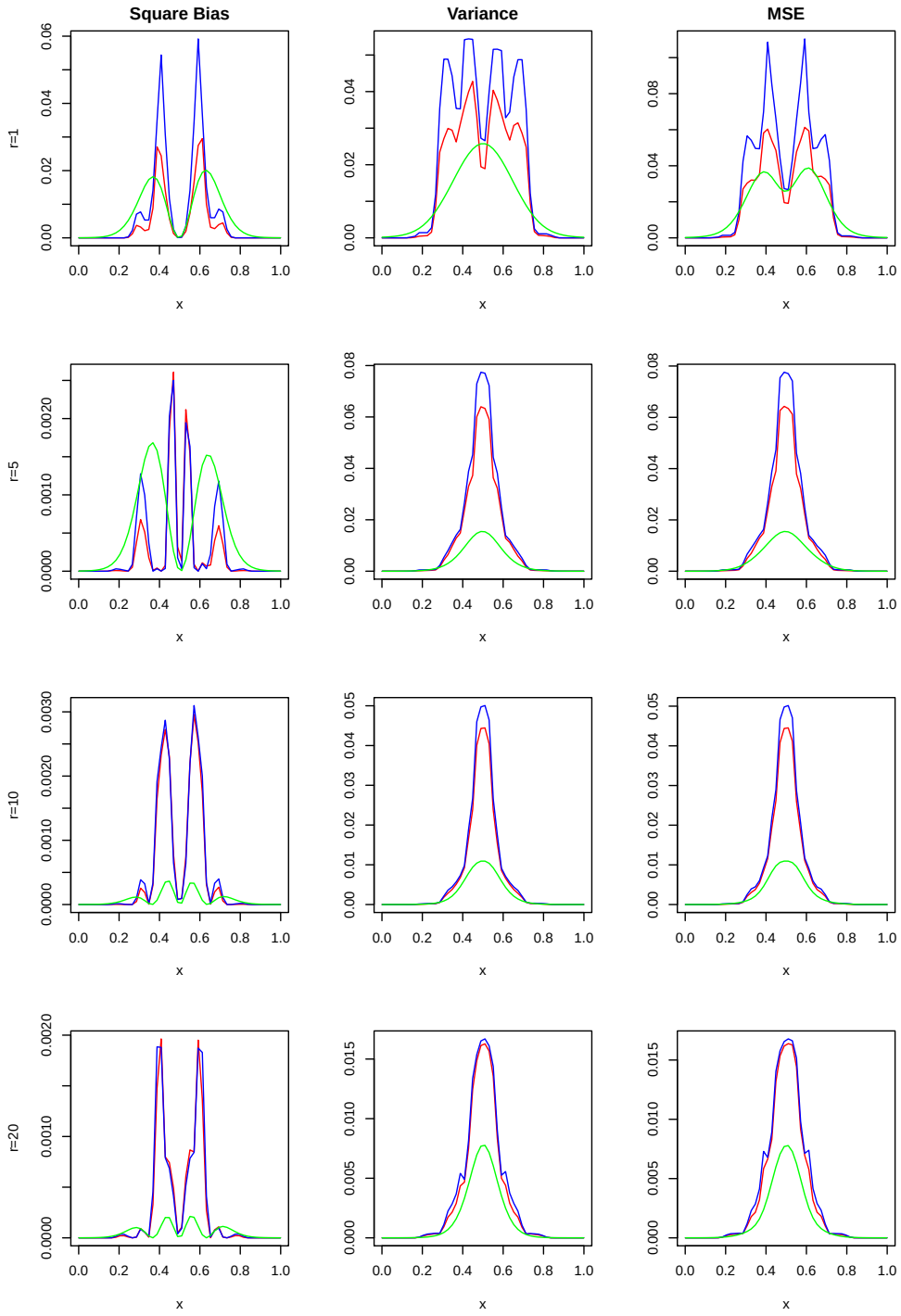

Figure 4.15: A comparison of the square bias, variance and MSE of the three monotone estimates of $p_4(x)$ with $n = 10$ and h_r set as $h_{P,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

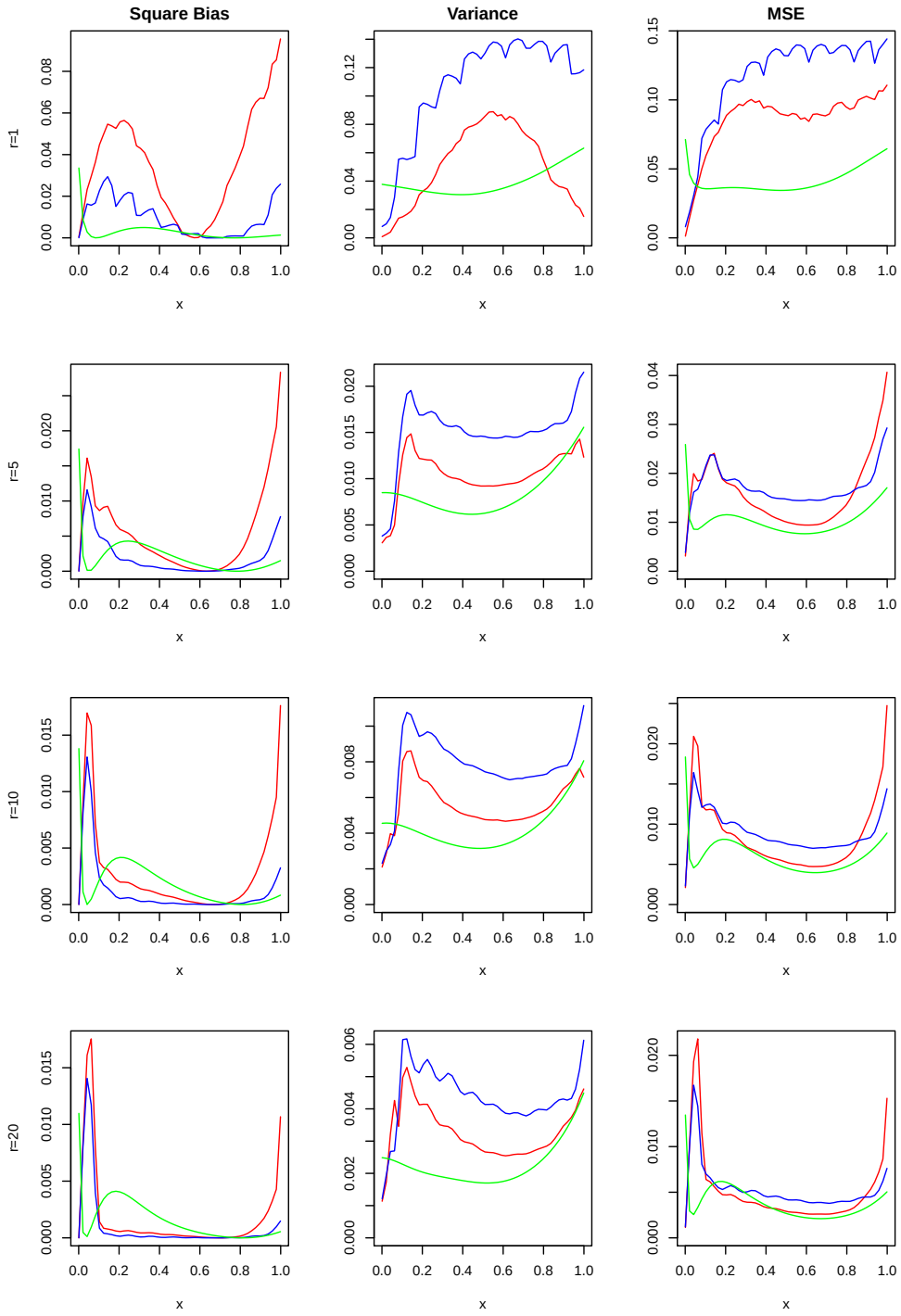

Figure 4.16: A comparison of the square bias, variance and MSE of the three monotone estimates of $p_5(x)$ with $n = 10$ and h_r set as $h_{P,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

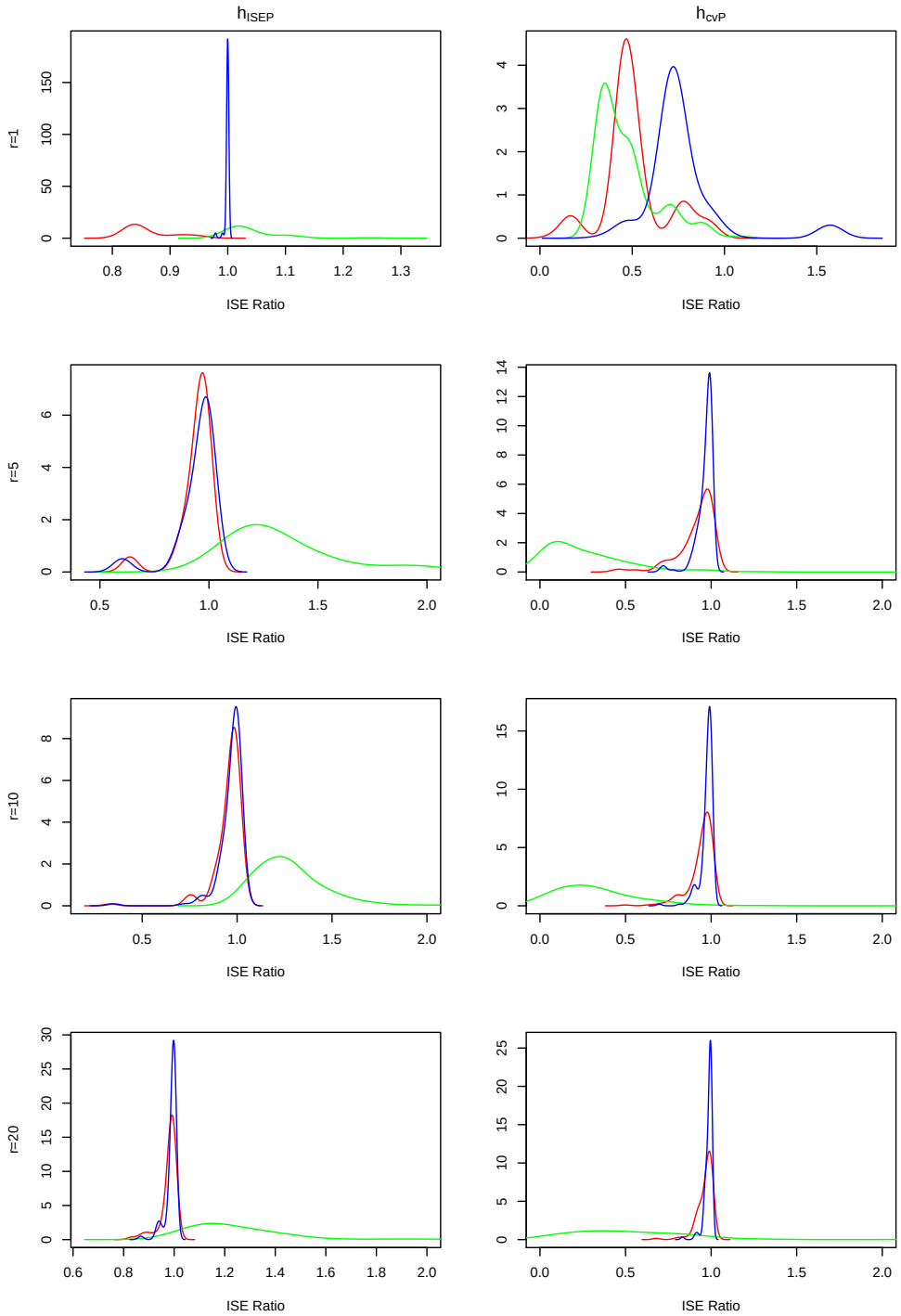

Figure 4.17: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_1(x)$ with $n = 10$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

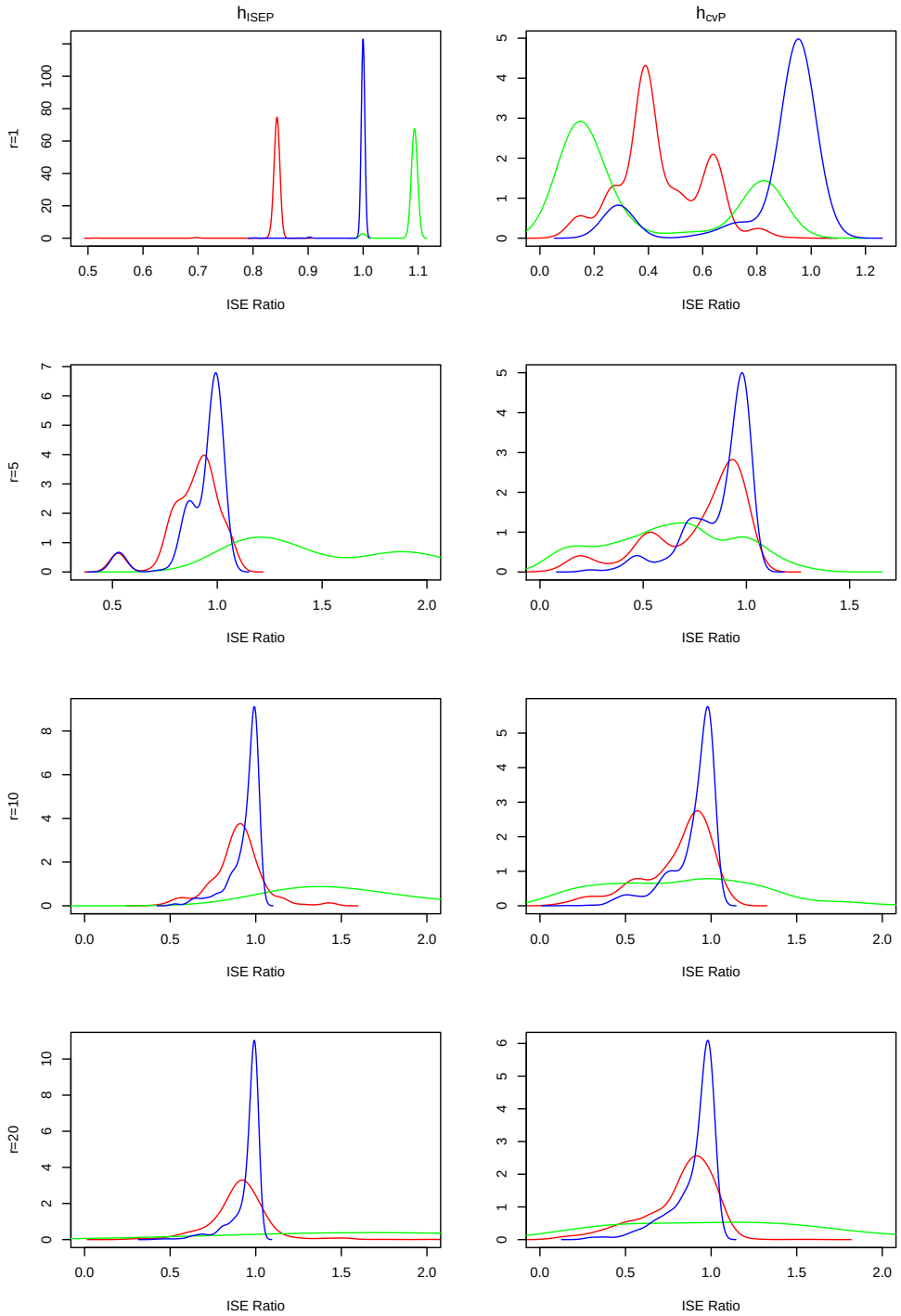

Figure 4.18: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_2(x)$ with $n = 10$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

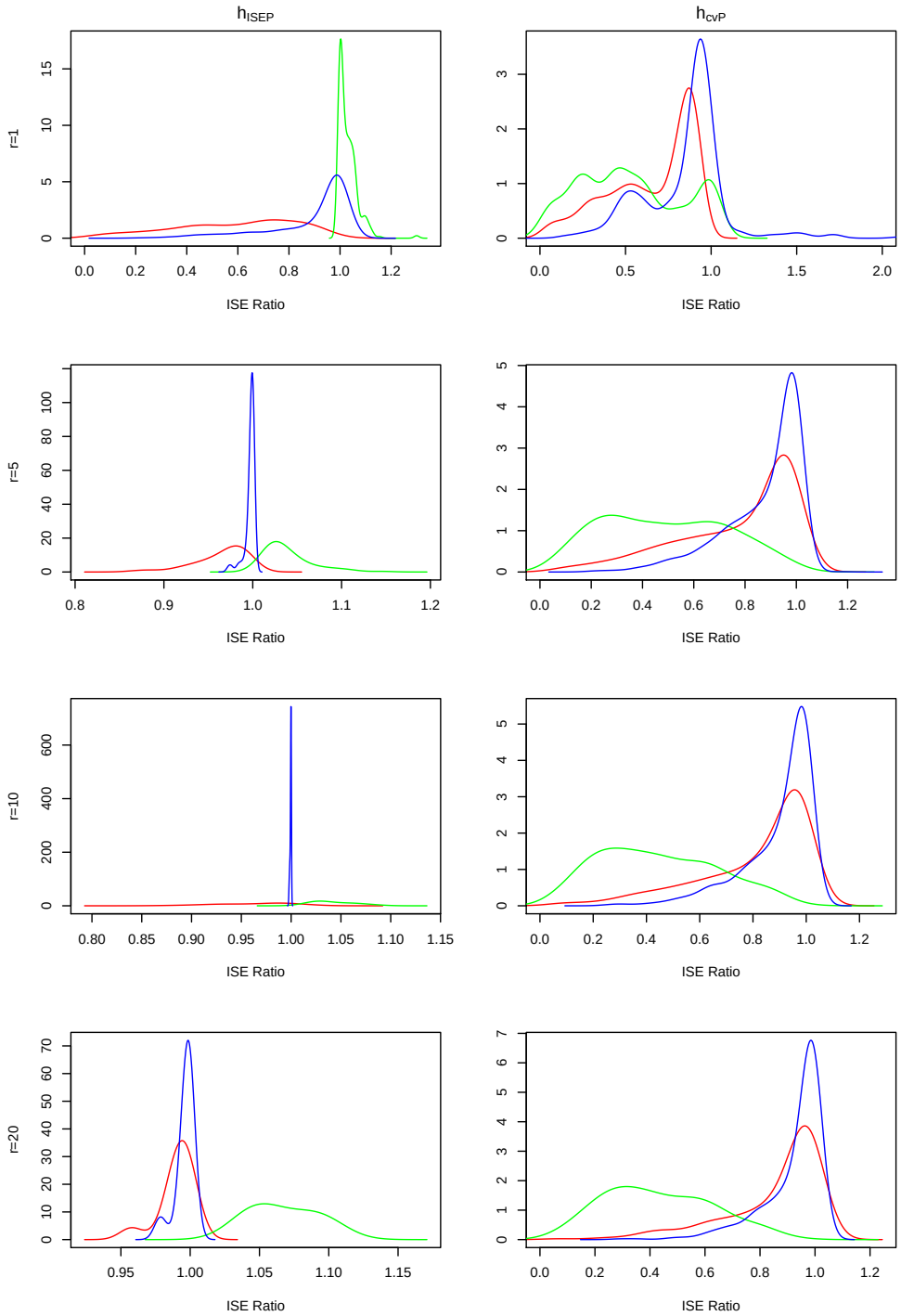

Figure 4.19: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_3(x)$ with $n = 10$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

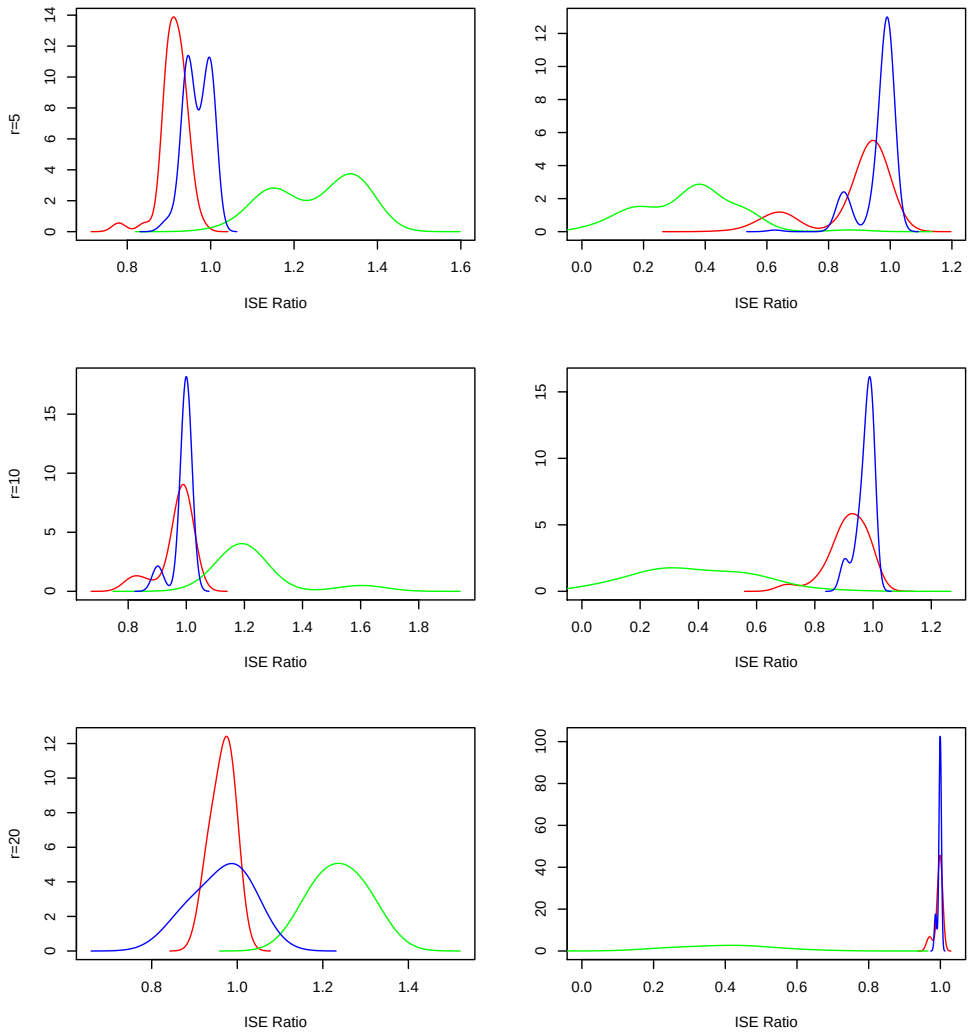

Figure 4.20: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_4(x)$ with $n = 10$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $r = 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

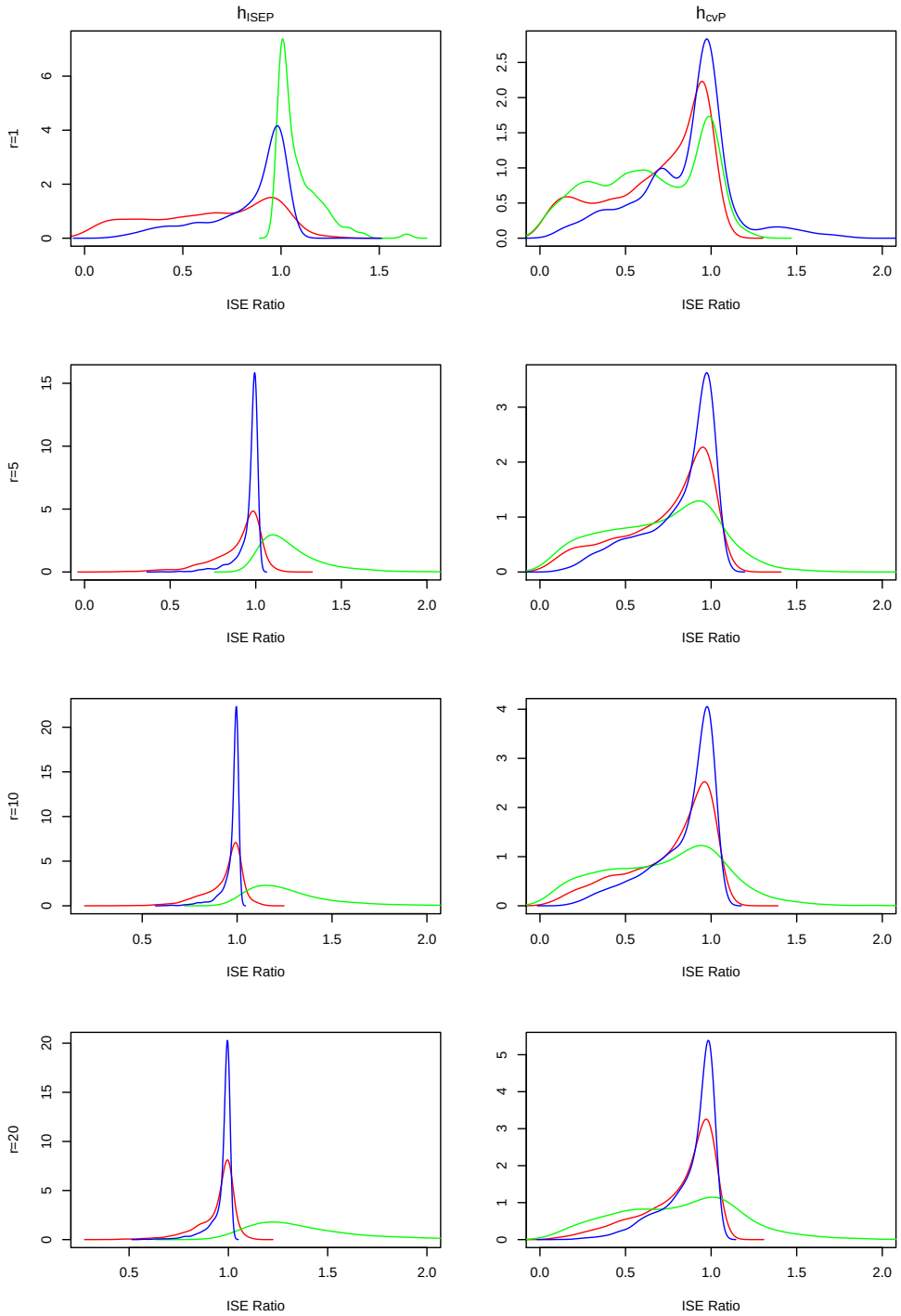

Figure 4.21: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_5(x)$ with $n = 10$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

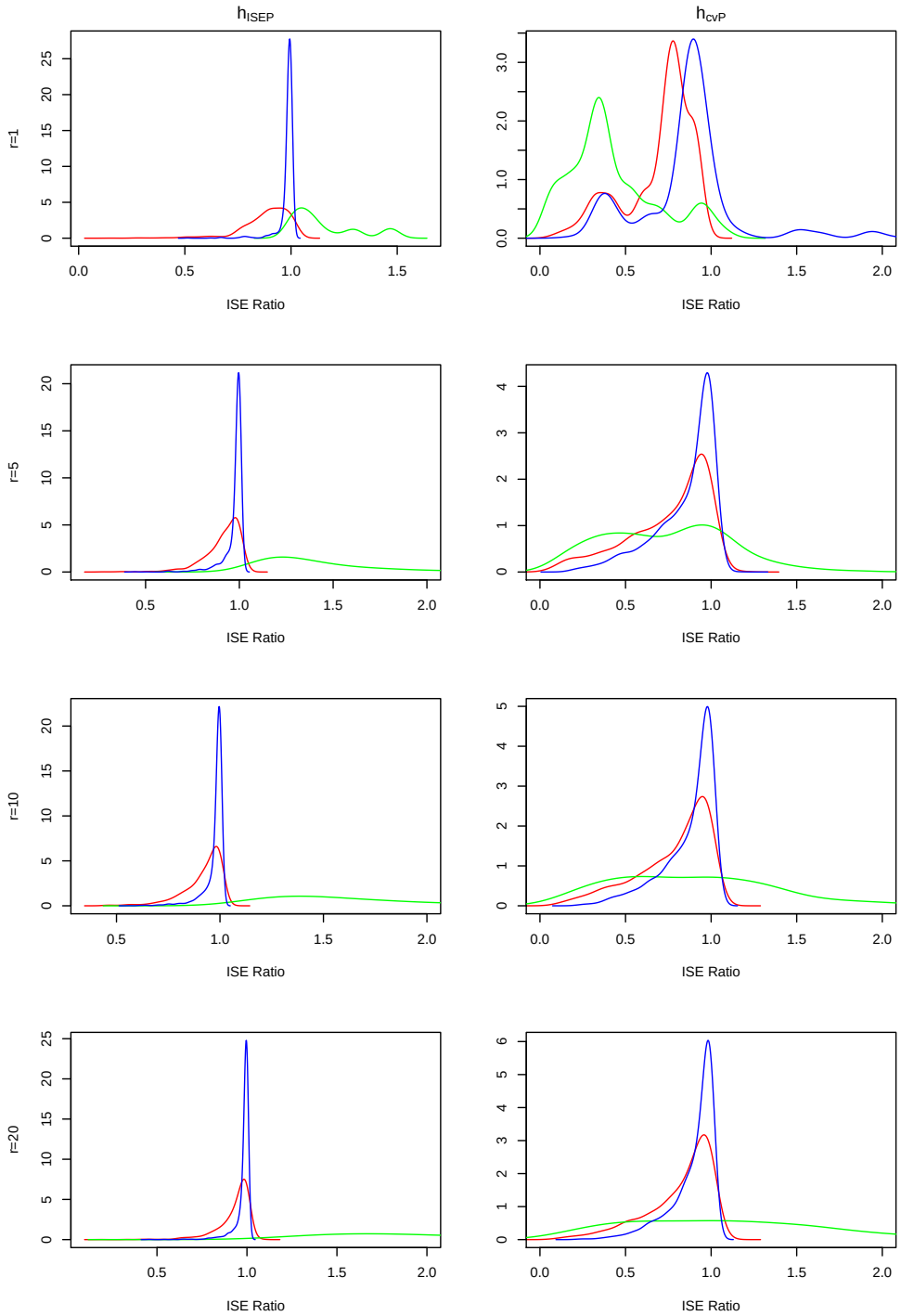

Figure 4.22: A comparison of the ISE ratio statistics for each of the three monotone estimates of $p_6(x)$ with $n = 10$ and h_r set as $h_{P,ISE}, h_{P,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

Chapter 5

A comparison of monotone estimates of Transducer functions

5.1 Introduction

In Chapter 4 I have considered the estimation of monotone psychometric functions and compared the LDNP method with the PAVA method and the bandwidth method. In this chapter, I will generalise this study and investigate how well the LDNP method estimates a monotone transducer function when the response function is no longer binomial. In particular, I will consider Poisson response models and Exponential response models. This means that I will then have considered the performance of the LDNP method on both discrete response models (binomial and Poisson) and continuous response models (exponential).

