University of California
Los Angeles

Essays on Discrete Choice Models

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Joonmo Kang

2016

Abstract of the Dissertation

Essays on Discrete Choice Models

by

Joonmo Kang

Doctor of Philosophy in Economics

University of California, Los Angeles, 2016

Professor Rosa Liliana Matzkin, Co-Chair

Professor Jinyong Hahn, Co-Chair

This dissertation consists of three essays divided into chapters. In chapter 1, I analyze the identification of a simultaneous binary response model without nonadditive unobservable random terms, and suggest an estimation method. In particular, the derivatives of structural equations are identified and estimated. The identification relies on a special regressor, which enters the underlying structural equation linearly. All other exogenous variables held constant, variation on this special regressor generates variation on the structural equation which determines the latent endogenous variable in a known way, so we can recover the conditional distribution of the structural equations. The estimator can be constructed using a least-squares method, after replacing the elements of a matrix with kernel density and density derivative estimates. The estimator is shown to be consistent and asymptotically normal.

In chapter 2, I examine the determinants smartphone adoption among the elderly in South Korea. The advent of smartphones has caused a dramatic change in access to information and media, leading to a super-connected world of real-time services. Meanwhile, the constant dissemination of new technologies makes the digital divide multi-layered. In particular, older persons fall far behind the overall population in the access and use of new devices. To understand the technological environment following the introduction of smartphones and other smart mobile devices, I examine individual, household, and regional factors that can

influence the preferences of the elderly with regard to obtaining a smartphone. I find that smartphone ownership among the elderly is mainly determined by personal rather than family characteristics. A smartphone is an individual mobile device that is not shared by other household members, and therefore personal preferences are most integral to the decision. Also, I find that the area where a person lives has a significant effect on the probability of their owning a smartphone. Living in the areas located far from Seoul Metropolitan Area has significant effect in the probability of having a smartphone, which suggest that regional imbalance may play a role in usage of smartphones and other ICT, as well as informatization.

In chapter 3, I analyze the evolution of preferences for brands in digital camera market. A consumer considers the value of a brand, as well as product characteristics when deciding which product to buy. One way to capture this effect is to use brand-specific dummy variables, as in Nevo (2001). However, including brand-specific dummy variables does not fully account for the variation of the unit sales of compact digital cameras, since the preference for digital camera brands evolves over time. Assuming that the brand preference is affected by the advertising expenditure of each brand and the reputation among consumers, I suggest a method to capture the time-varying brand preference under the specification of BLP model.

The dissertation of Joonmo Kang is approved.

Sarah J. Reber

Zhipeng Liao

Jinyong Hahn, Committee Co-Chair

Rosa Liliana Matzkin, Committee Co-Chair

University of California, Los Angeles

2016

To Soo and Lydia

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am particularly indebted to my advisors Rosa Matzkin and Jinyong Hahn for their invaluable support and guidance. I am also greatful to other members of the committee, Zhipeng Liao and Sarah Reber. I have also benefited from conversations with seminar participants at UCLA for their helpful comments and suggestions.

Many colleagues have supported and encouraged me. Keunwoo Kim, Hyo Sang Kim, Keyyong Park, Jinwook Hur and Ruoyao Shi are greatly acknowledged. They shared all the moments of anxiety, stress and joy with me. I also appreciate financial support from Ilju Foundation in Korea during the first four years of the doctoral course.

Finally, for begin there every step of the way, for her unconditional love, and for never giving up on me, I thank my wife, Soo Yun Jun. Each word in this work is a result of all her support and love.

Vita

2009	B.A. (Economics), Seoul National University, *Summa Cum Laude*.
2011	M.A. (Economics), University of California, Los Angeles
2012	C.Phil. (Economics), University of California, Los Angeles

CHAPTER 1

Nonparametric Estimation of Simultaneous Binary Response Model Without Additivity

1.1 Introduction

Imagine a situation where a married woman bears a child. After her parental leave is over, the family encounters a real problem: how do they take care of the baby while they are at work? As the demand for child care is strongly linked to maternal employment, the responsiveness of the labor supply of married women to the price of child care and changes in household income has attracted the attention of many scholars. One may consider three different child care benefit policies. If the government gives a certain amount of money to the household with a young child, it will change the household's overall income. Or, the government could provide more public child care centers. This policy would have the effect of lowering the average price of child care services. The third option is to subsidize only those low-income families with working mothers. Policymakers, as well as researchers, ask which policy is the most appropriate for achieving their policy objectives.

Estimating the parameters of interest and analyzing the counterfactuals to evaluate policies in this situation is not an easy task. Insufficient data exist regarding work hours and wages of the unemployed. Even for those who participate in the labor market, researchers often only observe qualitative responses, such as whether a person continues to work or exits from the workforce. The work participation decision is made by negotiation between the wife and the husband, which causes endogeneity in the model. Moreover, price and income effects are not universal; they depend on both observable and unobservable household characteristics. The standard approach for dealing with this problem is a linear binary choice model, but it offers only an approximation of the true parameters if the underlying decision process is expressed in a nonlinear way.

In this paper, we propose a nonparametric estimator for the binary response model. It is significantly different from other methods for binary response models with endogenous regressors in the literature. First, it does not impose any assumptions regarding the functional form of the structural equations or distributions of the unobserved variables, so it is free from the misspecification error. The unobserved random terms enter the model in a nonadditive way. Unlike linear models with additive errors, each reduced form function in the nonadditive model depends on the value of each of the unobservable variables in the system. Second, the endogeneity is controlled by full simultaneous structural equations in contrast to the triangular models where the endogenous variable has an explicit reduced form relationship. Though the instrument variable and control function approaches are the most common, they have some flaws in a nonparametric framework, as discussed below. To our knowledge, the estimator suggested in the present paper is the first estimator for a nonparametric binary response model that allows for full simultaneity. Third, the estimator relies on the pointwise identification conditions of the structural model, so it does not require the large support condition on the special regressor, which is a potential obstacle to empirical applications. As long as the conditional density is well defined on the neighborhood of a point of interest, we can identify and estimate elements of the structural equation.

The estimation is based on a two-step procedure. The first step consists of estimating functionals of the conditional densities of transformed variables, which can be done by kernel density and density derivative estimations from observable variables. In the second step, we integrate the functionals obtained in the first step and solve a minimization problem with an integrated quadratic loss function. The particular model structure, and the choice of quadratic loss function, enables us to express the estimated feature in a least-squares form. In this sense, the estimator is an extension of the average derivative methods of Stoker (1986) and Powell et al. (1989).

The binary response model plays an important role in applied fields of economics and is therefore the focus of extensive literature. Even when a rich dataset is available, researchers often only observe qualitative responses, and variables underlying the decision process such as subutility and willingness-to-pay are always latent by nature. Though the use of experimental and quasi-experimental methods has been preferred in empirical work in the last decade, analyses of counterfactuals, policy evaluation, and prediction of the evolution of markets require knowledge of the underlying structure and distributions in the economy, which can often only be estimated using structural models.[1] Empirical use of the structural binary response model has included the model of decision for post-secondary vocational school training from Nelson et al. (1978), the analysis of demand for deductibles in private health insurance from Van de Ven and Van Praag (1981), a variation of the labor force participation decision model offered by Blundell and Powell (2004), and the estimation of willingness-to-pay in Lewbel et al. (2011).

The standard method for estimating the linear binary response model involves logit and probit procedures, assuming that the distribution of the unobservable term follows a specific parametric law and that exogenous variables and error terms are statistically independent. However, neither of these assumptions might be appropriate in many empirical applications. First, the structural econometric model provides no guidance as to the functional form of the distribution of the error term, but misspecification generally results in inconsistent estimates. A number of semiparametric estimators have been proposed that do not impose parametric restrictions on the distribution of the error term, including semiparametric least squares (Ichimura, 1993), semiparametric maximum likelihood (Klein et al., 1993), average derivative estimators (Stoker, 1986; Powell et al., 1989), and the semiparametric estimator for discrete regressors (Horowitz and Härdle, 1996).

Second, when the structure arises in a system of simultaneous equations, some components of the independent variables will be endogenous, violating the independence assump-

[1] There had been much debate on the role of structural models in the analyses of economic data. For recent discussion on the topic, see Angrist and Pischke (2010), Nevo and Whinston (2010), Imbens (2010), Deaton (2010), and Heckman and Urzúa (2010).

tion. In regard to parametric specification of the distribution of the error term, Smith and Blundell (1986), Rivers and Vuong (1988), and Newey (1987) proposed estimators that are able to deal with endogeneity. In semiparametric methods, two general approaches exist for overcoming the endogeneity problem. One approach utilizes instrumental variables, which are independent of the unobservable term but functionally dependent on the endogenous variable. Lewbel (2000) used this approach to estimate the index coefficients in a linear binary response model, and Honoré and Lewbel (2002) expanded this to panel data models. The control function approach offers another method for dealing with endogeneity. The general idea of this approach is to use residuals from a reduced form of regressors to account for endogeneity. Blundell and Powell (2004) and Rothe (2009) proposed estimators using the control function approach in a binary response model. Lee (2012) reviewed semiparametric estimation methods for general limited dependent variable models with endogenous regressors.

In addition to the linear model, nonparametric models with nonseparable unobservables have received considerable attention in econometrics. These models are important because they can accommodate general forms of unobserved heterogeneity, but the nonseparability of unobserved heterogeneity complicates the identification and estimation of structural features relative to standard models with additively separable disturbances. Similar to the linear case, various methods can be considered to treat the endogeneity. The instrumental variable approach has been considered by Chernozhukov and Hansen (2005), Chernozhukov et al. (2007), Chen and Pouzo (2012), and Chen et al. (2014). However, estimators based on this approach may suffer from ill-posed inverse problems. Another approach comes from trying to describe the source of simultaneity by specifying the relation between endogenous variables. One way to achieve this is through the control function approach (Florens et al., 2008; Imbens and Newey, 2009; Torgovitsky, 2015). While it is possible to avoid an ill-posed inverse problem, the class of simultaneous structural equations that allows for the use of the control function approach is restrictive. Blundell and Matzkin (2014) presented the conditions required for using the control function approach.

The third approach to achieve the identification and estimation of nonseparable models is allowing for full simultaneity. This approach dates back to Brown (1983) and Roehrig (1988), but it has received much less attention in recent econometrics literature. One complication of this approach is that the change of variables involves the Jacobian of the transformation. Identification results under full simultaneity have been discussed by Matzkin (2007, 2008, 2012), Berry and Haile (2010, 2014), and Chiappori and Komunjer (2009). Matzkin (2012) considered the identification of general limited dependent variable models, but her identification strategy is not constructive. And the only existing estimation procedure for this approach with nonseparable models is proposed by Matzkin (2015).

The structure of the paper is as follows. In the next section, we formally introduce the model and provide preliminary assumptions. In addition, some empirical examples to which the model can be applied are investigated. In Section 3, we present the constructive identification result and the support condition. In Section 4, we describe our estimator and analyze its asymptotic properties. In addition, we compare the asymptotic properties with that of Matzkin (2015) when all the endogenous variables are observed. In Section 5, we report the results of our simulation study. Section 6 concludes the paper.

1.2 The Model

We consider the model

$$\begin{aligned} Y_1^* &= m^1(Y_2, Z, X_1, \epsilon_1) - W, \qquad Y_1 = 1(Y_1^* > 0) \\ Y_2 &= m^2(Y_1^* + W, Z, X_2, \epsilon_2), \end{aligned} \tag{1.1}$$

In this model, (Y_1, Y_2) is a vector of observable endogenous variables and Y_1^* is a latent variable generating binary response Y_1. (Z, X, W) is a vector of observable exogenous variables, and (ϵ_1, ϵ_2) is a vector of unobservable exogenous variables. The observable vector Z does not add any complications if we consider any distribution conditioning on Z, so it will be suppressed in the following discussion for simplicity. The simultaneity occurs between two continuous variables, $Y_1^* + W$ and Y_2, and the dummy variable Y_1 does not directly enter the model. In this sense, this model can be considered as a version of the hybrid model without structural shift offered by Heckman (1978).

One of the crucial assumption in this model is the additive separability between W and m^1. Along with the statistical independence between the observable variables (X, W) and the unobservable ϵ, it provides a mapping between the observable choice probability and the distribution of the latent variable. This special regressor method is a standard strategy in the literature on limited dependent variable models.[2] Note that we require full independence between the exogenous, observed explanatory variables and the unobservables, but the independence between X and W is not necessary.

Suppose that Y_1^* is observable, so that we can directly observe $Y_1^* + W$ rather than Y_1. Denote $B = (B_1, B_2) = (Y_1^* + W, Y_2)$, and rewrite the model as

$$\begin{aligned} B_1 &= m^1(B_2, X_1, \epsilon_1) \\ B_2 &= m^2(B_1, X_2, \epsilon_2) \end{aligned} \tag{1.2}$$

Since the structural equation m^j is unknown and the nonadditive error term ϵ_j is unobservable, we can identify features of the structural equation only up to an invertible transformation (Matzkin, 2007). We may assume that m^j is strictly monotone in ϵ_j as a normalization. Without loss of generality, suppose that m_j is strictly increasing in the last coordinate for $j = 1, 2$. This enables us to find a unique value of ϵ_j for each realization of (b, x). Denote such mapping by $\epsilon_j = r^j(b_1, b_2, x_j)$. Then, the system of indirect structural equations can be expressed as

$$\begin{aligned} \epsilon_1 &= r^1(B_1, B_2, X_1) \\ \epsilon_2 &= r^2(B_1, B_2, X_2) \end{aligned} \tag{1.3}$$

Choose a point $(\bar{b}_1, \bar{b}_2, \bar{x}_1, \bar{x}_2, \bar{\epsilon}_1, \bar{\epsilon}_2)$ on the graph of (1.3), and suppose the function r^1 is twice continuously differentiable. If the derivative of the function r^1 with respect to b_1 is

[2] Lewbel (2010) provides a brief overview of the special regressor method.

invertible at the point, then the implicit function theorem states that there is an open neighborhood U of the fixed point $(\bar{b}_1, \bar{b}_2, \bar{x}_1, \bar{x}_2, \bar{\epsilon}_1, \bar{\epsilon}_2)$ and a unique continuously differentiable function $b_1 = m^1(b_2, x_1, \epsilon_1)$ such that $\epsilon_1 = r^1(m^1(b_2, x_1, \epsilon_1), x_1, \epsilon_1)$ on U. The uniqueness of the function $m^1(b_2, x_1, \epsilon_1)$ guarantees that it matches up with $m^1(b_2, x_1, \epsilon_1)$ on U. Moreover, we can express the derivatives of the function m^1 in terms of r^1,

$$\begin{aligned} \left.\frac{\partial m^1(b_2, x_1, \epsilon_1)}{\partial b_2}\right|_{\epsilon_1 = r^1(b_1, b_2, x_1)} &= -\left[\frac{\partial r^1(b_1,b_2,x_1)}{\partial b_1}\right]^{-1}\left[\frac{\partial r^1(b_1,b_2,x_1)}{\partial b_2}\right] \\ \left.\frac{\partial m^1(b_2, x_1, \epsilon_1)}{\partial x_1}\right|_{\epsilon_1 = r^1(b_1, b_2, x_1)} &= -\left[\frac{\partial r^1(b_1,b_2,x_1)}{\partial b_1}\right]^{-1}\left[\frac{\partial r^1(b_1,b_2,x_1)}{\partial x_1}\right] \end{aligned} \tag{1.4}$$

and similar result holds between m^2 and r^2.

The identification and estimation method in the nonparametric simultaneous equations system rely on the change-of-variable technique. Under appropriate conditions, the conditional density function of B given $X = x$ is given by

$$f_{B|X=x}(b) = f_\epsilon(r(b, x)) \left|\frac{\partial r(b, x)}{\partial b}\right| \tag{1.5}$$

If all the terms in (1.5) are twice continuously differentiable, it can be transformed into a system of linear equations with derivatives of known functions on one side and derivatives of unknown functions on the other side. This allows us to identify some features of the function r. See Matzkin (2008, 2012, 2015) for a detailed discussion on the identification and estimation of the nonparametric simultaneous equation system.

