Information Theoretic Estimation of Econometric Functions

Millie Yi Mao*and Aman Ullah†

November, 2019

Abstract

This chapter introduces an information theoretic approach to specify econometric functions as an alternative to avoid parametric assumptions. We investigate the performances of the information theoretic method in estimating the regression (conditional mean) and response (derivative) functions. We have demonstrated that they are easy to implement, and are advantageous over parametric models and nonparametric kernel techniques.

Key Words: Information theory, Maximum entropy distributions, Econometric functions, Conditional mean.

*Department of Mathematics, Physics and Statistics, Azusa Pacific University, Azusa, CA 91702. E-mail: ymao@apu.edu

†Department of Economics, University of California, Riverside, CA 92521. E-mail: aman.ullah@ucr.edu

‡An earlier version of this work was first presented at a conference organized by Info-Metrics Institute, American University in November 2016, and then in its economics department in September 2017. The authors are thankful to Duncan Foley and other participants for their valuable comments. We are grateful to Amos Golan for many constructive and helpful suggestions. The comments from co-editors and a referee were also helpful.

1 Introduction

In the literature of estimation, specification, and testing of econometric models, many parametric assumptions have been made. Firstly, parametric functional forms of the relationship between independent and dependent variables are usually assumed to be known. For example, a regression function is often considered to be linear. Secondly, the variance of the error terms conditional on the independent variables is specified to have a parametric form. Thirdly, the joint distribution of the independent and dependent variables are conventionally assumed to be normal. Last but not least, in many econometric studies, the independent variables are considered to be non-stochastic. However, parametric econometrics has drawbacks since particular specifications may not capture the true data generating process. As a matter of fact, the true functional forms of econometric models are hardly known. Misspecification of parametric econometric models may therefore result in invalid conclusions and implications. Alternatively, data-based econometric methods can be adopted to avoid the disadvantages of parametric econometrics and implemented into practice. One widely-used approach is the nonparametric kernel technique, see Ullah (1988), Pagan and Ullah (1999), Li and Racine (2007) and Henderson and Parmeter (2015). However, nonparametric kernel procedures have some deficiencies, such as the "curse of dimensionality" and a lack of efficiency due to a slower rate of convergence of the variance to zero. In view of this, we propose a new information theoretic (IT) procedure for econometric model specification by using classical maximum entropy formulation. This is consistent, efficient, and based on minimal distributional assumptions.

Shannon (1948) derived the entropy (information) measure which is similar to that of Boltzmann (1872) and Gibbs (1902) . Using Shannon's entropy measure Jaynes (1957a, 1957b) developed the maximum entropy principle to infer probability distribution. Entropy is a measure of a variable's average information content, and its maximization subject to some moments and normalization provides a probability distribution of the variable. The resulting distribution is known as the maximum entropy distribution; see more on this in Zellner and Highfield (1988), Ryu (1993), Golan et al. (1996), Harte et al. (2008), Judge and Mittelhammer (2011) and Golan (2018). We note that the joint probability distribution based on the maximum entropy approach is a purely data-driven distribution where parametric assumptions are avoided, and this distribution can be used to determine the regression function (conditional mean) and its response function (derivative

function) which are of interest to empirical researchers. This is the main goal of this chapter.

We organize this chapter in the following order. In Section 2, we present the IT based regression and response functions using a bivariate maximum entropy distribution. A recursive integration process is developed for their implementations. In Section 3 we carry out simulation examples to illustrate the small sample efficiency of our methods, and then present an empirical example of the Canadian high school graduate earnings. In Section 4, we present asymptotic theory on our IT based regression and response function estimators. In Section 5, we draw conclusions and provide potential future extensions. The mathematical details of the algorithm used in Section 2, and the proofs of asymptotic properties of the IT based estimators, are shown in the Appendix.

2 Estimation of Distribution, Regression, and Response Functions

We consider $\{y_i, x_i\}$, $i = 1, \ldots, n$ independent and identically distributed observations from an absolutely continuous bivariate distribution $f(y, x)$. Suppose the conditional mean of y given x exists and it provides a formulation for the regression model as

$$\begin{aligned} y &= E(y|x) + u \\ &= m(x) + u, \end{aligned} \tag{1}$$

where the error term u is such that $E(u|x) = 0$, and the regression function (conditional mean) is

$$E(y|x) = m(x) = \int_y y \frac{f(y, x)}{f(x)} dy. \tag{2}$$

When the joint distribution of y and x is not known, which is often the case, we propose the IT based maximum entropy method to estimate the densities of the random variables and introduce a recursive integration method to solve the conditional mean of y given x.

2.1 Maximum Entropy Distribution Estimation: Bivariate and Marginal

Suppose x is a scalar and the marginal density of it is unknown. Our objective is to approximate the marginal density $f(x)$ by maximizing the information measure (Shannon's entropy) subject to some constraints. That is

$$\underset{f}{\text{Max}}\ H(f) = -\int_x f(x) \log f(x) dx,$$

subject to

$$\int_x \phi_m(x) f(x) dx = \mu_m = E\phi_m(x),\ m = 0, 1, ..., M,$$

where $\phi_m(x)$ are known functions of x. $\phi_0(x) = \mu_0 = 1$. See, for example, Jaynes (1957a, 1957b) and Golan (2018). The total number of constraints is $M+1$. In particular, $\phi_m(x)$ can be moment functions of x. We construct the Lagrangian

$$\mathcal{L}(\lambda_0, \lambda_1, \ldots, \lambda_M) = -\int_x f(x) \log f(x) dx + \sum\nolimits_{m=0}^{M} \lambda_m \left(\mu_m - \int_x \phi_m(x) f(x) dx \right),$$

where $\lambda_0, \lambda_1, \ldots, \lambda_M$ represent Lagrange multipliers. The solution has the form

$$f(x) = \exp\left[-\sum\nolimits_{m=0}^{M} \lambda_m \phi_m(x) \right] = \frac{\exp\left[-\sum_{m=1}^{M} \lambda_m \phi_m(x) \right]}{\int_x \exp\left[-\sum_{m=1}^{M} \lambda_m \phi_m(x) \right] dx} \equiv \frac{\exp\left[-\sum_{m=1}^{M} \lambda_m \phi_m(x) \right]}{\Omega(\lambda_m)},$$

where λ_m is the Lagrange multiplier corresponding to constraint $\int_x \phi_m(x) f(x) dx = \mu_m$, and λ_0 (with $m = 0$) is the multiplier associated with the normalization constraint. With some simple algebra, it can be easily shown that $\lambda_0 = \log \Omega(\lambda_m)$ is a function of other multipliers. Replacing $f(x)$ and λ_0 into $\mathcal{L}(\lambda_0, \lambda_1, \ldots, \lambda_M) = \mathcal{L}(\boldsymbol{\lambda})$, we get

$$\mathcal{L}(\boldsymbol{\lambda}) = \sum\nolimits_{m=1}^{M} \lambda_m E\phi_m(x) + \lambda_0.$$

The Lagrange multipliers are solved by maximizing $\mathcal{L}(\boldsymbol{\lambda})$ with respect to λ_m's. The above inferred density is based on minimal information and assumptions. It is the flattest density according to the constraints. In this case, the Lagrange multipliers are not only the inferred parameters characterizing the density function, but also capture the amount of information conveyed in each one of the constraints relative to rest of the constraints used. They measure strength of the constraints.

In particular, when $M = 0$, $f(x)$ is a constant and hence x follows a uniform distribution. When the first moment of x is known, $f(x)$ has the form of an exponential distribution. When the first two moments of x are known, $f(x)$ has the form of a normal distribution. Furthermore, if more moment information is given, i.e. $M \geq 3$, to estimate the Lagrange multipliers, we use the Newton method considered in the literature. See Mead and Papanicolaou (1984) and Wu (2003).

In the bivariate case, the joint density of y and x is obtained from maximizing the information criterion $H(f)$ subject to some constraints. Here, we assume the moment conditions up to 4th

order are known. Then

$$\underset{f}{\text{Max}}\ H(f) = -\int_x \int_y f(y,x)\log f(y,x)dydx \tag{3}$$

subject to

$$\int_x \int_y y^{m_1}x^{m_2}f(y,x)dydx = \mu_{m_1m_2} = E\left(y^{m_1}x^{m_2}\right),\ 0 \leq m_1+m_2 \leq 4. \tag{4}$$

We construct the Lagrangian

$$\mathcal{L}\left(\boldsymbol{\lambda},\lambda_{00}\right) = -\int_x \int_y f(y,x)\log f(y,x)dydx + \sum\nolimits_{m_1=0}^{4}\sum\nolimits_{m_2=0}^{4}\lambda_{m_1m_2}\left(\mu_{m_1m_2} - \int_x\int_y y^{m_1}x^{m_2}f(y,x)dydx\right), \tag{5}$$

where $\boldsymbol{\lambda} = (\lambda_{m_1m_2})_{14\times 1}$ for all $1 \leq m_1+m_2 \leq 4$. The solution of the joint density distribution yields the form

$$\begin{aligned} f\left(y,x\right) &= \exp\left[-\sum\nolimits_{m_1+m_2=0}^{4}\lambda_{m_1m_2}y^{m_1}x^{m_2}\right] \qquad (6)\\ &= \frac{\exp\left[-\sum_{m_1+m_2=1}^{4}\lambda_{m_1m_2}y^{m_1}x^{m_2}\right]}{\int_x\int_y \exp\left[-\sum_{m_1+m_2=1}^{4}\lambda_{m_1m_2}y^{m_1}x^{m_2}\right]dydx} \equiv \frac{\exp\left[-\sum_{m_1+m_2=1}^{4}\lambda_{m_1m_2}y^{m_1}x^{m_2}\right]}{\Omega\left(\lambda_{m_1m_2}\right)}, \end{aligned}$$

where $\lambda_{m_1m_2}$ is the Lagrange multiplier that corresponds to the constraint $\int_x\int_y y^{m_1}x^{m_2}f(y,x)dydx = \mu_{m_1m_2}$, and $\lambda_{00} = \log\Omega\left(\lambda_{m_1m_2}\right)$ (with $m_1+m_2=0$) is the multiplier associated with the normalization constraint which is a function of other multipliers. See, e.g., Golan (1988, 2018) and Ryu (1993).