In the Chapter 1 of this thesis, I have outlined various methods of estimating monotone transducer functions. I have considered the bandwidth method of Kappenman [29] which consists of gradually increasing the bandwidth used in the regression until the resulting estimate is monotone, the PAVA method (see Brunk [4] and [3], Mammen [32] and Mukerjee [34]), which 'averages' out any points causing bumps in the response estimate, and the DNP method (see Dette, Neumeyer and Pilz [9] and [8]) which is a combination of non-parametric regression and non-parametric density estimation. All three methods are widely used to obtain monotone

non-parametric regression estimates. All three methods are explained in more detail in previous chapters.

In this thesis, I have developed an alteration to the DNP method which I have labelled the LDNP method, which takes into account the likelihood. Without this adaptation the DNP method has problems in estimating monotone psychometric functions. In particular, the main problem was that the DNP method does not constrain the resulting monotone estimate to lie in the region $[0, 1]$, which is a requirement of a psychometric function, since it is a probability. This is largely due to the fact that a link function is not used and information about the changing variance is not used in the DNP estimation procedure. This problem has been overcome in the LDNP method. I have developed some of the asymptotic properties of the LDNP estimator and have shown that it is first order asymptotically equivalent to the other methods I have outlined above. More generally, a problem of the DNP method was that the resulting estimates were not constrained to lie within the correct bounds. This need not necessarily be the region $[0, 1]$ as is the case for psychometric functions. If the responses were from the Poisson distribution then the estimates would need to be in the region $[0, \infty)$. The correct link function used in this case would enforce that this would be the case. This is why the LDNP method has an advantage over the DNP method.

However, in many studies of transducer functions there is rarely a large sample size. This can be due to a number of factors, such as the cost of obtaining data, ethical issues and time constraints. Typically transducer function studies are carried out with small sample sizes. In this chapter, I compare the performance of monotone estimates of finite sample transducer functions.

There are a number of factors that affect the final estimate of the transducer function. In a typical study, a scientist will examine n stimulus levels. Clearly as the number of stimulus levels increases we have more information about the response and so should be able to estimate the true function more accurately.

In addition, when estimating the transducer function, a bandwidth, h_r, must be chosen to be used in the unconstrained step. For the LDNP estimate there is also a second bandwidth,

h_d, used in the density estimation step. Changing the value of h_r will clearly affect the ability of the estimating procedure to correctly estimate the true function. Too small a choice of bandwidth would give too much significance to individual points which may pull the estimate away from the true value. Too large a choice of h_r would result in oversmoothing and we may miss important features of the true response since the resulting estimate had been flattened too much. The choice of h_d has been considered in Chapter 3 and is not dealt with in this chapter.

A final factor to consider is the effect of the initial magnitude of the mean response. If a larger mean response is used then, with a Poisson response function for example, the variance will increase. I will investigate how this affects the final estimate of the transducer function. We will alter the other three factors in turn.

5.2 Design of Simulation Study

In this chapter I will consider four basic transducer functions. These will be the functions for both the Poisson response studies and the Exponential response studies. The term r in each of these functions is included as a means of controlling the size of the mean response. I will investigate what effect increasing this has on the performance of each monotonicity constraint. A plot of these functions is shown in Figure 5.2, whilst a plot of the the corresponding η functions is shown in Figure 5.1. These functions were chosen so that, for $x \in [0, 1]$, they take values in the range $(0, \infty)$,

$$
\begin{aligned}
T_1(x) &= r(1 + 0.2x) \\
T_2(x) &= 5 + rx^6 \\
T_3(x) &= \frac{1}{10} + rx^{1/5} \\
T_4(x) &= \frac{1}{10} + \frac{r}{2}\left(\Phi\left(\frac{x-\mu_1}{\tau}\right) + \Phi\left(\frac{x-\mu_2}{\tau}\right)\right) \\
&\qquad \mu_1 = 0.2, \mu_2 = 0.8, \tau = .05
\end{aligned}
$$

The function $T_1(x)$ is linear, $T_2(x)$ is concave with a flatter section on the left, $T_3(x)$ is

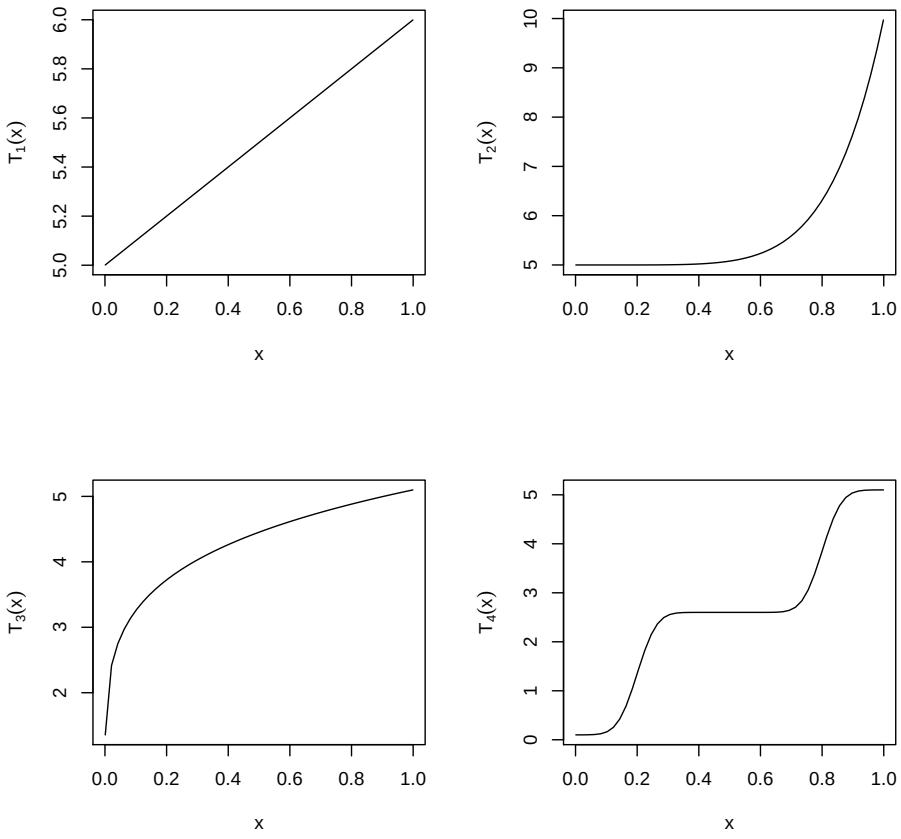

Figure 5.1: Transducer functions used in simulations

convex with a flatter section on the right. Finally, $T_4(x)$ is close to a step function. It should be noted that although in Figure 5.1, the function $\eta_1(x)$ looks linear it is not. This is a result of the scale of the plot. The function $\eta_2(x)$ looks very similar to the function $T_2(x)$ but this is only because of the range of x values. For $x > 1$, the shape of $\eta(x)$ becomes more 'S-shaped'.

In all of the following simulations I will generate n equally spaced stimulus levels, $x_1 \dots x_n$. For each model in turn, at each stimulus level I will simulate results from a Poisson distribution with mean response $T_j(x_i)$. For each of these data sets I will calculate a regression estimate using the PAVA method, the bandwidth method and the LDNP method. I will calculate the LDNP estimate using four different options for the choice of h_d, as described in section 3.4. I calculate the MISE for each of these options and use the option that produces the lowest MISE

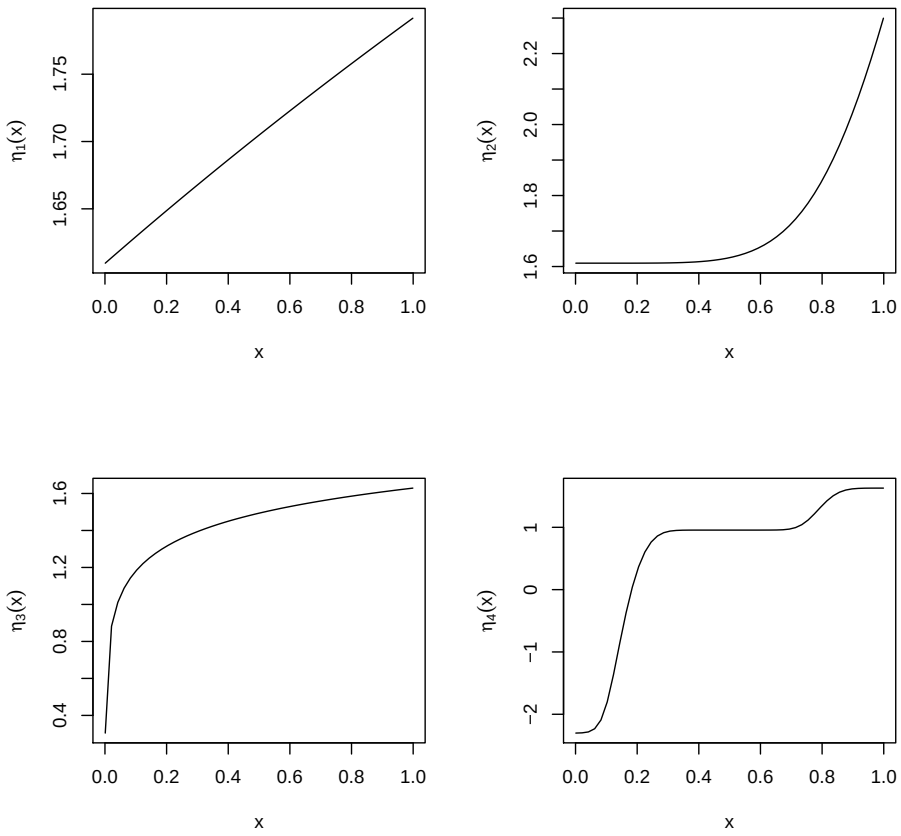

Figure 5.2: The functions $\eta(x)$ used in simulations obtained using the canonical log link.

for comparison with the PAVA and bandwidth methods. I will do the same for each model but generate results from an exponential distribution with rate $1/T_j(x_i)$. In all simulations, the Gaussian kernel defined in Table 3.1 is used for K_r and the Epanechnikov kernel for K_d.

For the unconstrained estimate I will use three different types of bandwidth as in Chapter 4. The first type of bandwidth is an MISE optimal bandwidth. This bandwidth will be dependent upon the number of stimulus levels n, but not upon the individual samples. So for a fixed value of n this bandwidth will be kept constant for all samples. I will consider a bandwidth $h_{T,MISE}$, given in (5.1), which optimises the MISE in terms of the transducer function $T(x)$, and also a bandwidth $h_{\eta,MISE}$, given in (5.2), which is the optimal bandwidth for the MISE in terms of $\eta(x)$.

$$h_{T,MISE} = \left[\frac{1}{2n\sqrt{\pi}} \frac{\int_0^1 Var[Y|X=x]dx}{\int_0^1 (\eta''(x))^2/[g'\{T(x)\}]^2 dx}\right]^{1/5} \tag{5.1}$$

$$h_{\eta,MISE} = \left[\frac{1}{2n\sqrt{\pi}} \frac{\int_0^1 \frac{1}{[g'\{T(x)\}]^2 Var[Y|X=x]}dx}{\int_0^1 (\eta''(x))^2 dx}\right]^{1/5}. \tag{5.2}$$

The second type of bandwidth I will use is an ISE optimal bandwidth. This bandwidth is the best possible bandwidth that could be used for each individual sample. I will use a T optimal ISE bandwidth, $h_{T,ISE}$ and also an η optimal bandwidth, $h_{\eta,ISE}$. These bandwidths are obtained by minimising the integrated square errors (5.3) and (5.4) respectively.

$$ISE_T = \int_0^1 (T(x) - \hat{T}(x))^2 dx \tag{5.3}$$

$$ISE_\eta = \int_0^1 (\eta(x) - \hat{\eta}(x))^2 dx. \tag{5.4}$$

Finally I will use cross-validation bandwidths denoted $h_{T,cv}$, and $h_{\eta,cv}$. These are calculated using formulas (5.5) and (5.6) given in Fan and Gijbels ([16] page 150)

$$CV_T(h) = \int \hat{T}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{T}_{-i}(X_i) \tag{5.5}$$

$$CV_\eta(h) = \int \hat{\eta}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{\eta}_{-i}(X_i). \tag{5.6}$$

where $\hat{T}_{-i}(X_i)$ and $\hat{\eta}_{-i}(X_i)$ are the estimates obtained by leaving out the i^{th} observation and calculating a regression estimate using the remaining observations. The effect of the bandwidth choice is discussed in section 4.3.3.

I will compare the methods for a range of different values of n. I will use the values $n = 5, 10, 20, 50$. As for the psychometric experiments discussed in Chapter 4, 50 stimulus levels would be larger than usually experienced in studies with the other three values much more common. Each of the methods I am comparing are first order asymptotically equivalent. The point of this investigation is to compare how well each method does for smaller numbers of data points. I will also compare the effect of changing the value of r. The values that I have considered are $r = 5, 10, 50$.

For each model and values of r and n we generated 10,000 samples. In the following analysis of these simulations I will report how many of the simulated samples required monotonicity constraints to be used. Only unconstrained estimates that required monotonisation were kept and analysed, for the reasons explained in Chapter 4.

Once I have obtained the simulated samples, I will follow the example of Dette and Scheder [11], and compare the square bias, variance and mean square error (MSE) for each of the models where a constant bandwidth has been used for all of the samples. This means that I will use this method for the MISE optimal bandwidths. In the cases where a T optimal bandwidth is used, I will compare the MSE in terms of estimating $T(x)$ and in the case where I use an η optimal bandwidth, I will compare the MSE in terms of the estimation of $\eta(x)$. These are standard ways of comparing the performance of the estimates and will give a good idea of whether any of the estimates are consistently better than others. Obviously, the closer to 0 each of these values is, the better the estimator. I will also compare the mean integrated square error (MISE), to give an idea of the performance of the estimator over the whole range (as compared to at individual points within the range).

When a different bandwidth has been used for each sample, this kind of analysis is no longer appropriate as all the results cannot really be combined in this way. In these cases, where ISE optimal bandwidths, or cross validation bandwidths have been used, then I will compare the methods in a different way. In this case, as in Chapter 4, I will calculate the ISE for both the unconstrained fit and the monotone fit for each sample and then calculate the ratio,

$$R = \frac{ISE_{monotone}}{ISE_{unconstrained}} \tag{5.7}$$

and report this value for each of the methods. I will plot these values as density estimates and show these plots as a means of comparing the performance of each of the three values. Obviously the method with the smallest R values will be the best method.

5.3 Results

5.3.1 Number of samples requiring monotonisation

The numbers of samples out of the original 10000 that required monotonisation for each of the six possible bandwidth choices for Poisson responses, are shown in Tables 5.1-5.4. It should be mentioned that not all of the 10000 samples will be unique. Inevitably some samples will be repeated. I record the number of unique samples for each trial for the sake of interest and completeness. In contrast to the binomial cases, most samples are unique and only for very small r and n there are some repeated samples. Recall that only samples which required monotonisation were considered in the following analysis.

There is no common pattern for the effect that n and r have on the number of samples requiring monotonisation. For the function $T_1(x)$, as n and r increase, the number of samples requiring monotonisation decrease. For the function $T_2(x)$, the number of samples requiring monotonisation increases with increasing n and decreases with increasing r. For the function $T_3(x)$ the number of samples requiring monotonisation increases with n in general (although not for η optimal ISE bandwidths). It is less clear what happens with an increase in r. This pattern is also true for function $T_4(x)$. I think the uniqueness of the function $T_1(x)$ maybe due to the shape of this function and the fact that it is virtually a flat line but I discuss this more in the coming analysis.

The numbers of samples requiring monotonisation for exponential responses are shown in Tables 5.5-5.8. For the exponential responses there were no repeated samples due to the fact that the response was continuous. It is less easy to distinguish a pattern in these table but it seems that as the value of n increases the number of samples requiring monotonisation increases.

5.3.2 Mean Integrated Square Error

For each of the sets of investigations that involved MISE optimal bandwidths, I calculated the MISE values. I calculated these using numerical integration of the Mean Square Error. These values for the Poisson responses are given in Tables 5.9-5.12. In each table the best performing method has been highlighted in red. In general it can be seen from these tables that as the value of n increases the MISE values decrease. This is to be expected. As the sample size increases we would expect a more accurate estimate. As the value of r increases then the MISE values also increase. This is because the larger values of $T(x)$ mean there is more variability in the Poisson responses that are generated.

For the function $T_1(x)$ the MISE values for the LDNP method are always very poor and the best values are always achieved by the PAVA estimates. For the function $T_2(x)$, the best MISE values are achieved by the LDNP estimate with choice of $h_d = h_r^3$ initially and then as the value of r increases, the PAVA estimate does better. The bandwidth adjustment method seems to consistently achieve the best MISE values for the function $T_3(x)$. Finally for the function $T_4(x)$ the LDNP method with either $h_d = h_r^2$ or $h_d = h_r^3$ generally achieves the best MISE values.

When it comes to the exponential responses, the MISE values are given in Tables 5.13-5.16. It is clear that for function $T_1(x)$ all of the LDNP estimates perform badly. The best method in terms of MISE performance is always the PAVA estimate. For function $T_2(x)$, the PAVA method again does well for most combinations of n and r. As n increases the LDNP estimates with $h_d = h_r^3$ do very well. It is clear from these tables that the choice of h_d does have a large effect on the performance of the estimator in these cases, which is in contrast to the investigations on psychometric functions in Chapter 3. However, the principle that one should choose $h_d = o(h_r)$ is backed up by these findings. That is to say that these investigations into transducer functions support the asymptotic findings of Chapter 3. For function $T_3(x)$ there is no difference produced by change of h_d choice. The PAVA method is best here. Finally for $T_4(x)$, the LDNP estimates with $h_d = h_r^3$ do well for high values of n but the PAVA estimate is best at low values of n.

5.4 Poisson response models

5.4.1 Changing the number of stimulus levels

In this section, I investigate the effect that the number of stimulus levels has on the estimation of monotone transducer functions. I keep the value of r fixed at $r = 5$. I considered the cases when $n = 5, 10, 20, 50$. I also only consider the T optimal versions of the MISE, ISE and cross-validation bandwidths although mention here that results for the other bandwidths followed a similar pattern. It is clear that the magnitude of the square bias, variance and MSE decreases as n increases. I also consider in this section the density plots of ISE ratios.

The results for the comparison of changing n for function $T_1(x)$ can be seen in Figure 5.3 and Figure 5.7, whilst the results for comparing the effect of changing r can be seen in Figure 5.11 and Figure 5.15. It is clear from these plots that the LDNP method performs badly for the function $T_1(x)$. The MISE values are clearly a lot higher as well which reinforces this point. I think that the reason for this is due to the shape of the function. The function $T_1(x)$ gives a line that is close to being horizontal. When the Poisson samples were generated in many cases there was not a lot of difference between the magnitude of the first point (corresponding to the value at x_1) and the last point (corresponding to the value at x_n). In fact in some cases the value of y_1 was greater than the value of y_n. This led to decreasing functions being estimated for the unconstrained regression. The three monotonicity constraint procedures are looking to calculate monotone increasing estimates. Clearly it was hard for the LDNP method to estimate a monotone increasing function from an unconstrained estimate that was monotone decreasing. This can be seen from the very poor square bias, variance and MSE scores.

I investigated this a little further. In the case where $n = 10$ and $r = 5$ I calculated the number of samples for which the unconstrained regression was monotone decreasing. As a note here, it should be said that in general the values in Table 5.1 and similar tables would be the number of non monotone samples. Strictly speaking they are really the number of non monotone increasing samples. In most cases this is the same as the number of non monotone samples but in this case this is not true. In fact, when $h_{T,MISE}$ or $h_{\eta,MISE}$ was used for the

unconstrained regression 3158 out of 3431 samples that are recorded as non monotone were in fact monotone decreasing. This indicates that this is a frequent problem. When $h_{T,ISE}$ was used, 1963 out of 5130 samples were monotone decreasing and when $h_{T,cv}$ was used 1973 out of 5491 samples were monotone decreasing.

Clearly this is an unexpected result. In some sense it does seem strange to try to force a monotone decreasing sample to be monotone increasing. However, in a case where the true function is close to being horizontal it is not surprising that some data samples produce regression estimates that are monotone decreasing. The problem in this case is not a problem with the LDNP method as such, but rather a problem with the DNP method. I believe in principle that it should still be able to estimate a monotone increasing curve that is reasonable for the data. However, in my investigations I found that the DNP code does not work well in this case. In fact, the larger the value of h_d the more steep the monotone estimate becomes. This is clearly not desirable

Work could be done to try to overcome this problem. I think it is a problem with the method used for calculation in R rather than the method in theory, but have not conclusively demonstrated this. Overcoming this problem is beyond the scope of these simulations. It would need some modification of the method to allow it to deal with these more unusual cases.

The PAVA estimate can be seen to perform well for all the choices of bandwidth for $T_1(x)$. This makes sense. The PAVA method essentially averages out 'bad point' and so if struggling to find a monotone increasing function would simply plot a flat line roughly in the middle of the scatter of points. Clearly, although not a true reflection of the real function a flat line is not a terrible description of the samples in this case and results in square bias, variance and MSE scores that are not very high compared to the other methods. This suggests that in the case of estimating transducer functions that are close to being flat the PAVA method is the best method to use. I omit the discussion of the function $T_1(x)$ from the rest of this section.

Figure 5.4 and Figure 5.8 show the results for function $T_2(x)$. It is clear from inspecting these plots that there is little to choose between the three methods. The variances are almost identical for each method whilst the bandwidth adjustment method has larger bias. As n

increases the LDNP method and the PAVA method become more similar. It does seem that as the number of stimulus levels increases, the bandwidth adjustment method performs less well than the other two. As n increases, the bandwidth method perform worse for both the cross-validation and ISE optimal bandwidths. This can be seen by the density estimates getting flatter as n increases. In contrast the LDNP and PAVA estimates stay closely bound about 1, which indicates better performance with little to choose between the two methods.

Figure 5.5 and Figure 5.9 show a similar pattern for the function $T_3(x)$. The bandwidth adjustment method performs well at low values of n for MISE optimal bandwidths although as the value of n increases the differences between the three methods become very small. At low values of n the LDNP method and the PAVA method have high bias at low values of x which contributes to their poor scores. It looks as if the LDNP method performs slightly better than the other two methods when the ISE optimal bandwidths or cross-validation bandwidths are used, although again there is not a lot to choose between the three methods.

The same kind of pattern is shown for the function $T_4(x)$ in Figure 5.6 and Figure 5.10. The bandwidth adjustment method works well with small n but less well as n increases. Figure 5.10 shows that although there is little to choose between the PAVA method and the LDNP method in terms of the mode of the density, the PAVA method has a far smaller range of ISE ratio values which might be considered an advantage in that there are not many 'bad' estimates.

5.4.2 Changing the value of r

In this section I compare the effect of changing the value of r. This was a way of investigating what effect the magnitude of the mean response has on the estimating procedure. I keep the number of stimulus levels constant at $n = 10$. As when comparing the change in n I compare only the T optimal bandwidths. I consider the values $r = 5, 10, 20$. The problems in estimation of the function $T_1(x)$ were discussed in section 5.4.1 and it will not be further considered here.

Figures 5.12- 5.14 show the use of MISE optimal bandwidths on functions $T_2(x)$, $T_3(x)$ and $T_4(x)$ respectively. It is clear from these three figures that the bandwidth adjustment method performs progressively worse as the value of r increases. The LDNP method and the PAVA

method appear far more resistant to changes in the value of r, and so could be considered better methods to use for this reason. Although all three methods have similar variances for the function $T_2(x)$, the larger bias of the bandwidth method means that its MSE performance is worse than the other two methods. This is more evident as r increases. This is also the case for $T_4(x)$.

This finding is repeated for ISE optimal bandwidths as can be seen from Figure 5.16- 5.18. It is also true for cross-validation bandwidths although the decrease in performance of the bandwidth adjustment method appears less obvious here. A clear picture from these figures is that as the value of r increases the density estimate for the bandwidth method becomes flatter, a fact that indicates poor performance due to increased variability. It is evident that when a good bandwidth is chosen for the unconstrained regression the LDNP method performs as well as or better than any other method. In fact the LDNP method and the PAVA method have the ability to improve the performance of the unconstrained estimator with ISE optimal bandwidth, as indicated by values of $R < 1$. That is, for some samples, the monotonised version is better (in ISE sense) than the best possible unconstrained estimate.

5.5 Exponential response models

5.5.1 Changing the number of stimulus levels

Figures 5.19-5.22 show the comparisons of square bias, variance and MSE for each of the three methods when the MISE optimal bandwidth was used for the unconstrained regression for each of $T_1(x)$-$T_4(x)$ respectively.

As with the Poisson response simulations, the LDNP method again is very bad at estimating the function $T_1(x)$. This has been discussed in section 4.3.4 and is not repeated here. I omit the discussion of the function $T_1(x)$ from the following analysis. The effect of changing the value of r for $T_1(x)$ can be seen in Figure 5.27 and Figure 5.31, whilst the comparison of ISE ratios for changing numbers of stimulus levels can be seen in Figure 5.23.

For the function $T_2(x)$, each of the three methods perform similarly. This is increasingly the

case as the number of stimulus levels increases although the bandwidth method suffers from high bias and variance at the boundaries. The LDNP method has higher bias at low values of x, but lower bias at high values of x. The LDNP method is not very good at estimating function $T_3(x)$. Although the biases are similar for each of the three methods the LDNP method has a much higher variance causing a poor MSE performance. For the function $T_4(x)$, the three methods perform similarly at small number of stimulus levels, although the LDNP method has high bias at values close to $x = 1$. As the number of stimulus levels increases the bandwidth method performs increasingly poorly, whilst the LDNP method and the PAVA method become more similar. The PAVA method seems to perform better in this case.

Figures 5.24-5.26 show the comparisons of the ISE ratios for each method for functions $T_2(x)$-$T_4(x)$ respectively. In the first column of the plots the results shown are for when an ISE optimal bandwidth is chosen for the unconstrained regression. For functions $T_3(x)$ and $T_4(x)$ the LDNP method perform poorly at $n = 5$ stimulus levels. This is indicated by the almost flat density estimates which suggest high variability. However, their performance improves for higher values of n. As in the case of Poisson responses, the PAVA method and the LDNP method often perform fairly similarly although the LDNP method is often slightly better. As the number of stimulus levels increases then the bandwidth method performs worse and worse. The density estimates become flatter. This suggests, as before that if a good bandwidth is chosen for the unconstrained regression then the bandwidth method performs poorly whilst the LDNP method is a little better than the PAVA method.

The second column of Figures 5.24-5.26 show the comparison of ISE ratios for each method when a cross validation method is used for the unconstrained regression. Here the LDNP method performs poorly for small numbers of stimulus levels in comparison to the other two methods. The bandwidth method does well at small numbers of stimulus levels but performs less well at higher values of n. This suggests that the cross-validation bandwidth is too low and the monotonicity is best fixed by simply increasing the bandwidth. As the number of stimulus levels increases then the bandwidth method is less good and the LDNP method and PAVA method become more comparable.