We encounter a problem when directly applying the method above to a binary response model. Since Y_1^* is not observed, we cannot directly observe $B_1 = Y_1^* + W$. However, the existence of the special regressor, W, enables us to recover the conditional probability of $(B_1, B_2) = (Y_1^* + W, Y_2)$ given (X_1, X_2, W) from the data. Suppose that ϵ is independently distributed of (X, W) and the structural equation system allows for the unique reduced-form expression, $B_j = h^j(X_1, X_2, \epsilon_1, \epsilon_2)$, for each j. Conditional on X, (B_1, B_2) is a function of ϵ, so they are independent of W conditional on X. Hence, for any b,

$$F_{B|X=x}(b_1, b_2) = P[B_1 \leq b_1, B_2 \leq b_2 | X = x] = P[B_1 \leq b_1, B_2 \leq b_2 | W = b_1, X = x] \tag{1.6}$$

$$
\begin{aligned}
&= P[Y_1^* + W \le b_1, Y_2 \le b_2 | W = b_1, X = x] \\
&= P[Y_1^* + b_1 \le b_1, Y_2 \le b_2 | W = b_1, X = x] \\
&= P[Y_1^* \le 0, Y_2 \le b_2 | W = b_1, X = x] \\
&= P[Y_1 = 0, Y_2 \le b_2 | W = b_1, X = x]
\end{aligned}
$$

The second equality in the first line comes from the conditional independence of B and W given X. The conditional probability density function of B at a certain value (b_1, b_2) given $X = x$ is obtained by differentiating (1.6) with respect to (b_1, b_2). More precisely,

$$
\begin{aligned}
F_{B|X=x}(b_1, b_2) &= P[B_1 \le b_1, B_2 \le b_2 | X = x] \\
&= P[Y_1^* \le 0, Y_2 \le b_2 | W = b_1, X = x] \\
&= \int_{-\infty}^{b_2} \int_{-\infty}^{0} \frac{f_{Y_1^*, Y_2, W, X}(y_1^*, y_2, b_1, x)}{f_{W,X}(b_1, x)} dy_1^* dy_2 \\
&= \int_{-\infty}^{b_2} \frac{f_{Y_1, Y_2, W, X}(0, y_2, b_1, x)}{f_{W,X}(b_1, x)} dy_2
\end{aligned}
$$

where $f_{Y_1,Y_2,W,X}(y_1, y_2, w, x)$ is the joint pdf of binary random variable Y_1 and continuous random vector (Y_2, W, X). From the definition, we know that $f_{Y_1,Y_2,W,X}(0, y_2, b_1, x) = \int_{-\infty}^{0} f_{Y_1^*,Y_2,W,X}(y_1^*, y_2, b_1, x) dy_1^*$ and $f_{Y_1,Y_2,W,X}(1, y_2, b_1, x) = \int_{0}^{\infty} f_{Y_1^*,Y_2,W,X}(y_1^*, y_2, b_1, x) dy_1^*$. Diffentiating with respect to b_1 and b_2, we obtain

$$
\begin{aligned}
f_{B|X=x}(b_1, b_2) &= \frac{\partial^2}{\partial b_1 \partial b_2} F_{B|X=x}(b_1, b_2) \qquad (1.7) \\
&= \frac{\partial^2}{\partial b_1 \partial b_2} \int_{-\infty}^{b_2} \frac{f_{Y_1^*, Y_2, W, X}(y_1^*, y_2, b_1, x)}{f_{W,X}(b_1, x)} dy_2 \\
&= \frac{\partial}{\partial b_1} \frac{f_{Y_1, Y_2, W, X}(0, b_2, b_1, x)}{f_{W,X}(b_1, x)} \\
&= \frac{f_w \widetilde{f} - f \widetilde{f}_w}{\widetilde{f}^2}
\end{aligned}
$$

where $f_{W,X}(w, x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y_1^*, y_2, w, x) dy_1^* dy_2 = \sum_{y_1=0}^{1} \int_{-\infty}^{\infty} f_{Y_1,Y_2,W,X}(y_1, y_2, w, x) dy_2$ is the joint pdf of (W, X), $f = f_{Y_1,Y_2,W,X}(0, b_2, b_1, x)$,
$f_w = \frac{\partial}{\partial w} f_{Y_1,Y_2,W,X}(y_1, y_2, w, x) \Big|_{y_1=0, y_2=b_2, w=b_1, x=x}$, $\widetilde{f} = f_{W,X}(b_1, x)$, and
$\widetilde{f}_w = \frac{\partial}{\partial w} f_{W,X}(w, x) \Big|_{w=b_1, x=x}$.[3] Since the last expression contains functions related to the

[3]The arguments in the density function and its derivatives are suppressed for notational convenience.

observable variables (Y_1, Y_2, W, X) only, they can be recovered from the data.

Below is a set of conditions required for identification.

Assumption 1 *(ϵ_1, ϵ_2) is distributed independently of (X, W), with an everywhere positive and twice continuously differentiable density f_ϵ.*

Assumption 2 *(X, W) possesses a differentiable density.*

Assumption 3 *For the model, there exist unique twice continuously differentiable functions r as in (1.3). r^j is invertible in b_j and the derivative of r^j with respect to x_j is bounded away from zero. Conditional on (X_1, X_2), the function r is 1-1, onto $\mathbb{R}^2$, and as a function of x, its Jacobian determinant is positive and bounded away from zero.*

Assumption 4 *For the model, there exists unique reduced-form system h such that for each $(x_1, x_2, \epsilon_1, \epsilon_2)$,*

$$\begin{aligned} b_1 &= h_1(x_1, x_2, \epsilon_1, \epsilon_2) \\ b_2 &= h_2(x_1, x_2, \epsilon_1, \epsilon_2) \end{aligned}$$

Remark *The conditional independence of B and W given X is crucial in deriving the conditional density of B given X from (1.6). Without the conditional independence, not only the value of B_1 but also the value of W affect the derivative with respect to b_1, so we cannot recover the density of the (possibly) latent variables $(Y_1^* + W, Y_2)$ from the observables. Assumption 4 guarantees conditional independence under the statistical independence of (X, W) and ϵ.*

Remark *Assumption 4 requires that the endogenous variables are uniquely determined by exogenous variables (X, ϵ), say, the model is coherent and complete. One of the sufficient condition for the existence of the unique reduced-form system is strict monotonicity of m in ϵ and crossing property such that $(\partial m^1(b_2, x_1, \epsilon_1)/\partial b_2)(\partial m^2(b_1, x_2, \epsilon_2)/\partial b_1) < 1$, as discussed in Blundell and Matzkin (2014).*

Remark *As shown in the next section, the identified feature of* r *is the ratios of derivatives,* $r^k_{b_j}(b, x_k)/r^k_{x_k}(b, x_k)$ *at a certain point* (b_1, b_2, x_1, x_2)*. However, since* b_1 *is not directly observed, the special regressor* W *is used to recover the distribution of the structural equation* m*. This approach requires a large support of* W *such that the support of* $s^1(Y_2, X_1, \epsilon_1)$ *is included in the support of* W*. However, since the identification result is pointwise, we need rather a relaxed support condition, in accordance with the choice of the point* (b_1, b_2)*. This condition will be discussed in the next section.*

1.2.1 Empirical Examples

The binary response model is widely used in applied microeconomics to analyze the impact of a policy change. We provide two empirical examples where our model is appropriate. The first example is the labor participation decision of a married woman with children, which was introduced at the beginning of this paper. The second example is the purchase of private health insurance after the implementation of a policy preventing the insurer from prescreening.

Example 1 (Full-time Employment after Childbirth) Parents in the labor force face numerous decisions when balancing their work and home lives. One decision is choosing the type of child care to provide for their children while they work. In the past, mothers were considered to be the caregiving parents, but maternal employment has become the norm rather than the exception. According to the 2010 Census, less than one-third of families with children have a full-time, stay-at-home parent. Hence, child care arrangements and their costs are a significant issue for parents, relatives, care providers, and policymakers. Many researchers have studied the effect of child care costs on the labor force participation of married women, including Heckman (1974), Blau and Robins (1988), and Connelly (1992). A detailed survey of this issue was provided by Blau and Currie (2006).

Parents decide on the optimal time arrangement for work, parental care, and care services. However, full-time employment generally requires a fixed amount of working hours, and parents cannot freely determine office schedules. Hence, a full-time worker with children may consider transferring from full-time to part-time employment. This decision depends on how strongly an individual wants to work, the accessibility of private child care services, and other household income. If the desired work hours of an individual are greater than the hours required by the full-time job, she might keep the job and find a way to take care of her child. Otherwise, she will look for another job with more flexible work hours, typically a part-time job, or exit the labor market. Moreover, the elasticity of the desired work hours varies with different levels of household income. In model (1.1), $Y_1^* + W$ is an individual's desired work hours, where W is exogenously determined work hours in the current full-time employment and Y_2 is other household income, including a spouse's income and various government benefits. With an appropriate choice of individual characteristics X that does not directly affect each equation, we can recover the underlying decision structure for determining the employment schedule, and this can be used to evaluate child care subsidy policies.

Example 2 (Private Health Insurance) During the early- and mid-1990s, many U.S. states introduced regulations on the small and non-group health insurance markets. Such regulations included rating restrictions, guaranteed issue requirements, limits on exclusions for pre-existing conditions, reinsurance requirements, minimum loss ratio requirements, and premium rate restrictions. Rating restrictions limited the insurers' ability to utilize certain predictors of healthcare use in setting premiums while guaranteed issue laws banned the denial of policies. Pre-existing condition exclusion laws and portability laws improved the continuity of access to health insurance for individuals working for small firms. The intent of the regulations was to generate transfers from the healthy to the sick, but such efforts may be fruitless due to adverse selection since the healthy can escape these transfers by reducing or dropping their coverage Rothschild and Stiglitz (1976); Buchmueller and Dinardo (2002).

The model for purchasing private health insurance that we may consider is a modification of the model used by Clemens et al. (2015). The optimal choice of insurance coverage for an individual depends on the shape of her utility function, her level of risk aversion, and her private information on the type, either healthy or sick. Solving the utility maximization problem leads to a relationship between the willingness-to-pay for health insurance and the optimal level of medical expenditure. Under the community-rating restriction, the insurance premium is determined exogenously, not based on individual characteristics. If the willingness-to-pay exceeds the premium proposed by the insurer, she will buy the plan. In model (1.1), $Y_1^* + W$ is an individual's willingness-to-pay for health insurance, where W is the community-rated premium determined in the market and Y_2 is the (expected) medical expenditure of an individual, which partially reflects the expected health cost and hence the type of individual. Since W is determined independently of an individual's characteristics, it would be independent of the unobserved individual characteristic ϵ. With an appropriate choice of socio-economic variable X, we can recover the likelihood that an individual buys a community-rated health insurance policy given her type of expected medical expenditure, and thus we can estimate the degree of adverse selection in a community.

1.3 Identification

Identification of the model comes from the observational equivalence results on nonparametric simultaneous equation models in Matzkin (2015), using the conditional density of transformed variables B given X.

Let us begin with introducing notations used throughout the paper. Let $f_{B|X=x}(b)$ denote the conditional density of the transformed variables $(B_1, B_2) = (Y_1^* + W, Y_2)$, which is derived in (1.7). Denote the derivative of the log of $f_{B|X=x}(b)$ with respect to b and x by $\mathbf{g}_b(b, x) = \left[\frac{\partial \log f_{B|X=x}(b)}{\partial(b_1, b_2)}\right]$ and $\mathbf{g}_x(b, x) = \left[\frac{\partial \log f_{B|X=x}(b)}{\partial(x_1, x_2)}\right]$, respectively. The derivatives of r^g with respect to b_j and x_s will be denoted by $r^g_{b_j}(b, x_g)$ and $r^g_{x_g}(b, x_g)$. The ratio of derivatives of r with respect to b_j and x_s, $r^g_{b_j}(b, x_g)/r^g_{x_g}(b, x_g)$, will be denoted by $\overline{r}^g_{b_j}(b, x_g)$. For fixed x, the Jacobian determinant of r as a function of b, $|\partial r(b, x)/\partial b|$, will be denoted by $|r_b(b, x)|$, and the derivatives of the Jacobian determinant with respect to its arguments $t \in \{b_1, b_2, x_1, x_2\}$

will be denoted by $|r_b(b,x)|_t$. Lastly, define functions $d_{b_g}(b,x)$ for $g=1,2$ by

$$d_{b_g}(b,x)=\frac{|r_b(b,x)|_{b_g}}{|r_b(b,x)|}-\frac{|r_b(b,x)|_{x_1}}{|r_b(b,x)|}\frac{r^1_{b_g}(b,x_1)}{r^1_{x_1}(b,x_1)}-\frac{|r_b(b,x)|_{x_2}}{|r_b(b,x)|}\frac{r^2_{b_g}(b,x_2)}{r^2_{x_2}(b,x_2)}.$$

Let Γ be the class of functions r that satisfy Assumption 3, and let Φ be the class of densities that satisfy Assumption 1. Define observational equivalence within Γ over a subset M in the interior of the support of the vector of transformed variables $(B,X)=(Y_1^*+W,Y_2,X_1,X_2)$. The definition comes from the change-of-variable formula in (1.5).

Definition 1 (Observational Equivalence over M) *Let M denote an open subset of the support of (B,X) such that for all $(b,x)\in M$, $f_{B,X}(b,x)>\delta$ for some positive constant δ. A function $\widetilde{r}\in\Gamma$ is observationally equivalent to $r\in\Gamma$ if there exist densities f_ϵ, $f_{\widetilde{\epsilon}}\in\Phi$ such that for all $(b,x)\in M$,*

$$f_{B|X=x}(b)=f_\epsilon(r(b,x))\left|\frac{\partial r(b,x)}{\partial b}\right|=f_{\widetilde{\epsilon}}(\widetilde{r}(b,x))\left|\frac{\partial\widetilde{r}(b,x)}{\partial b}\right|$$

Suppose (r,f_ϵ) and $(\widetilde{r},f_{\widetilde{\epsilon}})$ in $\Gamma\times\Phi$ generate the same density of (B,X) derived from the observational variables (Y_1,Y_2,X,W). The following theorem provides a characterization of observational equivalence on M in terms of r and $f_{B|X}$ only.

Theorem 1.3.1 *Suppose that (r,f_ϵ) generates $f_{B|X}$ on M and that Assumptions 1-4 are satisfied. A function $\widetilde{r}\in\Gamma$ is observationally equivalent to r on M if and only if for all $(b,x)\in M$,*

$$0 = \left(\overline{r}^1_{b_1}-\widetilde{\overline{r}}^1_{b_1}\right)g_{x_1}+\left(\overline{r}^2_{b_1}-\widetilde{\overline{r}}^2_{b_1}\right)g_{x_2}+\left(d_{b_1}-\widetilde{d}_{b_1}\right) \tag{1.8}$$

$$0 = \left(\overline{r}^1_{b_2}-\widetilde{\overline{r}}^1_{b_2}\right)g_{x_1}+\left(\overline{r}^2_{b_2}-\widetilde{\overline{r}}^2_{b_2}\right)g_{x_2}+\left(d_{b_2}-\widetilde{d}_{b_2}\right) \tag{1.9}$$

Proof. The theorem is analogous to Theorem 2.1 in Matzkin (2015) if we let G=2 and substitute $f_{Y|X}$ in her paper with $f_{B|X}$. □ □

Again, the change-of-variable formula in (1.5) can be used to develop constructive identification results for features of r. Taking logs and differentiating (1.5) with respect to b_j

gives

$$\frac{\partial \log f_{B|X=x}(b)}{\partial b_j} = \sum_{g=1}^{2} \frac{\partial \log f_\epsilon(r(b,x))}{\partial \epsilon_g} r^g_{b_j}(b,x_g) + \frac{|r_b(b,x)|_{b_j}}{|r_b(b,x)|} \tag{1.10}$$

and taking logs and differentiating (1.5) with respect to x_g gives

$$\frac{\partial \log f_{B|X=x}(b)}{\partial x_g} = \frac{\partial \log f_\epsilon(r(b,x))}{\partial \epsilon_g} r^g_{x_g}(b,x_g) + \frac{|r_b(b,x)|_{x_g}}{|r_b(b,x)|} \tag{1.11}$$

Solving for $\partial \log f_\epsilon(r)/\partial \epsilon_g$ in (1.11) and plugging the result into (1.10), we obtain

$$\begin{aligned} \frac{\partial \log f_{B|X=x}(b)}{\partial b_j} =& \sum_{g=1}^{2} \frac{\partial \log f_{B|X=x}(b)}{\partial x_g} \frac{r^g_{b_j}(b,x_g)}{r^g_{x_g}(b,x_g)} \\ &+ \frac{|r_b(b,x)|_{b_j}}{|r_b(b,x)|} - \frac{|r_b(b,x)|_{x_1}}{|r_b(b,x)|} \frac{r^1_{b_j}(b,x_1)}{r^1_{x_1}(b,x_1)} - \frac{|r_b(b,x)|_{x_2}}{|r_b(b,x)|} \frac{r^2_{b_g}(b,x_2)}{r^2_{x_2}(b,x_2)} \\ =& \sum_{g=1}^{2} \frac{\partial \log f_{B|X=x}(b)}{\partial x_g} \frac{r^g_{b_j}(b,x_g)}{r^g_{x_g}(b,x_g)} + d_{b_j}(b,x) \end{aligned} \tag{1.12}$$

Equation (1.12) implies that $(\overline{r}, d)$ satisfies the following system of equations

$$\begin{aligned} g_{b_1} &= \overline{r}^1_{b_1} g_{x_1} + \overline{r}^2_{b_1} g_{x_2} + d_{b_1} \\ g_{b_2} &= \overline{r}^1_{b_2} g_{x_1} + \overline{r}^2_{b_2} g_{x_2} + d_{b_2} \end{aligned} \tag{1.13}$$

For a certain point of (b_1, b_2, x_1, x_2), the derivatives of the log of conditional density $\mathbf{g}_b$ and $\mathbf{g}_x$ are either observable or estimable, while the ratios of derivatives of indirect structural function $\overline{r}^g_{b_j}$ and the terms d_{b_j} are unknown. We may interpret (1.13) as a linear equation system of 2 equations and 6 unknowns. Hence, we cannot solve (1.13) for the unknowns without further assumptions. Roughly speaking, if three distinct points of (b_1, b_2, x_1, x_2) give different values for $\mathbf{g}_b$ and $\mathbf{g}_x$ but $(\overline{r}, d)$ are the same over these three points, we can solve for $(\overline{r}, d)$ on such points. This leads to additional restrictions on the form of indirect equation r and the values of $\mathbf{g}_x$.