For deriving our results in Section 2, we rearrange the terms in $f(y,x)$ and write

$$\begin{aligned} f(y,x) = &\exp\left[-\left(\lambda_{04}x^4+\lambda_{03}x^3+\lambda_{02}x^2+\lambda_{01}x+\lambda_{00}\right)\right] \qquad (7)\\ &\times\exp\left\{-\left[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)\right]\right\}y \end{aligned}$$

where

$$\begin{aligned} &\lambda_{30}(x) = \lambda_{30}+\lambda_{31}x,\ \lambda_{20}(x) = \lambda_{20}+\lambda_{21}x+\lambda_{22}x^2,\\ &\lambda_{10}(x) = \lambda_{10}+\lambda_{11}x+\lambda_{12}x^2+\lambda_{13}x^3. \end{aligned}$$

Replacing $f(y,x)$ and λ_{00} into $\mathcal{L}\left(\boldsymbol{\lambda},\lambda_{00}\right) = \mathcal{L}\left(\boldsymbol{\lambda}\right)$, we obtain the Lagrange multipliers by maximizing

$$\mathcal{L}\left(\boldsymbol{\lambda}\right) = \sum\nolimits_{m_1+m_2=1}^{4}\lambda_{m_1m_2}\mu_{m_1m_2} + \lambda_{00}. \tag{8}$$

The marginal density of x is computed by integrating $f(y,x)$ over the support of y,

$$\begin{aligned} f(x) &= \int_y f(y,x)dy \\ &= \exp\left[-\left(\lambda_{04}x^4+\lambda_{03}x^3+\lambda_{02}x^2+\lambda_{01}x+\lambda_{00}\right)\right] \\ &\times \int_y \exp\left\{-\left[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)\right]\right\} y. \end{aligned} \tag{9}$$

We note that $f(x) = f(x,\boldsymbol{\lambda})$ and $f(y,x) = f(y,x,\boldsymbol{\lambda})$. When the Lagrange multipliers $\boldsymbol{\lambda}$ are estimated as $\hat{\boldsymbol{\lambda}}$ from (8), we get $\hat{f}(x) = f(x,\hat{\boldsymbol{\lambda}})$ and $\hat{f}(y,x) = f(y,x,\hat{\boldsymbol{\lambda}})$.

Although the above results are written under fourth order moment conditions in (4), they can be easily written when $0 \leq m_1 + m_2 \leq M$. We have considered fourth order moment conditions without any loss of generality since they capture data information on skewness and kurtosis.

2.2 Regression and Response Functions

Based on the bivariate maximum entropy joint distribution (7) and the marginal density (9), the conditional mean (regression function) of y given x is represented as

$$\begin{aligned} m(x) = E(y \mid x) &= \int_y y\frac{f(y,x)}{f(x)}dy \\ &= \frac{\int_y y\exp\left\{-\left[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y\right]\right\}dy}{\int_y \exp\left\{-\left[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y\right]\right\}dy}. \end{aligned} \tag{10}$$

Given the values of the Lagrange multipliers, we define

$$F_r(x) \equiv \int_y y^r \exp\left\{-\left[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y\right]\right\}dy. \tag{11}$$

where $r = 0,1,2,\ldots$. The regression function $m(x)$ thus takes the form

$$m(x) = m(x,\boldsymbol{\lambda}^*) = \frac{F_1(x)}{F_0(x)} = \frac{F_1(x,\boldsymbol{\lambda}^*)}{F_0(x,\boldsymbol{\lambda}^*)}, \tag{12}$$

where $\boldsymbol{\lambda}^* = (\lambda_{m_1m_2})_{10\times 1}$ for all $1 \leq m_1+m_2 \leq 4$ except λ_{0m_2} for $m_2 = 1,\ldots,4$. When the Lagrange multipliers are estimated from (8) by Newton method,

$$\hat{m}(x) = m\left(x,\hat{\boldsymbol{\lambda}}^*\right) = \frac{F_1\left(x,\hat{\boldsymbol{\lambda}}^*\right)}{F_0\left(x,\hat{\boldsymbol{\lambda}}^*\right)}. \tag{13}$$

This is the IT nonparametric regression function estimator. Furthermore, the response function $\beta(x) = \frac{dm(x)}{dx}$ (derivative) can be written as

$$\beta(x) = \beta(x, \boldsymbol{\lambda}^*) = \frac{F_1'(x, \boldsymbol{\lambda}^*) F_0(x, \boldsymbol{\lambda}^*) - F_1(x, \boldsymbol{\lambda}^*) F_0'(x, \boldsymbol{\lambda}^*)}{F_0^2(x, \boldsymbol{\lambda}^*)}, \tag{14}$$

and its estimator is given by

$$\hat{\beta}(x) = \beta\left(x, \hat{\boldsymbol{\lambda}}^*\right) \tag{15}$$

We note that $F_r'(x)$ represents the first derivative of $F_r(x)$ with respect to x, $r = 0, 1, 2, \ldots$.

2.3 Recursive Integration

It is unlikely to solve out the exponential polynomial integrals in the numerator and denominator from (10) in explicit forms. Numerical methods can be used to solve the problem by integrating the exponential polynomial function at each value of x. However, for large sample size, numerical methods are quite computationally expensive and hence are not satisfactory. We develop a recursive integration method which can not only solve the conditional mean $m(x)$ but also reduce the computational cost significantly.

According to the definition of $F_r(x)$ in (11), the changes in F_0, F_1 and F_2 are given by

$$\begin{aligned} F_0' &= -\lambda_{30}'(x)F_3 - \lambda_{20}'(x)F_2 - \lambda_{10}'(x)F_1 \\ F_1' &= -\lambda_{30}'(x)F_4 - \lambda_{20}'(x)F_3 - \lambda_{10}'(x)F_2 \\ F_2' &= -\lambda_{30}'(x)F_5 - \lambda_{20}'(x)F_4 - \lambda_{10}'(x)F_3, \end{aligned} \tag{16}$$

where $\lambda'(x)$ denotes the first derivative of $\lambda(x)$ with respect to x. Due to the special properties of (11), integrals of higher order exponential polynomial functions can be represented by those of lower orders. Based on this fact, F_3, F_4 and F_5 in (16) are replaced by the linear combinations of F_0, F_1 and F_2, resulting in a system of linear equations

$$\begin{aligned} F_0'(x) &= \Lambda_{00}(x)F_0(x) + \Lambda_{01}(x)F_1(x) + \Lambda_{02}(x)F_2(x) \\ F_1'(x) &= \Lambda_{10}(x)F_0(x) + \Lambda_{11}(x)F_1(x) + \Lambda_{12}(x)F_2(x) \\ F_2'(x) &= \Lambda_{20}(x)F_0(x) + \Lambda_{21}(x)F_1(x) + \Lambda_{22}(x)F_2(x). \end{aligned} \tag{17}$$

The derivations of (16) and (17) are provided in the Appendix A.1. Starting from an initial value

x_0, for a very small increment h, we trace out $F_0(x)$, $F_1(x)$ and $F_2(x)$ over the entire range of x

$$F_0(x_0+h) \approx F_0(x_0) + F_0'(x_0)h \tag{18}$$
$$F_1(x_0+h) \approx F_1(x_0) + F_1'(x_0)h$$
$$F_2(x_0+h) \approx F_2(x_0) + F_2'(x_0)h$$

The IT estimators $\hat{m}(x)$ in (13) and $\hat{\beta}(x)$ in (15) are thus evaluated using (17) and (18) with $\boldsymbol{\lambda}^*$ replaced by $\hat{\boldsymbol{\lambda}}^*$. The results for the finite domain integration are similar to the above, which are provided in Appendix A.2.

3 Simulation and Empirical Examples

Here we first consider two data generating processes (DGP) to evaluate the performance of our proposed IT estimator of response function in Sections 3.1 and 3.2. Then we present our illustrative empirical example to study regression and response functions in Section 3.3.