5.5.2 Changing the value of r

Figures 5.28-5.30 show the comparison of square bias, variance and MSE for each of the methods for functions $T_2(x)$-$T_4(x)$. I omit discussion of the function $T_1(x)$ for reasons explained in section 5.4.1. It can be seen from these that for functions $T_2(x)$ and $T_4(x)$ the three methods perform similarly when an MISE optimal bandwidth is used for the unconstrained regression. As the values of r increase then the differences between the bias become less noticeable. AT low values of r the bandwidth method performs poorly in terms of the bias, but on all other measures the three methods perform very similarly.

By examining the first column of Figures 5.32-5.34 the ISE ratios of each method for functions $T_2(x)$-$T_4(x)$, when using an ISE optimal bandwidth for the unconstrained regression can be compared. It is again clear here that the bandwidth method performs badly as the value of r increases, shown by the increasingly flat density estimates. The LDNP method and the PAVA method perform similarly to each other, although the LDNP method seems to show more variability. However the majority of the density estimates for both of these methods are close to the value 1 indicating reasonably good performance.

The second column of Figures 5.32-5.34 shows the same comparisons but with a cross-validation bandwidth for the unconstrained regression. As has often been the case the bandwidth method performs well at low values of r when the unconstrained regression bandwidth is too small. As the value of r increases the advantage of the bandwidth method decreases. In fact for high values of r the LDNP method performs at least as well, if not better than the competitors. As in the case of Poisson responses, the LDNP method and the PAVA method have the ability to improve the performance of the unconstrained estimator with ISE optimal bandwidth, as indicated by values of $R < 1$. That is, for some samples, the monotonised version is better (in ISE sense) than the best possible unconstrained estimate.

5.6 Figures and Tables

Table 5.1: Number of samples requiring monotonisation for $T_1(x)$ with Poisson responses for each of six different bandwidths. The number of unique samples is given in brackets.

Model	$h_{T,MISE}$	$h_{\eta,MISE}$	$h_{T,cv}$	$h_{\eta,cv}$	$h_{T,ISE}$	$h_{\eta,ISE}$
$T_1(x)$ **{r=5 n=5}**	3559 (3226)	3559 (3226)	5101 (4598)	5020 (4518)	4825 (4349)	4205 (3799)
$T_1(x)$ **{r=5 n=10}**	3431 (3431)	3431 (3431)	5491 (5491)	5495 (5945)	5130 (5130)	4255 (4255)
$T_1(x)$ **{r=5 n=20}**	2967 (2967)	2969 (2969)	5144 (5144)	5217 (5217)	4967 (4967)	3768 (3768)
$T_1(x)$ **{r=5 n=50}**	2122 (2122)	2122 (2122)	4305 (4305)	4520 (4520)	4243 (4243)	2785 (2785)
$T_1(x)$ **{r=10 n=5}**	3096 (3035)	3097 (3036)	4648 (4557)	4600 (4510)	4395 (4310)	3768 (3696)
$T_1(x)$ **{r=10 n=10}**	2835 (2835)	2837 (2837)	4987 (4987)	4960 (4960)	4552 (4552)	3615 (3615)
$T_1(x)$ **{r=10 n=20}**	2284 (2284)	2285 (2285)	4544 (4544)	4559 (4559)	4188 (4188)	2998 (2998)
$T_1(x)$ **{r=10 n=50}**	1251 (1251)	1252 (1252)	3427 (3427)	3477 (3477)	3127 (3127)	1702 (1702)
$T_1(x)$ **{r=50 n=5}**	1422 (1422)	1422 (1422)	2907 (2907)	2901 (2901)	2595 (2595)	1883 (1883)
$T_1(x)$ **{r=50 n=10}**	1000 (1000)	1001 (1001)	3156 (3156)	3139 (3139)	2263 (2263)	1414 (1414)
$T_1(x)$ **{r=50 n=20}**	453 (453)	454 (454)	2522 (2522)	2542 (2542)	1681 (1681)	708 (708)
$T_1(x)$ **{r=50 n=50}**	55 (55)	55 (55)	1758 (1758)	1757 (1757)	733 (733)	112 (112)

Table 5.2: Number of samples requiring monotonisation for $T_2(x)$ with Poisson responses for each of six different bandwidths. The number of unique samples is given in brackets.

Model	$h_{T,MISE}$	$h_{\eta,MISE}$	$h_{T,cv}$	$h_{\eta,cv}$	$h_{T,ISE}$	$h_{\eta,ISE}$
$T_2(x)$ **{r=5 n=5}**	6927 (6359)	6602 (6070)	3110 (2864)	2981 (2737)	5335 (4842)	2195 (2047)
$T_2(x)$ **{r=5 n=10}**	8456 (8456)	8052 (8052)	3923 (3923)	4062 (4062)	5819 (5819)	2076 (2076)
$T_2(x)$ **{r=5 n=20}**	8863 (8863)	8527 (8527)	4390 (4390)	4285 (4285)	6332 (6332)	1621 (1621)
$T_2(x)$ **{r=5 n=50}**	9174 (9174)	8837 (8837)	5463 (5463)	4686 (4686)	7355 (7355)	976 (976)
$T_2(x)$ **{r=10 n=5}**	6037 (5671)	5603 (5261)	2399 (2220)	2194 (2032)	5204 (4848)	936 (900)
$T_2(x)$ **{r=10 n=10}**	8118 (8118)	7586 (7586)	4172 (4172)	4245 (4245)	7100 (7100)	960 (960)
$T_2(x)$ **{r=10 n=20}**	8716 (8716)	8231 (8231)	5454 (5454)	5395 (5395)	7487 (7487)	723 (723)
$T_2(x)$ **{r=10 n=50}**	9121 (9121)	8703 (8703)	7071 (7071)	6863 (6863)	8237 (8237)	201 (201)
$T_2(x)$ **{r=50 n=5}**	3641 (3567)	3731 (3654)	2586 (2502)	2252 (2183)	3539 (3458)	4 (4)
$T_2(x)$ **{r=50 n=10}**	6544 (6544)	6752 (6752)	4416 (4416)	4906 (4906)	6713 (6713)	1 (1)
$T_2(x)$ **{r=50 n=20}**	7717 (7717)	7876 (7876)	5906 (5906)	6753 (6753)	7509 (7509)	0
$T_2(x)$ **{r=50 n=50}**	8285 (8285)	8476 (8476)	7406 (7406)	7917 (7917)	7894 (7894)	0

Table 5.3: Number of samples requiring monotonisation for $T_3(x)$ with Poisson responses for each of six different bandwidths. The number of unique samples is given in brackets.

Model	$h_{T,MISE}$	$h_{\eta,MISE}$	$h_{T,cv}$	$h_{\eta,cv}$	$h_{T,ISE}$	$h_{\eta,ISE}$
$T_3(x)$ **{r=5 n=5}**	9096 (6653)	9096 (6653)	3400 (2597)	3462 (2638)	3180 (2377)	1522 (1281)
$T_3(x)$ **{r=5 n=10}**	9874 (9874)	9876 (9876)	3886 (3886)	4148 (4148)	3843 (3843)	1501 (1501)
$T_3(x)$ **{r=5 n=20}**	9993 (9993)	9993 (9993)	3868 (3868)	4518 (4518)	4537 (4537)	1241 (1241)
$T_3(x)$ **{r=5 n=50}**	10000 (10000)	10000 (10000)	3502 (3502)	4659 (4659)	4837 (4837)	524 (524)
$T_3(x)$ **{r=10 n=5}**	8612 (8027)	8612 (8027)	2820 (2634)	2918 (2726)	3008 (2797)	754 (730)
$T_3(x)$ **{r=10 n=10}**	9777 (9777)	9778 (9778)	3483 (3483)	3767 (3767)	4072 (4072)	797 (797)
$T_3(x)$ **{r=10 n=20}**	9961 (9961)	9961 (9961)	3606 (3606)	4163 (4163)	4664 (4664)	527 (527)
$T_3(x)$ **{r=10 n=50}**	10000 (10000)	10000 (10000)	3804 (3804)	4757 (4757)	5588 (5588)	178 (178)
$T_3(x)$ **{r=50 n=5}**	7029 (7026)	6939 (6936)	2318 (2316)	2461 (2460)	3265 (3264)	122 (122)
$T_3(x)$ **{r=50 n=10}**	9234 (9234)	9239 (9239)	3245 (3245)	3720 (3720)	5021 (5021)	82 (82)
$T_3(x)$ **{r=50 n=20}**	9887 (9887)	9887 (9887)	4354 (4354)	5206 (5206)	6172 (6172)	16 (16)
$T_3(x)$ **{r=50 n=50}**	10000 (10000)	10000 (10000)	5637 (5637)	6939 (6939)	7581 (7581)	1 (1)

Table 5.4: Number of samples requiring monotonisation for $T_4(x)$ with Poisson responses for each of six different bandwidths. The number of unique samples is given in brackets.

Model	$h_{T,MISE}$	$h_{\eta,MISE}$	$h_{T,cv}$	$h_{\eta,cv}$	$h_{T,ISE}$	$h_{\eta,ISE}$
$T_4(x)$ **{r=5 n=5}**	7332 (2724)	9136 (3076)	1467 (715)	1967 (843)	2464 (813)	351 (217)
$T_4(x)$ **{r=5 n=10}**	9649 (9637)	9894 (9882)	1775 (1775)	5338 (5331)	3676 (3670)	55 955)
$T_4(x)$ **{r=5 n=20}**	9709 (9709)	9954 (9954)	2176 (2176)	6883 (6883)	5325 (5325)	9 (9)
$T_4(x)$ **{r=5 n=50}**	9860 (9860)	9973 (9973)	4078 (4078)	8292 (8292)	7617 (7617)	0
$T_4(x)$ **{r=10 n=5}**	7007 (4569)	8555 (5287)	832 (663)	1460 (1082)	2998 (1686)	49 (48)
$T_4(x)$ **{r=10 n=10}**	9320 (9316)	9831 (9827)	1260 (1260)	6685 (6681)	5650 (5648)	7 (7)
$T_4(x)$ **{r=10 n=20}**	9549 (9549)	9967 (9967)	2637 (2637)	8217 (8217)	7478 (7478)	0
$T_4(x)$ **{r=10 n=50}**	9781 (9781)	9991 (9991)	6559 (6559)	9198 (9198)	9249 (9249)	0
$T_4(x)$ **{r=50 n=5}**	5089 (4978)	6483 (6306)	16 (16)	95 (95)	3274 (3170)	0
$T_4(x)$ **{r=50 n=10}**	7672 (7672)	9493 (9493)	859 (859)	8353 (8353)	9091 (9091)	0
$T_4(x)$ **{r=50 n=20}**	8616 (8616)	9935 (9935)	6265 (6265)	9400 (9400)	9877 (9877)	0
$T_4(x)$ **{r=50 n=50}**	9327 (9327)	9992 (9992)	9541 (9541)	9745 (9745)	9982 (9982)	0

Table 5.5: Number of samples requiring monotonisation for $T_1(x)$ with exponential responses for each of six different bandwidths. No number is given in brackets since all the samples were unique.

Model	$h_{T,MISE}$	$h_{\eta,MISE}$	$h_{T,cv}$	$h_{\eta,cv}$	$h_{T,ISE}$	$h_{\eta,ISE}$
$T_1(x)$ **{r=5 n=5}**	3730	3736	4892	5325	5074	4464
$T_1(x)$ **{r=5 n=10}**	4204	4204	5905	6183	6019	5019
$T_1(x)$ **{r=5 n=20}**	4156	4156	6045	6112	5986	4851
$T_1(x)$ **{r=5 n=50}**	3890	3890	5829	5817	5578	4486
$T_1(x)$ **{r=10 n=5}**	3256	3258	4274	4703	4427	3903
$T_1(x)$ **{r=10 n=10}**	4010	4010	5693	5886	5821	4829
$T_1(x)$ **{r=10 n=20}**	4172	4172	5941	6092	5874	4843
$T_1(x)$ **{r=10 n=50}**	3847	3847	5823	5802	5527	4453
$T_1(x)$ **{r=50 n=5}**	2311	3037	3081	3210	3157	2778
$T_1(x)$ **{r=50 n=10}**	3133	3783	4433	4549	4495	3833
$T_1(x)$ **{r=50 n=20}**	3663	3956	5202	5315	5120	4327
$T_1(x)$ **{r=50 n=50}**	3677	3713	5526	5560	5313	4279

Table 5.6: Number of samples requiring monotonisation for $T_2(x)$ with exponential responses for each of six different bandwidths. No number is given in brackets since all the samples were unique.

Model	$h_{T,MISE}$	$h_{\eta,MISE}$	$h_{T,cv}$	$h_{\eta,cv}$	$h_{T,ISE}$	$h_{\eta,ISE}$
$T_2(x)$ **{r=5 n=5}**	5986	5820	3946	4309	4141	3468
$T_2(x)$ **{r=5 n=10}**	7585	7336	4820	5080	4944	3824
$T_2(x)$ **{r=5 n=20}**	8067	7800	4597	5029	4659	3498
$T_2(x)$ **{r=5 n=50}**	8362	8054	3496	5002	3960	2593
$T_2(x)$ **{r=10 n=5}**	5252	5135	3153	3627	3378	2661
$T_2(x)$ **{r=10 n=10}**	6905	6707	3868	4462	3955	2739
$T_2(x)$ **{r=10 n=20}**	7579	7271	3691	4746	3678	2220
$T_2(x)$ **{r=10 n=50}**	7821	7514	2681	5357	3386	1339
$T_2(x)$ **{r=50 n=5}**	2355	2919	1194	1632	1329	762
$T_2(x)$ **{r=50 n=10}**	3385	4513	1880	2581	1744	415
$T_2(x)$ **{r=50 n=20}**	4069	5315	2822	3478	896	180
$T_2(x)$ **{r=50 n=50}**	4485	5867	4182	4259	2426	12

Table 5.7: Number of samples requiring monotonisation for $T_3(x)$ with exponential responses for each of six different bandwidths. No number is given in brackets since all the samples were unique.

Model	$h_{T,MISE}$	$h_{\eta,MISE}$	$h_{T,cv}$	$h_{\eta,cv}$	$h_{T,ISE}$	$h_{\eta,ISE}$
$T_3(x)$ **{r=5 n=5}**	2590	2590	3664	4481	3959	2977
$T_3(x)$ **{r=5 n=10}**	2998	2998	4794	5318	5047	3356
$T_3(x)$ **{r=5 n=20}**	2632	2632	5027	5188	4880	2977
$T_3(x)$ **{r=5 n=50}**	1628	1628	4735	4809	4116	2022
$T_3(x)$ **{r=10 n=5}**	2342	2342	3298	4041	3501	2731
$T_3(x)$ **{r=10 n=10}**	2753	2753	4504	4898	4758	3111
$T_3(x)$ **{r=10 n=20}**	2422	2422	4715	4975	4603	2831
$T_3(x)$ **{r=10 n=50}**	1597	1597	4740	4777	4005	2034
$T_3(x)$ **{r=50 n=5}**	2070	2070	3026	3716	3150	2436
$T_3(x)$ **{r=50 n=10}**	2543	2543	4111	4665	4324	2908
$T_3(x)$ **{r=50 n=20}**	2295	2295	4468	4757	4338	2667
$T_3(x)$ **{r=50 n=50}**	1563	1563	4575	4636	3919	1950

Table 5.8: Number of samples requiring monotonisation for $T_4(x)$ with exponential responses for each of six different bandwidths. No number is given in brackets since all the samples were unique.

Model	$h_{T,MISE}$	$h_{\eta,MISE}$	$h_{T,cv}$	$h_{\eta,cv}$	$h_{T,ISE}$	$h_{\eta,ISE}$
$T_4(x)$ **{r=5 n=5}**	4336	3726	1829	2196	2014	1075
$T_4(x)$ **{r=5 n=10}**	6254	5627	2754	2311	2570	541
$T_4(x)$ **{r=5 n=20}**	7654	7520	3777	2288	2586	184
$T_4(x)$ **{r=5 n=50}**	9050	9364	5028	3329	2252	18
$T_4(x)$ **{r=10 n=5}**	3211	2743	1347	1387	1445	691
$T_4(x)$ **{r=10 n=10}**	4320	4412	2209	1621	1902	395
$T_4(x)$ **{r=10 n=20}**	5619	6145	3229	1821	2023	135
$T_4(x)$ **{r=10 n=50}**	7877	8619	5173	2983	2154	6
$T_4(x)$ **{r=50 n=5}**	1372	1907	1043	803	1065	390
$T_4(x)$ **{r=50 n=10}**	1634	2454	1386	812	964	170
$T_4(x)$ **{r=50 n=20}**	2745	4232	2680	1249	1430	81
$T_4(x)$ **{r=50 n=50}**	4664	6952	4860	2441	1674	3

Table 5.9: MISE values for model $T_1(x)$ with Poisson responses for various monotonicity constraints with the two MISE optimal unconstrained bandwidths.

	$T_1(x)$ {r=5 n=5}		$T_1(x)$ {r=5 n=10}		$T_1(x)$ {r=5 n=20}		$T_1(x)$ {r=5 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	1.206411	1.206411	0.626384	0.626383	0.35647	0.35629	0.190001	0.190001
Band	2.05323	2.05323	1.165437	1.165437	0.67364	0.673246	0.329491	0.329491
LDNP	17.66635	17.59149	13.63523	13.57669	11.02418	10.97589	8.343201	8.307534
LDNP1	85.16097	84.16308	50.38479	49.83885	31.76779	31.44505	17.82765	17.66087
LDNP2	28775.51	27171.85	3211.034	3074.543	608.095	587.8968	112.1732	109.3606
LDNP3	2.77E+11	2.17E+11	42799636	36483347	152750.1	137851.7	1643.616	1549.365
	$T_1(x)$ {r=10 n=5}		$T_1(x)$ {r=10 n=10}		$T_1(x)$ {r=10 n=20}		$T_1(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	2.577765	2.578257	1.419583	1.420268	0.852581	0.853886	0.531631	0.531733
Band	4.341961	4.34185	2.56498	2.564785	1.50825	1.509273	0.820333	0.820208
LDNP	54.51082	54.26876	43.78094	43.60565	35.526	35.36771	27.47294	27.35263
LDNP1	201.7044	199.4914	126.3257	125.0829	81.48793	80.698	47.62675	47.19503
LDNP2	12864.37	12316.16	2421.856	2342.003	647.7794	630.3817	162.7743	159.3117
LDNP3	1.72E+08	1.46E+08	608692.3	549452.2	15924.61	14889.78	755.2496	724.3631
	$T_1(x)$ {r=50 n=5}		$T_1(x)$ {r=50 n=10}		$T_1(x)$ {r=50 n=20}		$T_1(x)$ {r=50 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	18.36251	18.36244	13.33279	13.35061	10.41749	10.46004	8.472066	8.472075
Band	28.65409	28.65409	19.50347	19.51457	13.74418	13.77902	9.707814	9.707814
LDNP	826.0831	822.5418	681.8245	678.762	568.1157	566.1406	443.1205	441.2685
LDNP1	1768.038	1751.471	1182.595	1171.735	808.6293	802.2065	492.2162	488.0927
LDNP2	11146.96	10867.27	4045.299	3958.761	1701.052	1670.399	608.3651	598.0979
LDNP3	163510.1	154131.8	18783.69	18013.43	3877.988	3758.575	754.12	734.8277

Table 5.10: MISE values for model $T_2(x)$ with Poisson responses for various monotonicity constraints with the two MISE optimal unconstrained bandwidths.

	$T_2(x)$ {r=5 n=5}		$T_2(x)$ {r=5 n=10}		$T_2(x)$ {r=5 n=20}		$T_2(x)$ {r=5 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	2.115053	2.074763	1.266335	1.244463	0.747724	0.73426	0.364534	0.357338
Band	2.355762	2.348694	1.392216	1.388055	0.851371	0.850871	0.497746	0.500116
LDNP	2.673838	2.674258	1.76704	1.778532	1.158985	1.174282	0.712716	0.73328
LDNP1	2.190963	2.144987	1.368083	1.343605	0.839892	0.825263	0.445669	0.441563
LDNP2	2.124315	2.074154	1.307251	1.281469	0.787695	0.773479	0.395502	0.388701
LDNP3	2.083679	2.03775	1.269286	1.242654	0.764077	0.747666	0.38412	0.3755
	$T_2(x)$ {r=10 n=5}		$T_2(x)$ {r=10 n=10}		$T_2(x)$ {r=10 n=20}		$T_2(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	2.895256	2.837236	1.757153	1.741967	1.037114	1.017077	0.500344	0.491937
Band	3.624902	3.638497	2.311455	2.327426	1.553368	1.562219	1.063504	1.076907
LDNP	3.420544	3.457905	2.336718	2.386816	1.549699	1.577904	0.933506	0.966528
LDNP1	3.038697	2.995948	1.976196	1.97398	1.237059	1.218933	0.649211	0.645229
LDNP2	2.961017	2.92153	1.851204	1.848014	1.112388	1.097428	0.54302	0.535537
LDNP3	2.886461	2.835445	1.797154	1.781615	1.083577	1.061242	0.530224	0.519574
	$T_2(x)$ {r=50 n=5}		$T_2(x)$ {r=50 n=10}		$T_2(x)$ {r=50 n=20}		$T_2(x)$ {r=50 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	7.036718	7.096376	4.305567	4.311893	2.668929	2.680916	1.32898	1.333074
Band	12.30409	12.31983	8.724165	8.695157	6.794323	6.81605	5.15092	5.193997
LDNP	9.965838	9.935326	7.178789	7.097714	5.211081	5.16776	3.435725	3.394524
LDNP1	8.422395	8.448708	5.808818	5.778899	4.020659	4.034886	2.478837	2.460336
LDNP2	7.339798	7.381262	4.553907	4.545549	2.838711	2.845174	1.424363	1.425771
LDNP3	7.119906	7.177007	4.40776	4.412329	2.760696	2.77305	1.387863	1.392343

Table 5.11: MISE values for model $T_3(x)$ with Poisson responses for various monotonicity constraints with the two MISE optimal unconstrained bandwidths.

	$T_3(x)$ {r=5 n=5}		$T_3(x)$ {r=5 n=10}		$T_3(x)$ {r=5 n=20}		$T_3(x)$ {r=5 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	2.449804	2.614701	1.180026	1.313031	0.637445	0.681642	0.302017	0.318461
Band	1.555553	1.560598	0.874036	0.878631	0.492791	0.496074	0.235704	0.23841
LDNP	145499.2	3.05E+09	2.330467	2.76913	0.8794	1.001824	0.410444	0.448054
LDNP1	13267.72	5553218	1.431201	1.738821	0.737913	0.868728	0.326097	0.375013
LDNP2	3.729544	3.575433	1.36545	1.648215	0.70337	0.830285	0.309208	0.35647
LDNP3	2.538385	2.651451	1.354593	1.638638	0.698034	0.825972	0.307075	0.354768
	$T_3(x)$ {r=10 r=5}		$T_3(x)$ {r=10 n=10}		$T_3(x)$ {r=10 n=20}		$T_3(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	5.007685	5.092647	2.610904	2.805954	1.41602	1.497989	0.684621	0.718393
Band	3.723121	3.737933	1.960122	1.971644	1.096981	1.105571	0.562883	0.568852
LDNP	1974220	42892873	5.174386	4.397511	1.966941	2.148522	0.94659	0.997294
LDNP1	11872.62	118697.6	3.037612	3.502777	1.577845	1.827689	0.709919	0.800543
LDNP2	12.50601	11.58086	2.897868	3.339215	1.496826	1.737864	0.673001	0.760069
LDNP3	5.248758	5.326574	2.875779	3.321118	1.485826	1.729154	0.668942	0.75686
	$T_3(x)$ {r=50 n=5}		$T_3(x)$ {r=50 n=10}		$T_3(x)$ {r=50 n=20}		$T_3(x)$ {r=50 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	42.80082	42.7607	19.71545	19.97513	10.29325	10.69203	4.958319	5.143068
Band	39.04381	39.02389	17.45809	17.49808	9.49796	9.514766	5.696348	5.644891
LDNP	62.75915	58.46944	28.92103	27.01656	16.30897	15.60158	8.553355	7.977011
LDNP1	45.02502	44.18598	20.38231	20.53559	11.04977	12.03599	5.30505	5.817885
LDNP2	41.5039	41.39637	19.28416	19.52474	10.61101	11.51844	5.145942	5.611007
LDNP3	41.29056	41.2301	19.17046	19.4574	10.56092	11.47939	5.128955	5.597775

Table 5.12: MISE values for model $T_4(x)$ with Poisson responses for various monotonicity constraints with the two MISE optimal unconstrained bandwidths.

	$T_4(x)$ {r=5 n=5}		$T_4(x)$ {r=5 n=10}		$T_4(x)$ {r=5 n=20}		$T_2(x)$ {r=5 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	1.626561	1.718223	0.862309	0.912255	0.516904	0.536231	0.264304	0.271036
Band	1.183944	1.168333	0.804691	0.808556	0.558004	0.551471	0.378078	0.365249
LDNP	201.7526	57171.94	153.9757	59.31713	741.4429	56.13633	2.102522	3.929566
LDNP1	2.709927	16017.6	1.85512	1.36657	0.690521	0.741054	0.380671	0.416871
LDNP2	1.285867	357.051	0.824466	0.891243	0.509753	0.547782	0.279674	0.303776
LDNP3	1.270868	1.335036	0.808114	0.872183	0.505458	0.542472	0.277506	0.301059
	$T_4(x)$ {r=10 n=5}		$T_4(x)$ {r=10 n=10}		$T_4(x)$ {r=10 n=20}		$T_4(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	3.429342	3.366631	1.902225	2.016681	1.180199	1.221415	0.588223	0.592227
Band	3.072618	2.98911	2.14863	2.1026	1.663959	1.591379	1.248196	1.153543
LDNP	176.3891	31813.98	398.7778	192.0207	4.673763	159.5375	2.049661	7.599781
LDNP1	3.798321	3579.764	2.559814	3.916771	1.383835	1.90633	0.78436	1.130486
LDNP2	2.913804	5.131807	1.845988	2.105387	1.199079	1.338505	0.633958	0.724158
LDNP3	2.918413	2.943169	1.833934	2.044754	1.192453	1.324858	0.629585	0.715969
	$T_4(x)$ {r=50 r=5}		$T_4(x)$ {r=50 n=10}		$T_4(x)$ {r=50 n=20}		$T_4(x)$ {r=50 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	32.68697	30.07755	14.59551	13.66871	9.433486	8.050504	5.233598	3.946199
Band	40.31826	35.63224	33.70733	29.83211	29.72494	25.34182	26.13857	21.35084
LDNP	33.46748	483952.2	29.87509	379.5394	25.10338	180.3126	22.64933	179.461
LDNP1	29.32446	833.8559	17.42902	49.57347	12.27809	22.20455	8.117625	16.53098
LDNP2	31.01161	36.29861	14.98921	16.71305	9.912248	10.36583	5.699915	6.090306
LDNP3	31.66445	31.52732	14.96223	16.11201	9.862479	10.15483	5.647852	5.927534

Table 5.13: MISE values for model $T_1(x)$ with exponential responses for various monotonicity constraints with the two MISE optimal unconstrained bandwidths.

	$T_1(x)$ {r=5 n=5}		$T_1(x)$ {r=5 n=10}		$T_1(x)$ {r=5 n=20}		$T_1(x)$ {r=5 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	5.329285	5.32873	2.8617	2.861697	1.482165	1.482165	0.66731	0.66731
Band	9.817844	2.60E+278	5.615154	5.615154	2.956539	2.956539	1.315424	1.315424
LDNP	22.28008	22.27821	14.9662	14.96334	10.3993	10.39706	7.072708	7.07101
LDNP1	49.70956	49.69194	30.48744	30.47135	18.87275	18.86286	10.55481	10.54932
LDNP2	1202.134	1199.57	294.4178	293.8683	92.1707	92.03073	26.11485	26.08274
LDNP3	1407494	1396509	17459.73	17368.99	968.0017	964.6386	77.7003	77.52487
	$T_1(x)$ {r=10 n=5}		$T_1(x)$ {r=10 n=10}		$T_1(x)$ {r=10 n=20}		$T_1(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	22.20045	22.20041	11.52404	11.52403	6.190596	6.190594	2.777616	2.777616
Band	39.74895	3.37E+56	21.95228	21.95228	12.38399	12.38399	5.421234	5.421234
LDNP	91.14649	91.1322	59.60062	59.58915	42.86543	42.8562	28.63309	28.62622
LDNP1	203.6103	203.5021	122.1318	122.0669	77.3941	77.35368	42.64185	42.61978
LDNP2	4923.417	4911.721	1184.693	1182.482	375.3487	374.7802	105.2226	105.0935
LDNP3	5758901	5712526	70294.23	69928.84	3927.677	3914.042	312.5833	311.8783
	$T_1(x)$ {r=10 n=5}		$T_1(x)$ {r=10 n=10}		$T_1(x)$ {r=10 n=20}		$T_1(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	508.9205	508.9196	281.9647	281.9644	153.2467	153.2466	65.47362	65.45715
Band	553.4249	947.5864	364.2915	555.3476	241.4703	296.8803	127.4424	132.8614
LDNP	2254.516	2254.149	1520.424	1520.141	1036.754	1036.529	697.9252	697.7269
LDNP1	5132.071	5129.283	3106.572	3104.932	1881.174	1880.185	1041.114	1040.538
LDNP2	126017.7	125718.1	29986.43	29930.5	9186.591	9172.64	2576.529	2573.307
LDNP3	1.48E+08	1.46E+08	1776098	1766867	96499.14	96163.84	7672.85	7655.417

Table 5.14: MISE values for model $T_2(x)$ with exponential responses for various monotonicity constraints with the two MISE optimal unconstrained bandwidths.