Assumption 5 *For $j = 1, 2$ the inverse function r^j is such that for some function $s^j : \mathbb{R}^2 \to \mathbb{R}$ and all (b, x_j), $r^j(b, x_j) = s^j(b) + x_j$.*

The identification analysis says that we can identify only the ratios of derivatives, $\overline{r}$, which is equivalent to the statement that r is identified up to an invertible transformation.

The restrction of Γ satisfying Assumptions 3 and 5 guarantees that no two functions in Γ are invertible transformations of each other.[4]

At this point, we need to discuss the choice of b. All the results above assume that we analyze the identification of the model at a fixed point (b_1, b_2, x_1, x_2) using log of the conditional density of transformed variable B given X. If the density vanishes around that point, it is invalid to take the log of the density. This suggests a restriction on the choice of b.

Assumption 6 *The following support conditions hold:*

(a) Let $b = (b_1, b_2)$ be given. There exists a convex and compact subset $\overline{M}_b$ that is strictly included in the support of (W, Y_2) such that b is an interior point of $\overline{M}_b$.

(b) For all $(b_1', b_2') \in \overline{M}_b$, there exists a convex and compact subset $\overline{M}_x$ in the interior of supp$(X|W = b_1', Y_2 = b_2')$ which does not depend on the choice of (b_1', b_2')

(c) For all $(b_1', b_2') \in \overline{M}_b$ and $x \in \overline{M}_x$, $0 < P(Y_1 = 0, Y_2 \leq b_2' | W = b_1', X = x) < 1$.

Assumption 6 connects the support condition on M of transformed variables with the support of the observable variables. Condition (a) and (b) guarantee that the conditional density function $f_{B|X=x}(b)$ is well-defined, and condition (c) enables us to obtain non-vanishing density, and thus the differentiation in equations (1.10)–(1.12) is valid. Note that Assumption 6 does not require the full support condition on W.

The following proposition summarizes the consequence of the two assumptions,

Proposition 1.3.2 *Let b be a fixed value satisfying Assumption 6. When Assumptions 3 and 5 are satisfied, $(\overline{r}, d)$ is constant over the set $\overline{M} = \{(b, x_1, x_2) | (x_1, x_2) \in \overline{M}_x\}$.*

The last assumption is on the value of $\mathbf{g}_x$. Due to the additive separable nature of the model, the restriction on $\mathbf{g}_x$ is equivalent to the restriction imposed on the density of ϵ,

[4]For more discussion on the normalization of the class of structural functions where no two functions are invertible transformations of each other, see Matzkin (2008).

through (1.11). Observe that $r^g_{x_g}(b, x_g) = 1$ and $|r_b(b,x)|_{x_g} = 0$. Hence, we obtain a relation between $\mathbf{g}_x$ and f_ϵ:

$$g_{x_s} = \frac{\partial \log f_{B|X=x}(b)}{\partial x_s} = \frac{\partial \log f_\epsilon(r(b,x))}{\partial \epsilon_s} \tag{1.14}$$

As stated above, we need three distinct points of $t^{(1)}, t^{(2)}, t^{(3)}$ in $\overline{M}$, which makes the equation (1.13) solvable. Denote the values of g_{x_s} evaluated at $t^{(j)}$ by $g^{(j)}_{x_s}$ and the values of g_{y_g} evaluated at $t^{(j)}$ by $g^{(j)}_{y_g}$. The condition is given that the matrix

$$\begin{pmatrix} g^{(1)}_{x_1} & g^{(1)}_{x_2} & 1 \\ g^{(2)}_{x_1} & g^{(2)}_{x_2} & 1 \\ g^{(3)}_{x_1} & g^{(3)}_{x_2} & 1 \end{pmatrix} \tag{1.15}$$

is invertible. Combining with (1.14), the condition can be satisfied under the following two assumptions.

Assumption 7 *There exist three values $\epsilon^{(1)}, \epsilon^{(2)}, \epsilon^{(3)}$ of ϵ, not necessarily known, such that the matrix*

$$\begin{pmatrix} \frac{\partial \log f_\epsilon(\epsilon^{(1)})}{\partial \epsilon_1} & \frac{\partial \log f_\epsilon(\epsilon^{(1)})}{\partial \epsilon_2} & 1 \\ \frac{\partial \log f_\epsilon(\epsilon^{(2)})}{\partial \epsilon_1} & \frac{\partial \log f_\epsilon(\epsilon^{(2)})}{\partial \epsilon_2} & 1 \\ \frac{\partial \log f_\epsilon(\epsilon^{(3)})}{\partial \epsilon_1} & \frac{\partial \log f_\epsilon(\epsilon^{(3)})}{\partial \epsilon_2} & 1 \end{pmatrix} \tag{1.16}$$

is invertible.

Assumption 8 *There exist three points $t^{(1)}, t^{(2)}, t^{(3)}$ in the set $\overline{M}$, not necessarily known, where $(\overline{r}, d)$ is constant, such that for each $j = 1, 2, 3$, $\epsilon^{(j)} = r(t^{(j)})$, where $\epsilon^{(j)}$ is as in Assumption 7.*

Theorem 1.3.3 *Suppose that Assumptions 1-8 are satisfied on $\overline{M}$. Then, matrix (1.15) is invertible, and thus $(\overline{r}, d)$ are identified on $\overline{M}$.*

Remark *We may consider an alternative set of functions as a normalization in Assumption 5. One possible choice is that function r is homogeneous of degree one. This normalization is equivalent to assuming that the structural function m is homogeneous of degree one, and the identification occurs on a ray of (b_1, b_2, x_1, x_2) through the origin.*

1.4 Average Derivative Estimator and Asymptotic Properties

Next, we develop a least-squares type estimator for $\overline{r}$. Proposition 1.3.2 says that for a given b satisfying Assumption 6, (1.13) holds for all values of $(\mathbf{g}_b, \mathbf{g}_x)$ over any subset of $\overline{M}$. Let $\overline{M}_x$ be a compact subset of the support of X given $W = b_1$ such that $\{b\} \times \overline{M}_x \subset \overline{M}$. Since $(\overline{r}, d)$ is constant over $\overline{M}$, the choice of an integrated quadratic loss function with some weight function $\mu(x)$ gives a characterization of $(\overline{r}, d)$ as the unique solution of a minimization problem. Choose a continuously differentiable, non-negative weight function $\mu(x)$ defined on $\mathbb{R}^2$ such that $\int_{\overline{M}_x} \mu(x)dx = 1$ and $\mu(x)$ vanishes outside $\overline{M}_x$. For any vector $(\widetilde{\overline{r}}, \widetilde{d})$ generated from a function $\widetilde{r}$ satisfying the same assumptions as the true function r, define the loss function by

$$L(\widetilde{\overline{r}}, \widetilde{d}) = \int_{\overline{M}_x} \left[\sum_{j=1}^{2} \left(g_{b_j} - g_{x_1} \widetilde{\overline{r}}^1_{b_j} - g_{x_2} \widetilde{\overline{r}}^2_{b_j} - d_{b_j} \right)^2 \right] \mu(x)dx \tag{1.17}$$

and consider the minimization problem $\min_{(\widetilde{\overline{r}}, \widetilde{d})} L(\widetilde{\overline{r}}, \widetilde{d})$. It is clear that $L\left(\widetilde{\overline{r}}, \widetilde{d}\right) \geq 0$ and $L(\overline{r}, d) = 0$, so $(\overline{r}, d)$ is a minimizer of the criterion function. And, when $\mu(x)$ is strictly positive at $t^{(1)}, t^{(2)}, t^{(3)}$ defined in Assumption 8, $(\overline{r}, d)$ is the unique minimizer of $L(\cdot)$. The quadratic form of the loss function allows us to use the least-squares method. Especially, we can use partial linear regression to ignore the nuisance term d.

To explain the estimation method, we need to define additional notation regarding the weighted average of related terms. The weighted average of g_{b_j} and g_{x_s} over $\overline{M}_x$ will be denoted by

$$\begin{aligned}
\int_{\overline{M}_x} g_{b_j} &= \int_{\overline{M}_x} g_{b_j}(b, x)dx = \int_{\overline{M}_x} \frac{\partial \log f_{B|X=x}(b)}{\partial b_j} \mu(x)dx \\
\int_{\overline{M}_x} g_{x_s} &= \int_{\overline{M}_x} g_{x_s}(b, x)dx = \int_{\overline{M}_x} \frac{\partial \log f_{B|X=x}(b)}{\partial x_s} \mu(x)dx
\end{aligned}$$

The averaged centered cross products between g_{b_j} and g_{x_s} and between g_{x_j} and g_{x_s} will be denoted by

$$\begin{aligned}
T_{b_j, x_s} &= \int_{\overline{M}_x} (g_{b_j}(b, x) - \int_{\overline{M}_x} g_{b_j})(g_{x_s} - \int_{\overline{M}_x} g_{x_s}) \mu(x)dx \\
T_{x_j, x_s} &= \int_{\overline{M}_x} (g_{x_j}(b, x) - \int_{\overline{M}_x} g_{x_j})(g_{x_s} - \int_{\overline{M}_x} g_{x_s}) \mu(x)dx
\end{aligned}$$

Note that we focus on the estimation of $\overline{r}$ by factoring out intercept d. The matrices of centered cross products T_{XX} and T_{BX} will be defined by

$$T_{XX} = \begin{pmatrix} T_{x_1,x_2} & T_{x_2,x_1} \\ T_{x_1,x_2} & T_{x_2,x_2} \end{pmatrix} \text{ and } T_{BX} = \begin{pmatrix} T_{b_1,x_1} & T_{b_2,x_1} \\ T_{b_1,x_2} & T_{b_2,x_2} \end{pmatrix}.$$

The matrix of ratios of derivatives will be defined by

$$R(\overline{r}) = \begin{pmatrix} \overline{r}^1_{b_1} & \overline{r}^1_{b_2} \\ \overline{r}^2_{b_1} & \overline{r}^2_{b_2} \end{pmatrix}$$

The solution of the First Order Conditions for $\overline{r}$ is expressed as $T_{XX}R(\overline{r}) = T_{BX}$. Since $(\overline{r}, d)$ is the unique minimizer, the matrix T_{XX} must be invertible. Hence, $R(\overline{r})$ will be given by

$$R(\overline{r}) = T^{-1}_{XX} T_{BX} \tag{1.18}$$

The following lemma shows that $(\overline{r}, d)$ is the unique solution of the minimization problem with the loss function $L(\widetilde{\overline{r}}, \widetilde{d})$.

Lemma 1.4.1 (Theorem 2.3 in Matzkin (2015)) *Let b be a point satisfying Assumption 6 and let the compact set $\overline{M}_x$ defined as above. Suppose that Assumptions 1–8 are satisfied and that $\mu(x)$ is strictly positive at least at one set of points $w^{(1)}, w^{(2)}, w^{(3)}$ defined in Assumption 8. Then, $(\overline{r}, d)$ is the unique minimizer of*

$$L\left(\widetilde{\overline{r}}, \widetilde{d}\right) = \int_{\overline{M}_x} \left[\sum_{j=1}^{2} \left(g_{b_j} - \widetilde{\overline{r}}^1_{b_j} g_{x_1} - \widetilde{\overline{r}}^2_{b_j} g_{x_2} - \widetilde{d}_{b_j} \right)^2 \right] \mu(x) dx$$

and $R(\overline{r})$ is given by (1.18).

The estimator for $R(\overline{r})$ can be obtained by replacing $f_{B|X=x}(b)$ in all the expressions with a nonparametric estimator $\widehat{f}_{B|X=x}(b)$. Hereafter, the terms with hat, $\widehat{(\cdot)}$, mean that the conditional density or density derivatives are replaced by their sample analogues. Then, the estimator for the matrix of ratios of derivatives $R(\overline{r})$ is defined as

$$\widehat{R(\overline{r})} = \widehat{T_{XX}}^{-1} \widehat{T_{BX}}$$

One of the nonparametric methods widely used to estimate $f_{B|X=x}(b)$ is the kernel method. Let $\{Y_1^i, Y_2^i, W^i, X^i\}_{i=1}^N$ denote N iid observations generated from the model. The kernel estimators for the density and density derivatives are

$$\begin{aligned}
\widehat{f}_{Y_1,Y_2,W,X}(y_1,y_2,w,x) &= (N\sigma_N^4)^{-1}\sum_{i=1}^N 1(Y_1^i = y_1)K\left(\frac{Y_2^i - y_2}{\sigma_N}, \frac{W^i - w}{\sigma_N}, \frac{X^i - x}{\sigma_N}\right) \quad (1.19)\\
\partial_w \widehat{f}_{Y_1,Y_2,W,X}(y_1,y_2,w,x) &= (N\sigma_N^5)^{-1}\sum_{i=1}^N 1(Y_1^i = y_1)K_w\left(\frac{Y_2^i - y_2}{\sigma_N}, \frac{W^i - w}{\sigma_N}, \frac{X^i - x}{\sigma_N}\right)\\
\widehat{\widetilde{f}}_{W,X}(w,x) &= (N\sigma_N^3)^{-1}\sum_{i=1}^N K\left(\frac{W^i - w}{\sigma_N}, \frac{X^i - x}{\sigma_N}\right)\\
\partial_w \widehat{\widetilde{f}}_{W,X}(w,x) &= (N\sigma_N^4)^{-1}\sum_{i=1}^N K_w\left(\frac{W^i - w}{\sigma_N}, \frac{X^i - x}{\sigma_N}\right)
\end{aligned}$$

where K is a kernel function and σ_N is a bandwidth, $\partial_w g = \frac{\partial g}{\partial w}$ and $K_w(\cdot) = \frac{\partial K(\cdot)}{\partial w}$. For simple notation, denote the estimates evaluated at $(y_1, y_2, w, x) = (0, b_2, b_1, x)$ by $\widehat{f} = \widehat{f}_{Y_1,Y_2,W,X}(0, b_2, b_1, x)$, $\widehat{f}_w = \partial_w \widehat{f}_{Y_1,Y_2,W,X}(0, b_2, b_1, x)$, $\widehat{\widetilde{f}} = \widehat{\widetilde{f}}_{W,X}(b_1, x)$, and $\widehat{\widetilde{f}}_w = \partial_w \widehat{\widetilde{f}}_{W,X}(b_2, x)$, respectively. From (1.7), we can estimate the conditional density of the latent variable B given X by plugging estimates (1.19) into (1.7),

$$\widehat{f}_{B|X=x}(b) = \frac{\widehat{f}_w\widehat{\widetilde{f}} - \widehat{f}\widehat{\widetilde{f}}_w}{\widehat{\widetilde{f}}^2}$$

The element in the k-th row and i-th column of our estimator for T_{XX}, $\widehat{T}_{XX}$ is

$$\int_{\overline{M}_x} (\widehat{g}_{x_i}(b,x) - \int_{\overline{M}_x} \widehat{g}_{x_i})(\widehat{g}_{x_k} - \int_{\overline{M}_x} \widehat{g}_{x_k})\mu(x)dx$$

where

$$\widehat{g}_{x_k}(b,x) = \frac{\partial \log \widehat{f}_{B|X=x}(b)}{\partial x_k} \text{ and } \int_{\overline{M}_x} \widehat{g}_{x_k} = \int_{\overline{M}_x} \frac{\partial \log \widehat{f}_{B|X=x}(b)}{\partial x_k}\mu(x)dx$$

Similarly, the element in the k-th row and i-th column of our estimator for T_{BX}, $\widehat{T}_{BX}$ is

$$\int_{\overline{M}_x} (\widehat{g}_{b_i}(b,x) - \int_{\overline{M}_x} \widehat{g}_{b_i})(\widehat{g}_{x_k} - \int_{\overline{M}_x} \widehat{g}_{x_k})\mu(x)dx$$

where

$$\widehat{g}_{b_i}(b,x) = \frac{\partial \log \widehat{f}_{B|X=x}(b)}{\partial b_i} \text{ and } \int_{\overline{M}_x} \widehat{g}_{b_i} = \int_{\overline{M}_x} \frac{\partial \log \widehat{f}_{B|X=x}(b)}{\partial b_i}\mu(x)dx$$

1.4.1 Asymptotic Properties of the Estimator

To describe the asymptotic properties of the estimator, we need additional assumptions that control the behavior of the kernel estimator. For $b = (b_1, b_2)$ satisfying Assumption 6, let $\overline{M}_b$ be a convex and compact subset of the support of (W, Y_2) such that (b_1, b_2) is an interior point of $\overline{M}_b$, and $\overline{M}_x$ be a convex and compact set that is in the interior of the support of X given $W = b_1$.