3.1 Data Generating Process 1: Nonlinear Function

The true model considered is a nonlinear function[1]

$$y_i = -\frac{1}{5}\log\left(e^{-2.5} + 2e^{-5x_i}\right) + u_i \tag{19}$$

where $i = 1, 2, \ldots, n$, the variables y_i and x_i are in log values, and x_i are independent and identically drawn from uniform distribution with mean 0.5 and variance $\frac{1}{12}$. The error term u_i follows independent and identical normal distribution with mean 0 and variance 0.01.

The goal is to estimate the response coefficient $\beta(x) = \frac{\partial y}{\partial x}$. Two parametric approximations considered are

$$\begin{aligned} \textit{Linear} &: y_i = \beta_0 + \beta_1 x_i + u_i \\ \textit{Quadratic} &: y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i. \end{aligned}$$

These two parametric models are not correctly specified. Thus, one can expect that the estimation of the response coefficients may be biased. Besides these two parametric models, local constant nonparametric estimation of the response coefficient is also of our interest as a comparison with

[1] This simulation example is similar to Rilstone and Ullah (1989).

our IT method estimator. The local constant (Nadaraya-Watson) nonparametric kernel estimator is $\tilde{m}(x) = \sum y_i w_i(x)$, where $w_i(x) = \frac{K((x_i - x)/b)}{\sum K((x_i - x)/b)}$ in which $K(\cdot)$ is a kernel function and b is the bandwidth, for example, see Pagan and Ullah (1999). We have used normal kernel and cross-validated bandwidth. The bias and root mean square error (RMSE) results from linear function, quadratic approximation, local constant nonparametric method and IT method are reported in Table 1, averaged over 1000 replications of sample size 200. The values of the response coefficients shown are evaluated at the population mean of x, which is 0.5. Standard errors are given in the parentheses. True value of the response coefficient $\beta(x = 0.5) = 0.6667$.

Table 1

	Linear	Quadratic	Nonparametric	IT
$\beta(x) = \frac{\partial y}{\partial x}$	0.6288	0.6296	0.6468	0.6550
	(0.0276)	(0.0263)	(0.0904)	(0.0268)
Bias	0.0379	0.0371	0.0199	**0.0117**
RMSE	0.0469	0.0455	0.0926	**0.0292**

The biases for nonparametric kernel and IT estimators are smaller than those under linear and quadratic approximations. However, nonparametric estimation yields a larger RMSE compared with the three other methods. Even though nonparametric and IT estimations both have the advantage of avoiding the difficulties associated with the functional forms, results have indicated that the IT method outperforms the nonparametric method. This may be because the rate of convergence for MSE to zero for the IT estimator is n^{-1} whereas that of nonparametric kernel estimator is known to be $(nb)^{-1}$ where b is small (Li and Racine (2007)).

3.2 Data Generating Process 2: Linear Function

Now the true data generating process is a linear function:

$$y_i = 2 + x_i + u_i \tag{20}$$

where $i = 1, 2, \ldots, n$, x_i and u_i follow the same distributions as in DGP 1. Comparisons are made with linear, quadratic approximations and nonparametric estimation. Results on bias and RMSE are averaged over 1000 replications of sample size 200. The values of the response coefficients shown

are evaluated at $x = 0.5$.

Table 2

	Linear	Quadratic	Nonparametric	IT
$\beta(x) = \frac{\partial y}{\partial x}$	1.0009	1.0009	1.0105	1.0014
	(0.0250)	(0.0251)	(0.1133)	(0.0284)
Bias	0.0009	0.0009	0.0105	0.0014
RMSE	0.0250	0.0251	0.1138	0.0284

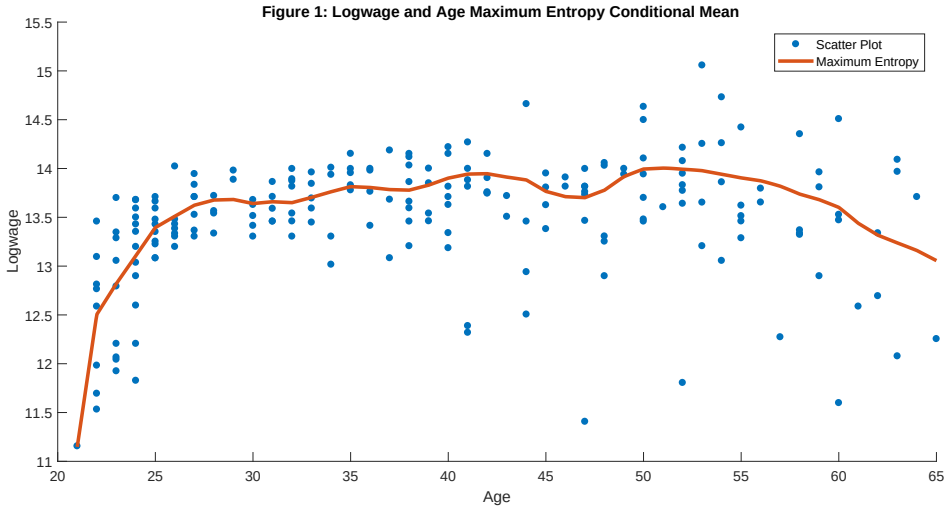
When the true DGP is linear in x, it is not surprising that linear approximation has the smallest bias and RMSE. The IT method has much smaller bias and RMSE than that under the nonparametric kernel estimation. Even though the true relationship between x and y is linear, considering IT estimator based on first four moments is still successful in capturing the linearity. For example, $\beta(x)$ based on IT, compared to nonparametric kernel, is closer to $\beta(x)$ based on the true linear DGP. This is similar to the result in DGP 1. Since true DGP in practice is not known, IT provides a better option to be used.

3.3 Empirical Study: Canadian High School Graduate Earnings

To further illustrate the superiority of maximum entropy method, we conduct the study of the average logwage conditional on the age using the 1971 dataset of 205 Canadian high school graduate earnings. According to (10), the variable y denotes the logwage of high school graduates and x denotes the age. As a comparison, local constant and local linear nonparametric estimations, as well as quadratic and quartic approximations are considered

$$\begin{aligned} \textit{Quadratic} &: \quad y = \beta_0 + \beta_1 x + \beta_2 x^2 + u \\ \textit{Quartic} &: \quad y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + u. \end{aligned}$$

The local linear nonparametric kernel estimators $m^*(x)$ and $\beta^*(x)$ are obtained by minimizing the local linear weighted squared losses $\sum (y_i - m(x) - (x_i - x)\beta(x))^2 K((x_i - x)/b)$ with respect to $m(x)$ and $\beta(x)$. We note that minimizing local constant weighted squared losses $\sum (y_i - m(x))^2 K((x_i - x)/b)$ with respect to $m(x)$ provides local constant nonparametric kernel estimator $\tilde{m}(x)$ used in Section 3.1 and 3.2. We use our IT method to show the plot of the estimated $\hat{m}(x)$ of logwage in Figure 1. An illustration of numerical calculations based on the IT method is given in Appendix A.3.



From Figure 1, log earning grows rapidly from age 21 to age 25. Before around age 45, the growth speed of logwage is slowed down. A depth at age 47 is displayed. From age 47 to age 65, log earning rises and then declines smoothly. It is shown that the IT estimation captures the tail observations very well, see Appendix A.3 for numerical calculations in tails. The average response coefficient $\hat{\beta}$ over the range of age is approximately 0.0434. The average response coefficients under local constant and local linear nonparametric kernel methods are 0.0315 and 0.0421 respectively. Under quadratic approximation, the average response coefficient is 0.0195. We use the average response coefficient under quartic approximation as a benchmark, which is 0.0461. Average $\hat{\beta}$ under IT method is large than that under quadratic approximation, local constant and local linear estimations, which shows that IT method is advantageous over the rest considered.

4 Asymptotic Properties of IT Estimators and Test for Normality

4.1 Asymptotic Normality

First, we define

$$\begin{aligned}
\underset{14\times 1}{\mathbf{Z}_i} &= \left(y_i, x_i, y_i^2, x_i^2, y_i^3, x_i^3, y_i^4, x_i^4, y_i x_i, y_i x_i^2, y_i^2 x_i, y_i x_i^3, y_i^3 x_i, y_i^2 x_i^2\right)^T, \qquad (21)\\
\underset{14\times 1}{\hat{\boldsymbol{\mu}}} &= \left(\hat{\mu}_{10}, \hat{\mu}_{01}, \hat{\mu}_{20}, \hat{\mu}_{02}, \hat{\mu}_{30}, \hat{\mu}_{03}, \hat{\mu}_{40}, \hat{\mu}_{04}, \hat{\mu}_{11}, \hat{\mu}_{12}, \hat{\mu}_{21}, \hat{\mu}_{13}, \hat{\mu}_{31}, \hat{\mu}_{22}\right)^T,\\
\underset{14\times 1}{\boldsymbol{\mu}} &= \left(\mu_{10}, \mu_{01}, \mu_{20}, \mu_{02}, \mu_{30}, \mu_{03}, \mu_{40}, \mu_{04}, \mu_{11}, \mu_{12}, \mu_{21}, \mu_{13}, \mu_{31}, \mu_{22}\right)^T,
\end{aligned}$$

where $\hat{\mu}_{m_1 m_2} = \frac{1}{n}\sum_{i=1}^{n} y_i^{m_1} x_i^{m_2}$, $\mu_{m_1 m_2} = E\left(y_i^{m_1} x_i^{m_2}\right)$, $m_1, m_2 = 0, 1, 2, 3, 4$ and $1 \leq m_1 + m_2 \leq 4$, and all the bold letters represent vectors. Suppose the following assumptions hold.