	$T_2(x)$ {r=5 n=5}		$T_2(x)$ {r=5 n=10}		$T_2(x)$ {r=5 n=20}		$T_2(x)$ {r=5 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	7.599314	7.561595	5.389267	5.314188421	3.12343	3.053847	1.677073	1.635755
Band	9.837699	NA	6.140509	1.10E+93	3.369037	3.365788	1.667026	1.666949
LDNP	9.459028	5.56E+61	6.693869	6.580549907	3.764444	3.677251	2.058004	2.011134
LDNP1	8.231448	1.99E+44	5.842302	5.718803191	3.21453	3.109526	1.694651	1.62834
LDNP2	7.852137	1.13E+14	5.660414	5.522875075	3.1409	3.041074	1.646595	1.596338
LDNP3	7.758362	72628.38	5.52039	5.40947007	3.054872	2.96855	1.598009	1.550182
	$T_2(x)$ {r=10 n=5}		$T_2(x)$ {r=10 n=10}		$T_2(x)$ {r=10 n=20}		$T_2(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	12.25518	12.13594	8.974071	8.997412286	5.753322	5.698605	3.047764	2.999623
Band	14.25421	14.56566	9.226711	2.44736E+11	5.601124	5.625428	3.099679	3.115836
LDNP	11.67498	3.99E+49	8.959848	9.060396373	5.831735	5.732836	3.326704	3.261565
LDNP1	11.05059	2.60E+39	8.411053	8.474154803	5.492708	5.382667	3.101515	3.025898
LDNP2	11.09186	1.4E+14	8.416486	8.522657713	5.470036	5.413002	3.003717	2.961596
LDNP3	10.97176	337.5858	8.192042	8.305777992	5.319975	5.259514	2.930359	2.883629
	$T_2(x)$ {r=10 n=5}		$T_2(x)$ {r=10 n=10}		$T_2(x)$ {r=10 n=20}		$T_2(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	124.4101	120.6851	82.56617	85.87304171	51.56527	55.29811	25.2096	28.58422
Band	123.2987	115.5426	77.77952	1.60E+201	50.47763	48.20378	31.1964	29.36496
LDNP	99.66583	8.84E+16	79.91264	86.44035524	56.26182	61.46014	29.67179	34.31107
LDNP1	105.5905	2.18E+12	80.38865	86.13564254	54.6001	59.86164	27.99034	32.79755
LDNP2	111.4504	55158.31	81.94965	84.11697229	53.23315	55.76212	26.02261	29.20698
LDNP3	112.7385	113.0504	80.13862	82.47630935	51.53954	54.57897	25.42561	28.71242

Table 5.15: MISE values for model $T_3(x)$ with exponential responses for various monotonicity constraints with the two MISE optimal unconstrained bandwidths.

	$T_3(x)$ {r=5 n=5}		$T_3(x)$ {r=5 n=10}		$T_3(x)$ {r=5 n=20}		$T_3(x)$ {r=5 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	3.807761	3.807761	2.231334	2.231334	1.368351	1.368351	0.804958	0.804958
Band	5.613576	5.613576	3.459629	3.459629	2.127352	2.127352	1.158136	1.158136
LDNP	5.513265	5.513265	3.59273	3.59273	2.389482	2.389482	1.791031	1.791031
LDNP1	5.513265	5.513265	3.59273	3.59273	2.389482	2.389482	1.791031	1.791031
LDNP2	5.513265	5.513265	3.59273	3.59273	2.389482	2.389482	1.791031	1.791031
LDNP3	5.513265	5.513265	3.59273	3.59273	2.389482	2.389482	1.791031	1.791031
	$T_3(x)$ {r=10 n=5}		$T_3(x)$ {r=10 n=10}		$T_3(x)$ {r=10 n=20}		$T_3(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	15.7191	15.7191	8.403696	8.403696	5.326183	5.326183	3.258642	3.258642
Band	3.80E+129	3.80E+129	13.30897	13.30897	8.07122	8.07122	4.544695	4.544695
LDNP	21.88	21.88	13.72008	13.72008	9.314132	9.314132	6.879427	6.879427
LDNP1	21.88	21.88	13.72008	13.72008	9.314132	9.314132	6.879427	6.879427
LDNP2	21.88	21.88	13.72008	13.72008	9.314132	9.314132	6.879427	6.879427
LDNP3	21.88	21.88	13.72008	13.72008	9.314132	9.314132	6.879427	6.879427
	$T_3(x)$ {r=50 n=5}		$T_3(x)$ {r=50 n=10}		$T_3(x)$ {r=50 n=20}		$T_3(x)$ {r=50 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	336.2891	336.2891	208.0563	208.0563	128.5122	128.5122	81.25144	81.25144
Band	509.8155	509.8155	316.4733	316.4733	199.3579	199.3579	113.2871	113.2871
LDNP	528.9493	528.9493	322.4039	322.4039	224.9415	224.9415	163.8335	163.8335
LDNP1	528.9493	528.9493	322.4039	322.4039	224.9415	224.9415	163.8335	163.8335
LDNP2	528.9493	528.9493	322.4039	322.4039	224.9415	224.9415	163.8335	163.8335
LDNP3	528.9493	528.9493	322.4039	322.4039	224.9415	224.9415	163.8335	163.8335

Table 5.16: MISE values for model $T_4(x)$ with exponential responses for various monotonicity constraints with the two MISE optimal unconstrained bandwidths.

	$T_4(x)$ {r=5 n=5}		$T_4(x)$ {r=5 n=10}		$T_4(x)$ {r=5 n=20}		$T_4(x)$ {r=5 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	2.721742	3.337829	2.045096	2.35135	1.412733	1.496283	0.741693	0.793215
Band	3.034254	3.374943	2.092711	2.289296	1.24913	1.321143	0.700951	0.714303
LDNP	2.97E+67	2.43E+59	4.46E+15	644337.9	7.191092	7.370402	0.818358	2.917816
LDNP1	8.22E+43	1.77E+20	234949.6	12.31607	1.451285	1.661458	0.746483	0.80419
LDNP2	6.58E+11	8.425659	1.981404	2.650605	1.343072	1.576656	0.700063	0.771109
LDNP3	37.08737	7.881792	1.928373	2.629266	1.331154	1.568906	0.695494	0.768879
	$T_4(x)$ {r=10 n=5}		$T_4(x)$ {r=10 n=10}		$T_4(x)$ {r=10 n=20}		$T_4(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	11.21839	12.59281	7.910831	8.887194	5.41222	5.953719	3.01389	NA
Band	12.59236	13.29027	8.327243	8.809141	4.951781	5.361721	2.853848	NA
LDNP	1.27E+59	6.88E+27	1.16E+26	8120.781	153.3943	14.51135	3.494006	1.32E+48
LDNP1	1.49E+35	1.01E+22	78441328	14.37779	5.522481	6.487274	3.060612	NA
LDNP2	50849268	43.68008	7.497888	9.969634	5.091917	6.18919	2.862734	NA
LDNP3	37.34741	22.33016	7.380398	9.930146	5.043025	6.166765	2.840801	NA
	$T_4(x)$ {r=10 n=5}		$T_4(x)$ {r=10 n=10}		$T_4(x)$ {r=10 n=20}		$T_4(x)$ {r=10 n=50}	
	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA	MISEP	MISEETA
PAVA	267.1016	279.1476	187.2884	230.0037	119.1169	140.6976	70.17869	78.84073
Band	380.3168	346.9927	208.3483	233.7563	116.9586	130.8722	68.47212	71.5403
LDNP	4.77E+56	1.17E+36	351.5953	226702.6	246.9653	223.9854	148.8023	94.45959
LDNP1	4.71E+44	4.52E+08	196.3658	294.3131	121.7635	148.119	73.00416	81.55932
LDNP2	3.33E+15	433.0218	172.2097	245.3691	107.7769	142.2195	66.69773	78.58842
LDNP3	1068.587	368.7103	170.1173	244.1748	107.2052	141.8118	66.22788	78.42494

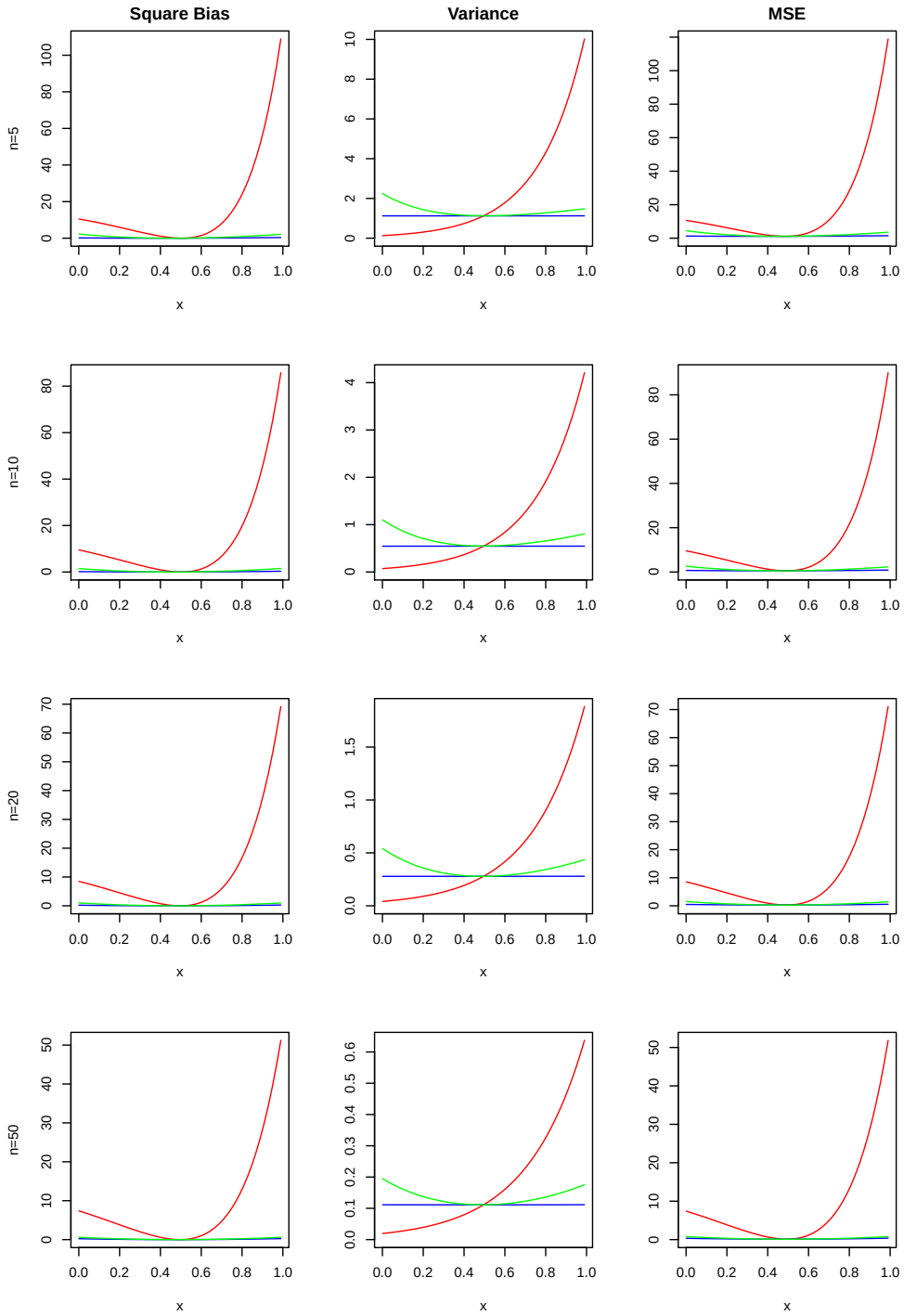

Figure 5.3: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_1(x)$, with Poisson responses, with $r = 5$ and h_r set as $h_{T,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

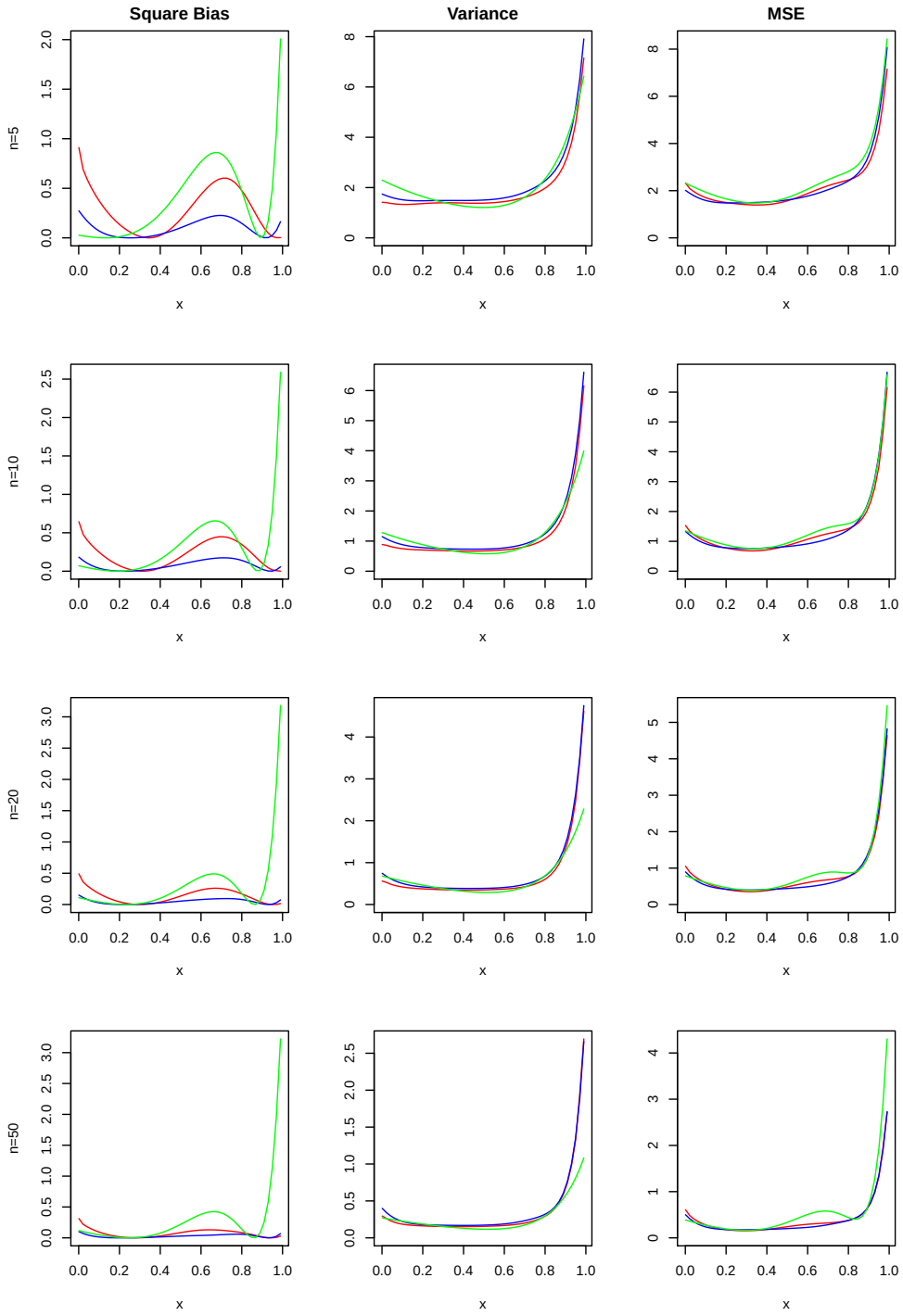

Figure 5.4: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_2(x)$, with Poisson responses, with $r = 5$ and h_r set as $h_{T,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

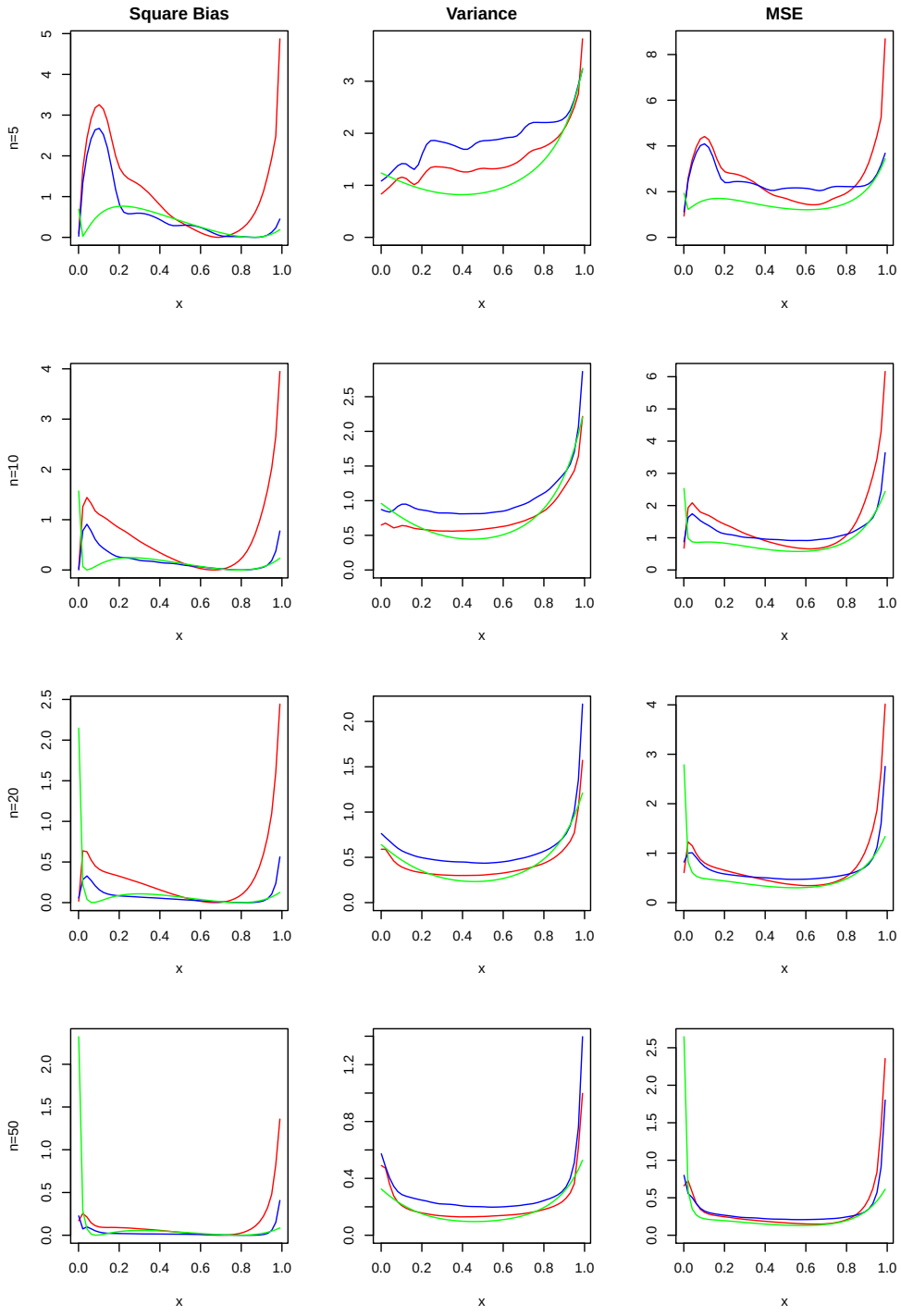

Figure 5.5: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_3(x)$, with Poisson responses, with $r = 5$ and h_r set as $h_{T,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

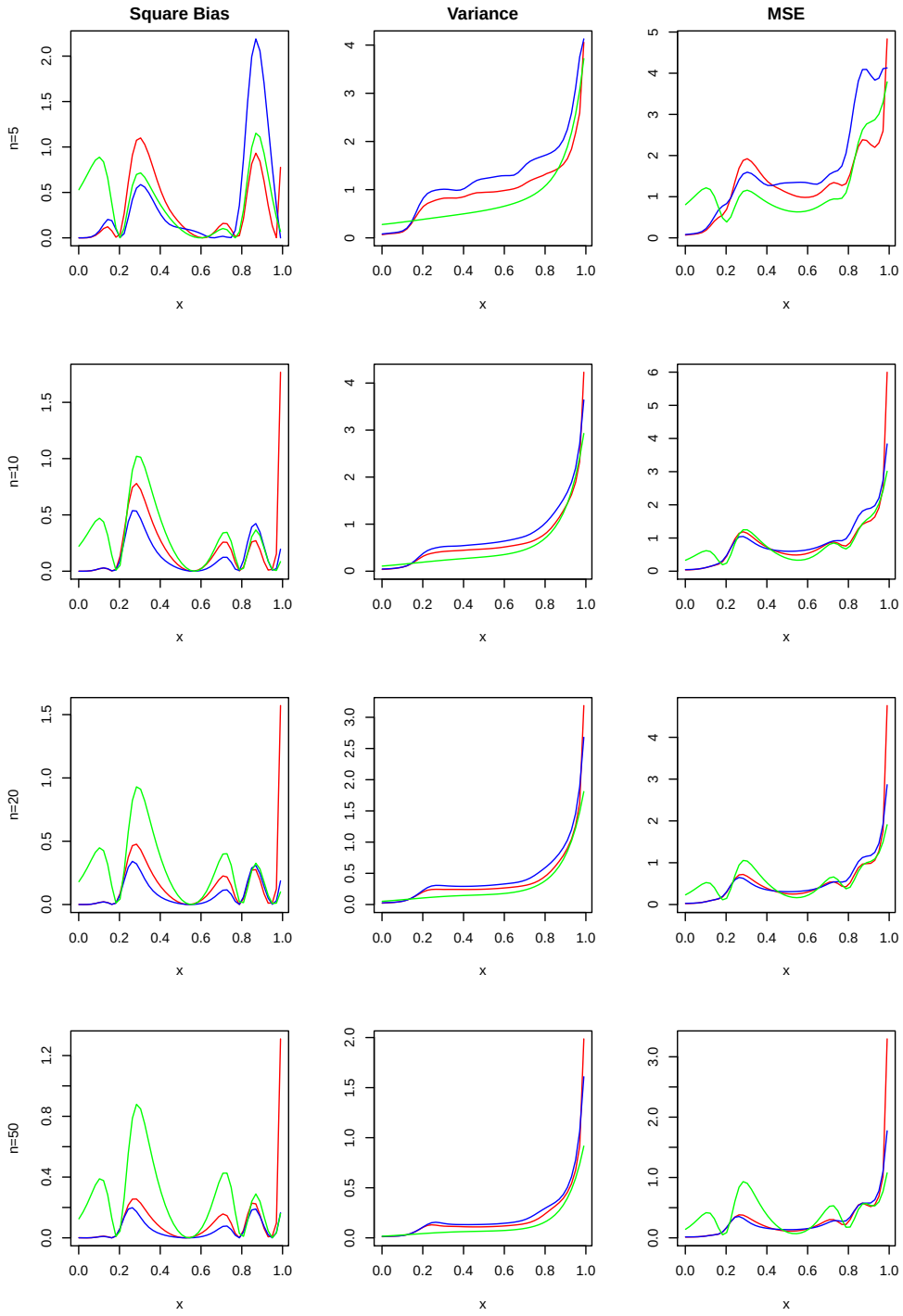

Figure 5.6: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_4(x)$, with Poisson responses, with $r = 5$ and h_r set as $h_{T,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

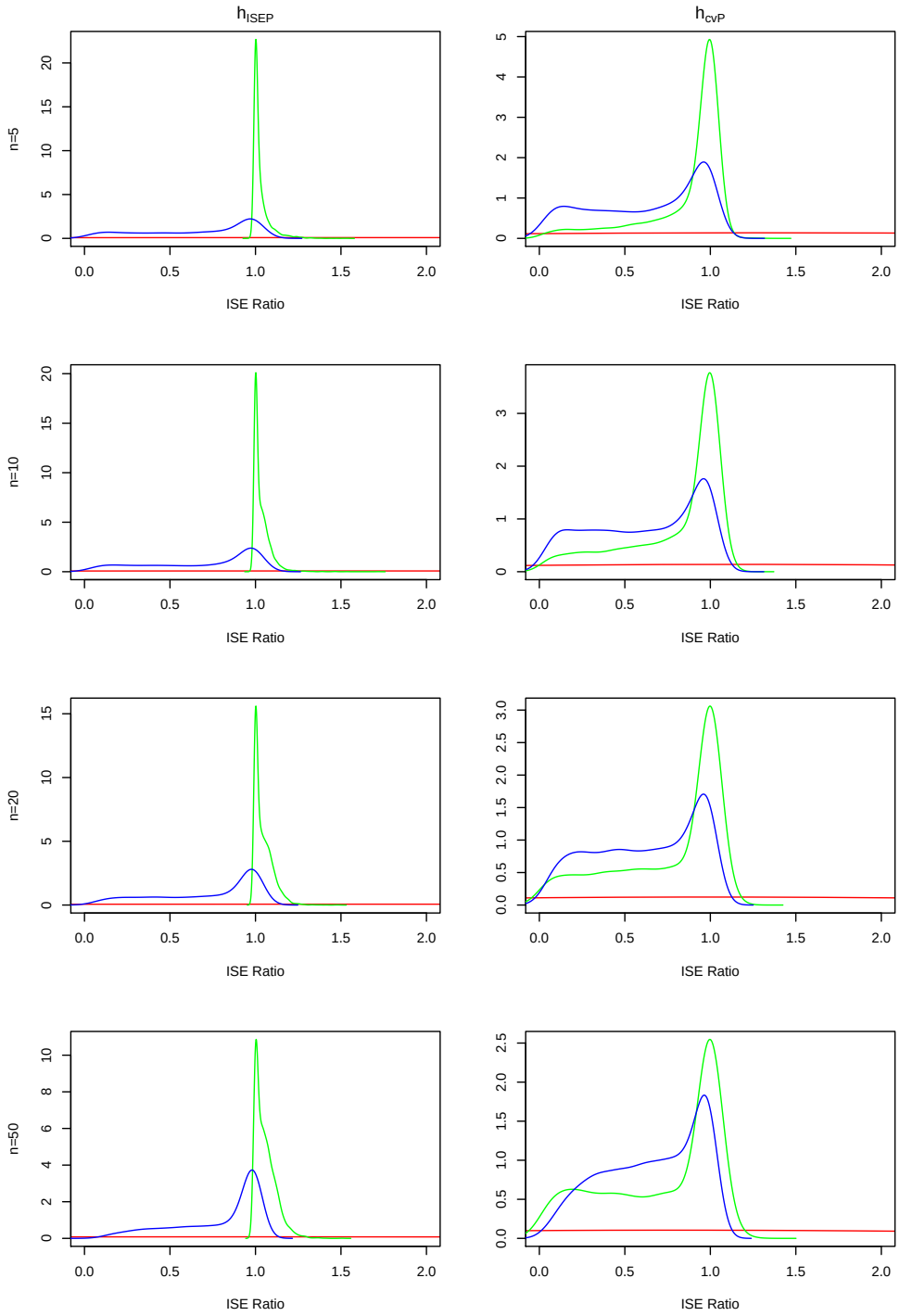

Figure 5.7: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_1(x)$, with Poisson responses, with $r = 5$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