Assumption 9 *The density $f_{Y_1,Y_2,W,X}$ generated by (f_ϵ, r) is bounded and continuously differentiable of order $d \geq s+2$, where s denotes the order of the kernel function. Moreover, there exists $\delta > 0$ such that for all $(y_2, w, x) \in \overline{M}_b \times \overline{M}_x$, $\widetilde{f}(w,x) > \delta$, $f(0, y_2, w, x) > \delta$, and $f_w(0, y_2, w, x)\widetilde{f}(w,x) - f(0, y_2, w, x)\widetilde{f}_w(w,x) > \delta$.*[5]

Assumption 10 *The kernel function K is of order s, where $s + 2 \leq d$. It attains the value zero outside a compact set, integrates to 1, is differentiable of order Δ, and its derivatives of order Δ are Lipschitz, where $\Delta \geq 3$.*

Assumption 11 *The sequence of bandwidths σ_N is such that $\sigma_N \to 0$, $N\sigma_N^6 \to \infty$, $\sqrt{N}\sigma_N^{3+s} \to 0$, $[N\sigma_N^8/\log(N)] \to \infty$, and $\sqrt{N\sigma^6}[\sqrt{\log(N)/N\sigma_N^8} + \sigma_N^s]^2 \to 0$.*

Assumption 12 *The weight function $\mu(x)$ is bounded and continuously differentiable, strictly positive over $\overline{M}_x$, with values and derivatives vanishing on the boundary and on the complement of $\overline{M}_x$. The set $\{b\} \times \overline{M}_x$ contains at least one set of points $t^{(1)}, t^{(2)}, t^{(3)}$ satisfying Assumption 8.*

Assumptions 9 and 10 are standard conditions in kernel-based estimation literature. With an appropriate choice of a norm, these conditions enable us to use the functional delta method in order to derive the asymptotic variance of the functional of the kernel density estimates. The restrictions on the kernel in Assumption 10 requires that the kernel be a higher order kernel of order s for bias reduction. It will be used here to guarantee that

[5] *The last expression is the numerator of the conditional density function B given X. By Assumption 6, the value of the conditional density function is bigger than 0, so we can find δ satisfying the inequality.*

the limiting distribution of the estimators are centered at the true value. Assumption 11 is for deriving uniform convergence rates for derivatives of kernel estimators. Assumption 12 guarantees that the estimator is well-defined and the unique minimizer of the previously discussed minimization problem.

To describe the asymptotic distribution of the estimator, we need slightly different notation by vectorization. Let $TT_{BX} = vec(T_{BX})$, $TT_{XX} = I_2 \otimes T_{XX}$ and $rr = vec\left(R(\overline{r})\right)$, where $vec(\cdot)$ is the vectorization operator and $\otimes$ is Kronecker product. $\widehat{TT}_{BX}$, $\widehat{TT}_{XX}$ and $\widehat{rr}$ denote the corresponding estimator for each term, respectively.

Theorem 1.4.2 *Under Assumptions 1–12, it holds that*

$$\sqrt{N\sigma_N^6}(\widehat{TT}_{BX} - TT_{BX}) \xrightarrow{d} N(0, V)$$

where

$$V = \int_{\overline{M}_x} W(x)\ \widetilde{KK}\ W(x)' f(y_2, w, x) dx$$

$$W(x) = \begin{bmatrix} \frac{\mu(x)\widetilde{f}}{f_w\widetilde{f} - f\widetilde{f}_w}\left(g_{x_1} - \int_{\overline{M}_x} g_{x_1}\mu(x)dx\right) & 0 \\ \frac{\mu(x)\widetilde{f}}{f_w\widetilde{f} - f\widetilde{f}_w}\left(g_{x_2} - \int_{\overline{M}_x} g_{x_2}\mu(x)dx\right) & 0 \\ 0 & \frac{\mu(x)\widetilde{f}}{f_w\widetilde{f} - f\widetilde{f}_w}\left(g_{x_1} - \int_{\overline{M}_x} g_{x_1}\mu(x)dx\right) \\ 0 & \frac{\mu(x)\widetilde{f}}{f_w\widetilde{f} - f\widetilde{f}_w}\left(g_{x_2} - \int_{\overline{M}_x} g_{x_2}\mu(x)dx\right) \end{bmatrix}$$

$$\widetilde{KK} = \int \widetilde{K}\widetilde{K}' d(y_2, w)$$

$$\widetilde{K} = \int \begin{pmatrix} \frac{\partial^2 K(y_2,w,x)}{\partial w^2} \\ \frac{\partial^2 K(y_2,w,x)}{\partial w \partial y_2} \end{pmatrix} dx$$

As a result, the asymptotic distribution of the average derivative estimator is

$$\sqrt{Nh^6}(\widehat{rr} - rr) \xrightarrow{d} N(0, (TT_{XX})^{-1} V (TT_{XX})^{-1})$$

The estimator for the asymptotic variance can be constructed in an apparent way by substituting components of the asymptotic variance V with the corresponding estimates. The estimator V is given by

$$\widehat{V} = \int_{\overline{M}_x} \widehat{W}(x)\widetilde{KK}\widehat{W}(x)'\widehat{f}(y_2, w, x)dx \tag{1.20}$$

where

$$\widehat{W}(x) = \begin{bmatrix} \frac{\mu(x)\widehat{\widetilde{f}}}{\widehat{f}_w\widehat{\widetilde{f}} - f\widehat{\widetilde{f}}_w}\left(\widehat{g}_{x_1} - \int_{\overline{M}_x} \widehat{g}_{x_1}\mu(x)dx\right) & 0 \\ \frac{\mu(x)\widehat{\widetilde{f}}}{\widehat{f}_w\widehat{\widetilde{f}} - f\widehat{\widetilde{f}}_w}\left(\widehat{g}_{x_2} - \int_{\overline{M}_x} \widehat{g}_{x_2}\mu(x)dx\right) & 0 \\ 0 & \frac{\mu(x)\widehat{\widetilde{f}}}{\widehat{f}_w\widehat{\widetilde{f}} - f\widehat{\widetilde{f}}_w}\left(\widehat{g}_{x_1} - \int_{\overline{M}_x} \widehat{g}_{x_1}\mu(x)dx\right) \\ 0 & \frac{\mu(x)\widehat{\widetilde{f}}}{\widehat{f}_w\widehat{\widetilde{f}} - f\widehat{\widetilde{f}}_w}\left(\widehat{g}_{x_2} - \int_{\overline{M}_x} \widehat{g}_{x_2}\mu(x)dx\right) \end{bmatrix}$$

and $\widetilde{KK}$ is as defined above.

Theorem 1.4.3 *Under Assumptions 1–12, the asymptotic variance estimator $\widehat{V}$ converges in probability to V. Moreover, $(\widehat{TT_{XX}})^{-1}\widehat{V}(\widehat{TT_{XX}})^{-1}$ is a consistent estimator for $(TT_{XX})^{-1}V(TT_{XX})^{-1}$.*

Remark *It is meaningful to compare the asymptotic property of the estimator with the result in Matzkin (2015). Assuming $(B_1, B_2) = (Y_1^* + W, Y_2)$ is observable, the asymptotic variance of her estimator is given by*

$$\sqrt{N\sigma_N^4}(\widehat{TT}_{BX} - TT_{BX}) \xrightarrow{d} N(0, \widetilde{V})$$

$$\widetilde{V} = \int_{\overline{M}_x} \widetilde{W}(x)\ \widetilde{KK}\ \widetilde{W}(x)' f(b, x)dx$$

$$\widetilde{W}(x) = \begin{bmatrix} \frac{\mu(x)}{f(b,x)}\left(g_{x_1} - \int_{\overline{M}_x} g_{x_1}\mu(x)dx\right) & 0 \\ \frac{\mu(x)}{f(b,x)}\left(g_{x_2} - \int_{\overline{M}_x} g_{x_2}\mu(x)dx\right) & 0 \\ 0 & \frac{\mu(x)}{f(b,x)}\left(g_{x_1} - \int_{\overline{M}_x} g_{x_1}\mu(x)dx\right) \\ 0 & \frac{\mu(x)}{f(b,x)}\left(g_{x_2} - \int_{\overline{M}_x} g_{x_2}\mu(x)dx\right) \end{bmatrix}$$

$$\widetilde{KK} = \int \widetilde{K}\widetilde{K}' d(b_1, b_2)$$

$$\widetilde{K} = \int \begin{pmatrix} \frac{\partial K(b_1,b_2,x)}{\partial b_1} \\ \frac{\partial K(b_1,b_2,x)}{\partial b_2} \end{pmatrix} dx$$

The differences of the asymptotic distribution come from recovering the conditional density of B given X. When the endogenous variables are observable as in Matzkin (2015), we can estimate the conditional density directly from the variation of B. On the other hand, when the endogenous variables are latent, we use the variation of W in order to estimate *the conditional density of the latent variables, so the first-order derivative of the density of observable variables are necessary. This affects the asymptotic properties in two ways. First, It slows down the rate of the convergence by the factor of the bandwidth. And second, since the order of the derivative of the kernel function increases, the roughness of the second-order derivative of the kernel function rather than the first-order derivative is used in the asymptotic variance. It increases the asymptotic variance.*

1.5 Simulation Study

In this section, we report the results of simulation experiments with two model specifications. The first design is from a linear model:

$$\begin{aligned} y_1^* &= 0.75y_2 - x_1 + \epsilon_1 - w \\ y_2 &= -0.5(y_1^* + w) - x_2 + \epsilon_2 \end{aligned}$$

The transformed model is

$$\begin{aligned} b_1 &= 0.75b_2 - x_1 + \epsilon_1 \\ b_2 &= -0.5b_1 - x_2 + \epsilon_2 \end{aligned}$$

and the inverse structural function is given by

$$\begin{aligned} \epsilon_1 &= r^1(b_1, b_2, x_1) = s^1(b_1, b_2) + x_1 = b_1 - 0.75b_2 + x_1 \\ \epsilon_2 &= r^2(b_1, b_2, x_2) = s^2(b_1, b_2) + x_2 = 0.5b_1 + b_2 + x_2 \end{aligned}$$

For the second design, we consider a nonlinear model:

$$\begin{aligned} y_1^* &= 10[1 + \exp(-2(y_2 - 5 - x_1 + \epsilon_1))]^{-1} - w \\ y_2^* &= 4.5 + 0.1(y_1^* + w) - x_2 + \epsilon_2 \end{aligned}$$

The transformed model is

$$b_1 = 10[1 + \exp(2(b_2 - 5 - x_1 + \epsilon_1))]^{-1}$$

$$b_2 = 4.5 + 0.1 b_1 - x_2 + \epsilon_2$$

and the corresponding inverse structural function is given by

$$\begin{aligned}
\epsilon_1 &= r^1(b_1, b_2, x_1) = s^1(b_1, b_2) + x_1 = 5 = 0.5 \log\left(\frac{10}{b_1} - 1\right) - b_2 + x_1 \\
\epsilon_2 &= r^2(b_1, b_2, x_2) = s^2(b_1, b_2) + x_1 = -0.1 b_1 + b_2 - 4.5 + x_2
\end{aligned}$$

The simulation samples are drawn from the distribution

$$X \sim N(0, 2I), \quad \epsilon \sim N(0, I)$$

and for the exogenous special regressor W, we use standard normal distribution for the linear model and $N(5,1)$ for the nonlinear case. The mean of W in the nonlinear case is chosen to guarantee that both $y_1 = 0$ and $y_1 = 1$ are observed enough.

The estimator involves choosing many parameters: the choice of bandwidth, the order of kernel function, the region of integration $\overline{M}_x$, and the weight function $\mu(x)$. Moreover, the estimator contains a term in the denominator, which may attain a value too close to zero, so trimming is desired to avoid a situation that a tiny value overwhelms the entire procedure. The results are rather stable for the linear case, but those for the nonlinear model are sensitive to the choice of these parameters; thus, we report the results for the various different choices of the parameters to show the difference. For the simulation, we use a product Gaussian kernel function of order 6 and 8. The numerical integration is performed over $\overline{M}_x = [-1, 1] \times [-1, 1]$. The weight function $\mu(x)$ used in the simulation is the pdf of a normal distribution with zero mean and a standard deviation of 1/3. In addition, the point of (x_1, x_2) where the estimated value of the conditional density $f_{B|X=x}(b)$ is below $1e-8$.

It is well known that the bandwidth choice determines overall performance of the kernel-based estimator. There is no established theory how to choose the bandwidth parameter for the functional of the kernel density with finite samples, so we use the bandwidth derived from the asymptotic property of the bandwidth. Typically, the rate of the bandwidth is expressed as $\sigma \sim n^{-k}$ where n is the sample size. Plugging this into Assumption 11, we obtain that the range of k should be $\frac{1}{6+2s} < k < \frac{1}{10}$, where s is the order of the kernel. Note that the optimal bandwidth for minimizing AMISE is $k = \frac{1}{4+2s+2j}$, where j is the order of the density derivative. It suggests that the choice of bandwidth should be substantially smaller than the typical bandwidth for density estimation. For the simulation, we use $k = \frac{1}{4+2\cdot6+4-3}$ for 6th order kernel and $k = \frac{1}{4+2\cdot8+4-4}$ for 8th order kernel. These choices are totally arbitrary.

We also compare the result with the estimates proposed in Matzkin (2015), assuming that (B_1, B_2) are observable. It should give similar results, but with smaller asymptotic variance.

1.5.1 Linear Model

For the linear model, we consider the sample size $n = 10000, 25000, 50000, 100000$ with 500 replications. The derivatives of the function r are evaluated at $(b_1, b_2) = (0, 0)$, which corresponds to the mean of the underlying distribution. The derivatives evaluated at this point are given by

$$r^1_{b_1} = 1;\ r^1_{b_2} = -0.75;\ r^2_{b_1} = 0.5;\ r^2_{b_2} = 1$$

The results are presented in Table 1.1 and 1.2, and Figures 1.1–4 in the appendix. The table consists of four parts; the first part denotes the number of observations in each sample, N. The second part is about the average derivative estimator. The first column denotes the sample mean, the second the sample median, the third the sample standard deviation, and the fourth the root-mean-square error. The third part is about the asymptotic standard deviation estimator. The first column denotes the sample mean, the second the root-mean-square error, and the third the proportion that the true value is within the estimated 95% confidence interval. The last part is about the result of the estimator in Matzkin (2015), assuming that (B_1, B_2) is observed. The first column denotes the sample mean of the average derivative estimates, the second the sample standard deviation, and the third the sample mean of the asymptotic standard deviation estimates. We can see that the average derivative estimates converge to the true values as the sample size increases. Also, the asymptotic standard deviation estimates converge to the sample standard deviation of the average derivative estimator. Also, note that the estimator in Matzkin gives similar convergence results, but the standard deviation is smaller than the estimator suggested in the paper so that narrow confidence interval can be constructed.

Figure 1 and 3 show the box-and-whisker plots of the average derivative estimator with 6th and 8th Gaussian kernel functions, respectively. The red line in each plot denotes the true values of the derivative. We can see the estimates converge to the true value and the dispersion become narrower. Figure 2 and 4 show the box-and-whisker plots of the asymptotic variance estimator. We can also see that the asymptotic variance estimates also converges to the true asymptotic variance, and the variation decreases as the sample size grows.

1.5.2 Nonlinear Model

The performance of the estimator in a nonlinear case is somewhat disappointing. A large sample size is required to obtain estimates close to the true value. We consider the sample size $n =, 100000, 200000, 500000, 1000000$ with 500 replications. The derivatives of the function r are evaluated at $(b_1, b_2) = (5, 5)$, which is a grid point closed to the sample mean of the simulation set. The derivatives evaluated at this point are given by

$$r^1_{b_1} = -0.2;\ r^1_{b_2} = -1;\ r^2_{b_1} = -0.1;\ r^2_{b_2} = 1$$

The results for the nonlinear case are presented in Table 1.3 and 1.4, and Figures 1.5–8 in the appendix. The bias of the estimator is still large even with one million samples. The result suggests that further bias reduction technique is necessary for precise estimation. In addition, we can see that Matzkin's estimator is still biased even with large samples. Our conjecture on this bias is because of the sub-optimal choice of bandwidths.

1.6 Concluding Remark

This paper proposes a new nonparametric estimator for a simultaneous binary response model without additivity. We discuss how the distribution of the latent variables can be recovered from the distribution of observed mixed variables, how identification is achieved via a change-of-variable technique. The identification strategy used in this paper does not require the full support condition on a special regressor. And then, we develop an estimator for the average derivatives of the indirect structural equations that are computed by simple matrix manipulation, and derive the asymptotic properties of the new estimator.

This paper has some limitations. First, it converges at a nonparametric rate, slower than the square root of the sample size. This implies that it is not possible to obtain meaningful estimates without a large number of observations, which hinders researchers from using the estimator in empirical works. Imposing functional restriction on the structural equations can be one way to recover the semiparametric rate of the estimator, which will be interesting future research. Second, the estimator involves undersmoothing, but there is no practical solution to the critical problem of choosing bandwidth. Furthermore, undersmoothing results in larger asymptotic variance, which in turn gives too wide confidence interval. We may consider a bootstrap-based bias correction by combining the method suggested in Hall and Horowitz (2013). This is another possible direction for future research.