1. $\mathbf{Z}_i$, $i = 1, \ldots, n$ are independent and identically distributed from $(\boldsymbol{\mu}, \Sigma)$.

2. $\Sigma = \mathrm{COV}(\mathbf{Z}_i)$ is assumed to be positive semi-definite, where the diagonals of Σ are

$$\mathrm{Var}\left(y_i^{m_1} x_i^{m_2}\right) = \mu_{(2m_1)(2m_2)} - \mu_{m_1 m_2}^2,$$

   and the off-diagonals of Σ are

$$\mathrm{Cov}\left(y_i^{m_1} x_i^{m_2}, y_i^{m_1^*} x_i^{m_2^*}\right) = \mu_{(m_1+m_1^*)(m_2+m_2^*)} - \mu_{m_1 m_2}\mu_{m_1^* m_2^*}.$$

3. $\mu_{(m_1+m_1^*)(m_2+m_2^*)} < \infty$, $\forall m_1, m_2, m_1^*, m_2^* = 0, 1, 2, 3, 4$, $m_1 + m_2 \leq 4, m_1^* + m_2^* \leq 4$.

Now we present the following proposition.

Proposition 1. *Under assumptions 1 to 3, as n goes to* ∞*,*

$$\sqrt{n}\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right) \sim N\left(\mathbf{0}, \Sigma\right). \tag{22}$$

The proof of this proposition is given in Appendix B.1.

Now, suppose the unique solution for each Lagrange multiplier exists. Then from (8), the vector $\boldsymbol{\lambda} = (\lambda_{10}, \lambda_{01}, \lambda_{20}, \lambda_{02}, \lambda_{30}, \lambda_{03}, \lambda_{40}, \lambda_{04}, \lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{13}, \lambda_{31}, \lambda_{22})^T$ can be expressed as a function of $\boldsymbol{\mu}$, i.e.

$$\boldsymbol{\lambda} = g\left(\boldsymbol{\mu}\right) \text{ and } \hat{\boldsymbol{\lambda}} = g\left(\hat{\boldsymbol{\mu}}\right). \tag{23}$$

Since from Proposition 1, $\sqrt{n}\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right) \sim N\left(\mathbf{0}, \Sigma\right)$ as $n \to \infty$, it follows that

$$\sqrt{n}\left(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\right) \sim N\left(\mathbf{0}, g^{(1)}\left(\boldsymbol{\mu}\right)\Sigma g^{(1)}\left(\boldsymbol{\mu}\right)^T\right) \text{ as } n \longrightarrow \infty, \tag{24}$$

where $g^{(1)}\left(\boldsymbol{\mu}\right) = \frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^T}$ is the first derivative of $g\left(\boldsymbol{\mu}\right)$ with respect to $\boldsymbol{\mu}$. See Appendix B.1.

Using the results in Proposition 1 and (24), $\sqrt{n}\left(\hat{\boldsymbol{\lambda}}^* - \boldsymbol{\lambda}^*\right) \sim N\left(\mathbf{0}, g^{*(1)}\left(\boldsymbol{\mu}\right)\Sigma g^{*(1)}\left(\boldsymbol{\mu}\right)^T\right)$ as $n \longrightarrow \infty$, where $\boldsymbol{\lambda}^* = g^*\left(\boldsymbol{\mu}\right)$ and $g^{*(1)}\left(\boldsymbol{\mu}\right) = \frac{\partial g^*(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^T}$. We get the following proposition for $\hat{m}\left(x\right)$ and $\hat{\beta}\left(x\right)$.

Proposition 2. *Under assumptions 1 to 3 and* (24)*, the asymptotic distributions of* $\hat{m}\left(x\right) =$

$m\left(x,\hat{\boldsymbol{\lambda}}^*\right)$ *and* $\hat{\beta}(x)=\beta\left(x,\hat{\boldsymbol{\lambda}}^*\right)$ *are given as* $n\to\infty$,

$$\sqrt{n}\left(m\left(x,\hat{\boldsymbol{\lambda}}^*\right)-m\left(x,\boldsymbol{\lambda}^*\right)\right)\sim N\left(\mathbf{0},m^{(1)}\left(x,\boldsymbol{\lambda}^*\right)g^{*(1)}\left(\boldsymbol{\mu}\right)\Sigma g^{*(1)}\left(\boldsymbol{\mu}\right)^T m^{(1)}\left(x,\boldsymbol{\lambda}^*\right)^T\right), \tag{25}$$

where $m^{(1)}\left(x,\boldsymbol{\lambda}^*\right)=\frac{\partial m(x,\boldsymbol{\lambda}^*)}{\partial\boldsymbol{\lambda}^{*T}}$ *is the first derivative of* $m\left(x,\boldsymbol{\lambda}^*\right)$ *with respect to* $\boldsymbol{\lambda}^*$. *And*

$$\sqrt{n}\left(\beta\left(x,\hat{\boldsymbol{\lambda}}^*\right)-\beta\left(x,\boldsymbol{\lambda}^*\right)\right)\sim N\left(\mathbf{0},\beta^{(1)}\left(x,\boldsymbol{\lambda}^*\right)g^{*(1)}\left(\boldsymbol{\mu}\right)\Sigma g^{*(1)}\left(\boldsymbol{\mu}\right)^T\beta^{(1)}\left(x,\boldsymbol{\lambda}^*\right)^T\right), \tag{26}$$

where $\beta^{(1)}\left(x,\boldsymbol{\lambda}^*\right)=\frac{\partial\beta(x,\boldsymbol{\lambda}^*)}{\partial\boldsymbol{\lambda}^{*T}}$ *is the first derivative of* $\beta\left(x,\boldsymbol{\lambda}^*\right)$ *with respect to* $\boldsymbol{\lambda}^*$.

The proof of Proposition 2 is given in Appendix B.1. Also, we note that the convergence rates of $m\left(x,\hat{\boldsymbol{\lambda}}^*\right)$ and $\beta\left(x,\hat{\boldsymbol{\lambda}}^*\right)$ are each $\sqrt{n}$.

4.2 Testing for Normality

When the true distribution $f(x,y)$ is normal, the Lagrange multipliers for moments with orders higher than two are equal to zero, i.e. $\lambda_{ij}=0,\ \forall\ i+j>2$. $\boldsymbol{\lambda}$ contains 9 elements with orders higher than two. Testing whether (x,y) are jointly normal is equivalent to testing the null

$$H_0: R\boldsymbol{\lambda}=\mathbf{0},$$

where R is a 9×14 matrix with elements $R(1,6),R(2,7),\cdots,R(9,14)=1$ and the rest elements $=0$. We develop a Wald test statistic

$$W=\left(R\hat{\boldsymbol{\lambda}}\right)'\left(V\left(R\hat{\boldsymbol{\lambda}}\right)\right)^{-1}\left(R\hat{\boldsymbol{\lambda}}\right),$$

where $V\left(R\hat{\boldsymbol{\lambda}}\right)=RV\left(\hat{\boldsymbol{\lambda}}\right)R'$ in which $V\left(\hat{\boldsymbol{\lambda}}\right)$ is the asymptotic variance. Since $R\hat{\boldsymbol{\lambda}}$ is asymptotically normal from (24), it follows that W is χ^2_9 asymptotically. Conclusion is drawn based on the comparison between calculated value of W and tabulated Chi-square distribution critical value. When the true distribution $f(x,y)$ is normal, the relationship between x and y is linear. Therefore, it is also a test for linearity. In our empirical example in Section 3.3, we compute the Wald statistic $W\approx 13548$. At 1% significance level, we reject the null hypothesis. Thus, we conclude that the relationship between age and logwage is nonlinear. Alternatively, one can test the null hypothesis using entropy-ratio test.

5 Conclusions

In this chapter, we have estimated the econometric functions through an IT method, which is nonparametric. Two basic econometric functions, regression and response, have been analyzed. The advantages of using IT method over parametric specifications and nonparametric kernel approaches have been explained by the simulation and empirical examples. It can be a useful tool for practitioners due to its simplicity and efficiency. Asymptotic properties are established. The IT based estimators are shown to be $\sqrt{n}$ consistent and normal. Thus, it has a faster rate of convergence compared to the nonparametric kernel procedures.