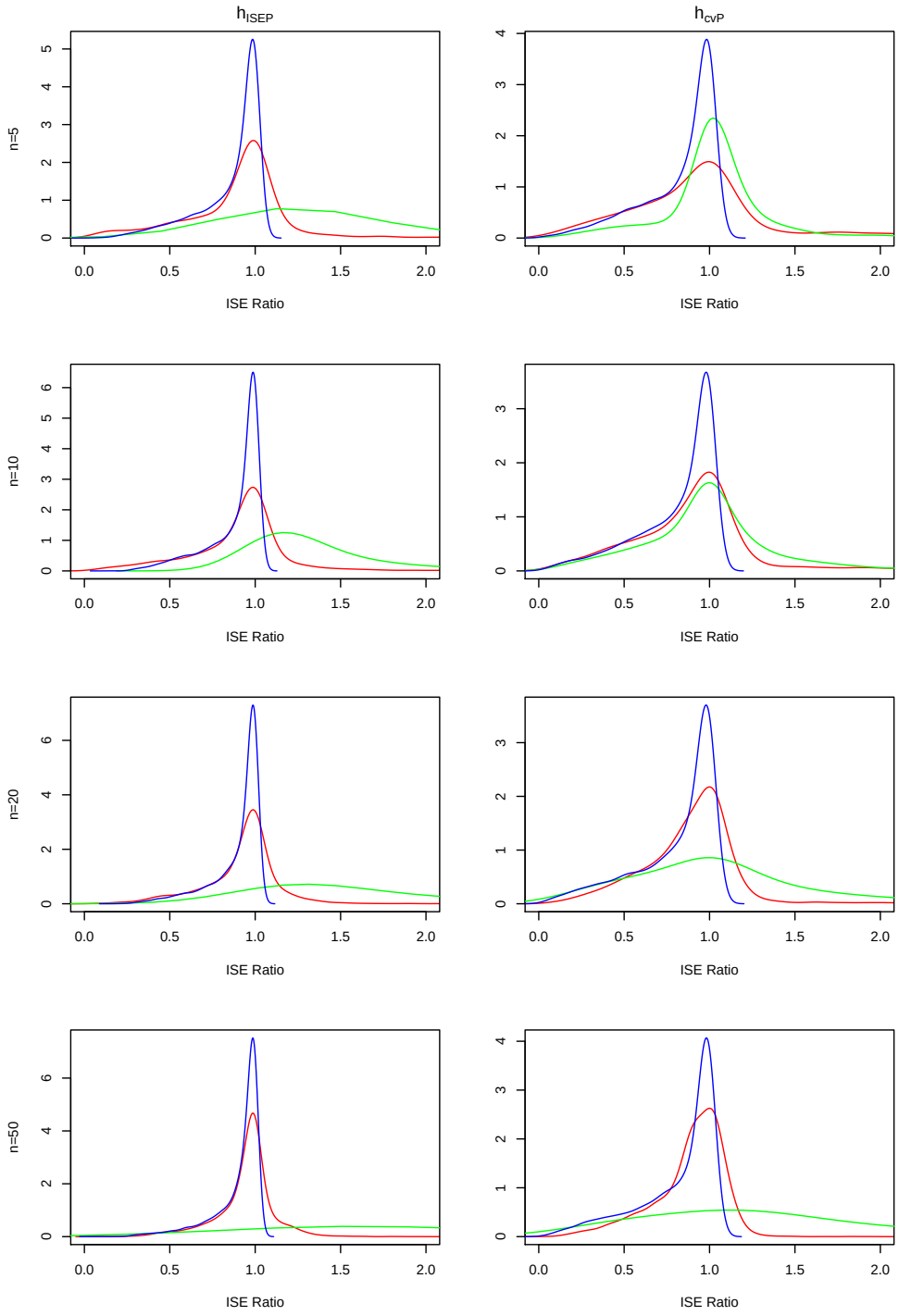

Figure 5.8: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_2(x)$, with Poisson responses, with $r = 5$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

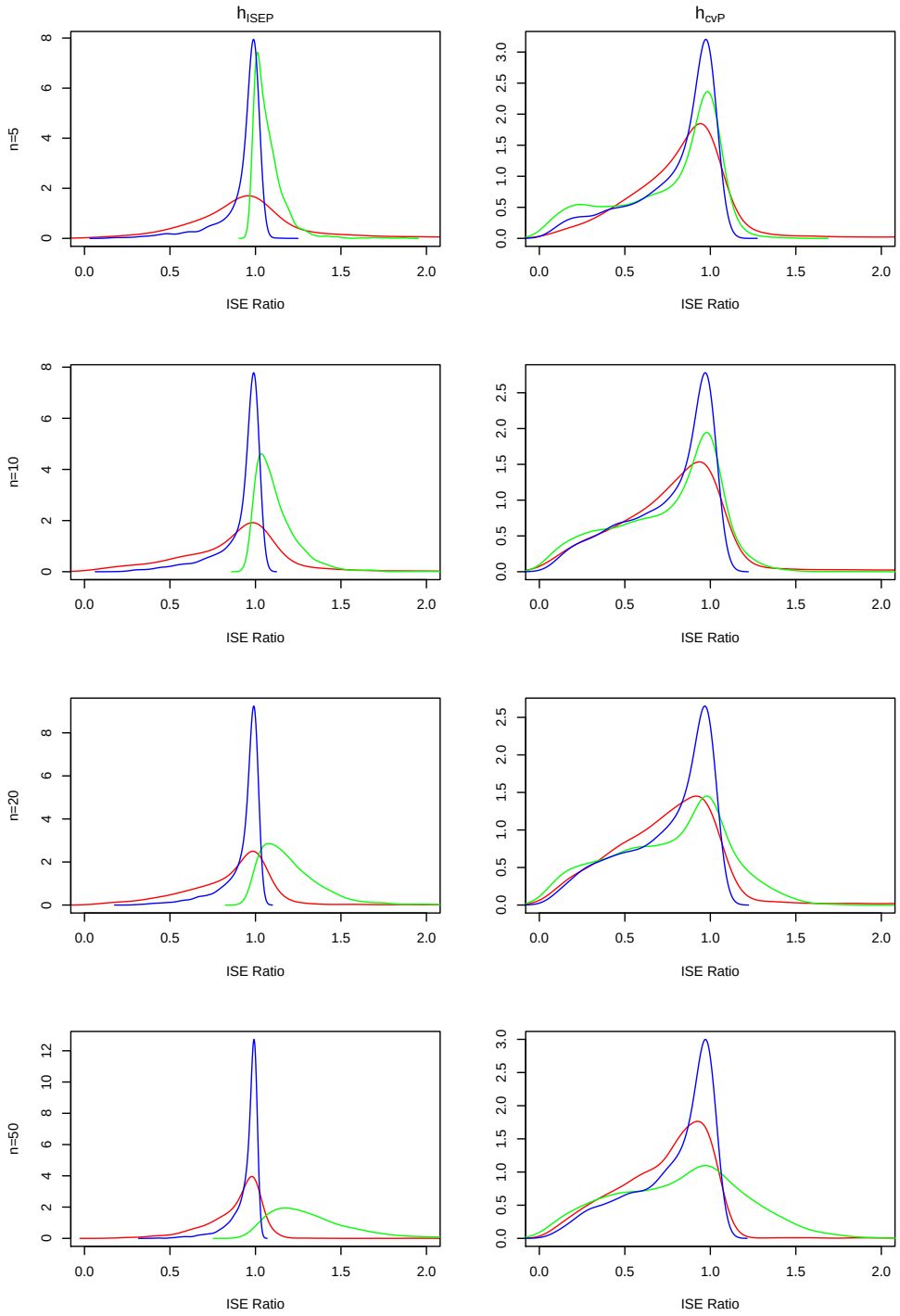

Figure 5.9: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_3(x)$, with Poisson responses, with $r = 5$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

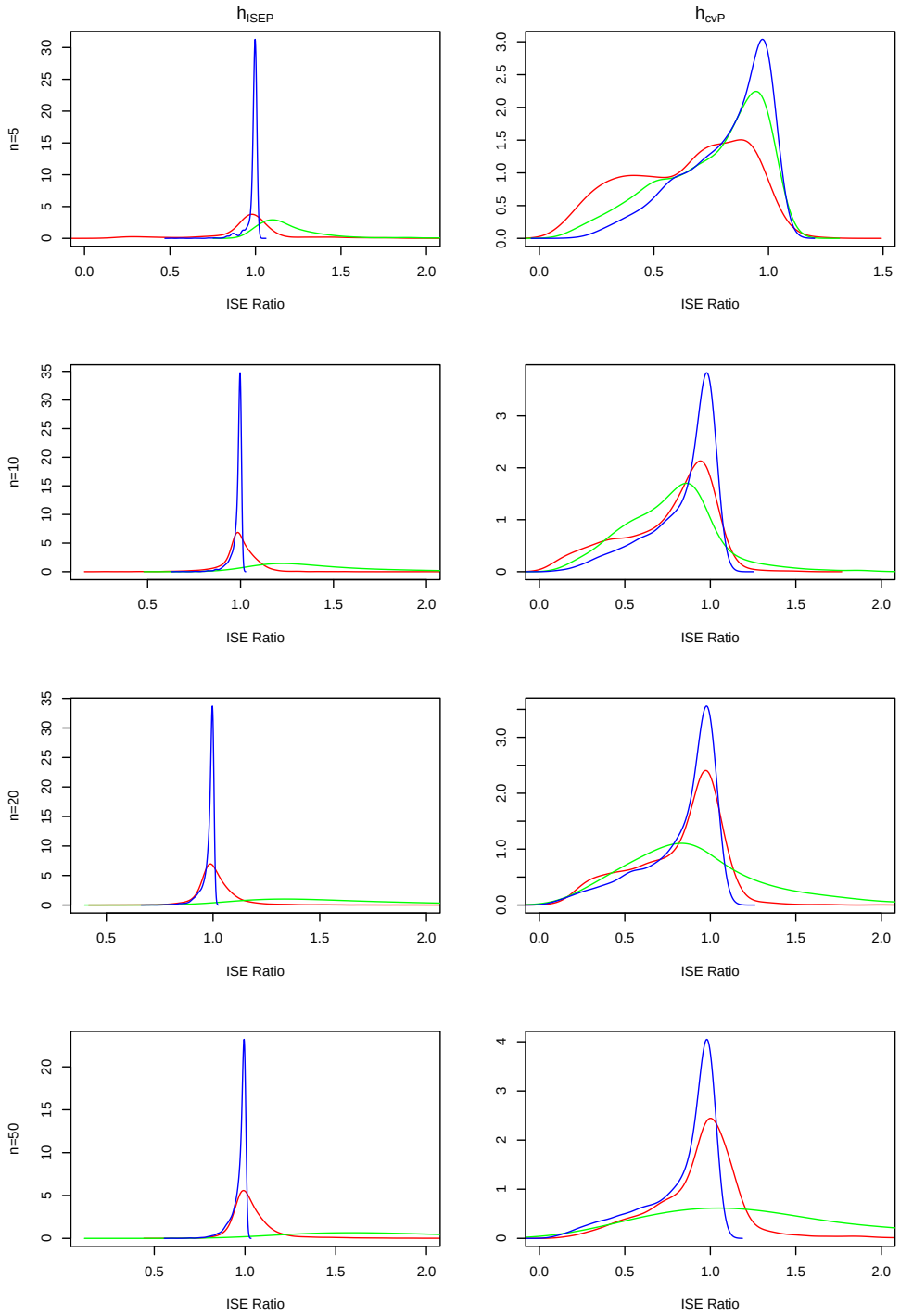

Figure 5.10: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_4(x)$, with Poisson responses, with $r = 5$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

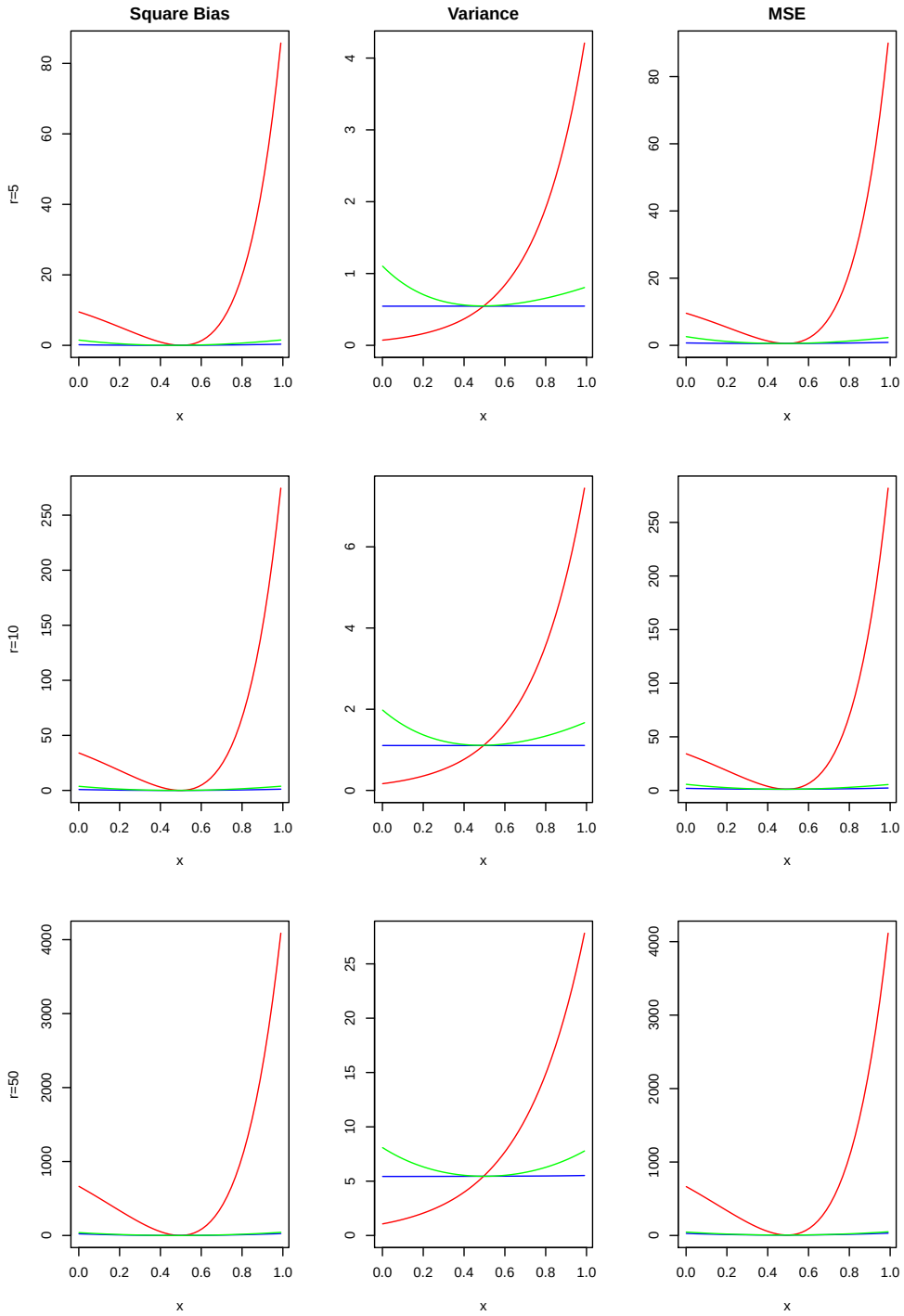

Figure 5.11: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_1(x)$, with Poisson responses, with $n = 10$ and h_r set as $h_{T,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

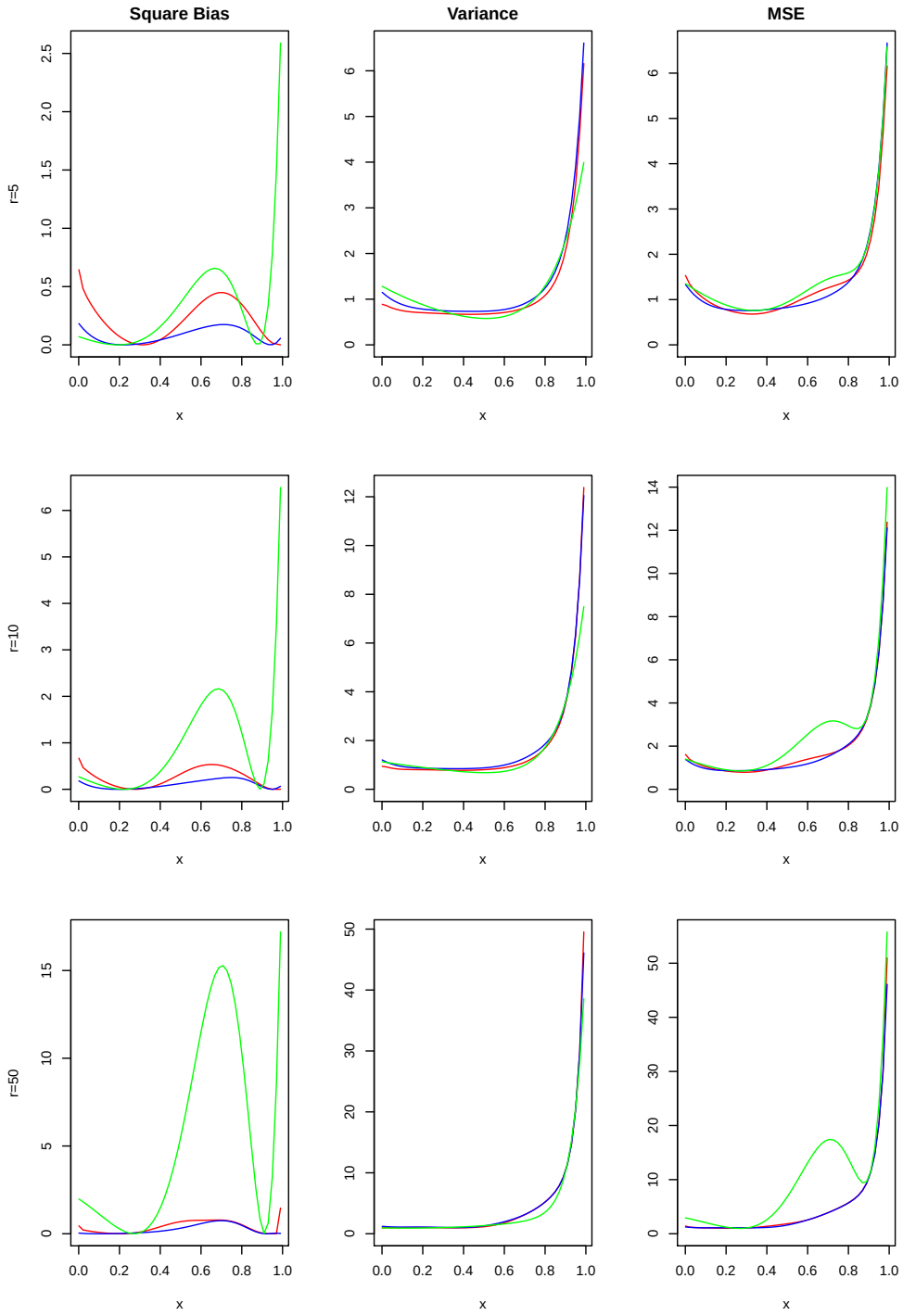

Figure 5.12: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_2(x)$, with Poisson responses, with $n = 10$ and h_r set as $h_{T,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

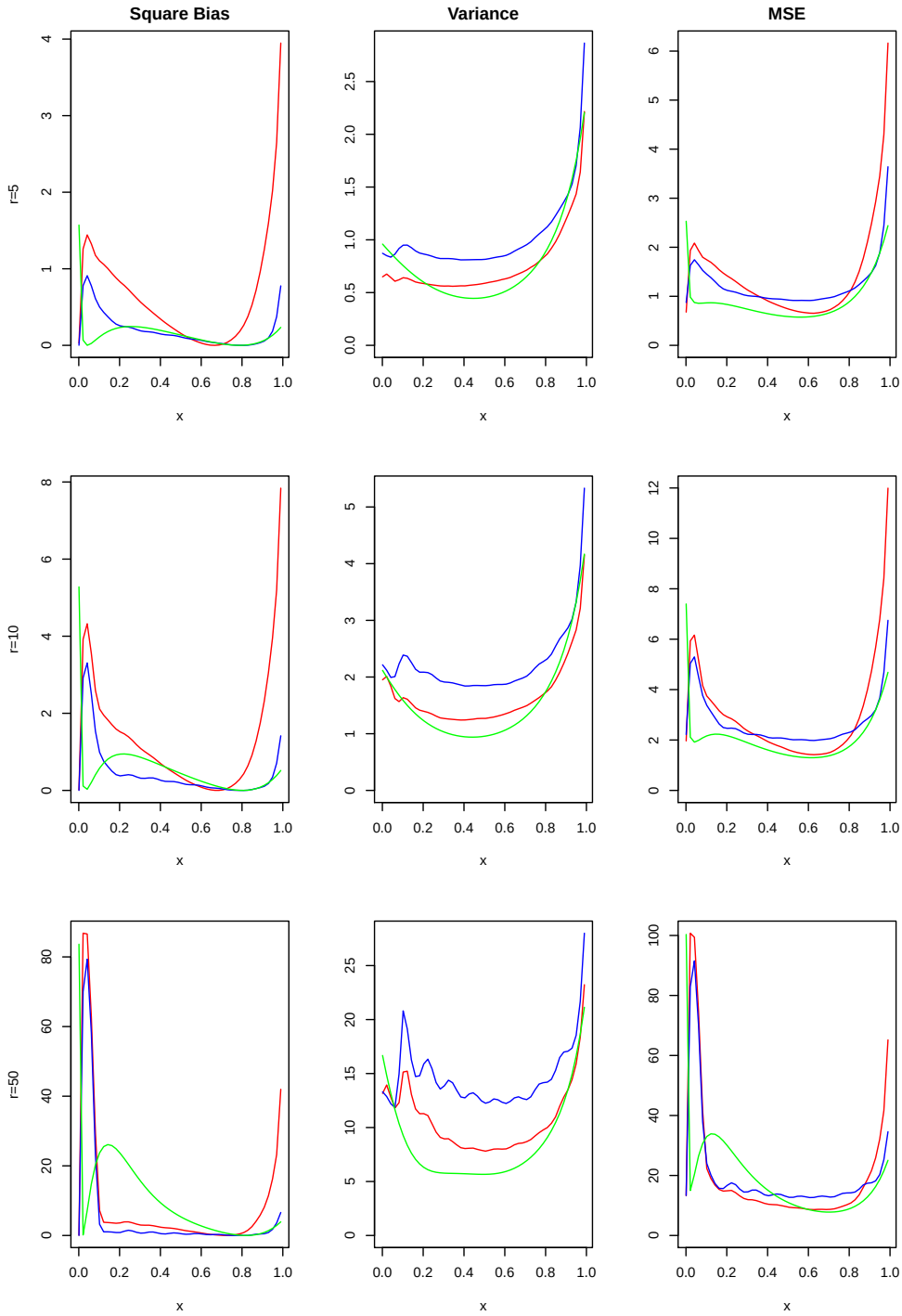

Figure 5.13: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_3(x)$, with Poisson responses, with $n = 10$ and h_r set as $h_{T,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

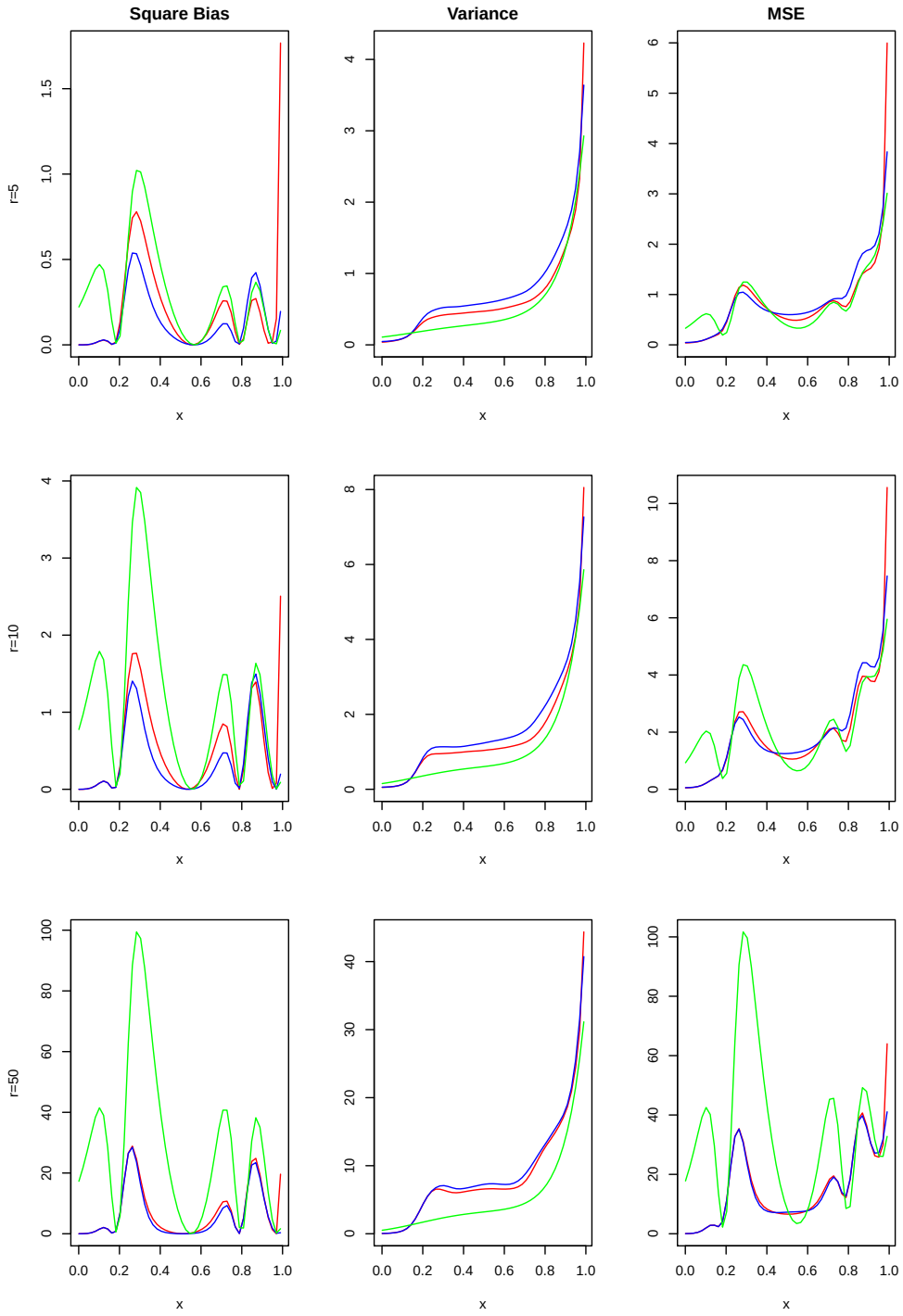

Figure 5.14: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_4(x)$, with Poisson responses, with $n = 10$ and h_r set as $h_{T,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

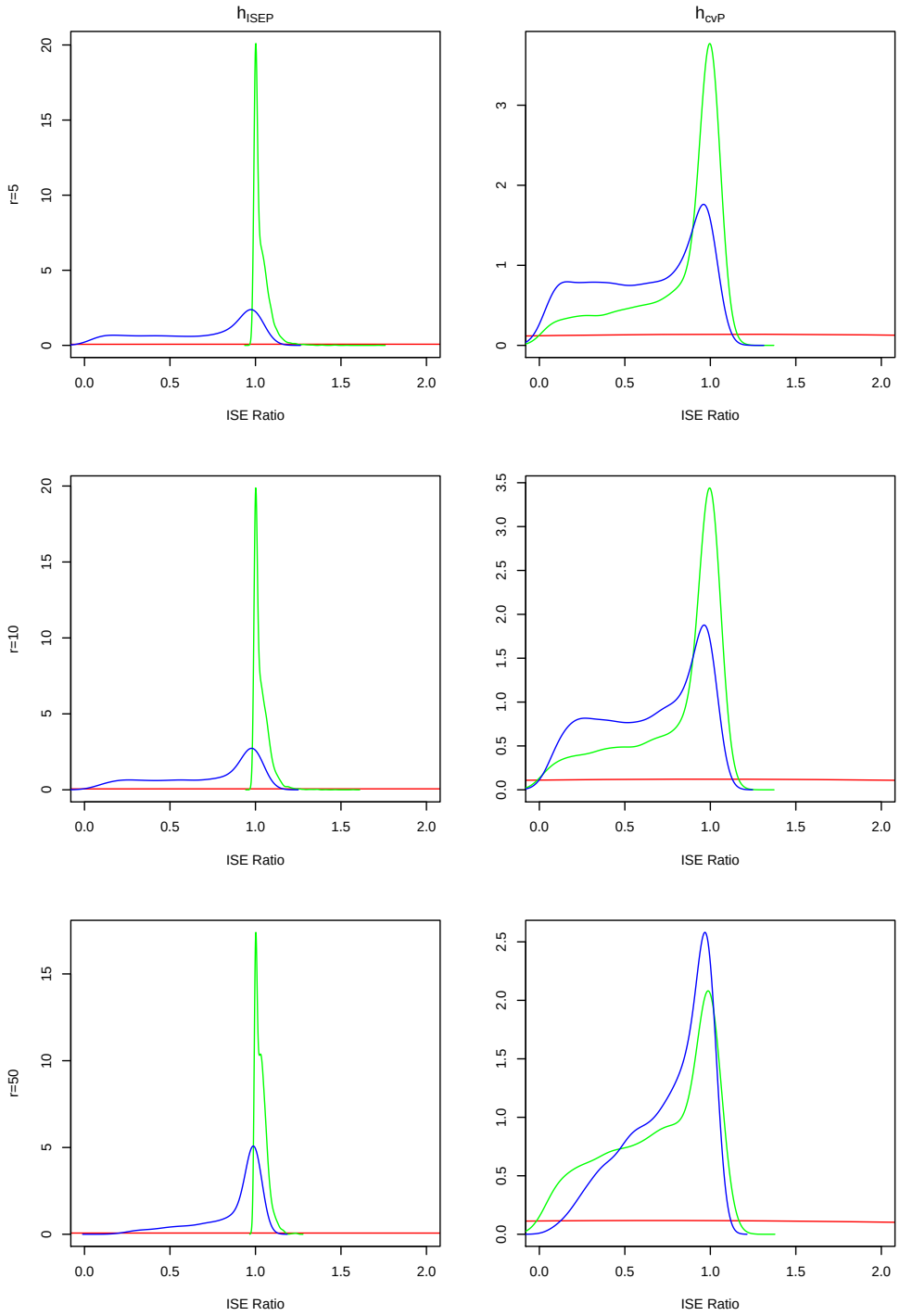

Figure 5.15: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_1(x)$, with Poisson responses, with $n = 10$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