1.7 Appendix

1.7.A Proofs

1.7.A.1 Proof of Lemmas

Lemma 1.7.1 *Let $\overline{M}_b$ and $\overline{M}_x$ be a compact and convex set defined as in Assumption 6. For any function $g(y_1, y_2, w, x) \in \{0,1\} \times \mathbb{R}^4$, define the marginal density functions $\widetilde{g}(w,x) = \sum_{y_1=0}^{1} \int g(y_1, y_2, w, x) dy_2$, and the conditional density function*

$$g_{B|X}(y_2, w|x) = \frac{g(0, y_2, w, x)\widetilde{g}_w(w, x) - g_w(0, y_2, w, x)\widetilde{g}(w, x)}{\widetilde{g}^2(w, x)},$$

whenever these functions exist. Let $\mathcal{F}$ denote the set of bounded and twice continuously differentiable functions g on the extension of the support to the whole space such that the marginal density function $\widetilde{g}$ is bounded and twice continuously differentiable, and that the conditional density function $g_{B|X}$ is bounded and continuously diffentiable. Define the norm $\|\cdot\|$ on $\mathcal{F}$ by

$$\begin{aligned}\|g\| = \max_{k\in\{0,1,2\}} \sup_{(b_1,b_2,x)\in\overline{M}_b\times\overline{M}_x} &\left\|\partial^k g(0, y_2, w, x)|_{w=b_1,y_2=b_2}\right\| \\ &+ \max_{k\in\{0,1,2\}} \sup_{(b_1,b_2,x)\in\overline{M}_b\times\overline{M}_x} \left\|\partial^k \widetilde{g}(w, x)|_{w=b_1,y_2=b_2}\right\| \\ &\qquad + \max_{k\in\{0,1\}} \sup_{(b_1,b_2,x)\in\overline{M}_b\times\overline{M}_x} \left\|\partial^k g_{B|X}(w, x)|_{w=b_1,y_2=b_2}\right\|.\end{aligned}$$

For simplicity, we leave the argument (y_1, y_2, w, x) implicit. Assume that the density $f(y_1, y_2, w, x)$ belongs to $\mathcal{F}$ and it is such that for $\delta > 0$ and all $(b, x) \in \overline{M}_b \times \overline{M}_x$, $f(0, b_2, b_1, x) > \delta$, $\widetilde{f}(b_1, x) > \delta$, and $f_w(0, b_2, b_1, x)\widetilde{f}(b_1, x) - f(0, b_2, b_1, x)\widetilde{f}_w(b_1, x) > \delta$.

Let $A(g)$ be a differentiation operator on g and $B(g)$ a differentiation operator on $\widetilde{g}$. Define the functional $S(g)$ by

$$S(g) = \frac{A(g)B(g)}{g_w\widetilde{g} - g\widetilde{g}_w}.$$

Then, for all $h \in \mathcal{F}$ such that $\|h\|$ is sufficiently small,

$$S(f + h) - S(f) = \frac{A(f + h)B(f + h)}{(f_w + h_w)(\widetilde{f} + \widetilde{h}) - (f + h)(\widetilde{f}_w + \widetilde{h}_w)} - \frac{A(f)B(f)}{f_w\widetilde{f} - f\widetilde{f}_w} = D(f; h) + R(f; h)$$

where

$$D(f;h) = \frac{A(f)f_w\widetilde{f}B(h) + f_wB(f)\widetilde{f}A(h) + A(f)fB(f)\widetilde{h}_w + A(f)B(f)\widetilde{f}_wh}{(f_w\widetilde{f} - f\widetilde{f}_w)^2} - \frac{A(f)f\widetilde{f}_wB(h) + fB(f)\widetilde{f}_wA(h) + A(f)B(f)\widetilde{f}h_w + A(f)f_wB(f)\widetilde{h}}{(f_w\widetilde{f} - f\widetilde{f}_w)^2}$$

$$\begin{aligned} R(f;h) = & \frac{(f_w\widetilde{f} - f\widetilde{f}_w)A(h)B(h) - A(f)B(f)(h_w\widetilde{h} - h\widetilde{h}_w)}{(f_w\widetilde{f} - f\widetilde{f}_w)^2} \\ & - \left[\frac{f_w\widetilde{h} - f\widetilde{h}_w + \widetilde{f}h_w - f_wh + h_w\widetilde{h} - h\widetilde{h}_w}{(f_w + h_w)(\widetilde{f} + \widetilde{h}) - (f+h)(\widetilde{f}_w + \widetilde{h}_w)}\right] \times \\ & \left[\frac{A(f)f_w\widetilde{f}B(h) + f_wB(f)\widetilde{f}A(h) + A(f)fB(f)\widetilde{h}_w + A(f)B(f)\widetilde{f}_wh}{(f_w\widetilde{f} - f\widetilde{f}_w)^2}\right. \\ & - \frac{A(f)f\widetilde{f}_wB(h) + fB(f)\widetilde{f}_wA(h) + A(f)B(f)\widetilde{f}h_w + A(f)f_wB(f)\widetilde{h}}{(f_w\widetilde{f} - f\widetilde{f}_w)^2} \\ & \left. + \frac{A(h)B(h)(f_w\widetilde{f} - f\widetilde{f}_w) - A(f)B(f)(h_w\widetilde{h} - h\widetilde{h}_w)}{(f_w\widetilde{f} - f\widetilde{f}_w)^2}\right]. \end{aligned}$$

Moreover, $D(f;h) = O(\|h\|)$ *and* $R(f;h) = O(\|h\|^2)$*.*

Proof. We will make use of the equality

$$\frac{N'}{D'} - \frac{N}{D} = \frac{N'D - ND'}{D^2} - \frac{(D'-D)(N'D - ND')}{D'D^2}.$$

Denote N', D', N, D by

$$\begin{aligned} N' &= A(f+h)B(f+h) = [A(f) + A(h)][B(f) + B(h)] \\ N &= A(f)B(f) \\ D' &= (f_w + h_w)(\widetilde{f} + \widetilde{h}) - (f+h)(\widetilde{f}_w + \widetilde{h}_w) \\ D &= f_w\widetilde{f} - f\widetilde{f}_w, \end{aligned}$$

where the last equality of N' holds since a differentiation operator is linear. Then,

$$\begin{aligned} N'D - ND' = & A(f)f_w\widetilde{f}B(h) + f_wB(f)\widetilde{f}A(h) + A(f)fB(f)\widetilde{h}_w + A(f)B(f)\widetilde{f}_wh \\ & - A(f)f\widetilde{f}_wB(h) - fB(f)\widetilde{f}_wA(h) - A(f)B(f)\widetilde{f}h_w - A(f)f_wB(f)\widetilde{h} \end{aligned}$$

$$
\begin{aligned}
& +(f_w\widetilde{f} - f\widetilde{f}_w)A(h)B(h) - A(f)B(f)(h_w\widetilde{h} - h\widetilde{h}_w) \\
D' - D &= f_w\widetilde{h} - f\widetilde{h}_w + \widetilde{f}h_w - f_wh + h_w\widetilde{h} - h\widetilde{h}_w.
\end{aligned}
$$

Define ω_a and ω_b as

$$
\begin{aligned}
\omega_a &= A(f)f_w\widetilde{f}B(h) + f_wB(f)\widetilde{f}A(h) + A(f)fB(f)\widetilde{h}_w + A(f)B(f)\widetilde{f}_wh \\
&\quad -A(f)f\widetilde{f}_wB(h) - fB(f)\widetilde{f}_wA(h) - A(f)B(f)\widetilde{f}h_w - A(f)f_wB(f)\widetilde{h} \\
\omega_b &= (f_w\widetilde{f} - f\widetilde{f}_w)A(h)B(h) - A(f)B(f)(h_w\widetilde{h} - h\widetilde{h}_w)
\end{aligned}
$$

so that $N'D - ND' = \omega_a + \omega_b$. Let $\widetilde{\delta}_0 = \min\{\delta/2, \delta/(16\|f\|), 1\}$. Choose $h \in \mathcal{F}$ such that $\|h\| \leq \widetilde{\delta}_0$. By assumptions and the definition of sup norms on $\mathcal{F}$, it follows that

$$
\begin{aligned}
|\omega_a| &= \Big|A(f)f_w\widetilde{f}B(h) + f_wB(f)\widetilde{f}A(h) + A(f)fB(f)\widetilde{h}_w + A(f)B(f)\widetilde{f}_wh \\
&\quad -A(f)f\widetilde{f}_wB(h) - fB(f)\widetilde{f}_wA(h) - A(f)B(f)\widetilde{f}h_w - A(f)f_wB(f)\widetilde{h}\Big| \\
&\leq 8\|f\|^3\|h\| \\
|\omega_b| &= \Big|A(h)B(h)(f_w\widetilde{f} - f\widetilde{f}_w) - A(f)B(f)(h_w\widetilde{h} - h\widetilde{h}_w)\Big| \\
&\leq 4\|f\|^2\|h\|^2 \\
|D' - D| &= \Big|f_w\widetilde{h} - f\widetilde{h}_w + \widetilde{f}h_w - f_wh + h_w\widetilde{h} - h\widetilde{h}_w\Big| \\
&\leq 4\|f\|\|h\| + 2\|h\|^2
\end{aligned}
$$

Moreover, $D' = (f_w\widetilde{f} - f\widetilde{f}_w) + (h_w\widetilde{h} - h\widetilde{h}_w) + (f_w\widetilde{h} + \widetilde{f}h_w - f\widetilde{h}_w - \widetilde{f}_wh)$. By the choice of $\widetilde{\delta}_0$, we have

$$
|f_w\widetilde{h} + \widetilde{f}h_w - f\widetilde{h}_w - \widetilde{f}_wh| \leq 4\|f\|\|h\| \leq 4\|f\|\widetilde{\delta}_0 \leq \delta/4
$$

and since $f_w\widetilde{f} - f\widetilde{f}_w > \delta$ and $|h_w\widetilde{h} - h\widetilde{h}_w| < \widetilde{\delta} \leq \delta/2$, it follows that $(f_w\widetilde{f} - f\widetilde{f}_w) + (h_w\widetilde{h} - h\widetilde{h}_w) > \delta/2$. Hence, $D' \geq \delta/4$. Then, it follows that

$$
\begin{aligned}
|D(f;h)| &= \left|\frac{\omega_a}{(f_w\widetilde{f} - f\widetilde{f}_w)^2}\right| \leq \frac{8\|f\|^3}{\delta^2}\|h\| \\
|R(f;h)| &= \left|\frac{\omega_b}{(f_w\widetilde{f} - f\widetilde{f}_w)^2} + \frac{D' - D}{D'}\frac{\omega_a + \omega_b}{(f_w\widetilde{f} - f\widetilde{f}_w)^2}\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \left|\frac{\omega_b}{(f_w\widetilde{f}-f\widetilde{f}_w)^2}\right| + \left|\frac{D'-D}{D'}\right|\left|\frac{\omega_a+\omega_b}{(f_w\widetilde{f}-f\widetilde{f}_w)^2}\right| \\
&\leq \frac{4\|f\|^2}{\delta^2}\|h\|^2 + \frac{16\|f\|\|h\|+8\|h\|^2}{\delta} \times \frac{8\|f\|^3\|h\|+4\|f\|^2\|h\|^2}{\delta^2} \\
&= \left[\frac{4\|f\|^2}{\delta^2} + \frac{32\|f\|^4}{\delta^3}\right]\|h\|^2 + \frac{32\|f\|^3}{\delta^3}\|h\|^3 + \frac{8\|f\|^2}{\delta^3}\|h\|^4 \\
&\leq \left[\frac{4\|f\|^2}{\delta^2} + \frac{32\|f\|^4+32\|f\|^3+8\|f\|^2}{\delta^3}\right]\|h\|^2.
\end{aligned}
$$

This completes the proof. □

Lemma 1.7.2 *Let $\overline{M}_b, \overline{M}_x, \mathcal{F}, \|\cdot\|, \widetilde{g}$ and $g_{B|X}$ be defined as in Lemma 1.7.1. Assume that the density $f(y_1, y_2, w, x)$ belongs to $\mathcal{F}$ and it is such that for $\delta > 0$ and all $(b, x) \in \overline{M}_b \times \overline{M}_x$, $f(0, b_2, b_1, x) > \delta$, $\widetilde{f}(b_1, x) > \delta$, and $f_w(0, b_2, b_1, x)\widetilde{f}(b_1, x) - f(0, b_2, b_1, x)\widetilde{f}_w(b_1, x) > \delta$.*

Let $C(g)$ be a differentiation operator on $\widetilde{g}$. Define the functional $T(g)$ on $\mathcal{F}$ by

$$
T(g) = \frac{g\widetilde{g}_w C(g)}{\widetilde{g}(g_w\widetilde{g} - g\widetilde{g}_w)}.
$$

Then, for all $h \in \mathcal{F}$ such that $\|h\|$ is sufficiently small,

$$
\begin{aligned}
T(f+h) - T(f) &= \frac{(f+h)(\widetilde{f}_w+\widetilde{h}_w)(C(f)+C(h))}{(\widetilde{f}+\widetilde{h})[(f_w+h_w)(\widetilde{f}+\widetilde{h})-(f+h)(\widetilde{f}_w+\widetilde{h}_w)]} - \frac{f\widetilde{f}_w C(f)}{\widetilde{f}(f_w\widetilde{f}-f\widetilde{f}_w)} \\
&= D(f;h) + R(f;h)
\end{aligned}
$$

where

$$
\begin{aligned}
D(f;h) = &\frac{ff_wC(h) + ff_wC(f)\widetilde{h}_w + f_wC(f)\widetilde{f}_wh + fC(f)\widetilde{f}_wh_w}{(f_w\widetilde{f}-f\widetilde{f}_w)^2} \\
&- \frac{f^2\widetilde{f}_w^2C(h) + ff_wC(f)\widetilde{f}_w\widetilde{h} + ff_wC(f)\widetilde{f}_w\widetilde{h}}{\widetilde{f}(f_w\widetilde{f}-f\widetilde{f}_w)^2} + \frac{f^2\widetilde{f}_w^2C(f)\widetilde{h}}{\widetilde{f}^2(f_w\widetilde{f}-f\widetilde{f}_w)^2} \\
R(f;h) = &\frac{f_wC(f)\widetilde{f}^2h\widetilde{h}_w + (f_w\widetilde{f}^2 - f\widetilde{f}\widetilde{f}_w)(fC(h)\widetilde{h}_w + \widetilde{f}_wC(h)h + hC(h)\widetilde{h}_w)}{\widetilde{f}^2(f_w\widetilde{f}-f\widetilde{f}_w)^2} \\
&- \frac{f\widetilde{f}_wC(f)(\widetilde{f}h_w\widetilde{h} + f_w\widetilde{h}^2 + \widetilde{f}h_w\widetilde{h} + h_w\widetilde{h}^2 - f\widetilde{h}\widetilde{h}_w - \widetilde{f}_wh\widetilde{h} - h\widetilde{h}\widetilde{h}_w)}{\widetilde{f}^2(f_w\widetilde{f}-f\widetilde{f}_w)^2} \\
&- \left[\frac{\widetilde{f}(f_w\widetilde{h} + \widetilde{f}h_w - f\widetilde{h}_w - \widetilde{f}_wh + h_w\widetilde{h} - h\widetilde{h}_w)}{(\widetilde{f}+\widetilde{h})[(f_w+h_w)(\widetilde{f}+\widetilde{h})-(f+h)(\widetilde{f}_w+\widetilde{h}_w)]}\right.
\end{aligned}
$$

$$
\left. + \frac{\widetilde{h}[(f_w + h_w)(\widetilde{f} + \widetilde{h}) - (f + h)(\widetilde{f}_w + \widetilde{h}_w)]}{(\widetilde{f} + \widetilde{h})[(f_w + h_w)(\widetilde{f} + \widetilde{h}) - (f + h)(\widetilde{f}_w + \widetilde{h}_w)]} \right]
$$

$$
\begin{aligned}
\times & \left[\frac{f f_w \widetilde{f}^2 C(h) + f f_w C(f) \widetilde{f}^2 \widetilde{h}_w + f_w C(f) \widetilde{f}^2 \widetilde{f}_w h + f^2 \widetilde{f}_w^2 C(f) \widetilde{h}}{\widetilde{f}^2 (f_w \widetilde{f} - f \widetilde{f}_w)^2} \right. \\
& - \frac{f^2 \widetilde{f} \widetilde{f}_w^2 C(h) + f f_w C(f) \widetilde{f} \widetilde{f}_w \widetilde{h} + f C(f) \widetilde{f}^2 \widetilde{f}_w h_w + f f_w C(f) \widetilde{f} \widetilde{f}_w \widetilde{h}}{\widetilde{f}^2 (f_w \widetilde{f} - f \widetilde{f}_w)^2} \\
& + \frac{f_w C(f) \widetilde{f}^2 h \widetilde{h}_w + (f_w \widetilde{f}^2 - f \widetilde{f} \widetilde{f}_w)(f C(h) \widetilde{h}_w + \widetilde{f}_w C(h) h + h C(h) \widetilde{h}_w)}{\widetilde{f}^2 (f_w \widetilde{f} - f \widetilde{f}_w)^2} \\
& \left. - \frac{f \widetilde{f}_w C(f)(\widetilde{f} h_w \widetilde{h} + f_w \widetilde{h}^2 + \widetilde{f} h_w \widetilde{h} + h_w \widetilde{h}^2 - f \widetilde{h} \widetilde{h}_w - \widetilde{f}_w h \widetilde{h} - h \widetilde{h} \widetilde{h}_w)}{\widetilde{f}^2 (f_w \widetilde{f} - f \widetilde{f}_w)^2} \right]
\end{aligned}
$$

Moreover, $D(f;h) = O(\|h\|)$ and $R(f;h) = O(\|h\|^2)$.