Based on what has been developed in this chapter, further work can be done in the future; for example, the bivariate regression approach introduced in this chapter can be potentially extended to the multivariate case. Next, we can extend our chapter's IT analysis for conditional variance and conditional covariance functions, among other econometric functions. Furthermore, the IT method may be carried over from Shannon's information theory to Kullback and Leibler (1951) divergence which has been discussed in the literature by Golan et al. (1996), Judge and Mittelhammer (2011), Golan (2018), Ullah (1996), etc. Chakrabarty et al. (2015), Maasoumi and Racine (2016), and Racine and Li (2017) have approached quantile estimation problems using nonparametric kernel methods. Similarly, the maximum entropy based probability distributions derived in our chapter may be adopted for nonparametric quantile estimation problems. In addition, our IT-based estimator of conditional mean can be applied to the nonparametric component in semiparametric models, such as the partial linear model. Along with all these, other future work may be to explore links between IT-based density and the log-spline density considered in Stone (1990). Moreover, it would be useful to establish connections of our asymptotically χ^2 distributed Wald's type normality test in Section 4.2 with those of Neyman's smooth test considered in Ledwina (1994) and Inglot and Ledwina (1996) and the entropy-ratio test on the lambdas which is 2-times the difference of the objective functions (with/without imposing the null hypothesis) and asymptotically distributed as χ^2, see for example Golan (2018, p.96). We feel the IT approach for specifying regression and response functions considered here may open a new path to address specification and other related issues in econometrics with many applications.

Appendix A: Calculations

A.1: Recursive Integration

When the range for y is from $-\infty$ to $+\infty$, define the following integrals as functions of x.

$$F_r(x) \equiv F_r \equiv \int_{-\infty}^{+\infty} y^r \exp\left\{-\left[\lambda_{40}y^4 + \lambda_{30}(x)y^3 + \lambda_{20}(x)y^2 + \lambda_{10}(x)y\right]\right\} dy,$$

where $r = 0, 1, 2, \ldots$. In particular,

$$\begin{aligned}
F_0(x) &\equiv F_0 \equiv \int_{-\infty}^{+\infty} \exp\left\{-\left[\lambda_{40}y^4 + \lambda_{30}(x)y^3 + \lambda_{20}(x)y^2 + \lambda_{10}(x)y\right]\right\} dy \\
F_1(x) &\equiv F_1 \equiv \int_{-\infty}^{+\infty} y \exp\left\{-\left[\lambda_{40}y^4 + \lambda_{30}(x)y^3 + \lambda_{20}(x)y^2 + \lambda_{10}(x)y\right]\right\} dy \\
F_2(x) &\equiv F_2 \equiv \int_{-\infty}^{+\infty} y^2 \exp\left\{-\left[\lambda_{40}y^4 + \lambda_{30}(x)y^3 + \lambda_{20}(x)y^2 + \lambda_{10}(x)y\right]\right\} dy.
\end{aligned}$$

Suppose that λ_{40} is positive. Firstly, solve for F_3.

$$\begin{aligned}
0 &= \int_{-\infty}^{+\infty} d\exp\left\{-\left[\lambda_{40}y^4 + \lambda_{30}(x)y^3 + \lambda_{20}(x)y^2 + \lambda_{10}(x)y\right]\right\} \\
&= \int_{-\infty}^{+\infty} \left(-4\lambda_{40}y^3 - 3\lambda_{30}(x)y^2 - 2\lambda_{20}(x)y - \lambda_{10}(x)\right) e^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]} dy \\
&= -4\lambda_{40}F_3 - 3\lambda_{30}(x)F_2 - 2\lambda_{20}(x)F_1 - \lambda_{10}(x)F_0
\end{aligned}$$

$$F_3 = -\frac{1}{4\lambda_{40}}(3\lambda_{30}(x)F_2 + 2\lambda_{20}(x)F_1 + \lambda_{10}(x)F_0)$$

Secondly, solve for F_4.

$$\begin{aligned}
F_0 &= \int_{-\infty}^{+\infty} \exp\left\{-[\lambda_{40}y^4 + \lambda_{30}(x)y^3 + \lambda_{20}(x)y^2 + \lambda_{10}(x)y]d\right\} y \\
&= ye^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]}|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} yde^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]} \\
&= \int_{-\infty}^{+\infty} (4\lambda_{40}y^4 + 3\lambda_{30}(x)y^3 + 2\lambda_{20}(x)y^2 + \lambda_{10}(x)y)e^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]} dy \\
&= 4\lambda_{40}F_4 + 3\lambda_{30}(x)F_3 + 2\lambda_{20}(x)F_2 + \lambda_{10}(x)F_1
\end{aligned}$$

$$\begin{aligned}
F_4 &= \frac{1}{4\lambda_{40}}(-3\lambda_{30}(x)F_3 - 2\lambda_{20}(x)F_2 - \lambda_{10}(x)F_1 + F_0) \\
&= -\frac{3\lambda_{30}(x)}{4\lambda_{40}}F_3 - \frac{\lambda_{20}(x)}{2\lambda_{40}}F_2 - \frac{\lambda_{10}(x)}{4\lambda_{40}}F_1 + \frac{1}{4\lambda_{40}}F_0.
\end{aligned}$$

Replace F_3 with $-\frac{1}{4\lambda_{40}}(3\lambda_{30}(x)F_2 + 2\lambda_{20}(x)F_1 + \lambda_{10}(x)F_0)$.

$$\begin{aligned} F_4 &= \left(\frac{9\lambda_{30}^2(x)}{16\lambda_{40}^2} - \frac{\lambda_{20}(x)}{2\lambda_{40}}\right)F_2 + \\ &\left(\frac{3\lambda_{30}(x)\lambda_{20}(x)}{8\lambda_{40}^2} - \frac{\lambda_{10}(x)}{4\lambda_{40}}\right)F_1 + \\ &\left(\frac{3\lambda_{30}(x)\lambda_{10}(x)}{16\lambda_{40}^2} + \frac{1}{4\lambda_{40}}\right)F_0 \end{aligned}$$

Thirdly, solve for F_5.

$$\begin{aligned} F_1 &= \int_{-\infty}^{+\infty} y \exp\left\{-[\lambda_{40}y^4 + \lambda_{30}(x)y^3 + \lambda_{20}(x)y^2 + \lambda_{10}(x)y]\right\} dy \\ &= \int_{-\infty}^{+\infty} \exp\left\{-[\lambda_{40}y^4 + \lambda_{30}(x)y^3 + \lambda_{20}(x)y^2 + \lambda_{10}(x)y]\right\} d\left(\frac{1}{2}y^2\right) \\ &= \frac{1}{2}y^2 e^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]}|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \frac{1}{2}y^2 de^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]} \\ &= \int_{-\infty}^{+\infty} \frac{1}{2}y^2(4\lambda_{40}y^3 + 3\lambda_{30}(x)y^2 + 2\lambda_{20}(x)y + \lambda_{10}(x))e^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]}dy \\ &= 2\lambda_{40}F_5 + \frac{3}{2}\lambda_{30}(x)F_4 + \lambda_{20}(x)F_3 + \frac{1}{2}\lambda_{10}(x)F_2 \end{aligned}$$

$$\begin{aligned} F_5 &= -\frac{1}{2\lambda_{40}}\left(\frac{3}{2}\lambda_{30}(x)F_4 + \lambda_{20}(x)F_3 + \frac{1}{2}\lambda_{10}(x)F_2 - F_1\right) \\ &= -\frac{3\lambda_{30}(x)}{4\lambda_{40}}F_4 - \frac{\lambda_{20}(x)}{2\lambda_{40}}F_3 - \frac{\lambda_{10}(x)}{4\lambda_{40}}F_2 + \frac{1}{2\lambda_{40}}F_1. \end{aligned}$$

Replace F_3 and F_4.

$$\begin{aligned} F_5 &= \left(-\frac{27\lambda_{30}^3(x)}{64\lambda_{40}^3} + \frac{3\lambda_{30}(x)\lambda_{20}(x)}{4\lambda_{40}^2} - \frac{\lambda_{10}(x)}{4\lambda_{40}}\right)F_2 + \\ &\left(-\frac{9\lambda_{30}^2(x)\lambda_{20}(x)}{32\lambda_{40}^3} + \frac{3\lambda_{30}(x)\lambda_{10}(x)}{16\lambda_{40}^2} + \frac{\lambda_{20}^2(x)}{4\lambda_{40}^2} + \frac{1}{2\lambda_{40}}\right)F_1 + \\ &\left(-\frac{9\lambda_{30}^2(x)\lambda_{10}(x)}{64\lambda_{40}^3} - \frac{3\lambda_{30}(x)}{16\lambda_{40}^2} + \frac{\lambda_{20}(x)\lambda_{10}(x)}{8\lambda_{40}^2}\right)F_0 \end{aligned}$$

Define

$$F_0' \equiv \frac{dF_0(x)}{dx}, F_1' \equiv \frac{dF_1(x)}{dx}, F_2' \equiv \frac{dF_2(x)}{dx}$$
$$\lambda_{30}'(x) \equiv \frac{d\lambda_{30}(x)}{dx}, \lambda_{20}'(x) \equiv \frac{d\lambda_{20}(x)}{dx}, \lambda_{10}'(x) \equiv \frac{d\lambda_{10}(x)}{dx}$$

Firstly, solve for F_0'.

$$\begin{aligned}
F_0' &\equiv \frac{d}{dx}\int_{-\infty}^{+\infty} \exp\left\{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]\right\}dy \\
&= \int_{-\infty}^{+\infty} \frac{d}{dx}\exp\left\{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]\right\}dy \\
&= \int_{-\infty}^{+\infty} (-\lambda_{30}'(x)y^3-\lambda_{20}'(x)y^2-\lambda_{10}'(x)y)e^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]}dy \\
&= -\lambda_{30}'(x)F_3-\lambda_{20}'(x)F_2-\lambda_{10}'(x)F_1
\end{aligned}$$