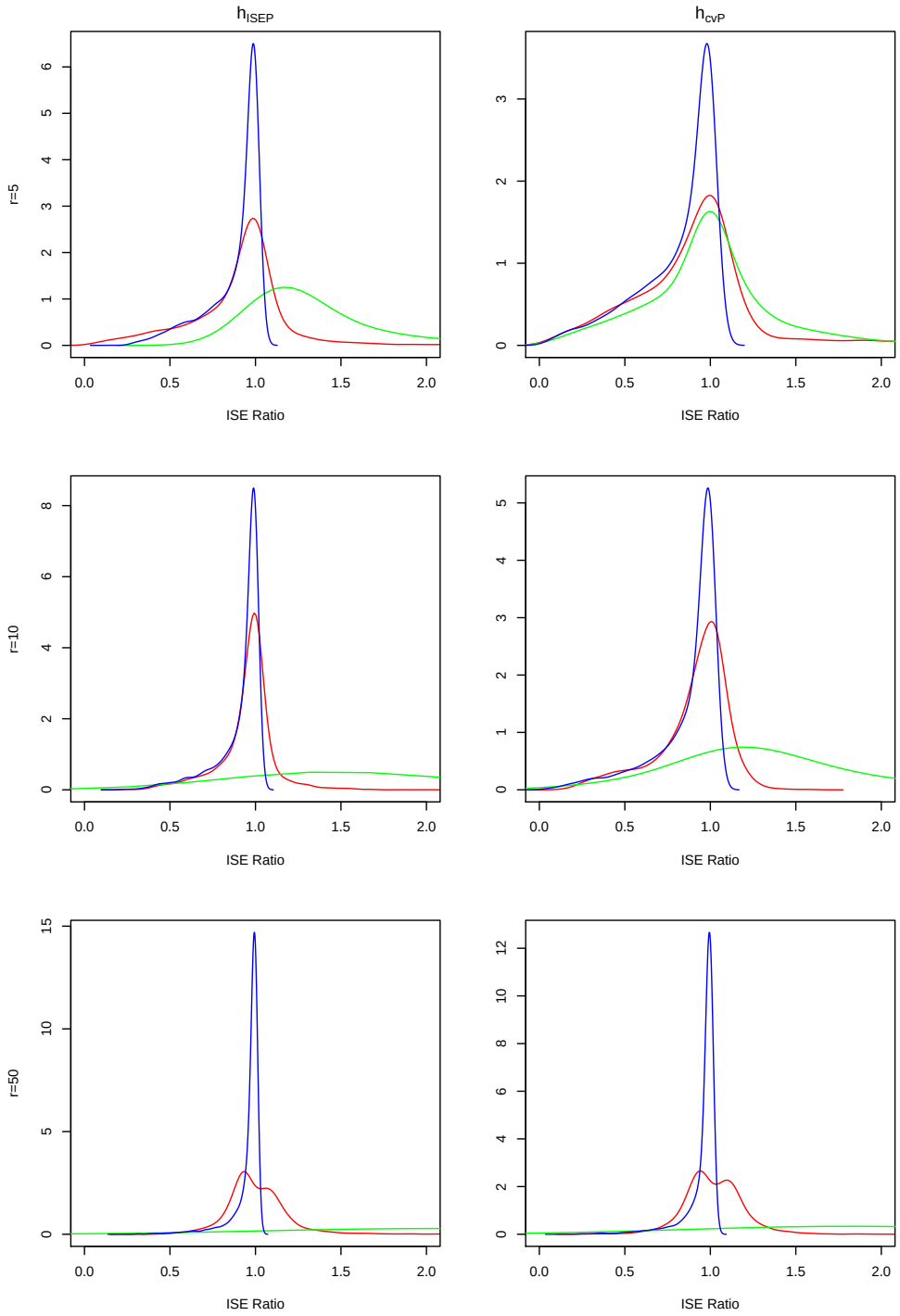

Figure 5.16: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_2(x)$, with Poisson responses, with $n = 10$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

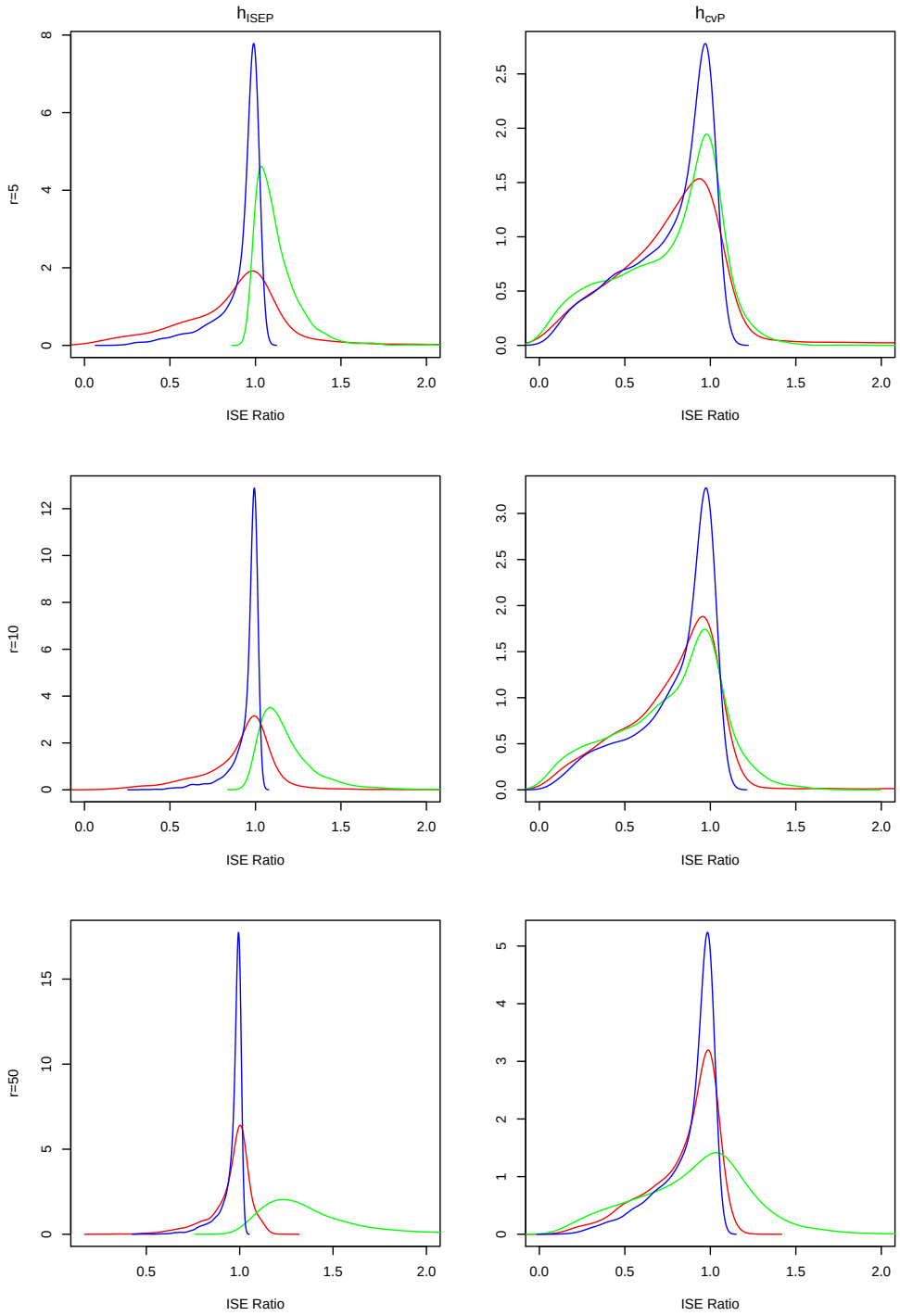

Figure 5.17: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_3(x)$, with Poisson responses, with $n = 10$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

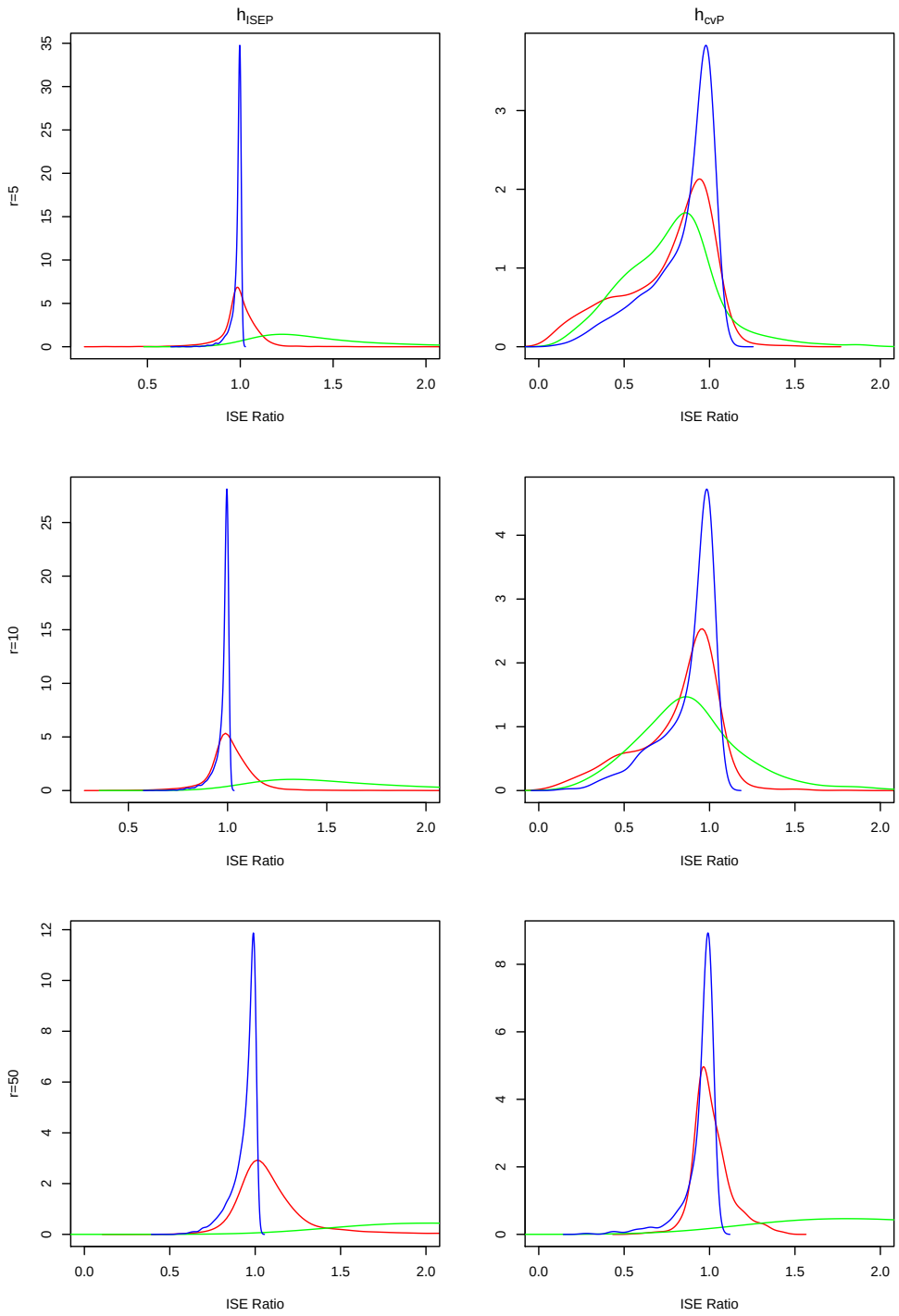

Figure 5.18: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_4(x)$, with Poisson responses, with $n = 10$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

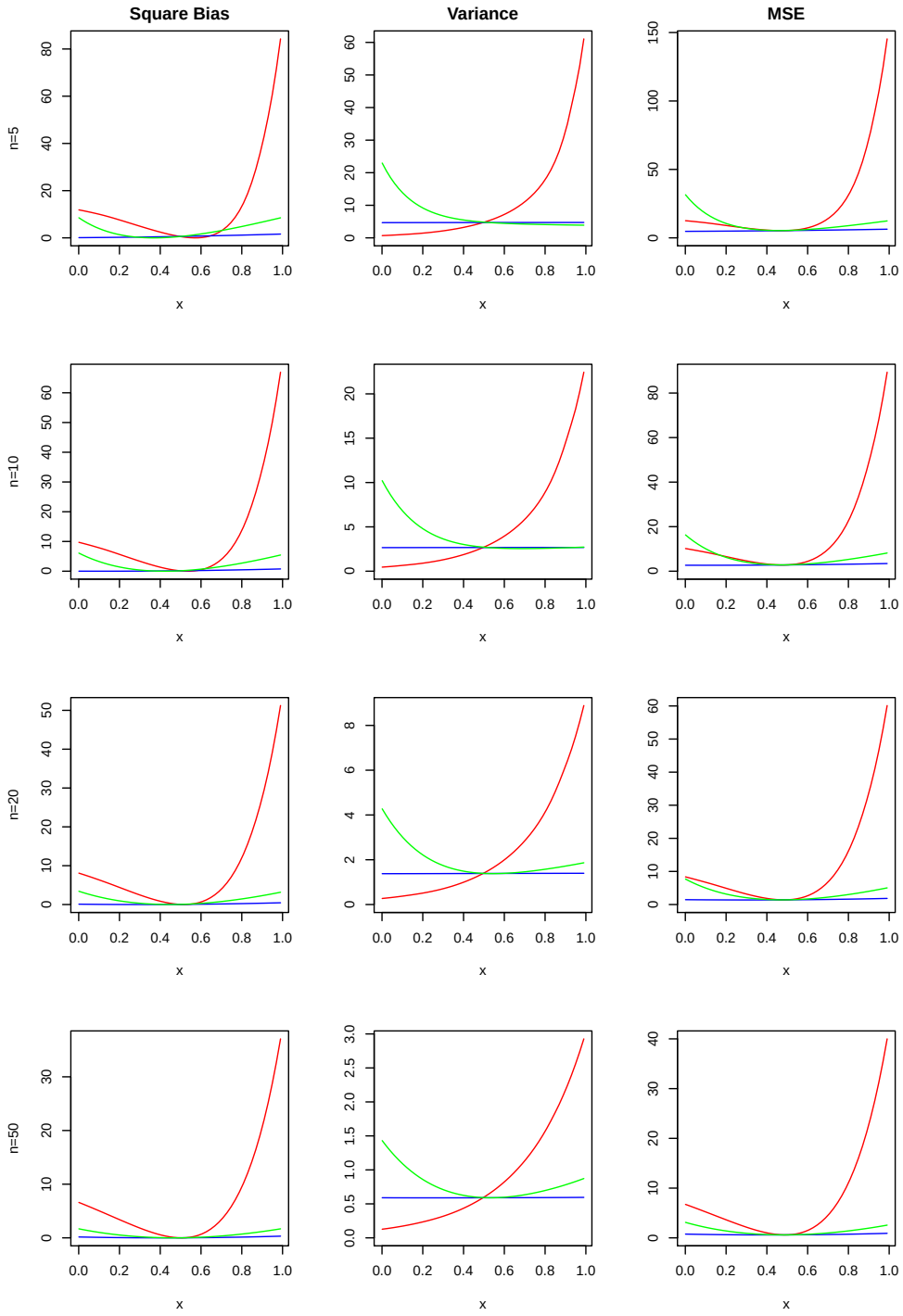

Figure 5.19: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_1(x)$, with exponential responses, with $r = 5$ and h_r set as $h_{T,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

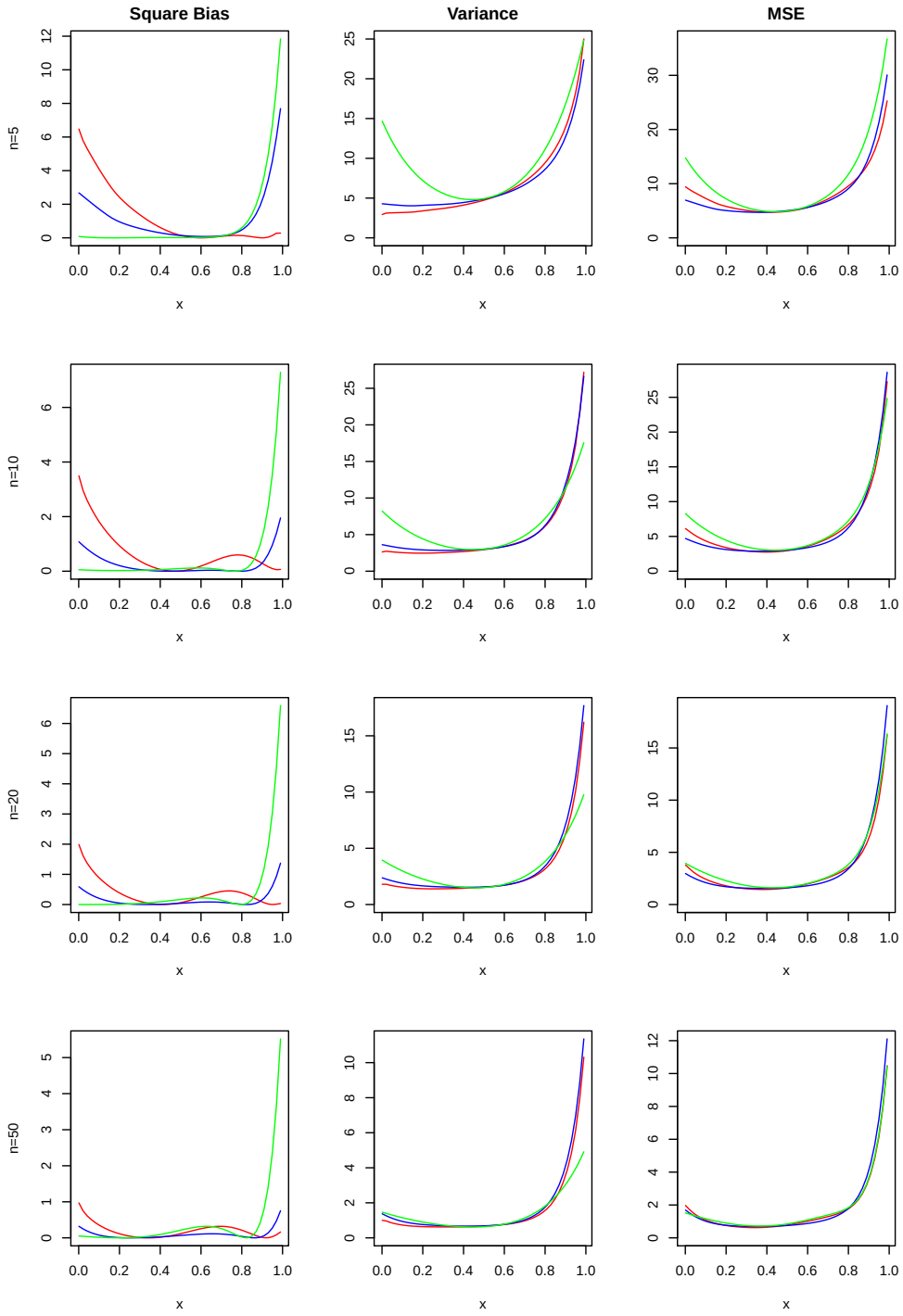

Figure 5.20: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_2(x)$, with exponential responses, with $r = 5$ and h_r set as $h_{T,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

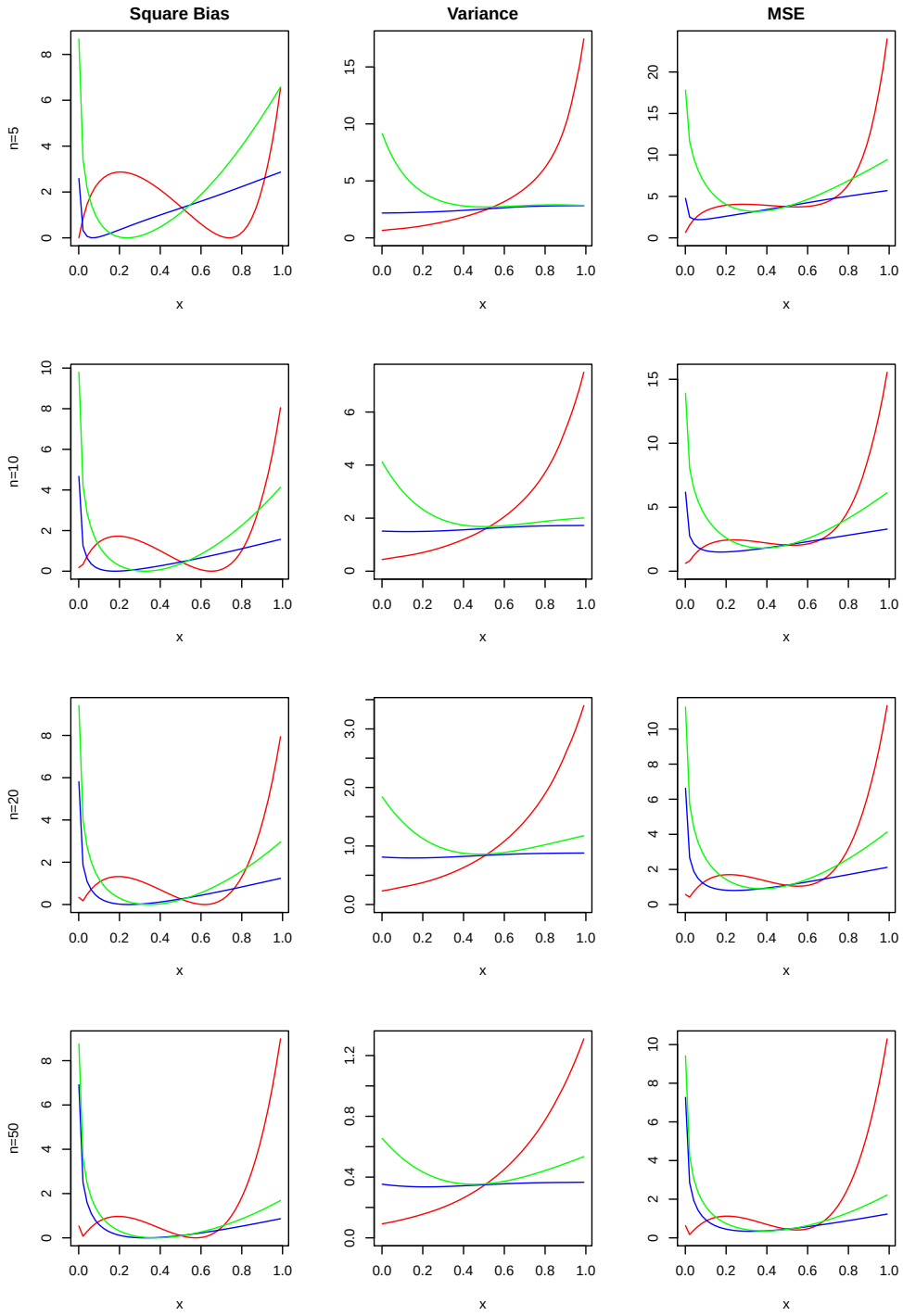

Figure 5.21: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_3(x)$, with exponential responses, with $r = 5$ and h_r set as $h_{T,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

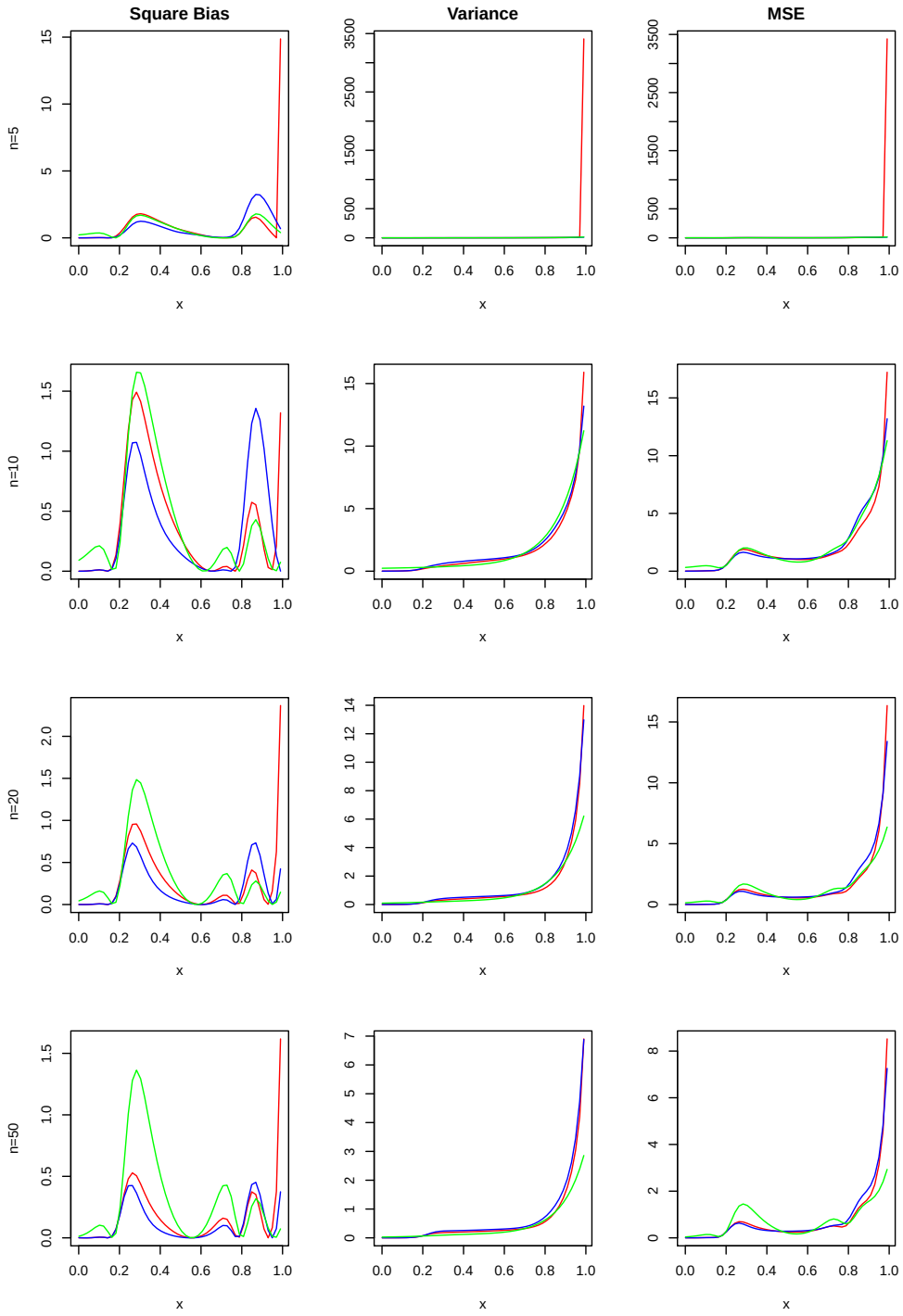

Figure 5.22: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_4(x)$, with exponential responses, with $r = 5$ and h_r set as $h_{T,MISE}$ and $n = 5, 10, 20, 50$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