Proof. The arguments are almost identical to those in the proof of Lemma 1.7.1, so omitted.

□

Lemma 1.7.3 *Let $\overline{M}_b, \overline{M}_x, \mathcal{F}, \|\cdot\|, \widetilde{g}$ and $g_{B|X}$ be defined as in Lemma 1.7.1. Assume that the density $f(y_1, y_2, w, x)$ belongs to $\mathcal{F}$ and it is such that for $\delta > 0$ and all $(b, x) \in \overline{M}_b \times \overline{M}_x$, $f(0, b_2, b_1, x) > \delta$, $\widetilde{f}(b_1, x) > \delta$, and $f_w(0, b_2, b_1, x)\widetilde{f}(b_1, x) - f(0, b_2, b_1, x)\widetilde{f}_w(b_1, x) > \delta$.*

Define the functionals α_{b_j} and β_{x_s} on $\mathcal{F}$ by

$$
\begin{aligned}
\alpha_{b_1}(g) &= \frac{g_{ww}\widetilde{g}^2 - g\widetilde{g}\widetilde{g}_{ww} - 2g_w\widetilde{g}\widetilde{g}_w + 2g\widetilde{g}_w^2}{\widetilde{g}(g_w\widetilde{g} - g\widetilde{g}_w)} = \frac{g_{ww}\widetilde{g} - g\widetilde{g}_{ww} - 2g_w\widetilde{g}_w}{g_w\widetilde{g} - g\widetilde{g}_w} + \frac{2g\widetilde{g}_w^2}{\widetilde{g}(g_w\widetilde{g} - g\widetilde{g}_w)} \\
\alpha_{b_2}(g) &= \frac{g_{wy_2}\widetilde{g} - g_{y_2}\widetilde{g}_w}{g_w\widetilde{g} - g\widetilde{g}_w} \\
\beta_{x_s}(g) &= \frac{g_{wx_s}\widetilde{g}^2 - g_{x_s}\widetilde{g}\widetilde{g}_w - g\widetilde{g}\widetilde{g}_{wx_s} - g_w\widetilde{g}\widetilde{g}_{x_s} + 2g\widetilde{g}_w\widetilde{g}_{x_s}}{\widetilde{g}(g_w\widetilde{g} - g\widetilde{g}_w)} \\
&= \frac{g_{wx_s}\widetilde{g} - g_{x_s}\widetilde{g}_w - g\widetilde{g}_{wx_s} - g_w\widetilde{g}_{x_s}}{g_w\widetilde{g} - g\widetilde{g}_w} + \frac{2g\widetilde{g}_w\widetilde{g}_{x_s}}{\widetilde{g}(g_w\widetilde{g} - g\widetilde{g}_w)}.
\end{aligned}
$$

Then, for all $h \in \mathcal{F}$ such that $\|h\|$ is sufficiently small, $\alpha_{b_j}(f+h)$ and $\beta_{x_s}(f+h)$ admit the first order Taylor expansion around $\alpha_{b_j}(f)$ and $\beta_{x_s}(f)$:

$$
\begin{aligned}
\alpha_{b_j}(f+h) - \alpha_{b_j}(f) &= D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h) \\
\beta_{x_s}(f+h) - \beta_{x_s}(f) &= D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h)
\end{aligned}
$$

where

$$
\begin{aligned}
D\alpha_{b_1}(f;h) &= \frac{(f_{ww}\widetilde{h} + h_{ww}\widetilde{f}) - (f\widetilde{h}_{ww} + h\widetilde{f}_{ww}) - 2(f_w\widetilde{h}_w + h_w\widetilde{f}_w)}{f_w\widetilde{f} - f\widetilde{f}_w} + \frac{2(2f\widetilde{h}_w + h\widetilde{f}_w^2)}{\widetilde{f}(f_w\widetilde{f} - f\widetilde{f}_w)} \\
&\quad - \frac{(f_{ww}\widetilde{f} - f\widetilde{f}_{ww} - 2f_w\widetilde{f}_w)}{(f_w\widetilde{f} - f\widetilde{f}_w)^2} \times (f_w\widetilde{h} + h_w\widetilde{f} - f\widetilde{h}_w - h\widetilde{f}_w) \\
&\quad - \frac{2f\widetilde{f}_w^2[\widetilde{h}(f_w\widetilde{f} - f\widetilde{f}_w) + \widetilde{f}(f_w\widetilde{h} + h_w\widetilde{f} - f\widetilde{h}_w - h\widetilde{f}_w)]}{\widetilde{f}^2(f_w\widetilde{f} - f\widetilde{f}_w)^2} \\
R\alpha_{b_1}(f;h) &= \alpha_{b_1}(f+h) - \alpha_{b_1}(f) - D\alpha_{b_1}(f;h) \\
D\alpha_{b_2}(f;h) &= \frac{(f_{wy_2}\widetilde{h} + \widetilde{f}h_{wy_2}) - (f_{y_2}\widetilde{h}_w + \widetilde{f}_w h_{y_2})}{f_w\widetilde{f} - f\widetilde{f}_w} \\
&\quad - \frac{(f_{wy_2}\widetilde{f} - f_{y_2}\widetilde{f}_w)(f_w\widetilde{h} + h_w\widetilde{f} - f\widetilde{h}_w - h\widetilde{f}_w)}{(f_w\widetilde{f} - f\widetilde{f}_w)^2} \\
R\alpha_{b_2}(f;h) &= \alpha_{b_2}(f+h) - \alpha_{b_2}(f) - D\alpha_{b_2}(f;h) \\
D\beta_{x_s}(f;h) &= \frac{(f_{wx_s}\widetilde{h} + \widetilde{f}h_{wx_s}) - (f\widetilde{h}_{wx_s} + h\widetilde{f}_{wx_s}) - (f_{x_s}\widetilde{h}_w + h_{x_s}\widetilde{f}_w) - (f_w\widetilde{h}_{x_s} + h_w\widetilde{f}_{x_s})}{f_w\widetilde{f} - f\widetilde{f}_w} \\
&\quad + \frac{2(f\widetilde{f}_w\widetilde{h}_{x_s} + f\widetilde{h}_w\widetilde{f}_{x_s} + h\widetilde{f}_w\widetilde{f}_{x_s})}{\widetilde{f}(f_w\widetilde{f} - f\widetilde{f}_w)} \\
&\quad - \frac{(f_{wx_s}\widetilde{f} - f\widetilde{f}_{wx_s} - f_{x_s}\widetilde{f}_w - f_w\widetilde{f}_{x_s})(f_w\widetilde{h} + h_w\widetilde{f} - f\widetilde{h}_w - h\widetilde{f}_w)}{(f_w\widetilde{f} - f\widetilde{f}_w)^2} \\
&\quad - \frac{2f\widetilde{f}_w\widetilde{f}_{x_s}[\widetilde{h}(f_w\widetilde{f} - f\widetilde{f}_w) + \widetilde{f}(f_w\widetilde{h} + h_w\widetilde{f} - f\widetilde{h}_w - h\widetilde{f}_w)]}{\widetilde{f}^2(f_w\widetilde{f} - f\widetilde{f}_w)^2} \\
R\beta_{x_s}(f;h) &= \beta_{x_s}(f+h) - \beta_{x_s}(f) - D\beta_{x_s}(f;h).
\end{aligned}
$$

Moreover, $D\alpha_{b_j}(f;h) = O(\|h\|)$, $D\beta_{x_s}(f;h) = O(\|h\|)$, $R\alpha_{b_j}(f;h) = O(\|h\|^2)$ *and* $R\beta_{x_s}(f;h) = O(\|h\|^2)$*.*

Proof. The terms in α_{b_j} and β_{x_s} have forms of $S(f)$ and $T(f)$. Hence, the result follows from Lemma 1.7.1 and Lemma 1.7.2. □

Lemma 1.7.4 *Let* $\overline{M}_b, \overline{M}_x, \mathcal{F}, \|\cdot\|, \widetilde{g}, g_{B|X}, \alpha_{b_j}$ *and* β_{x_s} *be defined as in Lemmas 1.7.1 and 1.7.3. Assume that the density* $f(y_1, y_2, w, x)$ *belongs to* $\mathcal{F}$ *and it is such that for* $\delta > 0$ *and all* $(b,x) \in \overline{M}_b \times \overline{M}_x$*,* $f(0, b_2, b_1, x) > \delta$*,* $\widetilde{f}(b_1, x) > \delta$*, and* $f_w(0, b_2, b_1, x)\widetilde{f}(b_1, x) - f(0, b_2, b_1, x)\widetilde{f}_w(b_1, x) > \delta$*. Let* μ *denote a bounded and continuously differentiable, strictly* *positive over $\overline{M}_x$, with values and derivatives vanishing on the boundary and on the complement of $\overline{M}_x$, and such that $\int_{\overline{M}_x} \mu(x)dx = 1$.*

Define the functional Φ_{b_j,x_s} on $\mathcal{F}$ by

$$\Phi_{b_j,x_s}(g) = \int_{\overline{M}_x} \alpha_{b_j}(g)\beta_{x_s}(g)\mu(x)dx - \left(\int_{\overline{M}_x} \alpha_{b_j}(g)\mu(x)dx\right)\left(\int_{\overline{M}_x} \beta_{x_s}(g)\mu(x)dx\right).$$

Then, for all $h \in \mathcal{F}$ such that $\|h\|$ is sufficiently small, $\Phi_{b_j,x_s}(f+h)$ admit the first order Taylor expansion around $\Phi_{b_j,x_s}(f)$:

$$\Phi_{b_j,x_s}(f+h) - \Phi_{b_j,x_s}(f) = D\Phi_{b_j,x_s}(f;h) + R\Phi_{b_j,x_s}(f;h)$$

where

$$\begin{aligned}
D\Phi_{b_j,x_s}(f;h) &= \int_{\overline{M}_x} D\alpha_{b_j}(f;h)\left(\beta_{x_s}(f) - \int_{\overline{M}_x} \beta_{x_s}(f)\mu dx\right)\mu dx \\
&\quad + \int_{\overline{M}_x} D\beta_{x_s}(f;h)\left(\alpha_{b_j}(f) - \int_{\overline{M}_x} \alpha_{b_j}(f)\mu dx\right)\mu dx \\
R\Phi_{b_j,x_s}(f;h) &= \int_{\overline{M}_x} R\alpha_{b_j}(f;h)\left(\beta_{x_s}(f) - \int_{\overline{M}_x} \beta_{x_s}(f)\mu dx\right)\mu dx \\
&\quad + \int_{\overline{M}_x} R\beta_{x_s}(f;h)\left(\alpha_{b_j}(f) - \int_{\overline{M}_x} \alpha_{b_j}(f)\mu dx\right)\mu dx \\
&\quad + \int_{\overline{M}_x} (D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h))(D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx \\
&\quad - \left(\int_{\overline{M}_x} (D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h))\mu dx\right) \\
&\qquad \times \left(\int_{\overline{M}_x} (D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx\right).
\end{aligned}$$

Moreover, $D\Phi_{b_j,x_s}(f;h) = O(\|h\|)$ and $R\Phi_{b_j,x_s}(f;h) = O(\|h\|^2)$.

Proof.

$$\begin{aligned}
\Phi_{b_j,x_s}(f+h) - \Phi_{b_j,x_s}(f) = \int_{\overline{M}_x} &\alpha_{b_j}(f+h)\beta_{x_s}(f+h)\mu(x)dx \\
&- \left(\int_{\overline{M}_x} \alpha_{b_j}(f+h)\mu(x)dx\right)\left(\int_{\overline{M}_x} \beta_{x_s}(f+h)\mu(x)dx\right) \\
&- \alpha_{b_j}(f)\beta_{x_s}(f)\mu(x)dx + \left(\int_{\overline{M}_x} \alpha_{b_j}(f)\mu(x)dx\right)\left(\int_{\overline{M}_x} \beta_{x_s}(f)\mu(x)dx\right)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\overline{M}_x} (\alpha_{b_j}(f+h) - \alpha_{b_j}(f)) \left(\beta_{x_s}(f) - \int_{\overline{M}_x} \beta_{x_s}(f)\mu dx \right) \mu dx \\
&\quad + \int_{\overline{M}_x} \left(\alpha_{b_j}(f) - \int_{\overline{M}_x} \alpha_{b_j}(f)\mu dx \right) (\beta_{x_s}(f+h) - \beta_{x_s}(f))\mu dx \\
&\quad + \int_{\overline{M}_x} (\alpha_{b_j}(f+h) - \alpha_{b_j}(f))(\beta_{x_s}(f+h) - \beta_{x_s}(f))\mu dx \\
&\quad - \left(\int_{\overline{M}_x} (\alpha_{b_j}(f+h) - \alpha_{b_j}(f))\mu dx \right) \left(\int_{\overline{M}_x} (\beta_{x_s}(f+h) - \beta_{x_s}(f))\mu dx \right).
\end{aligned}
$$

Employing the expressions of $\alpha_{b_j}(f+h)-\alpha_{b_j}(f)$ and $\beta_{x_s}(f+h)-\beta_{x_s}(f)$ in terms of $D\alpha_{b_j}(f;h)$, $\mathbb{R}\alpha_{b_j}(f;h)$, $D\beta_{x_s}(f;h)$ and $R\beta_{x_s}(f;h)$ in Lemma, it follows that

$$
\begin{aligned}
\Phi_{b_j,x_s}(f+h) - \Phi_{b_j,x_s}(f) &= \int_{\overline{M}_x} (D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h)) \left(\beta_{x_s}(f) - \int_{\overline{M}_x} \beta_{x_s}(f)\mu dx \right) \mu dx \\
&\quad + \int_{\overline{M}_x} \left(\alpha_{b_j}(f) - \int_{\overline{M}_x} \alpha_{b_j}(f)\mu dx \right) (D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx \\
&\quad + \int_{\overline{M}_x} (D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h))(D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx \\
&\quad - \left(\int_{\overline{M}_x} (D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h))\mu dx \right) \\
&\qquad \times \left(\int_{\overline{M}_x} (D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx \right) \\
&= \int_{\overline{M}_x} D\alpha_{b_j}(f;h) \left(\beta_{x_s}(f) - \int_{\overline{M}_x} \beta_{x_s}(f)\mu dx \right) \mu dx \\
&\quad + \int_{\overline{M}_x} D\beta_{x_s}(f;h) \left(\alpha_{b_j}(f) - \int_{\overline{M}_x} \alpha_{b_j}(f)\mu dx \right) \mu dx \\
&\quad + \int_{\overline{M}_x} R\alpha_{b_j}(f;h) \left(\beta_{x_s}(f) - \int_{\overline{M}_x} \beta_{x_s}(f)\mu dx \right) \mu dx \\
&\quad + \int_{\overline{M}_x} R\beta_{x_s}(f;h) \left(\alpha_{b_j}(f) - \int_{\overline{M}_x} \alpha_{b_j}(f)\mu dx \right) \mu dx \\
&\quad + \int_{\overline{M}_x} (D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h))(D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx \\
&\quad - \left(\int_{\overline{M}_x} (D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h))\mu dx \right) \\
&\qquad \times \left(\int_{\overline{M}_x} (D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx \right) \\
&= D\Phi_{b_j,x_s}(f;h) + R\Phi_{b_j,x_s}(f;h).
\end{aligned}
$$