Replace F_3 with $-\frac{1}{4\lambda_{40}}(3\lambda_{30}(x)F_2+2\lambda_{20}(x)F_1+\lambda_{10}(x)F_0)$.

$$F_0' = \left(\frac{3\lambda_{30}'(x)\lambda_{30}(x)}{4\lambda_{40}}-\lambda_{20}'(x)\right)F_2+\left(\frac{\lambda_{30}'(x)\lambda_{20}(x)}{2\lambda_{40}}-\lambda_{10}'(x)\right)F_1+\frac{\lambda_{30}'(x)\lambda_{10}(x)}{4\lambda_{40}}F_0$$

Secondly, solve for F_1'.

$$\begin{aligned}
F_1' &\equiv \frac{d}{dx}\int_{-\infty}^{+\infty} y\exp\left\{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]\right\}dy \\
&= \int_{-\infty}^{+\infty} \frac{d}{dx}y\exp\left\{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]\right\}dy \\
&= \int_{-\infty}^{+\infty} (-\lambda_{30}'(x)y^4-\lambda_{20}'(x)y^3-\lambda_{10}'(x)y^2)e^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]}dy \\
&= -\lambda_{30}'(x)F_4-\lambda_{20}'(x)F_3-\lambda_{10}'(x)F_2
\end{aligned}$$

Replace F_3 and F_4.

$$\begin{aligned}
F_1' = &\left(-\frac{9\lambda_{30}'(x)\lambda_{30}^2(x)}{16\lambda_{40}^2}+\frac{\lambda_{30}'(x)\lambda_{20}(x)}{2\lambda_{40}}+\frac{3\lambda_{20}'(x)\lambda_{30}(x)}{4\lambda_{40}}-\lambda_{10}'(x)\right)F_2+ \\
&\left(-\frac{3\lambda_{30}'(x)\lambda_{30}(x)\lambda_{20}(x)}{8\lambda_{40}^2}+\frac{\lambda_{30}'(x)\lambda_{10}(x)}{4\lambda_{40}}+\frac{\lambda_{20}'(x)\lambda_{20}(x)}{2\lambda_{40}}\right)F_1+ \\
&\left(-\frac{3\lambda_{30}'(x)\lambda_{30}(x)\lambda_{10}(x)}{16\lambda_{40}^2}-\frac{\lambda_{30}'(x)}{4\lambda_{40}}+\frac{\lambda_{20}'(x)\lambda_{10}(x)}{4\lambda_{40}}\right)F_0
\end{aligned}$$

Thirdly, solve for F_2'.

$$\begin{aligned}
F_2' &\equiv \frac{d}{dx}\int_{-\infty}^{+\infty} y^2\exp\left\{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]\right\}dy \\
&= \int_{-\infty}^{+\infty} \frac{d}{dx}y^2\exp\left\{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]\right\}dy \\
&= \int_{-\infty}^{+\infty} (-\lambda_{30}'(x)y^5-\lambda_{20}'(x)y^4-\lambda_{10}'(x)y^3)e^{-[\lambda_{40}y^4+\lambda_{30}(x)y^3+\lambda_{20}(x)y^2+\lambda_{10}(x)y]}dy \\
&= -\lambda_{30}'(x)F_5-\lambda_{20}'(x)F_4-\lambda_{10}'(x)F_3
\end{aligned}$$

Replace F_5, F_4 and F_3.

$$
\begin{aligned}
F_2' &= \begin{pmatrix} \frac{27\lambda_{30}'(x)\lambda_{30}^3(x)}{64\lambda_{40}^3} - \frac{3\lambda_{30}'(x)\lambda_{30}(x)\lambda_{20}(x)}{4\lambda_{40}^2} + \frac{\lambda_{30}'(x)\lambda_{10}(x)}{4\lambda_{40}} \\ -\frac{9\lambda_{20}'(x)\lambda_{30}^2(x)}{16\lambda_{40}^2} + \frac{\lambda_{20}'(x)\lambda_{20}(x)}{2\lambda_{40}} + \frac{3\lambda_{10}'(x)\lambda_{30}(x)}{4\lambda_{40}} \end{pmatrix} F_2 + \\
&\begin{pmatrix} \frac{9\lambda_{30}'(x)\lambda_{30}^2(x)\lambda_{20}(x)}{32\lambda_{40}^3} - \frac{3\lambda_{30}'(x)\lambda_{30}(x)\lambda_{10}(x)}{16\lambda_{40}^2} - \frac{\lambda_{30}'(x)\lambda_{20}^2(x)}{4\lambda_{40}^2} \\ -\frac{\lambda_{30}'(x)}{2\lambda_{40}} - \frac{3\lambda_{20}'(x)\lambda_{30}(x)\lambda_{20}(x)}{8\lambda_{40}^2} + \frac{\lambda_{20}'(x)\lambda_{10}(x)}{4\lambda_{40}} + \frac{\lambda_{10}'(x)\lambda_{20}(x)}{2\lambda_{40}} \end{pmatrix} F_1 + \\
&\begin{pmatrix} \frac{9\lambda_{30}'(x)\lambda_{30}^2(x)\lambda_{10}(x)}{64\lambda_{40}^3} + \frac{3\lambda_{30}'(x)\lambda_{30}(x)}{16\lambda_{40}^2} - \frac{\lambda_{30}'(x)\lambda_{20}(x)\lambda_{10}(x)}{8\lambda_{40}^2} \\ -\frac{3\lambda_{20}'(x)\lambda_{30}(x)\lambda_{10}(x)}{16\lambda_{40}^2} - \frac{\lambda_{20}'(x)}{4\lambda_{40}} + \frac{\lambda_{10}'(x)\lambda_{10}(x)}{4\lambda_{40}} \end{pmatrix} F_0
\end{aligned}
$$

Equations (16) and (17) are thus obtained.

A.2: Finite Integral Range

When the range for y $[a(x), b(x)]$ is varying based on x, define the following functions.

$$F_r(x) \equiv F_r \equiv \int_{a(x)}^{b(x)} y^r \exp\left\{-[\lambda_{40}y^4 + \lambda_{30}(x)y^3 + \lambda_{20}(x)y^2 + \lambda_{10}(x)y]\right\} dy$$

where $r = 0, 1, 2, \ldots$ Define the following functions of x.

$$
\begin{aligned}
A_0(x) &\equiv A_0 \equiv \exp\left\{-[\lambda_{40}a(x)^4 + \lambda_{30}(x)a(x)^3 + \lambda_{20}(x)a(x)^2 + \lambda_{10}(x)a(x)]\right\} \\
B_0(x) &\equiv B_0 \equiv \exp\left\{-[\lambda_{40}b(x)^4 + \lambda_{30}(x)b(x)^3 + \lambda_{20}(x)b(x)^2 + \lambda_{10}(x)b(x)]\right\} \\
A_1(x) &\equiv A_1 \equiv a(x)\exp\left\{-[\lambda_{40}a(x)^4 + \lambda_{30}(x)a(x)^3 + \lambda_{20}(x)a(x)^2 + \lambda_{10}(x)a(x)]\right\} \\
B_1(x) &\equiv B_1 \equiv b(x)\exp\left\{-[\lambda_{40}b(x)^4 + \lambda_{30}(x)b(x)^3 + \lambda_{20}(x)b(x)^2 + \lambda_{10}(x)b(x)]\right\} \\
A_2(x) &\equiv A_2 \equiv a(x)^2\exp\left\{-[\lambda_{40}a(x)^4 + \lambda_{30}(x)a(x)^3 + \lambda_{20}(x)a(x)^2 + \lambda_{10}(x)a(x)]\right\} \\
B_2(x) &\equiv B_2 \equiv b(x)^2\exp\left\{-[\lambda_{40}b(x)^4 + \lambda_{30}(x)b(x)^3 + \lambda_{20}(x)b(x)^2 + \lambda_{10}(x)b(x)]\right\} \\
L_{0,a}(x) &\equiv a'(x)\exp\left\{-[\lambda_{40}a(x)^4 + \lambda_{30}(x)a(x)^3 + \lambda_{20}(x)a(x)^2 + \lambda_{10}(x)a(x)]\right\} \\
L_{0,b}(x) &\equiv b'(x)\exp\left\{-[\lambda_{40}b(x)^4 + \lambda_{30}(x)b(x)^3 + \lambda_{20}(x)b(x)^2 + \lambda_{10}(x)b(x)]\right\} \\
L_{1,a}(x) &\equiv a'(x)a(x)\exp\left\{-[\lambda_{40}a(x)^4 + \lambda_{30}(x)a(x)^3 + \lambda_{20}(x)a(x)^2 + \lambda_{10}(x)a(x)]\right\} \\
L_{1,b}(x) &\equiv b'(x)b(x)\exp\left\{-[\lambda_{40}b(x)^4 + \lambda_{30}(x)b(x)^3 + \lambda_{20}(x)b(x)^2 + \lambda_{10}(x)b(x)]\right\} \\
L_{2,a}(x) &\equiv a'(x)a(x)^2\exp\left\{-[\lambda_{40}a(x)^4 + \lambda_{30}(x)a(x)^3 + \lambda_{20}(x)a(x)^2 + \lambda_{10}(x)a(x)]\right\} \\
L_{2,b}(x) &\equiv b'(x)b(x)^2\exp\left\{-[\lambda_{40}b(x)^4 + \lambda_{30}(x)b(x)^3 + \lambda_{20}(x)b(x)^2 + \lambda_{10}(x)b(x)]\right\}
\end{aligned}
$$