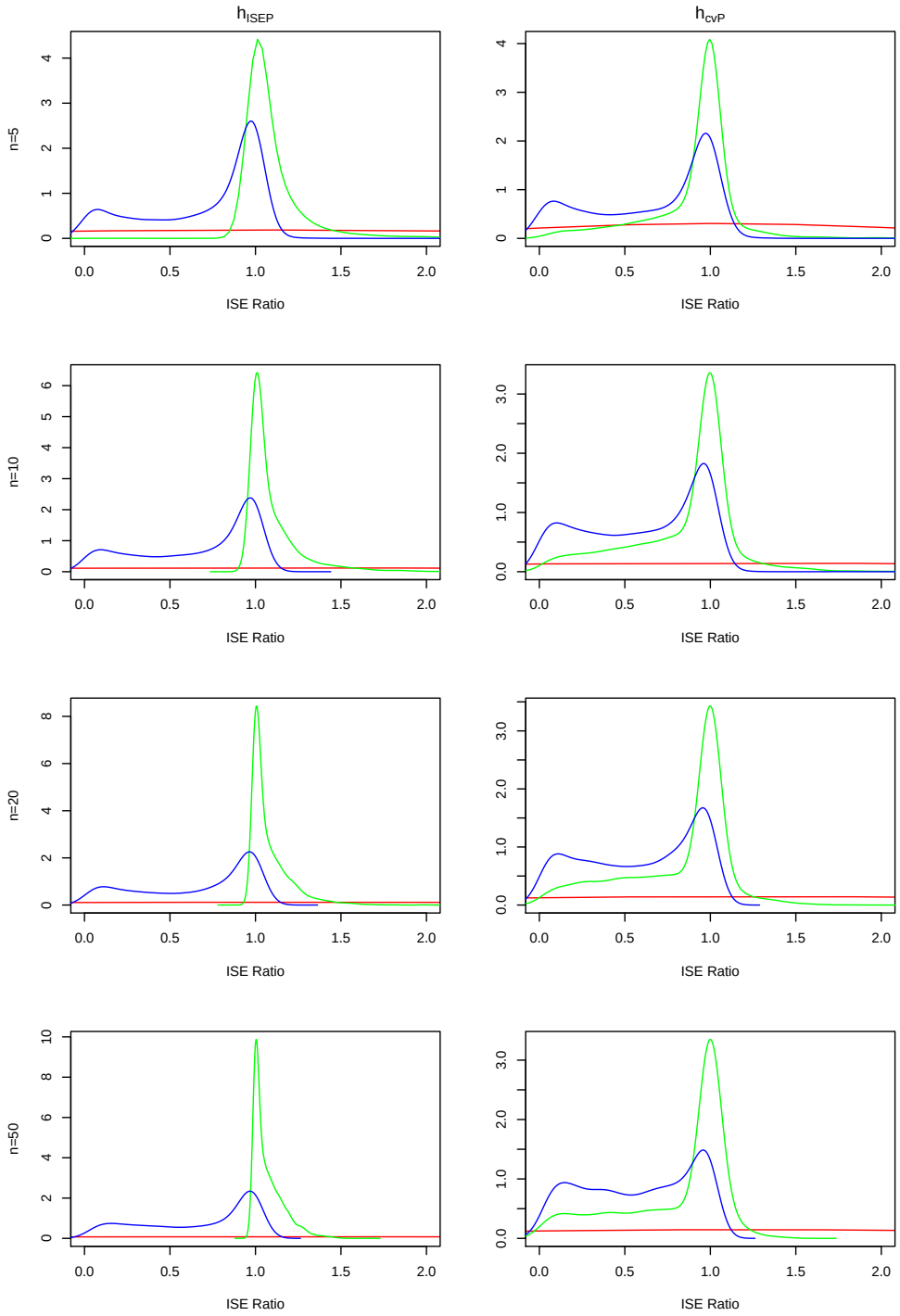

Figure 5.23: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_1(x)$, with exponential responses, with $r = 5$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

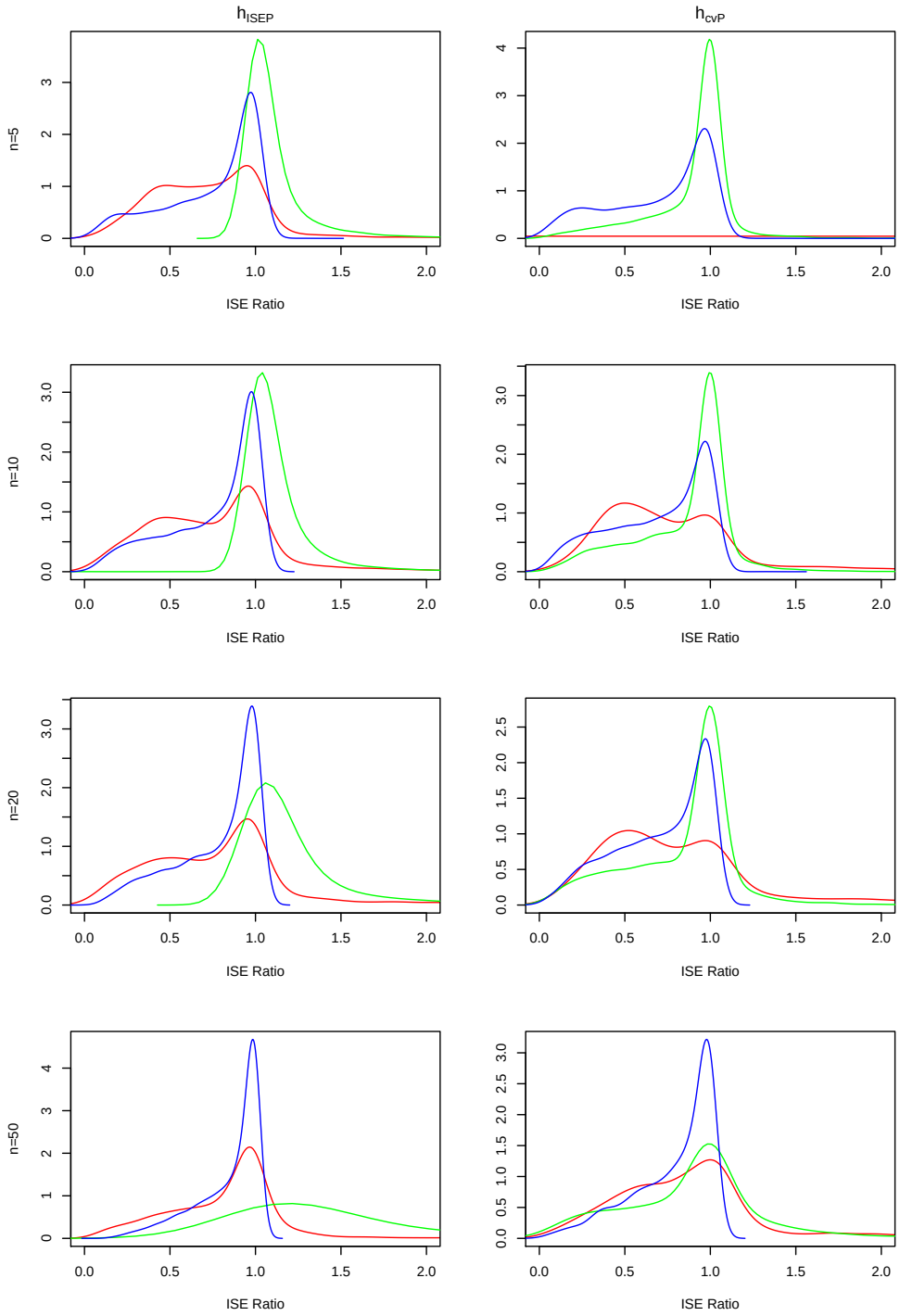

Figure 5.24: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_2(x)$, with exponential responses, with $r = 5$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

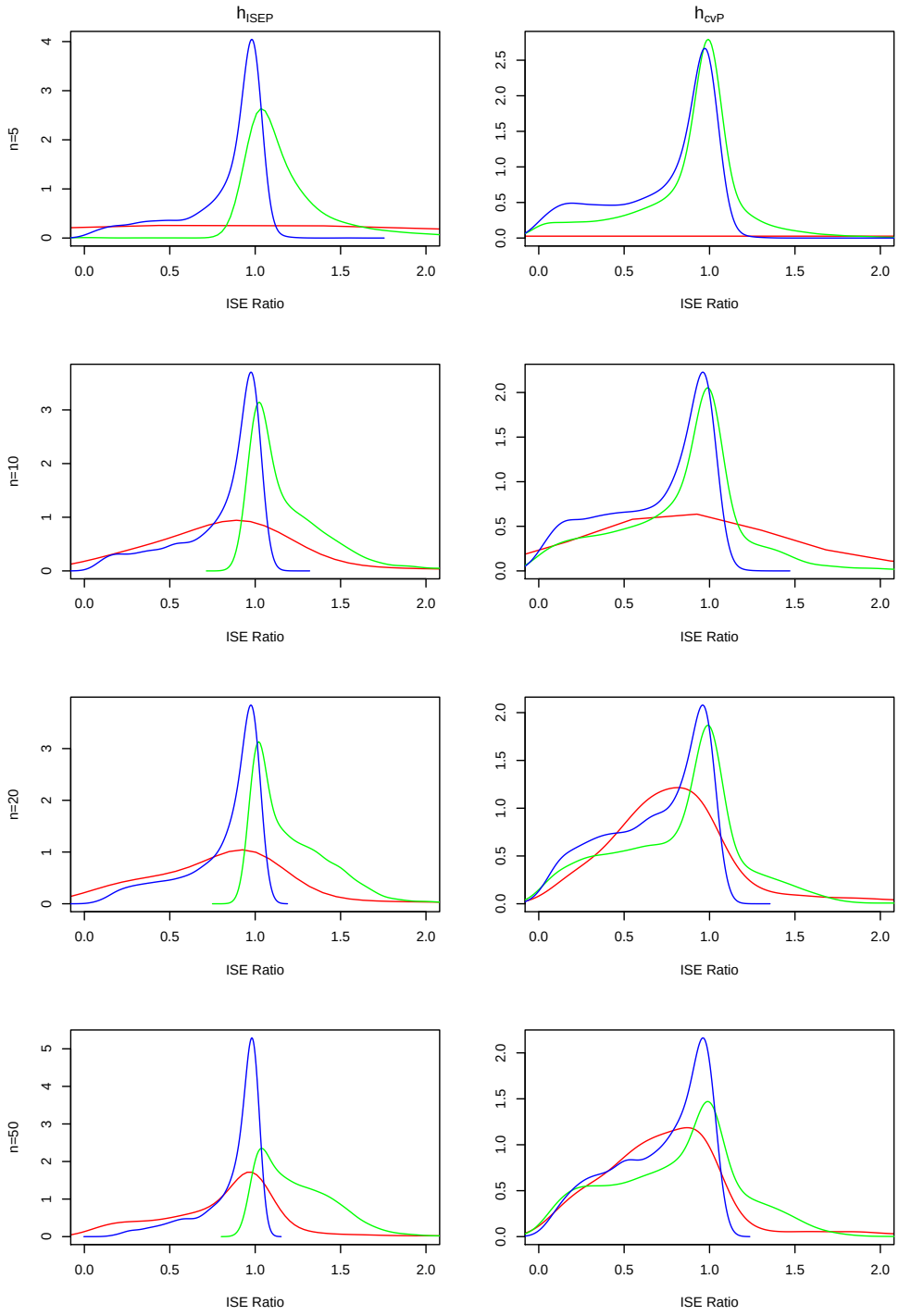

Figure 5.25: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_3(x)$, with exponential responses, with $r = 5$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

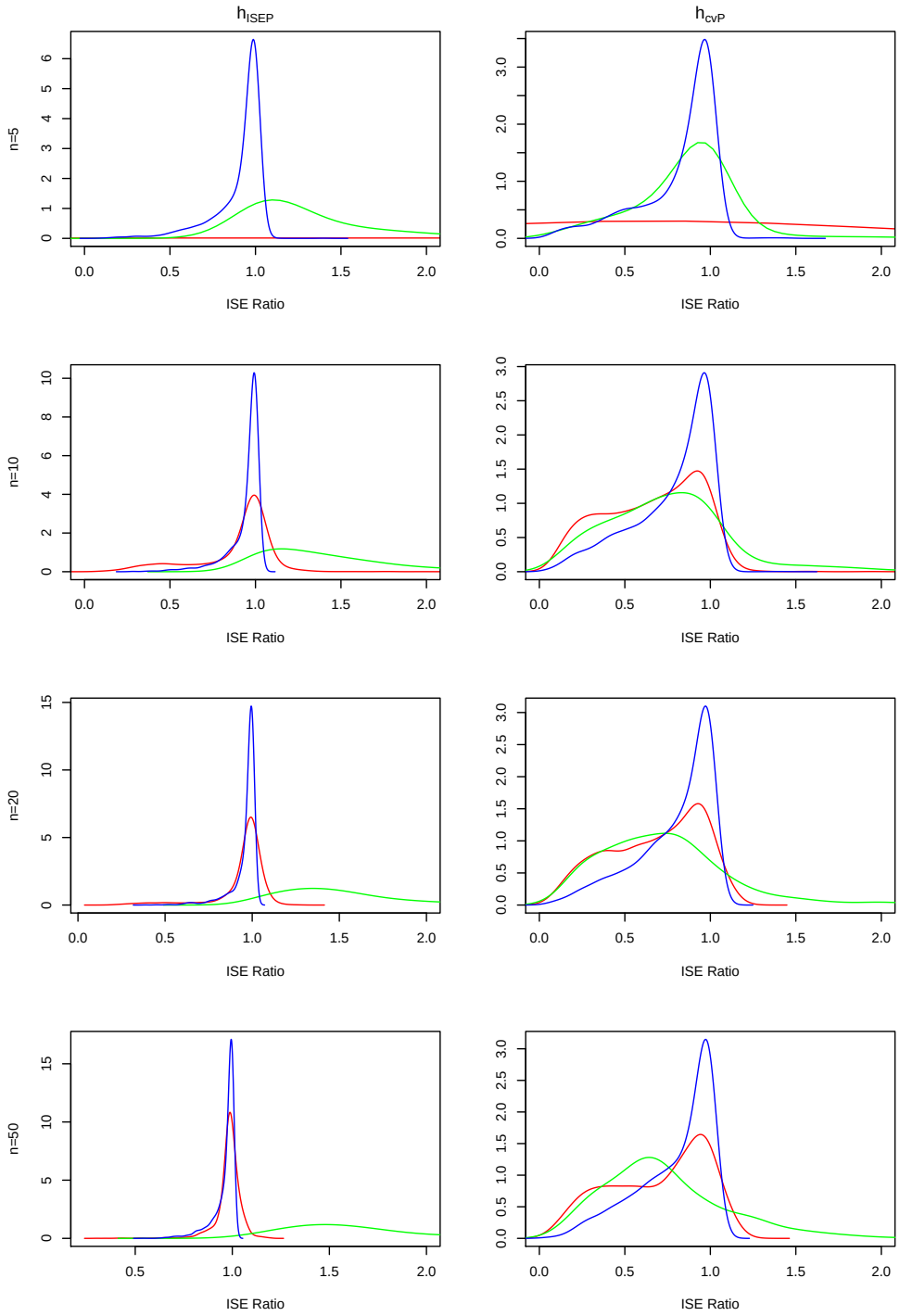

Figure 5.26: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_4(x)$, with exponential responses, with $r = 5$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $n = 5, 10, 20, 50$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

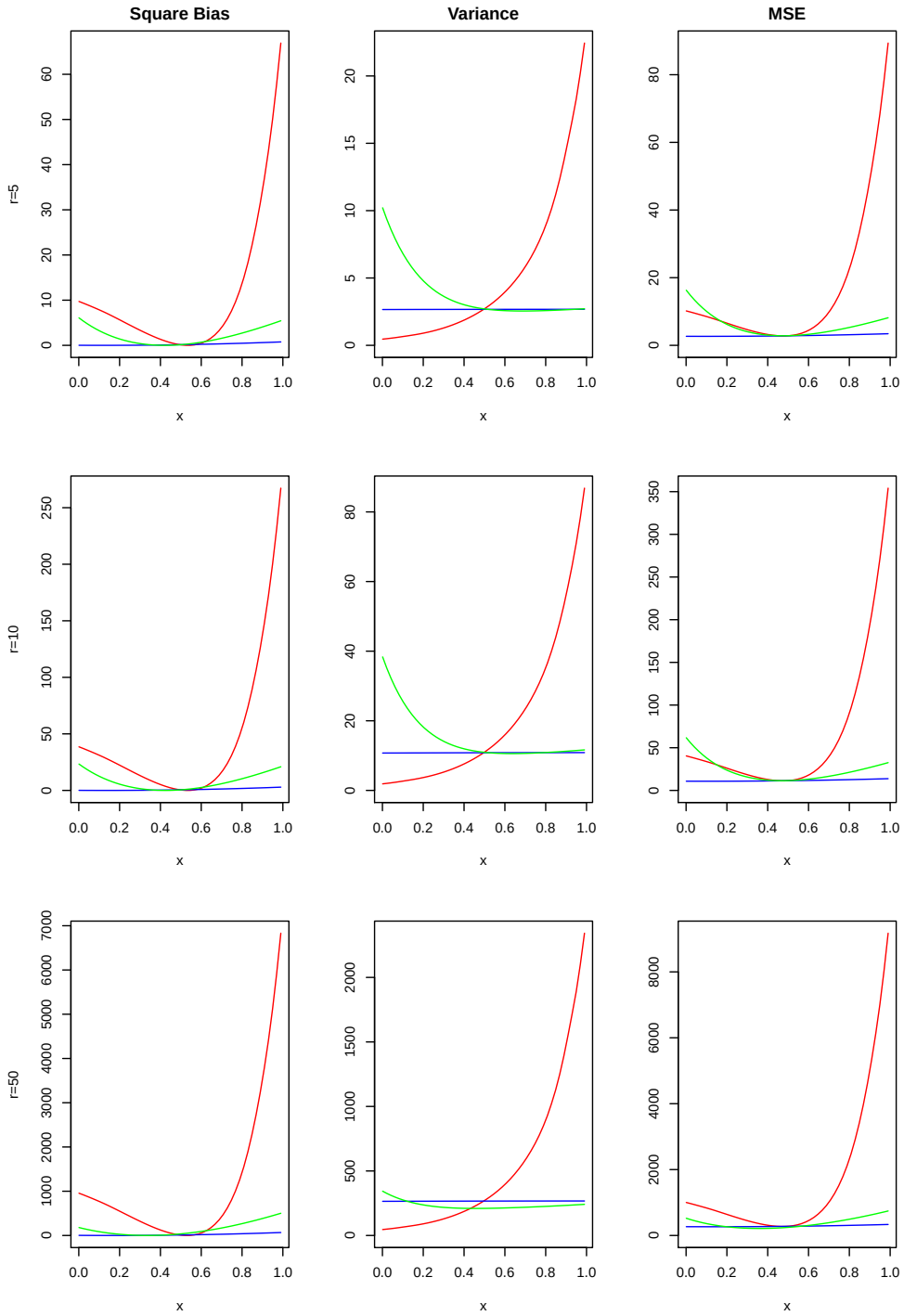

Figure 5.27: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_1(x)$, with exponential responses, with $n = 10$ and h_r set as $h_{T,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

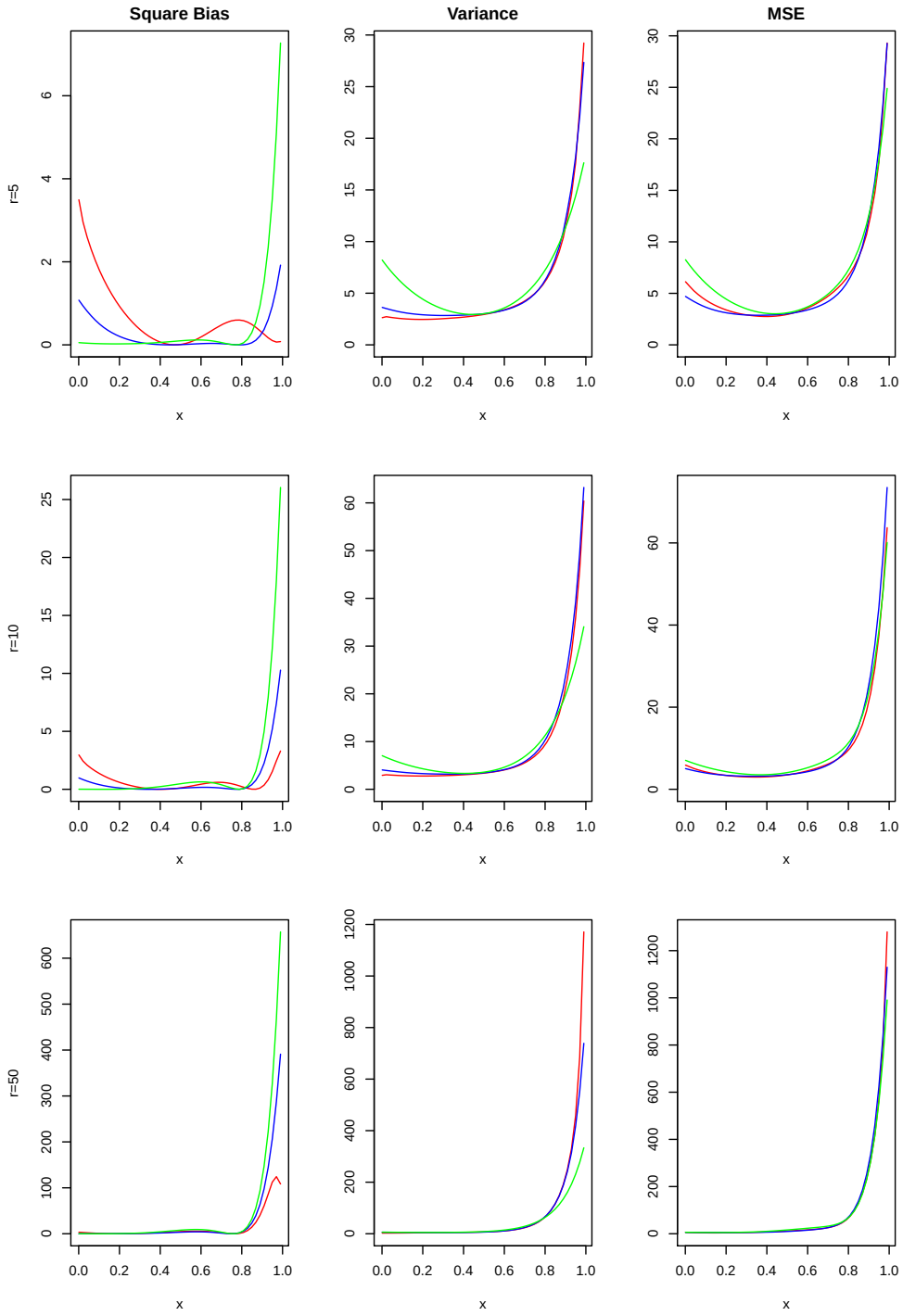

Figure 5.28: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_2(x)$, with exponential responses, with $n = 10$ and h_r set as $h_{T,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

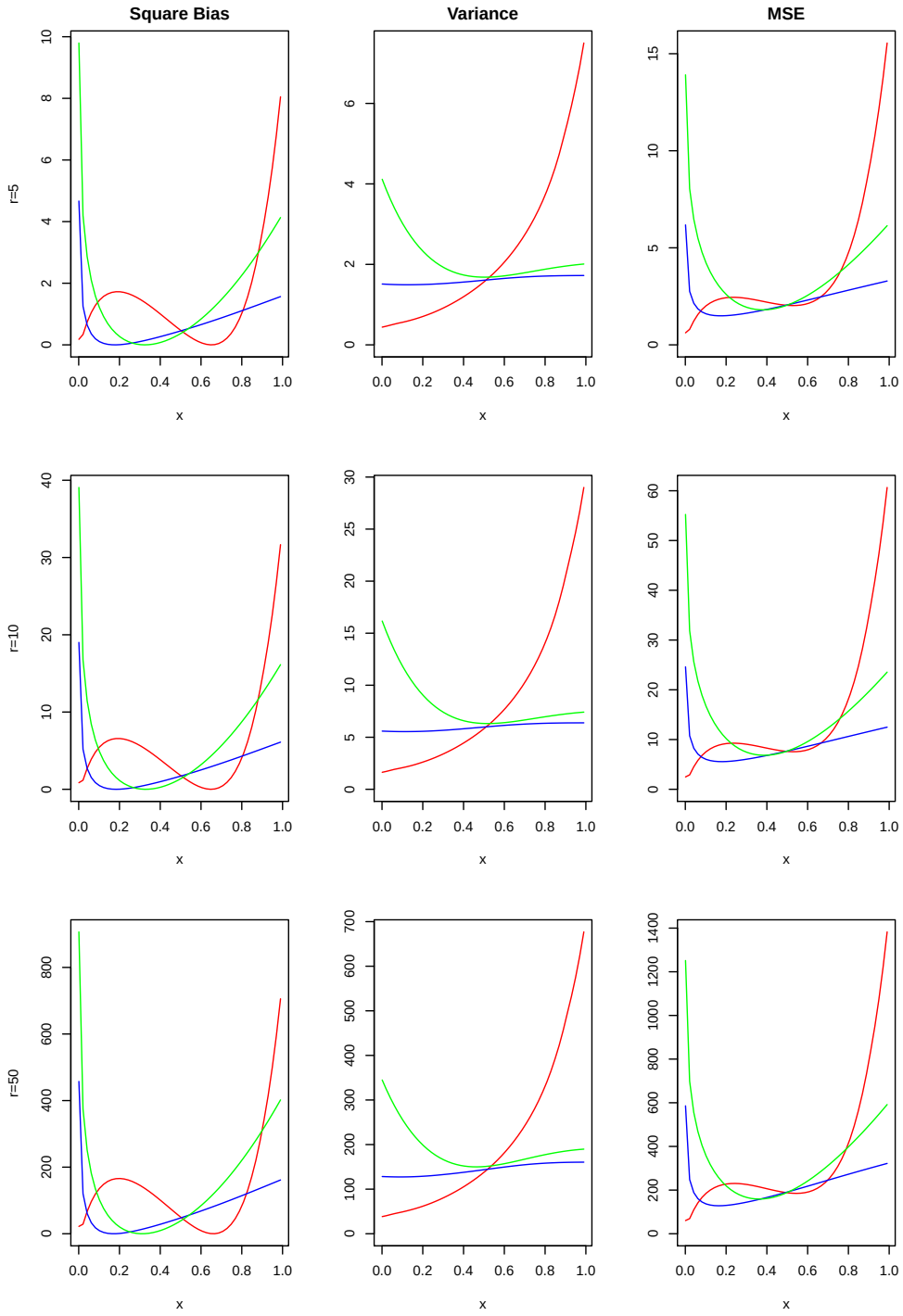

Figure 5.29: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_3(x)$, with exponential responses, with $n = 10$ and h_r set as $h_{T,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

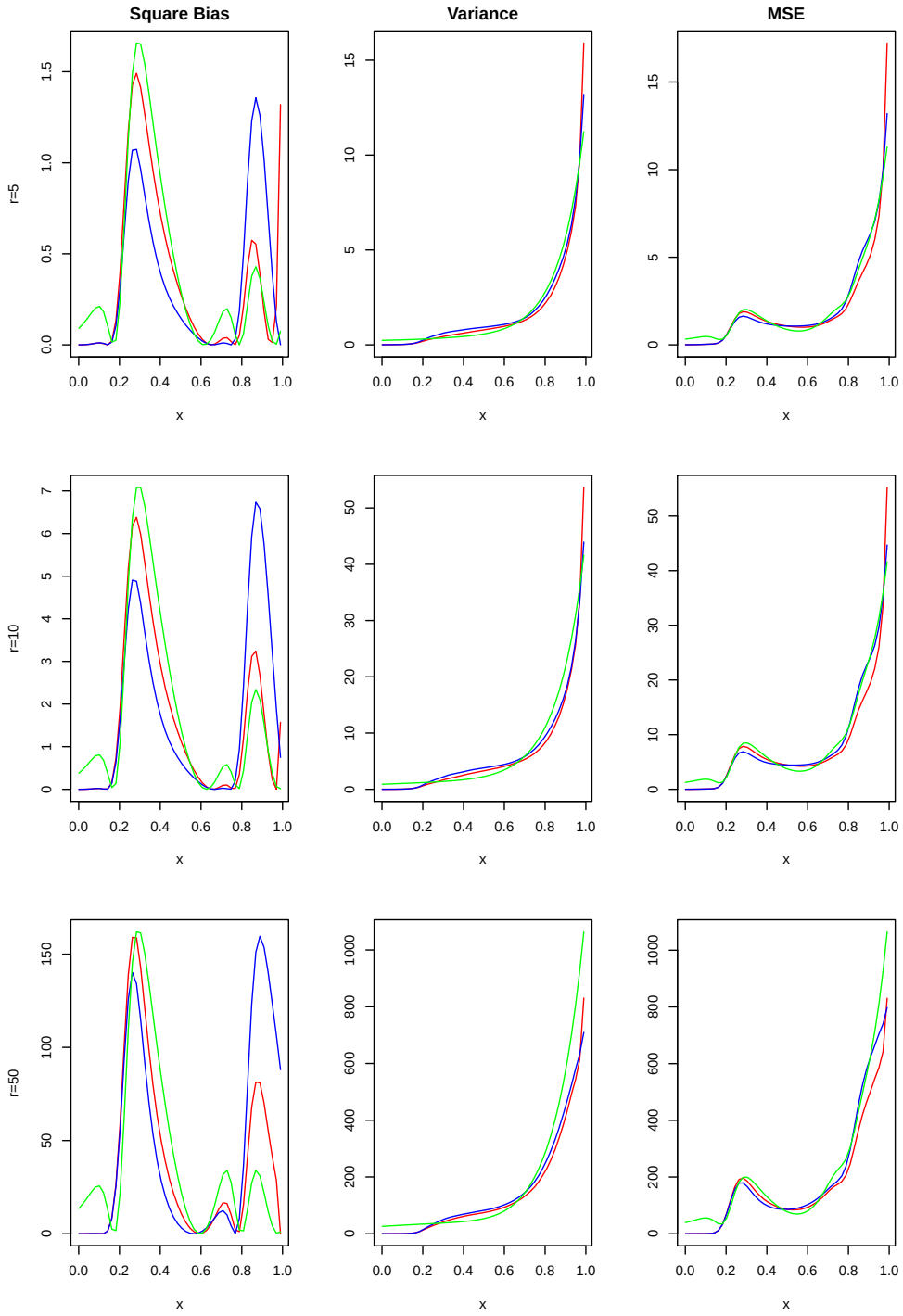

Figure 5.30: A comparison of the square bias, variance and MSE of the three monotone estimates of $T_4(x)$, with exponential responses, with $n = 10$ and h_r set as $h_{T,MISE}$ and $r = 1, 5, 10, 20$ respectively for each row of the plot. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