It remains to show that $D\Phi_{b_j,x_s}(f;h) = O(\|h\|)$ and $R\Phi_{b_j,x_s}(f;h) = O(\|h\|^2)$. Since we assume $f > \delta$, $\widetilde{f} > \delta$ and $f_w\widetilde{f} - f\widetilde{f}_w > \delta$,

$$\begin{aligned}
|\alpha_{b_1}(f)| &= \left| \frac{f_{ww}\widetilde{f} - f\widetilde{f}_{ww} - 2f_w\widetilde{f}_w}{f_w\widetilde{f} - f\widetilde{f}_w} + \frac{2f\widetilde{f}_w^2}{\widetilde{f}(f_w\widetilde{f} - f\widetilde{f}_w)} \right| \\
&\leq \frac{4\|f\|^2}{\delta} + \frac{2\|f\|^3}{\delta^2} \\
|\alpha_{b_2}(f)| &= \left| \frac{f_{wy_2}\widetilde{f} - f_{y_2}\widetilde{f}_w}{f_w\widetilde{f} - f\widetilde{f}_w} \right| \leq \frac{2\|f\|^2}{\delta} \\
|\beta_{x_s}(f)| &= \left| \frac{f_{wx_s}\widetilde{f} - f_{x_s}\widetilde{f}_w - f\widetilde{f}_{wx_s} - f_w\widetilde{f}_{x_s}}{f_w\widetilde{f} - f\widetilde{f}_w} + \frac{2f\widetilde{f}_w\widetilde{f}_{x_s}}{\widetilde{f}(f_w\widetilde{f} - f\widetilde{f}_w)} \right| \\
&\leq \frac{4\|f\|^2}{\delta} + \frac{2\|f\|^3}{\delta^2}.
\end{aligned}$$

Let $c_0 = \max\{\frac{4\|f\|^2}{\delta} + \frac{2\|f\|^3}{\delta^2}, \frac{2\|f\|^2}{\delta}\}$. Since $D\alpha_{b_j}(f;h) = O(\|h\|)$ and $D\beta_{x_s}(f;h) = O(\|h\|)$, we can find a constant c_1 such that $|D\alpha_{b_j}(f;h)| \leq c_1\|h\|$ and $|D\beta_{x_s}(f;h)| \leq c_1\|h\|$, and a constant c_2 such that $|R\alpha_{b_j}(f;h)| \leq c_2\|h\|^2$ and $|R\beta_{x_s}(f;h)| \leq c_2\|h\|^2$. By the assumption that $\int_{\overline{M}_x} \mu(x)dx = 1$, it follows that

$$\begin{aligned}
&\left| \int_{\overline{M}_x} D\alpha_{b_j}(f;h) \left(\beta_{x_s}(f) - \int_{\overline{M}_x} \beta_{x_s}(f)\mu dx \right) \mu dx \right| \\
\leq &\int_{\overline{M}_x} |D\alpha_{b_j}(f;h)| \left| \left(\beta_{x_s}(f) - \int_{\overline{M}_x} \beta_{x_s}(f)\mu dx \right) \right| \mu dx \\
\leq &\int_{\overline{M}_x} c_1\|h\|c_0\mu dx \leq c_0c_1\|h\|
\end{aligned}$$

and that

$$\begin{aligned}
&\left| \int_{\overline{M}_x} D\beta_{x_s}(f;h) \left(\alpha_{b_j}(f) - \int_{\overline{M}_x} \alpha_{b_j}(f)\mu dx \right) \mu dx \right| \\
\leq &\int_{\overline{M}_x} |D\beta_{x_s}(f;h)| \left| \left(\alpha_{b_j}(f) - \int_{\overline{M}_x} \alpha_{b_j}(f)\mu dx \right) \right| \mu dx \\
\leq &\int_{\overline{M}_x} c_1\|h\|c_0\mu dx \leq c_0c_1\|h\|
\end{aligned}$$

Similarly, for the terms in $R\Phi_{b_j,x_s}(f;h)$,

$$\begin{aligned}
&\left| \int_{\overline{M}_x} R\alpha_{b_j}(f;h) \left(\beta_{x_s}(f) - \int_{\overline{M}_x} \beta_{x_s}(f)\mu dx \right) \mu dx \right| \\
\leq &\int_{\overline{M}_x} c_2\|h\|^2c_0\mu dx \leq c_0c_2\|h\|^2
\end{aligned}$$

$$
\begin{aligned}
&\left| \int_{\overline{M}_x} R\beta_{x_s}(f;h)\left(\alpha_{b_j}(f) - \int_{\overline{M}_x} \alpha_{b_j}(f)\mu dx\right)\mu dx\right| \\
&\leq \int_{\overline{M}_x} c_2\|h\|^2 c_0 \mu dx \leq c_0 c_2 \|h\|^2
\end{aligned}
$$

$$
\begin{aligned}
&\left| \int_{\overline{M}_x} (D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h))(D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx\right| \\
&\leq \int_{\overline{M}_x} (c_1\|h\| + c_2\|h\|^2)^2 \mu dx \leq (c_1\|h\| + c_2\|h\|^2)^2 \leq (c_1^2 + 2c_1c_2 + c_2^2)\|h\|^2
\end{aligned}
$$

$$
\begin{aligned}
&\left| \left(\int_{\overline{M}_x} (D\alpha_{b_j}(f;h) + R\alpha_{b_j}(f;h))\mu dx\right)\left(\int_{\overline{M}_x} (D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx\right)\right| \\
&\leq (c_1\|h\| + c_2\|h\|^2)^2 \leq (c_1^2 + 2c_1c_2 + c_2^2)\|h\|^2
\end{aligned}
$$

This completes the proof. □

Lemma 1.7.5 *Let $\overline{M}_b, \overline{M}_x, \mathcal{F}, \|\cdot\|, \widetilde{g}, g_{B|X}, \alpha_{b_j}$ and β_{x_s} be defined as in Lemma 1.7.1 and 1.7.3. Assume that the density $f(y_1, y_2, w, x)$ belongs to $\mathcal{F}$ and it is such that for $\delta > 0$ and all $(b,x) \in \overline{M}_b \times \overline{M}_x$, $f(0, b_2, b_1, x) > \delta$, $\widetilde{f}(b_1, x) > \delta$, and $f_w(0, b_2, b_1, x)\widetilde{f}(b_1, x) - f(0, b_2, b_1, x)\widetilde{f}_w(b_1, x) > \delta$. Let μ denote a bounded and continuously differentiable, strictly positive over $\overline{M}_x$, with values and derivatives vanishing on the boundary and on the complement of $\overline{M}_x$, and such that $\int_{\overline{M}_x} \mu(x)dx = 1$.*

Define the functional Φ_{x_j,x_s} on $\mathcal{F}$ by

$$
\Phi_{x_j,x_s}(g) = \int_{\overline{M}_x} \alpha_{b_j}(g)\beta_{x_s}(g)\mu(x)dx - \left(\int_{\overline{M}_x} \alpha_{b_j}(g)\mu(x)dx\right)\left(\int_{\overline{M}_x} \beta_{x_s}(g)\mu(x)dx\right).
$$

Then, for all $h \in \mathcal{F}$ such that $\|h\|$ is sufficiently small, $\Phi_{x_j,x_s}(f+h)$ admit the first order Taylor expansion around $\Phi_{x_j,x_s}(f)$:

$$
\Phi_{x_j,x_s}(f+h) - \Phi_{x_j,x_s}(f) = D\Phi_{x_j,x_s}(f;h) + R\Phi_{x_j,x_s}(f;h)
$$

where

$$
\begin{aligned}
D\Phi_{x_j,x_s}(f;h) &= \int_{\overline{M}_x} D\beta_{x_j}(f;h)\left(\beta_{x_s}(f) - \int_{\overline{M}_x}\beta_{x_s}(f)\mu dx\right)\mu dx \\
&\quad + \int_{\overline{M}_x} D\beta_{x_s}(f;h)\left(\beta_{x_j}(f) - \int_{\overline{M}_x}\beta_{x_j}(f)\mu dx\right)\mu dx \\
R\Phi_{x_j,x_s}(f;h) &= \int_{\overline{M}_x} R\beta_{x_j}(f;h)\left(\beta_{x_s}(f) - \int_{\overline{M}_x}\beta_{x_s}(f)\mu dx\right)\mu dx \\
&\quad + \int_{\overline{M}_x} R\beta_{x_s}(f;h)\left(\beta_{x_j}(f) - \int_{\overline{M}_x}\beta_{x_j}(f)\mu dx\right)\mu dx \\
&\quad + \int_{\overline{M}_x} (D\beta_{x_j}(f;h) + R\beta_{x_j}(f;h))(D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx \\
&\quad - \left(\int_{\overline{M}_x} (D\beta_{x_j}(f;h) + R\beta_{x_j}(f;h))\mu dx\right) \\
&\qquad \times \left(\int_{\overline{M}_x} (D\beta_{x_s}(f;h) + R\beta_{x_s}(f;h))\mu dx\right).
\end{aligned}
$$

Moreover, $D\Phi_{x_j,x_s}(f;h) = O(\|h\|)$ *and* $R\Phi_{x_j,x_s}(f;h) = O(\|h\|^2)$.

Proof. The proof of Lemma 1.7.5 is analogous to that of Lemma 1.7.4, so omitted. □

Lemma 1.7.6 *Let* $\omega(x)$ *denote a bounded and continuously differentiable, strictly positive over* $\overline{M}_x$*, with values and derivatives vanishing on the boundary and on the complement of* $\overline{M}_x$*. Note that* $\int_{\overline{M}_x}\omega(x)dx$ *needs not be unity. Suppose that Assumptions 9–11 are satisfied. Then,*

$$
\begin{aligned}
\int_{\overline{M}_x} (\widehat{f} - f)\omega(x)dx &= O_p((N\sigma_N^2)^{-1/2}) \\
\int_{\overline{M}_x} (\widehat{f}_w - f_w)\omega(x)dx &= O_p((N\sigma_N^4)^{-1/2}) \\
\int_{\overline{M}_x} (\widehat{f}_{x_s} - f_{x_s})\omega(x)dx &= O_p((N\sigma_N^2)^{-1/2}) \\
\int_{\overline{M}_x} (\widehat{f}_{ww} - f_{ww})\omega(x)dx &= O_p((N\sigma_N^6)^{-1/2}) \\
\int_{\overline{M}_x} (\widehat{f}_{wy_2} - f_{wy_2})\omega(x)dx &= O_p((N\sigma_N^6)^{-1/2}) \\
\int_{\overline{M}_x} (\widehat{f}_{wx_s} - f_{wx_s})\omega(x)dx &= O_p((N\sigma_N^4)^{-1/2}) \\
\int_{\overline{M}_x} (\widehat{\widetilde{f}} - \widetilde{f})\omega(x)dx &= O_p((N\sigma_N)^{-1/2})
\end{aligned}
$$

$$\int_{\overline{M}_x} (\widehat{\widetilde{f}}_w - \widetilde{f}_w)\omega(x)dx = O_p((N\sigma_N^3)^{-1/2})$$
$$\int_{\overline{M}_x} (\widehat{\widetilde{f}}_{x_s} - \widetilde{f}_{x_s})\omega(x)dx = O_p((N\sigma_N)^{-1/2})$$
$$\int_{\overline{M}_x} (\widehat{\widetilde{f}}_{ww} - \widetilde{f}_{ww})\omega(x)dx = O_p((N\sigma_N^5)^{-1/2})$$
$$\int_{\overline{M}_x} (\widehat{\widetilde{f}}_{wy2} - \widetilde{f}_{wy2})\omega(x)dx = O_p((N\sigma_N^5)^{-1/2})$$
$$\int_{\overline{M}_x} (\widehat{\widetilde{f}}_{wx_s} - \widetilde{f}_{wx_s})\omega(x)dx = O_p((N\sigma_N^3)^{-1/2}).$$

Proof. First note that $f(y_1, y_2, w, x)$ is the joint pdf of mixed data where Y_1 is binary and the other variables are continuous, and $\widehat{f}(y_1, y_2, w, x)$ is a kernel density estimator for f. It is known that when estimating a density and density derivatives of mixed data using the kernel method, the discrete variable part does not affect the rate of convergence of the kernel estimates Li and Racine (2003).

Next, since x is the variable of integration, we may use integration by parts for the terms differentiated with respect to x. By the assumption imposed on $\omega(x)$, , the values and derivatives of $\omega(x)$ vanish on the boundary of $\overline{M}_x$. Hence,

$$\int_{\overline{M}_x} (\widehat{f}_{x_s} - f_{x_s})\omega(x)dx = -\int_{\overline{M}_x} (\widehat{f} - f)\left(\frac{\partial\omega(x)}{\partial x_s}\right)dx$$
$$\int_{\overline{M}_x} (\widehat{f}_{wx_s} - f_{wx_s})\omega(x)dx = -\int_{\overline{M}_x} (\widehat{f}_w - f_w)\left(\frac{\partial\omega(x)}{\partial x_s}\right)dx$$
$$\int_{\overline{M}_x} (\widehat{\widetilde{f}}_{x_s} - \widetilde{f}_{x_s})\omega(x)dx = -\int_{\overline{M}_x} (\widehat{\widetilde{f}} - \widetilde{f})\left(\frac{\partial\omega(x)}{\partial x_s}\right)dx$$
$$\int_{\overline{M}_x} (\widehat{\widetilde{f}}_{wx_s} - \widetilde{f}_{wx_s})\omega(x)dx = -\int_{\overline{M}_x} (\widehat{f}_w - \widetilde{f}_w)\left(\frac{\partial\omega(x)}{\partial x_s}\right)dx$$

The results follow by Lemma 5.3 in Newey (1994). For the equations about the joint density f, let $k_1 = 2$, $k_2 = 2$, l be the order of the derivative, $t = x$, $h_0 = f$, $m(\widehat{h}) = \left(\int_{\overline{M}_x}(\partial^l \widehat{f})\omega(x)dx\right)$, and $m(h_0) = \left(\int_{\overline{M}_x}(\partial^l f)\omega(x)dx\right)$. For the equations about the marginal density $\widetilde{f}$, let $k_1 = 1$, $k_2 = 2$, l be the order of the derivative, $t = x$, $h_0 = \widetilde{f}$, $m(\widehat{h}) = \left(\int_{\overline{M}_x}(\partial^l \widehat{\widetilde{f}})\omega(x)dx\right)$, and $m(h_0) = \left(\int_{\overline{M}_x}(\partial^l \widetilde{f})\omega(x)dx\right)$. Then, the asymptotic normality result in the Lemma 5.3 in Newey (1994) implies that the convergence rate is $\sqrt{N\sigma_N^{k_1+2l}}$. $\square$

1.7.A.2 Proofs of the Theorems

Proof of Theorem 1.3.3. By Proposition 1.3.2, $(\overline{r}, d)$ is constant over the set $\overline{M}$. The system of linear equations (1.13) can be expressed in a matrix form

$$\begin{pmatrix} g_{b_g}^{(1)} \\ g_{b_g}^{(2)} \\ g_{b_g}^{(3)} \end{pmatrix} = \begin{pmatrix} g_{x_1}^{(1)} & g_{x_2}^{(1)} & 1 \\ g_{x_1}^{(2)} & g_{x_2}^{(2)} & 1 \\ g_{x_1}^{(3)} & g_{x_2}^{(3)} & 1 \end{pmatrix} \begin{pmatrix} \overline{r}_{b_g}^{1} \\ \overline{r}_{b_g}^{2} \\ d_{b_g} \end{pmatrix}$$

for $g = 1, 2$. By Assumptions 7 and 8, the first matrix on the right-hand-side is invertible. Premultiplying the inverse of the matrix gives us the unique solution for $(\overline{r}, d)$, so it is identified. $\square$

Proof of Theorem 1.4.2. Let $\overline{M}_b, \overline{M}_x, \mathcal{F}, \|\cdot\|, \widetilde{g}, g_{B|X}, \alpha_{b_j}, \beta_{x_s}, \Phi_{b_j,x_s}$ and Φ_{x_j,x_s} be defined as in Lemmas 1.7.1–1.7.5. Suppose that Assumptions 9–12 are satisfied.