The expressions for $F_0'(x)$, $F_1'(x)$, and $F_2'(x)$ are modified as

$$\begin{aligned} F_0'(x) &= \Lambda_{00}(x)F_0(x) + \Lambda_{01}(x)F_1(x) + \Lambda_{02}(x)F_2(x) + C_0(x) \\ F_1'(x) &= \Lambda_{10}(x)F_0(x) + \Lambda_{11}(x)F_1(x) + \Lambda_{12}(x)F_2(x) + C_1(x) \\ F_2'(x) &= \Lambda_{20}(x)F_0(x) + \Lambda_{21}(x)F_1(x) + \Lambda_{22}(x)F_2(x) + C_2(x) \end{aligned}$$

where Λ's denote the corresponding coefficients. C's are defined as follows, which contain the age x and its logwage range $[a(x), b(x)]$.

$$\begin{aligned} C_0(x) &= -\frac{\lambda_{30}'(x)}{4\lambda_{40}}(A_0 - B_0) + L_{0,b}(x) - L_{0,a}(x) \\ C_1(x) &= -\frac{\lambda_{30}'(x)}{4\lambda_{40}}(A_1 - B_1) + \left(\frac{3\lambda_{30}'(x)\lambda_{30}(x)}{16\lambda_{40}^2} - \frac{\lambda_{20}'(x)}{4\lambda_{40}}\right)(A_0 - B_0) + L_{1,b}(x) - L_{1,a}(x) \\ C_2(x) &= -\frac{\lambda_{30}'(x)}{4\lambda_{40}}(A_2 - B_2) + \left(\frac{3\lambda_{30}'(x)\lambda_{30}(x)}{16\lambda_{40}^2} - \frac{\lambda_{20}'(x)}{4\lambda_{40}}\right)(A_1 - B_1) \\ &\quad + \begin{pmatrix} -\frac{9\lambda_{30}'(x)\lambda_{30}^2(x)}{64\lambda_{40}^3} + \frac{\lambda_{30}'(x)\lambda_{20}(x)}{8\lambda_{40}^2} \\ +\frac{3\lambda_{20}'(x)\lambda_{30}(x)}{16\lambda_{40}^2} - \frac{\lambda_{10}'(x)}{4\lambda_{40}} \end{pmatrix}(A_0 - B_0) + L_{2,b}(x) - L_{2,a}(x) \end{aligned}$$

Since

$$\begin{aligned} F_0(x_0 + h) &\approx F_0(x_0) + F_0'(x)h \\ F_1(x_0 + h) &\approx F_1(x_0) + F_1'(x)h \\ F_2(x_0 + h) &\approx F_2(x_0) + F_2'(x)h \end{aligned}$$

for a given initial value x_0 and a small increment h, the functions of x, $F_0(x)$, $F_1(x)$ and $F_2(x)$ can be traced out. At each x value, the logwage(y) limits $a(x)$, $b(x)$ can be obtained through Taylor expansion in the neighborhood of a certain data point x^*,

$$\begin{aligned} a(x) &\approx a(x^*) + a'(x^*)(x - x^*) + \frac{1}{2!}a''(x^*)(x - x^*)^2 + \frac{1}{3!}a'''(x^*)(x - x^*)^3 + \cdots \\ b(x) &\approx b(x^*) + b'(x^*)(x - x^*) + \frac{1}{2!}b''(x^*)(x - x^*)^2 + \frac{1}{3!}b'''(x^*)(x - x^*)^3 + \cdots . \end{aligned}$$

Derivatives of $a(x)$, $b(x)$ can be approximated by

$$\begin{aligned} a'(x) &= \frac{a(x+1) - a(x-1)}{2} \\ a''(x) &= \frac{a(x+2) - 2a(x) + a(x-2)}{4} \\ a'''(x) &= \frac{a(x+3) - 3a(x+1) + 3a(x-1) - a(x-3)}{8}. \end{aligned}$$

It is similar for the upper limit $b(x)$.

A.3: Empirical Study: An Illustration of Calculations

The logwage range is approximated by Taylor expansion. For example, at age 30, logwage range $[a(30), b(30)]$ is estimated twice in the neighborhood of age 29 and 31,

$$\begin{aligned} a(30) &\approx a(29) + a'(29)(30-29) + \frac{1}{2!}a''(29)(30-29)^2 \\ a(30) &\approx a(31) + a'(31)(30-31) + \frac{1}{2!}a''(31)(30-31)^2 \end{aligned}$$

$a(30)$ is the average of these two estimates. $b(30)$ is calculated in the same way. For ages x at the two tails, range $[a(x), b(x)]$ is averaged by Taylor expansions in the neighborhood of several data points. For example, the starting age is 21 in the data set. Logwage range $[a(21), b(21)]$ is estimated by averaging Taylor expansions in the neighborhood of age 22, 23 and 24. The initial values $F_0(21), F_1(21)$ and $F_2(21)$ are computed by integration with the approximated range $[a(21), b(21)]$, i.e. $x_0 = 21$ in the algorithm above. For small enough value h, sequences of $F_0(x), F_1(x)$ and $F_2(x)$ are obtained by the recursive algorithm. Thus, $m(x)$ is computed by taking ratios of $F_1(x)$ and $F_0(x)$ at every age x. Integration is needed only once, at the initial age 21.

Appendix B: Asymptotic Properties of IT Estimators

B.1: Proof of Proposition 1 and (24)

From (21),

$$\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i - \boldsymbol{\mu}\right) = \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}\mathbf{Z}_i - \boldsymbol{\mu}\right).$$

The multivariate characteristic function is

$$\begin{aligned} \varphi_{\sqrt{n}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})}(\mathbf{t}) &= \varphi_{\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}\mathbf{Z}_i-\boldsymbol{\mu}\right)}(\mathbf{t}) \\ &= \varphi_{\mathbf{Z}_1-\boldsymbol{\mu}}\left(\frac{\mathbf{t}}{\sqrt{n}}\right)\varphi_{\mathbf{Z}_2-\boldsymbol{\mu}}\left(\frac{\mathbf{t}}{\sqrt{n}}\right)\cdots\varphi_{\mathbf{Z}_n-\boldsymbol{\mu}}\left(\frac{\mathbf{t}}{\sqrt{n}}\right) \\ &= \left[\varphi_{\mathbf{Z}_1-\boldsymbol{\mu}}\left(\frac{\mathbf{t}}{\sqrt{n}}\right)\right]^n \\ &= E\left[e^{i\left(\frac{\mathbf{t}}{\sqrt{n}}\right)'(\mathbf{Z}_1-\boldsymbol{\mu})}\right], \end{aligned}$$

where $\mathbf{t}$ is a column vector. By Taylor's Theorem,

$$\varphi_{Z_1-\boldsymbol{\mu}}\left(\frac{\mathbf{t}}{\sqrt{n}}\right) = 1 - \frac{1}{2n}\mathbf{t}'\Sigma\mathbf{t} + O\left(\mathbf{t}^3\right),\ \mathbf{t} \longrightarrow \mathbf{0}.$$

Since $e^x = \lim_{n \longrightarrow \infty} \left(1 + \frac{x}{n}\right)^n$,

$$\varphi_{\sqrt{n}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})}(\mathbf{t}) = \left[1 - \frac{1}{2n}\mathbf{t}'\Sigma\mathbf{t} + O\left(\mathbf{t}^3\right)\right]^n \longrightarrow \exp\left(-\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right) \text{ as } n \longrightarrow \infty.$$

Thus, $\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \sim N(\mathbf{0}, \Sigma)$ as given in Proposition 1.

Now to obtain the result in (24), we write the first-order approximation of $\hat{\boldsymbol{\lambda}}$ as

$$\begin{aligned} \hat{\boldsymbol{\lambda}} &= g(\hat{\boldsymbol{\mu}}) \\ &\simeq g(\boldsymbol{\mu}) + \frac{\partial g(\hat{\boldsymbol{\mu}})}{\partial \hat{\boldsymbol{\mu}}^T}|_{\hat{\boldsymbol{\mu}}=\boldsymbol{\mu}} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\ &= g(\boldsymbol{\mu}) + g^{(1)}(\boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}). \end{aligned}$$

$$\begin{aligned} \sqrt{n}\left(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\right) &= \sqrt{n}\left(\left(g(\hat{\boldsymbol{\mu}}) - g(\boldsymbol{\mu})\right)\right) \\ &\simeq \sqrt{n}\left(g^{(1)}(\boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right). \end{aligned}$$

Since $\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \sim N(\mathbf{0}, \Sigma)$ as $n \longrightarrow \infty$,

$$\sqrt{n}\left(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\right) \sim N\left(\mathbf{0}, g^{(1)}(\boldsymbol{\mu})\,\Sigma g^{(1)}(\boldsymbol{\mu})^T\right).$$

The convergence rate of $\hat{\boldsymbol{\lambda}}$ is $n^{1/2}$. This is the result in (24).