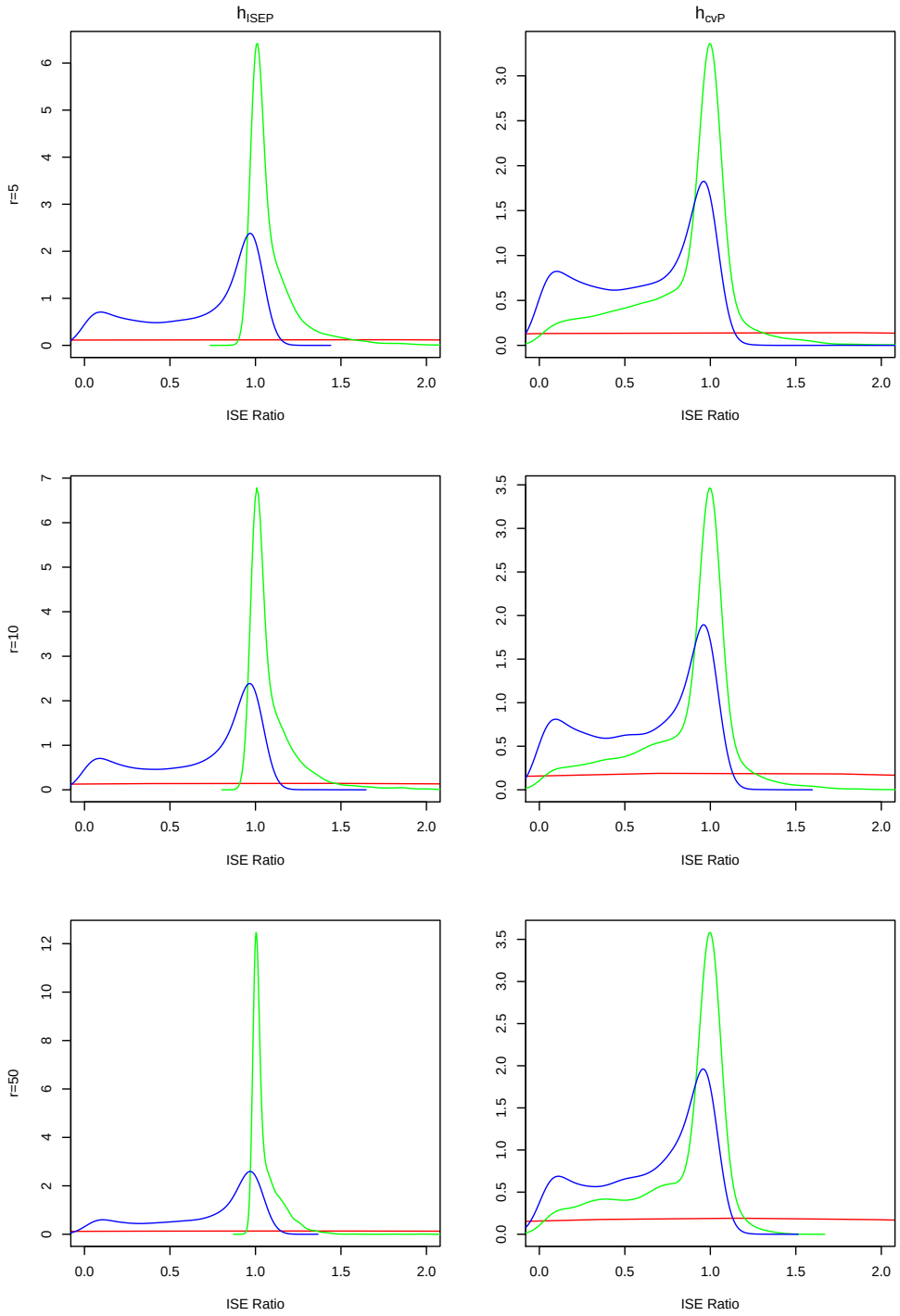

Figure 5.31: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_1(x)$, with exponential responses, with $n = 10$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

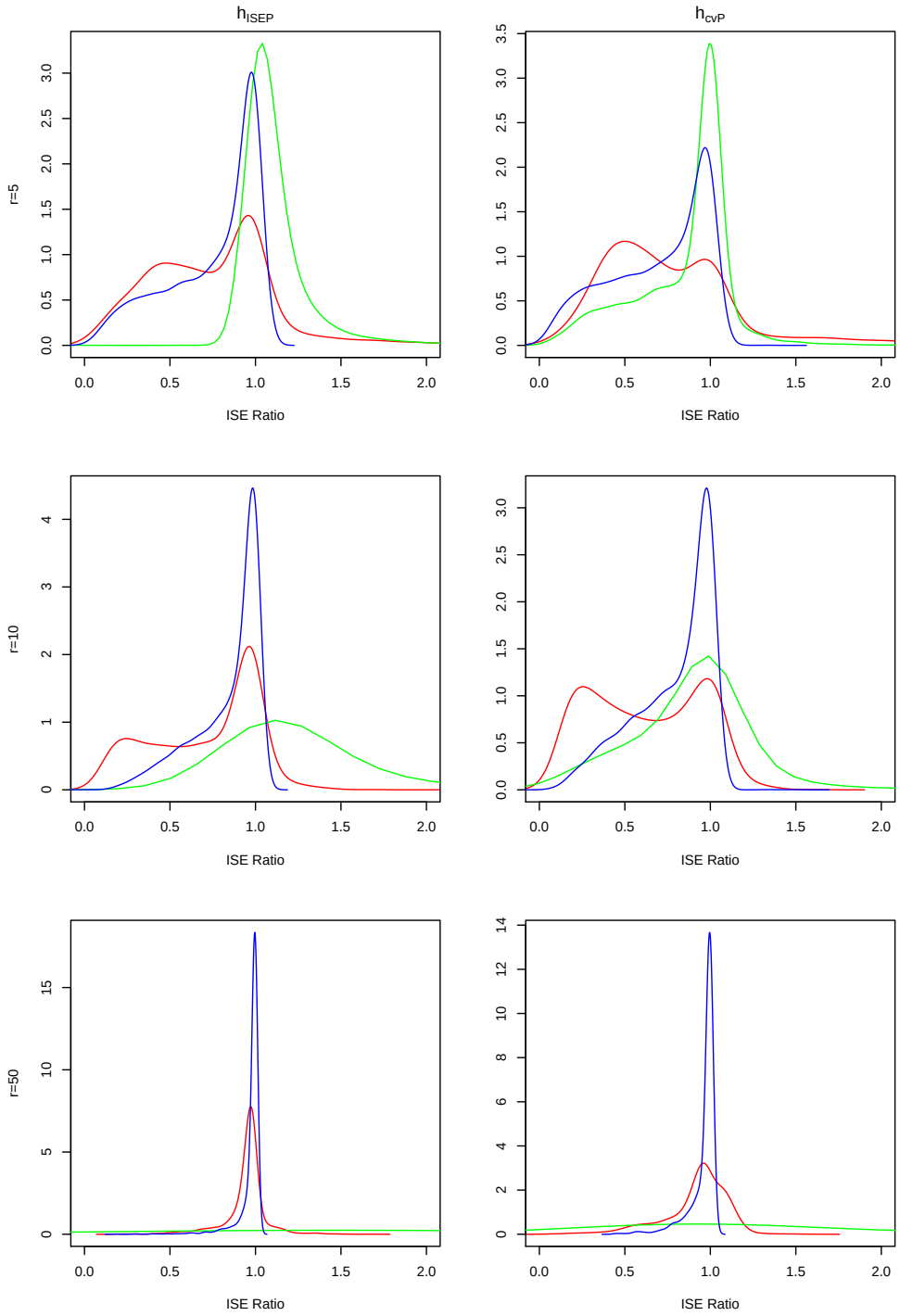

Figure 5.32: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_2(x)$, with exponential responses, with $n = 10$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

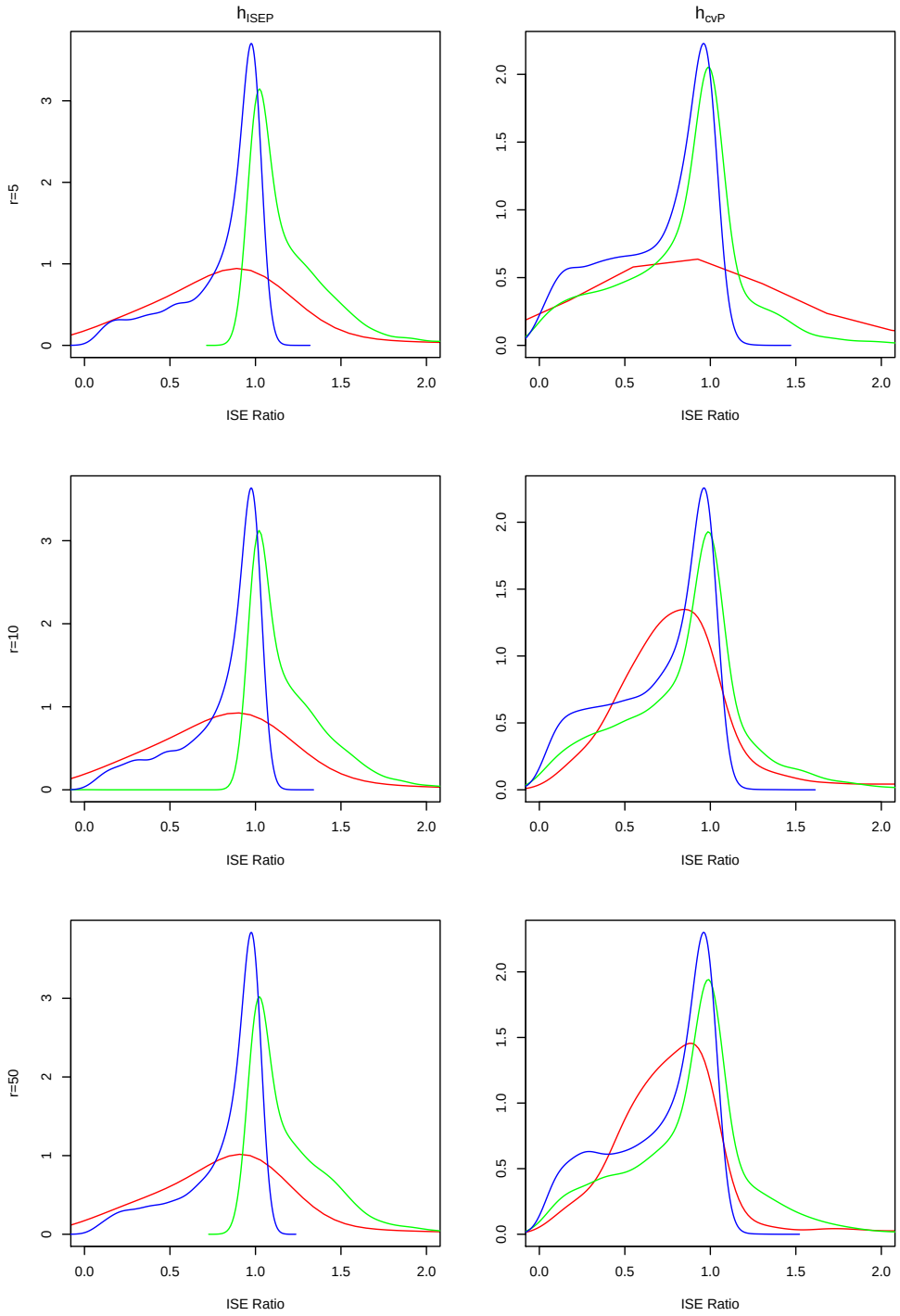

Figure 5.33: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_3(x)$, with exponential responses, with $n = 10$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

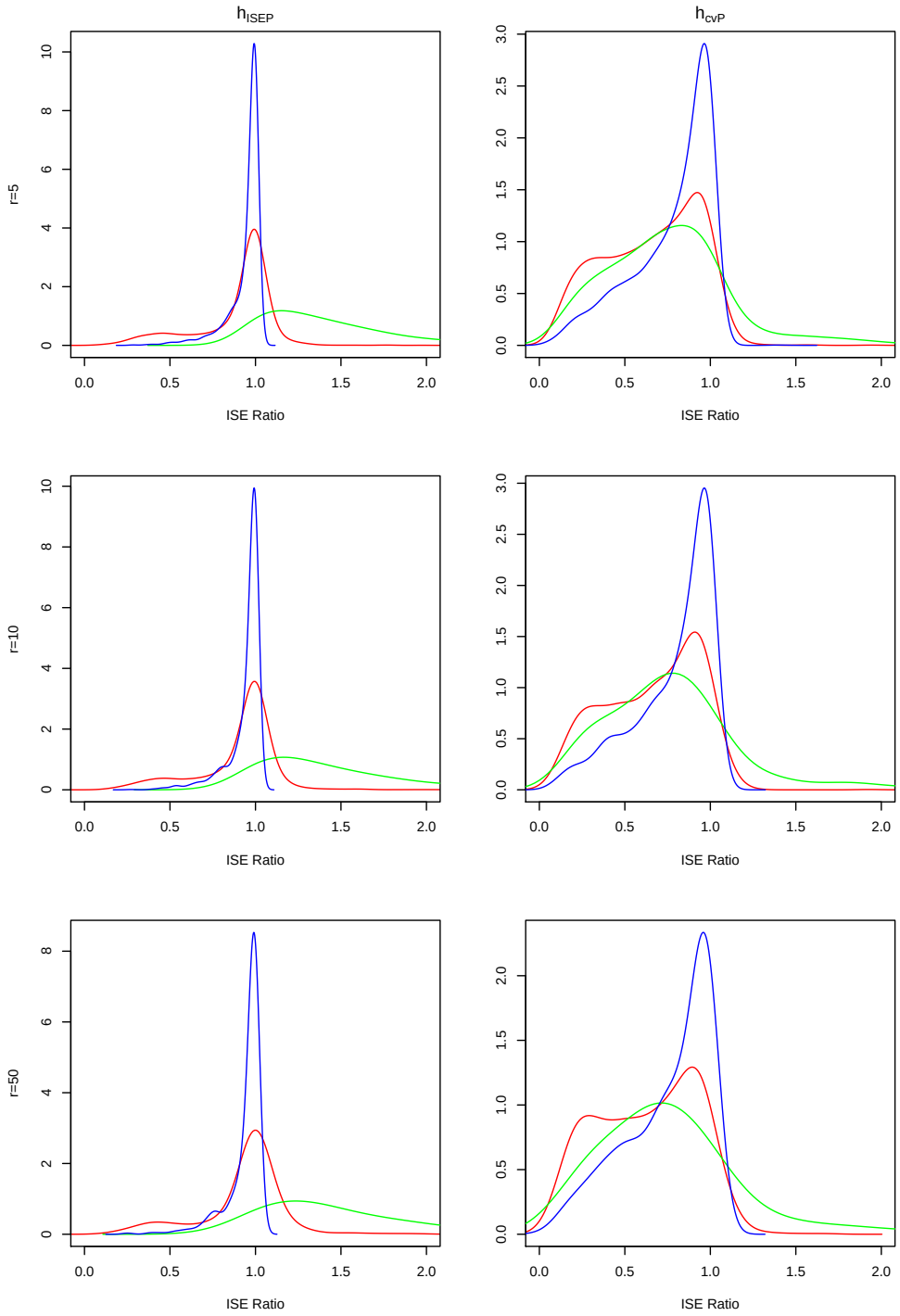

Figure 5.34: A comparison of the ISE ratio statistics for each of the three monotone estimates of $T_4(x)$, with exponential responses, with $n = 10$ and h_r set as $h_{T,ISE}, h_{T,cv}$ for each column respectively and $r = 1, 5, 10, 20$ for each row respectively. The PAVA estimate is shown in blue, the bandwidth estimate in green and the LDNP estimate in red.

Chapter 6

Conclusions and Further Work

In this thesis I have proposed a monotonicity constraint for use in nonparametric regression when used to estimate monotone transducer functions where the response is a member of the exponential family. This method, called here the LDNP method, is an adaptation of the DNP method of Dette Neumeyer and Pilz [9].

One of the main advantages of the LDNP method is that it constrains the estimate to be within the correct limits. It does so by incorporating the inverse link function into the estimation procedure. So long as an appropriate link function for the response being considered is used then the resulting estimate will lie within the correct limits. This is an immediate advantage over the DNP method.

Another advantage of the LDNP method is that the resulting monotone estimate is continuous. The PAVA method, one of the main competitors, often has 'flat points' that are a poor reflection of reality. The LDNP method does not have these flat points and thus provides a more appealing final estimate.

The LDNP method also incorporates the dependence of the variance of the response upon its mean into the estimation procedure. The DNP method, used with local polynomial regression does not do this, and so important information is not used.

I have shown that the LDNP method is first order asymptotically equivalent to the competing methods discussed in this thesis, namely the bandwidth adjustment method and the PAVA

method. I have calculated its bias and shown that in the case where $\lim_{n\to\infty} h_r/h_d = \infty$, then the bias is of order $O(h_r^2)$. The variance of the LDNP estimate is $O((nh_r)^{-1})$ which is again comparable with its competitors. In this case where $\lim_{n\to\infty} h_r/h_d = c \in [0, \infty)$, then the bias is of order $O(h_r^2 + h_d^2))$ and so so is still comparable to other methods as long as $h_d = o(h_r)$. I also show that the LDNP estimate is asymptotically normally distributed, with an MSE that is comparable to other methods.

I have investigated the choice of bandwidth, h_d, for the density estimation step. I have shown that asymptotically it is preferable to choose a bandwidth for which $h_d = o(h_r)$. A small simulation study has shown that this is not necessarily true in small samples. In this case the choice of h_d seems to be less significant in terms of the effect it has on the final estimate. However, the asymptotically optimal choice of h_d did not perform worse than other options in small sample simulations and so I suggest this is a good type of bandwidth to choose for use in the LDNP method. A more detailed study of the choice of h_d would be an area for further work. It would be interesting to know if bandwidths for which $h_d = o(h_r)$ are always the optimal choice or whether some functions would produce different conclusions.

I have tested the performance of the LDNP method against other monotonicity constraints. Firstly I have investigated its use in estimating monotone psychometric functions, since this was the setting that motivated the LDNP method. I have found that the LDNP method performs as well as the competing methods. It often improves with an increase in the number of stimulus levels or the number of repeats at each stimulus level, in contrast to the bandwidth adjustment method which got worse. I found that when ISE optimal bandwidths were used for the unconstrained regression then the LDNP method was clearly the best performing method. In other cases the LDNP method still performed roughly as well as its competitors.

I then tested the performance of the LDNP method when estimating transducer functions with Poisson responses. Here I have found that for most functions the LDNP method performed as well as its competitors. However, for functions that are almost flat, the LDNP method struggled to give good estimates of the true function. In this case the PAVA method worked much better. Further investigations into how the performance of LDNP method could be

improved for such functions are a topic for further research.

I have also investigated the performance of the LDNP method when estimating transducer functions with exponential respsonses. This allowed me to investigate how well the LDNP method could estimate continuous responses. The LDNP method again performed as well as its competitors in most cases.

In general, for any distribution, if non-monotonicity was caused by too small a bandwidth chosen for the uncononstrained regression, then the best method is to increase the bandwidth. However, if the lack of monotonicity was caused by something other than too small a bandwidth then the LDNP method performs as well as, and often better than other competing methods. An investigation of the LDNP method on other members of the exponential family would be an interesting area for further work.

A useful byproduct of this research was R-code that calculates monotone nonparamteric regression estimates of transducer functions for each of the methods discussed in this thesis. This code will be made available for wider use by psychometricians etc., and will allow them easily to calculate monotone psychometric functions using nonparametric regression.

In the introduction to this thesis I discuss a further monotonicity constraint method, namely the tilting method. I have tried to adapt this method to the setting where responses come from the exponential family but have had computational difficulties. In theory, this method is very appealing since it relies only on changing the weights used in order to get a monotone regression estimate. It is intuitively simple. However, the non linear constrained optimisation techniques it requires caused some problems. An interesting area of future work would be to develop this method as a tool for estimating monotone psychometric functions. I believe this should be possible.

Bibliography

[1] R. J. Baker, D. Jayewardene, C. Sayle, and S. Saeed. Failure to find asymmetry in auditory gap detection. *Laterality: Asymmetries of Body, Brain and Cognition*, 13(1):1–21, 2008.

[2] A. W. Bowman, M. C. Jones, and I. Gijbels. Testing monotonicity of regression. *Journal of Computational and Graphical Statistics*, 7(4):489–500, 1998.

[3] H. D. Brunk. Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics*, 26(4):607–616, 1955.

[4] H. D. Brunk. On the estimation of parameters restricted by inequalities. *The Annals of Mathematical Statistics*, 29:437–454, 1958.

[5] C. K. Chu and K. F. Cheng. Nonparametric regression estimates using misclassified binary responses. *Biometrika*, 82(2):315–325, 1995.

[6] G. Claeskens and I. Van Keilegom. Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.*, 31(6):1852–1884, 2003.

[7] N. Cressie and T. R. C. Read. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3):440–464, 1984.

[8] H. Dette, N. Neumeyer, and K. F. Pilz. A note on nonparametric estimation of the effective dose in quantal bioassay. *Journal of the American Statistical Association*, 100(470):503–510, 2005.

[9] H. Dette, N. Neumeyer, and K. F. Pilz. A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3):469–490, 2006.

[10] H. Dette and K. F. Pilz. A comparative study of monotone nonparametric kernel estimates. *Journal of Statistical Computation and Simulation*, 76(1):41–56, 2006.

[11] H. Dette and R. Scheder. A finite sample comparison of nonparametric estimates of the effective dose in quantal bioassay. *Journal of Statistical Computation and Simulation*, 80(5):527–544, 2010.

[12] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 301–369, 1995.

[13] J. Fan. Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87(420):998–1004, 1992.

[14] J. Fan. Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 21(1):196–216, 1993.

[15] J. Fan and I. Gijbels. Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(2):pp. 371–394, 1995.

[16] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London, 1996.

[17] J. Fan, N. E. Heckman, and M. P. Wand. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, 90(429):141–150, 1995.

[18] D. H. Foster and K. Żychaluk. Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics*, 71(6):1414–1425, 2009.

[19] D.H. Foster and K. Żychaluk. Nonparametric estimates of biological transducer functions. *Signal Processing Magazine, IEEE*, 24(4):49 –58, 2007.

[20] J. Friedman and R. Tibshirani. The monotone smoothing of scatterplots. *Technometrics*, 26(3):243–250, 1984.

[21] T. Gasser and H. Müller. Kernel estimation of regression functions. In *Smoothing techniques for curve estimation (Proc. Workshop, Heidelberg, 1979)*, volume 757 of *Lecture Notes in Math.*, pages 23–68. Springer, Berlin, 1979.

[22] I. Gijbels, P. Hall, M. C. Jones, and I. Koch. Tests for monotonicity of a regression mean with guaranteed level. *Biometrika*, 87(3):663–673, 2000.

[23] Peter J Green and Bernard W Silverman. *Nonparametric regression and generalized linear models: a roughness penalty approach*. CRC Press, London, 1993.

[24] S. Guidi, O. Parlangeli, S. Bettella, S. Roncato, et al. Features of the selectivity for contrast polarity in contour integration revealed by a novel tilt illusion. *Perception-London*, 40(11):1357, 2011.

[25] P. Hall and N. E. Heckman. Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.*, 28(1):20–39, 2000.

[26] P. Hall and L.-S. Huang. Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.*, 29(3):624–647, 2001.

[27] W. Härdle. *Applied nonparametric regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge, 1990.

[28] T. J. Hastie and R. J. Tibshirani. *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London, 1990.

[29] R. F. Kappenman. Nonparametric estimation of dose-response curves with application to ed50 estimation. *Journal of Statistical Computation and Simulation*, 28(1):13–28, 1987.

[30] H.L. Lee and U. Noppeney. Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences*, 108(51):E1441–E1450, 2011.

[31] D. M. Levi and S. P. Tripathy. Is the ability to identify deviations in multiple trajectories compromised by amblyopia? *Journal of Vision*, 6(12), 2006.

[32] E. Mammen. Estimating a smooth monotone regression function. *The Annals of Statistics*, 19(2):724–740, 1991.

[33] P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman & Hall, London, 1990.

[34] H. Mukerjee. Monotone nonparameteric regression. *Ann. Statist.*, 16(2):741–750, 1988.

[35] E. A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9:141–142, 1964.

[36] S. M. C. Nascimento, D. H. Foster, and K. Amano. "Psychophysical estimates of the number of spectral-reflectance basis functions needed to reproduce natural scenes". *Journal of the ical Society of America A-ics Image Science and Vision*, 22(6):1017–1022, 2005.

[37] D. Park and S. Park. Parametric and nonparametric estimators of ed100a. *Journal of Statistical Computation and Simulation*, 76(8):661–672, 2006.

[38] M. B. Priestley and M. T. Chao. Non-parametric function fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(3):pp. 385–392, 1972.

[39] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.

[40] D. Ruppert, S. J. Sheather, and M. P. Wand. An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, 90(432):1257–1270, 1995.

[41] A. J. Schofield, T. Ledgeway, and C. V. Hutchinson. Asymmetric transfer of the dynamic motion aftereffect between first- and second-order cues and among different second-order cues. *Journal of Vision*, 7(8), 2007.

[42] B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

[43] R. Tibshirani and T. Hastie. Local likelihood estimation. *Journal of the American Statistical Association*, 82(398):559–567, 1987.

[44] M. P. Wand and M. C. Jones. *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London, 1995.

[45] G. S. Watson. Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.

[46] N.A. Whitmal and K. DeRoy. Adaptive bandwidth measurements of importance functions for speech intelligibility prediction. *The Journal of the Acoustical Society of America*, 130(6):4032–4043, 2011.

[47] N.A. Whitmal and K. DeRoy. Use of an adaptive-bandwidth protocol to measure importance functions for simulated cochlear implant frequency channels. *The Journal of the Acoustical Society of America*, 131:1359, 2012.

[48] F. A. Wichmann and N. Hill. The psychometric function: I. fitting, sampling, and goodness of fit. *Attention, Perception, & Psychophysics*, 63:1293–1313, 2001.

[49] F. A. Wichmann and N. Hill. The psychometric function: Ii bootstrap-based confidence intervals and sampling. *Attention, Perception, & Psychophysics*, 63:1314–1329, 2001.

[50] Y. Xie and L. D. Griffin. A 'portholes' experiment for probing perception of small patches of natural images. *Perception*, 36:315, 2007.

[51] R. Yssaad-Fesselier and K. Knoblauch. Modeling psychometric functions in r. *Behavior Research Methods*, 38:28–41, 2006. 10.3758/BF03192747.

[52] L. Zhang, C. Cavaro-Ménard, P. Le Callet, LHK Cooper, G. Hunault, and J.Y. Tanguy. The effects of anatomical information and observer expertise on abnormality detection task. In *SPIE Medical Imaging*, pages 79661G–79661G. International Society for Optics and Photonics, 2011.