B.2: Asymptotic Normality of Maximum Entropy Joint Density, Regression Function and Response Function

Using first-order approximation of the estimated maximum entropy joint density,

$$\begin{aligned} f\left(y, x, \hat{\boldsymbol{\lambda}}\right) &\simeq f(y, x, \boldsymbol{\lambda}) + \frac{\partial f\left(y, x, \hat{\boldsymbol{\lambda}}\right)}{\partial \hat{\boldsymbol{\lambda}}^T}|_{\hat{\boldsymbol{\lambda}}=\boldsymbol{\lambda}} \left(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\right) \\ &= f(y, x, \boldsymbol{\lambda}) + f^{(1)}(y, x, \boldsymbol{\lambda})\left(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\right). \end{aligned}$$

$$\begin{aligned} \sqrt{n}\left(f\left(y, x, \hat{\boldsymbol{\lambda}}\right) - f(y, x, \boldsymbol{\lambda})\right) &= \sqrt{n}\left(f^{(1)}(y, x, \boldsymbol{\lambda})\left(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\right)\right) \\ &= \sqrt{n}\left(f^{(1)}(y, x, \boldsymbol{\lambda})\, g^{(1)}(\boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\right). \end{aligned}$$

Since $\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \sim N(\mathbf{0}, \Sigma)$ as $n \longrightarrow \infty$,

$$\sqrt{n}\left(f\left(y, x, \hat{\boldsymbol{\lambda}}\right) - f(y, x, \boldsymbol{\lambda})\right) \sim N\left(\mathbf{0}, f^{(1)}(y, x, \boldsymbol{\lambda})\, g^{(1)}(\boldsymbol{\mu})\,\Sigma g^{(1)}(\boldsymbol{\mu})^T f^{(1)}(y, x, \boldsymbol{\lambda})^T\right).$$

The convergence rate of $f\left(y, x, \hat{\boldsymbol{\lambda}}\right)$ is $n^{1/2}$.

The maximum entropy regression function of x and $\hat{\boldsymbol{\lambda}}^*$ is approximated by

$$\begin{aligned} m\left(x,\hat{\boldsymbol{\lambda}}^*\right) &\simeq m\left(x,\boldsymbol{\lambda}^*\right)+\frac{\partial m\left(x,\hat{\boldsymbol{\lambda}}^*\right)}{\partial\hat{\boldsymbol{\lambda}}^{*T}}|_{\hat{\boldsymbol{\lambda}}^*=\boldsymbol{\lambda}^*}\left(\hat{\boldsymbol{\lambda}}^*-\boldsymbol{\lambda}^*\right) \\ &= m\left(x,\boldsymbol{\lambda}^*\right)+m^{(1)}\left(x,\boldsymbol{\lambda}^*\right)\left(\hat{\boldsymbol{\lambda}}^*-\boldsymbol{\lambda}^*\right). \end{aligned}$$

$$\begin{aligned} \sqrt{n}\left(m\left(x,\hat{\boldsymbol{\lambda}}^*\right)-m\left(x,\boldsymbol{\lambda}^*\right)\right) &= \sqrt{n}\left(m^{(1)}\left(x,\boldsymbol{\lambda}^*\right)\left(\hat{\boldsymbol{\lambda}}^*-\boldsymbol{\lambda}^*\right)\right) \\ &= \sqrt{n}\left(m^{(1)}\left(x,\boldsymbol{\lambda}^*\right)g^{*(1)}\left(\boldsymbol{\mu}\right)\left(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}\right)\right). \end{aligned}$$

Since $\sqrt{n}\left(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}\right)\sim N\left(\mathbf{0},\Sigma\right)$ as $n\longrightarrow\infty$,

$$\sqrt{n}\left(m\left(x,\hat{\boldsymbol{\lambda}}^*\right)-m\left(x,\boldsymbol{\lambda}^*\right)\right)\sim N\left(\mathbf{0},m^{(1)}\left(x,\boldsymbol{\lambda}^*\right)g^{*(1)}\left(\boldsymbol{\mu}\right)\Sigma g^{*(1)}\left(\boldsymbol{\mu}\right)^T m^{(1)}\left(x,\boldsymbol{\lambda}^*\right)^T\right).$$

The convergence rate of $m\left(x,\hat{\boldsymbol{\lambda}}^*\right)$ is $n^{1/2}$.

Similarly, it can be shown that

$$\sqrt{n}\left(\beta\left(x,\hat{\boldsymbol{\lambda}}^*\right)-\beta\left(x,\boldsymbol{\lambda}^*\right)\right)\sim N\left(\mathbf{0},\beta^{(1)}\left(x,\boldsymbol{\lambda}^*\right)g^{*(1)}\left(\boldsymbol{\mu}\right)\Sigma g^{*(1)}\left(\boldsymbol{\mu}\right)^T\beta^{(1)}\left(x,\boldsymbol{\lambda}^*\right)^T\right).$$

References

Boltzmann, L. (1872). "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen" [Further Studies on the Thermal Equilibrium of Gas Molecules]. *Sitzungsberichte der Akedamie der Wissenschaften, Mathematische-Naturwissenschaftliche Klasse*, 275-370.

Chakrabarty, M., Majumder, A. and Racine, J.S. (2015). "Household Preference Distribution and Welfare Implication: An Application of Multivariate Distributional Statistics." *Journal of Applied Statistics* 42, 2754-2768.

Gibbs, J.W. (1902). "Elementary Principles in Statistical Mechanics." New Haven, CT: Yale University Press.

Golan, A. (1988). "A Discrete Stochastic Model of Economic Production and A Model of Fluctuations in Production–Theory and Empirical Evidence." Ph.D. thesis, University of California, Berkeley.

Golan, A., Judge, G. and Miller, D. (1996). "Maximum Entropy Econometrics: Robust Estimation with Limited Data." John Wiley & Sons.

Golan, A. (2018). "Foundations of Info-Metrics: Modeling, Inference, and Imperfect Information." Oxford University Press, New York.

Harte, J., Zillio, T., Conlisk, E. and Smith, A.B. (2008). "Maximum Entropy and the State-Variable Approach to Macroecology." *Ecology* 89, 2700-2711.

Henderson, D. and Parmeter, C. (2015). "Applied Nonparametric Econometrics." Cambridge University Press.

Inglot, T. and Ledwina, T. (1996). "Asymptotic Optimality of Data-Driven Neyman's Tests for Uniformity." *The Annuals of Statistics* 24, 1982-2019.

Jaynes, E.T. (1957a). "Information Theory and Statistical Mechanics." *Physical Review* 106, 620-630.

Jaynes, E.T. (1957b). "Information Theory and Statistical Mechanics II." *Physical Review* 108, 171-190.

Judge, G and Mittelhammer, R. (2011). "An Information Theoretic Approach to Econometrics." Cambridge University Press.

Kullback, S. and Leibler, R.A. (1951). "On Information and Sufficiency." *The Annals of Mathematical Statistics* 22, 79-86.

Ledwina, T. (1994). "Data-Driven Version of Neyman's Smooth Test of Fit." *Journal of the American Statistical Association* 89, 1000-1005.

Li, Q. and Racine, J. (2007). "Nonparametric Econometrics: Theory and Practice." Princeton University Press.

Maasoumi, E. and Racine, J.S. (2016). "A Solution to Aggregation and an Application to Multidimensional 'Well-Being' Frontiers." *Journal of Econometrics* 191, 374-383.

Mead, L.R. and Papanicolaou, N. (1984). "Maximum Entropy in the Problem of Moments." *Journal of Mathematical Physics* 25, 2404-2417.

Pagan, A. and Ullah, A. (1999). "Nonparametric Econometrics." Cambridge University Press.

Racine, J.S. and Li, K. (2017). "Nonparametric Conditional Quantile Estimation: A Locally Weighted Quantile Kernel Approach." *Journal of Econometrics* 201, 72-94.

Rilstone, P. and Ullah, A. (1989). "Nonparametric Estimation of Response Coefficients." *Communications in Statistics-Theory and Methods* 18, 2615-2627.

Ryu, H. K. (1993). "Maximum Entropy Estimation of Density and Regression Functions." *Journal of Econometrics* 56, 397-440.

Shannon, C.E. (1948). "A Mathematical Theory of Communications." *The Bell System Technical Journal* 27, 379–423, 623–656.

Stone, C. (1990). "Large-Sample Inference for Log-Spline Models." *The Annals of Statistics* 18, 717-741.

Wu, X. (2003). "Calculation of Maximum Entropy Densities with Application to Income Distribution." *Journal of Econometrics* 115, 347-354.

Ullah, A. (1988). "Non-Parametric Estimation of Econometric Functionals." *The Canadian Journal of Economics* 21, 625-658.

Ullah, A. (1996). "Entropy, Divergence and Distance Measures with Econometric Applications." *Journal of Statistical Planning and Inference* 49, 137-162.

Zellner, A. and Highfield, R.A. (1988). "Calculation of Maximum Entropy Distributions and Approximation of Marginalposterior Distributions." *Journal of Econometrics* 37, 195-209.