First note that by the assumptions and Lemma B.3 in Newey (1994),

$$\|\widehat{f} - f\| = O_p(\sigma_N^s + \sqrt{\log(N)/(N\sigma_N^8)}).$$

By Assumption 11, $\sqrt{N\sigma_N^6}[\sqrt{\log(N)/N\sigma_N^8} + \sigma_N^s]^2 \to 0$. Hence,

$$\sqrt{N\sigma_N^6}\|\widehat{f} - f\|^2 = o_p(1). \tag{T.1}$$

By the definition of Φ_{b_j,x_s}, $\widehat{T}_{b_j,x_s} = \Phi_{b_j,x_s}(\widehat{f})$ and $T_{b_j,x_s} = \Phi_{b_j,x_s}(f)$. Let $h = \widehat{f} - f$. By Lemma 1.7.4, for $h = \widehat{f} - f$ such that $\|\widehat{f} - f\|$ is sufficiently small,

$$\widehat{T}_{b_j,x_s} - T_{b_j,x_s} = \Phi_{b_j,x_s}(f + (\widehat{f} - f)) - \Phi_{b_j,x_s}(f) = D\Phi_{b_j,x_s}(f; \widehat{f} - f) + R\Phi_{b_j,x_s}(f; \widehat{f} - f)$$

where $D\Phi_{b_j,x_s} = O_p(\|\widehat{f} - f\|)$ and $R\Phi_{b_j,x_s} = O_p(\|\widehat{f} - f\|^2)$. (T.1) implies that

$$\sqrt{N\sigma_N^6}R\Phi_{b_j,x_s}(f; \widehat{f} - f) = o_p(1).$$

Lemma 1.7.6 shows that the second order derivatives of the joint density dominate the asymptotic distribution. Combining the definition of $D\Phi_{b_j,x_s}(f; \widehat{f} - f)$ in Lemma 1.7.4 with the convergence rate in Lemma 1.7.6, it can be shown that

$$\sqrt{N\sigma_N^6}D\Phi_{b_1,x_s}(f;\widehat{f}-f)=\sqrt{N\sigma_N^6}\int_{\overline{M}_x}(\widehat{f}_{ww}-f_{ww})\frac{\mu\widetilde{f}\left(\beta_{x_s}(f)-\int_{\overline{M}_x}\beta_{x_s}(f)\mu dx\right)}{f_w\widetilde{f}-f\widetilde{f}_w}dx+o_p(1)$$

$$\sqrt{N\sigma_N^6}D\Phi_{b_2,x_s}(f;\widehat{f}-f)=\sqrt{N\sigma_N^6}\int_{\overline{M}_x}(\widehat{f}_{wy_2}-f_{wy_2})\frac{\mu\widetilde{f}\left(\beta_{x_s}(f)-\int_{\overline{M}_x}\beta_{x_s}(f)\mu dx\right)}{f_w\widetilde{f}-f\widetilde{f}_w}dx+o_p(1)$$

Let $\omega_1(x)=\frac{\mu\widetilde{f}}{f_w\widetilde{f}-f\widetilde{f}_w}\left(\beta_{x_1}(f)-\int_{\overline{M}_x}\beta_{x_1}(f)\mu dx\right)$, $\omega_2(x)=\frac{\mu\widetilde{f}}{f_w\widetilde{f}-f\widetilde{f}_w}\left(\beta_{x_2}(f)-\int_{\overline{M}_x}\beta_{x_2}(f)\mu dx\right)$, and let $W(x)=I_2\otimes\binom{\omega_1(x)}{\omega_2(x)}$ as in Theorem 1.4.2. By Assumptions 9 and 12, $\omega_1(x)$ and $\omega_2(x)$ are bounded and continuous, and equal to zero on the boundary and the complement of $\overline{M}_x$. Hence, Assumption 5.1 in Newey (1994) is satisfied. Assumptions 9 and 12 imply that Assumptions K, H, and Y in Newey (1994) are also satisfied. In addition, the bandwidth condition of Lemma 5.3 in Newey (1994) is satisfied by Assumption 11. Hence, it follows that

$$\sqrt{N\sigma_N^6}(\widehat{TT}_{BX}-TT_{BX})\xrightarrow{d}N(0,V)$$

where V is as defined in Theorem 1.4.2.

Next, by the definition of Φ_{x_j,x_s}, $\widehat{T}_{x_j,x_s}=\Phi_{x_j,x_s}(\widehat{f})$ and $T_{x_j,x_s}=\Phi_{x_j,x_s}(f)$. By Lemma 1.7.4, for $h=\widehat{f}-f$ such that $\|\widehat{f}-f\|$ is sufficiently small,

$$\widehat{T}_{x_j,x_s}-T_{x_j,x_s}=\Phi_{x_j,x_s}(\widehat{f})-\Phi_{x_j,x_s}(f)=D\Phi_{x_j,x_s}(f;\widehat{f}-f)+R\Phi_{x_j,x_s}(f;\widehat{f}-f)$$

where $D\Phi_{x_j,x_s}=O_p(\|\widehat{f}-f\|)$ and $R\Phi_{x_j,x_s}=O_p(\|\widehat{f}-f\|^2)$. Again, combining the definition of $D\Phi_{x_j,x_s}(f;\widehat{f}-f)$ and $R\Phi_{x_j,x_s}(f;\widehat{f}-f)$ with the convergence rate in Lemma 1.7.6, we have $|\widehat{T}_{x_j,x_s}-T_{x_j,x_s}|\xrightarrow{p}0$. Hence $\widehat{TT}_{XX}$ converges in probability to TT_{XX}. Hence, the result follows by Slutsky's Theorem. □

Proof of Theorem 1.4.3. It suffices to show that $\widehat{V}\xrightarrow{p}V$. The matrix multiplication of the integrand of V and $\widehat{V}$ gives that the elements of $\widehat{V}$ and V corresponding to the covariance between T_{b_j,x_s} and T_{b_k,x_l} is, respectively,

$$\left\{\int\left(\widehat{g}_{x_s}-\int\widehat{g}_{x_s}\mu(x)dx\right)\left(\widehat{g}_{x_l}-\int\widehat{g}_{x_l}\mu(x)dx\right)\left(\frac{\mu(x)^2\widehat{f}(y_2,w,x)}{(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2}\right)dx\right\}\widetilde{KK}_{b_j,b_k}$$

and

$$\left\{\int_{\overline{M}_x}\left(g_{x_s}-\int_{\overline{M}_x} g_{x_s}\mu(x)dx\right)\left(g_{x_l}-\int_{\overline{M}_x} g_{x_l}\mu(x)dx\right)\left(\frac{\mu(x)^2 f(y_2,w,x)}{(f_w\widetilde{f}-f\widetilde{f}_w)^2}\right)dx\right\}\widetilde{KK}_{b_j,b_k}$$

where

$$\widetilde{KK}_{b_j,b_k}=\left\{\int\left[\int\left(\frac{\partial^2 K(b,x)}{\partial b_1\partial b_j}\right)\right]\left[\int\left(\frac{\partial^2 K(b,x)}{\partial b_1\partial b_k}\right)\right]\right\}.$$

Since the terms are integrated over the compact set $\overline{M}_x$, it is enough to show that the integrand of the first equation converges in probability to that of the second equation uniformly over $\overline{M}_x$. $\mu(x)$ is a bounded function over the compact set $\overline{M}_x$, so it is uniformly bounded. Hence, it remains to show that

$$\sup_{\overline{M}_x}\Bigg|\left(\widehat{g}_{x_s}-\int\widehat{g}_{x_s}\mu(x)dx\right)\left(\widehat{g}_{x_l}-\int\widehat{g}_{x_l}\mu(x)dx\right)\left(\frac{\widehat{f}(y_2,w,x)}{(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2}\right)$$
$$-\left(g_{x_s}-\int_{\overline{M}_x} g_{x_s}\mu(x)dx\right)\left(g_{x_l}-\int_{\overline{M}_x} g_{x_l}\mu(x)dx\right)\left(\frac{f(y_2,w,x)}{(f_w\widetilde{f}-f\widetilde{f}_w)^2}\right)\Bigg|\xrightarrow{p}0.$$

By the definition of the functional β_{x_s}, $\widehat{g}_{x_s}=\beta_{x_s}(\widehat{f})$ and $g_{x_s}=\beta_{x_s}(f)$. By Lemma 1.7.3, for sufficiently small $\|\widehat{f}-f\|$, we can find a constant c_1 and c_2 such that for all $x\in\overline{M}_x$,

$$|\beta_{x_s}(\widehat{f})-\beta_{x_s}(f)|\le|D\beta_{x_s}(f;\widehat{f}-f)|+|R\beta_{x_s}(f;\widehat{f}-f)|\le c_1\|\widehat{f}-f\|+c_2\|\widehat{f}-f\|^2.$$

Hence,

$$\sup_{\overline{M}_x}\left|\beta_{x_s}(\widehat{f})-\beta_{x_s}(f)\right|\le c_1\|\widehat{f}-f\|+c_2\|\widehat{f}-f\|^2.$$

Since μ is bounded, $\int_{\overline{M}_x}\mu(x)dx=1$, and $\overline{M}_x$ is compact, it follows that

$$\begin{aligned}&\left|\int_{\overline{M}_x}\beta_{x_s}(\widehat{f})\mu dx-\int_{\overline{M}_x}\beta_{x_s}(f)\mu dx\right|\\ &\le\int_{\overline{M}_x}|\beta_{x_s}(\widehat{f})-\beta_{x_s}(f)|\mu dx\\ &\le c_1\|\widehat{f}-f\|+c_2\|\widehat{f}-f\|^2.\end{aligned}$$

And then, for all $x\in\overline{M}_x$,

$$\left|\left(\beta_{x_s}(\widehat{f})-\int_{\overline{M}_x}\beta_{x_s}(\widehat{f})\mu dx\right)-\left(\beta_{x_s}(f)-\int_{\overline{M}_x}\beta_{x_s}(f)\mu dx\right)\right|\qquad\text{(T.2)}$$

$$\leq 2c_1\|\widehat{f}-f\|+2c_2\|\widehat{f}-f\|^2.$$

Next, by Assumption 9, $f>\delta$ and $f_w\widetilde{f}-f\widetilde{f}_w>\delta$. For all $\widehat{f}$ such that $\|\widehat{f}-f\|\leq\min\{\delta/2,1\}$, $\widehat{f}>\delta/2$ and $\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w>\delta/2$. Hence,

$$\left|\frac{\widehat{f}}{(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2}-\frac{f}{(f_w\widetilde{f}-f\widetilde{f}_w)^2}\right|\leq\frac{|\widehat{f}(f_w\widetilde{f}-f\widetilde{f}_w)^2-f(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2|}{|(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2||(f_w\widetilde{f}-f\widetilde{f}_w)^2|}\leq\frac{64\|f\|^3\|\widehat{f}-f\|}{\delta^4}.$$

Denote $\widehat{\gamma}_{x_s}=\beta_{x_s}(\widehat{f})-\int_{\overline{M}_x}\beta_{x_s}(\widehat{f})\mu dx$ and $\gamma_{x_s}=\beta_{x_s}(f)-\int_{\overline{M}_x}\beta_{x_s}(f)\mu dx$. Then,

$$\begin{aligned}\left|\frac{\widehat{\gamma}_{x_s}\widehat{\gamma}_{x_l}\widehat{f}}{(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2}-\frac{\gamma_{x_s}\gamma_{x_l}f}{(f_w\widetilde{f}-f\widetilde{f}_w)^2}\right|&=\frac{\left|\widehat{\gamma}_{x_s}\widehat{\gamma}_{x_l}\widehat{f}(f_w\widetilde{f}-f\widetilde{f}_w)^2-\gamma_{x_s}\gamma_{x_l}f(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2\right|}{|(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2||(f_w\widetilde{f}-f\widetilde{f}_w)^2|}\\&\leq\frac{|\widehat{\gamma}_{x_s}-\gamma_{x_s}||\widehat{\gamma}_{x_l}||\widehat{f}|+|\gamma_{x_s}||\widehat{\gamma}_{x_l}-\gamma_{x_l}||\widehat{f}|}{|(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2|}\\&\quad+\frac{|\gamma_{x_s}||\gamma_{x_l}||\widehat{f}(f_w\widetilde{f}-f\widetilde{f}_w)^2-f(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2|}{|(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2||(f_w\widetilde{f}-f\widetilde{f}_w)^2|}\end{aligned}$$

By (T.2), $|\widehat{\gamma}_{x_s}-\gamma_{x_s}|$ and $|\widehat{\gamma}_{x_s}-\gamma_{x_s}|$ are bounded by a constant times $\|\widehat{f}-f\|$. $|\widehat{\gamma}_{x_s}|\leq|\gamma_{x_s}|+|D\gamma_{x_s}|+|R\gamma_{x_s}|$ and all the three terms are bounded by a constant times $\|\widehat{f}-f\|$, as shown in the proof of Lemma 1.7.3 and 1.7.4. Boundedness of $|f|$ an $|\widehat{f}|$ is guaranteed by Assumption 9 and the choice of $\widehat{f}$. Since the norm defined on $\mathcal{F}$ is the sup norm and the choice of constants does not depend on the point, we can find a constant $\overline{c}>0$ such that for all $x\in\overline{M}_x$,

$$\begin{aligned}&\Bigg|\left(\widehat{g}_{x_s}-\int\widehat{g}_{x_s}\mu(x)dx\right)\left(\widehat{g}_{x_l}-\int\widehat{g}_{x_l}\mu(x)dx\right)\left(\frac{\widehat{f}(y_2,w,x)}{(\widehat{f}_w\widehat{\widetilde{f}}-\widehat{f}\widehat{\widetilde{f}}_w)^2}\right)\\&\qquad-\left(g_{x_s}-\int_{\overline{M}_x}g_{x_s}\mu(x)dx\right)\left(g_{x_l}-\int_{\overline{M}_x}g_{x_l}\mu(x)dx\right)\left(\frac{f(y_2,w,x)}{(f_w\widetilde{f}-f\widetilde{f}_w)^2}\right)\Bigg|\leq\overline{c}\|\widehat{f}-f\|.\end{aligned}$$

This completes the proof. □

1.7.B Tables

Table 1.1: Simulation Result: Linear Model

N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
$r^1_{b_1} = 1$										
$k = 6$										
10,000	0.6539	0.6538	1.0585	1.1127	2.0504	90.1636	0.99	1.0524	0.2395	0.2907
25,000	0.8314	0.8262	0.8041	0.8208	1.1174	6.1943	0.99	1.0482	0.1740	0.1936
50,000	0.9193	0.8715	0.7082	0.7120	0.8508	8.4054	0.98	1.0342	0.1401	0.1424
100,000	0.9688	0.9074	0.5604	0.5607	0.6055	0.5276	0.97	1.0238	0.1077	0.1057
$k = 8$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
10,000	0.6722	0.6508	0.8450	0.9056	1.7082	30.6498	0.99	1.0208	0.2620	0.3035
25,000	0.8733	0.8356	0.6468	0.6585	0.9044	1.5142	0.98	1.0272	0.1920	0.2020
50,000	0.9028	0.8671	0.5474	0.5555	0.6693	1.9770	0.97	1.0173	0.1543	0.1483
100,000	0.9658	0.9299	0.4314	0.4323	0.4787	0.1494	0.97	1.0088	0.1176	0.1097
$r^1_{b_2} = -0.75$										
$k = 6$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
10,000	-0.3737	-0.3877	0.6781	0.7749	1.4583	46.3371	1.00	-0.7838	0.2603	0.3251
25,000	-0.6143	-0.6137	0.5580	0.5737	0.7930	3.0809	1.00	-0.7930	0.1939	0.2164
50,000	-0.7015	-0.6689	0.4671	0.4692	0.6040	4.1788	0.99	-0.7788	0.1378	0.1592
100,000	-0.7477	-0.6995	0.3904	0.3900	0.4298	0.2675	0.98	-0.7760	0.1091	0.1182
$k = 8$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
10,000	-0.4805	-0.4939	0.6261	0.6811	1.2141	15.4853	0.99	-0.7599	0.2866	0.3395
25,000	-0.7168	-0.6859	0.4957	0.4963	0.6420	0.7529	0.99	-0.7788	0.2147	0.2258
50,000	-0.7576	-0.7184	0.4031	0.4028	0.4753	0.9952	0.98	-0.7667	0.1520	0.1658
100,000	-0.8056	-0.7593	0.3770	0.3807	0.3399	0.0721	0.95	-0.7661	0.1191	0.1226

Table 1.2: Simulation Result: Linear Model (Continued)

$r^2_{b_1} = 0.5$										
$k = 6$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
10,000	0.3397	0.3259	0.9605	0.9729	2.1893	145.4085	1.00	0.5245	0.2113	0.2838
25,000	0.4604	0.4232	0.7768	0.7770	1.1580	20.5989	0.99	-0.7930	0.1939	0.2164
50,000	0.5203	0.4930	0.6659	0.6655	0.8395	3.0865	0.98	-0.7788	0.1378	0.1592
100,000	0.5103	0.5150	0.5300	0.5296	0.5939	0.2701	0.99	-0.7760	0.1091	0.1182
$k = 8$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
10,000	0.3776	0.3514	0.8449	0.8529	1.7831	49.1549	1.00	0.5078	0.2341	0.2984
25,000	0.4723	0.4381	0.6535	0.6535	0.9145	3.7409	1.00	0.5120	0.1711	0.1993
50,000	0.5066	0.4917	0.5261	0.5256	0.6633	1.7387	0.99	0.5032	0.1272	0.1464
100,000	0.4960	0.4703	0.4338	0.4334	0.4667	0.1249	0.97	0.5092	0.1056	0.1085
$r^2_{b_2} = 1$										
$k = 6$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
10,000	0.5128	0.5443	0.7104	0.8608	1.5565	73.9231	1.00	1.0517	0.2700	0.3175
25,000	0.7978	0.7805	0.6110	0.6431	0.8216	10.2013	0.99	1.0489	0.1980	0.2119
50,000	0.9764	0.9017	0.5118	0.5118	0.5960	1.5293	0.97	1.0349	0.1495	0.1563
100,000	1.0332	0.9710	0.4051	0.4060	0.4217	0.1294	0.96	1.0283	0.1140	0.1163
$k = 8$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
10,000	0.7055	0.6890	0.7042	0.7626	1.2675	25.1447	1.00	1.0183	0.2962	0.3338
25,000	0.9324	0.8733	0.5406	0.5443	0.6491	1.8810	0.99	1.0278	0.2211	0.2227
50,000	1.0447	0.9936	0.4610	0.4627	0.4710	0.8681	0.95	1.0178	0.1655	0.1637
100,000	1.0783	1.0072	0.4016	0.4088	0.3314	0.0745	0.92	1.0142	0.1249	0.1213

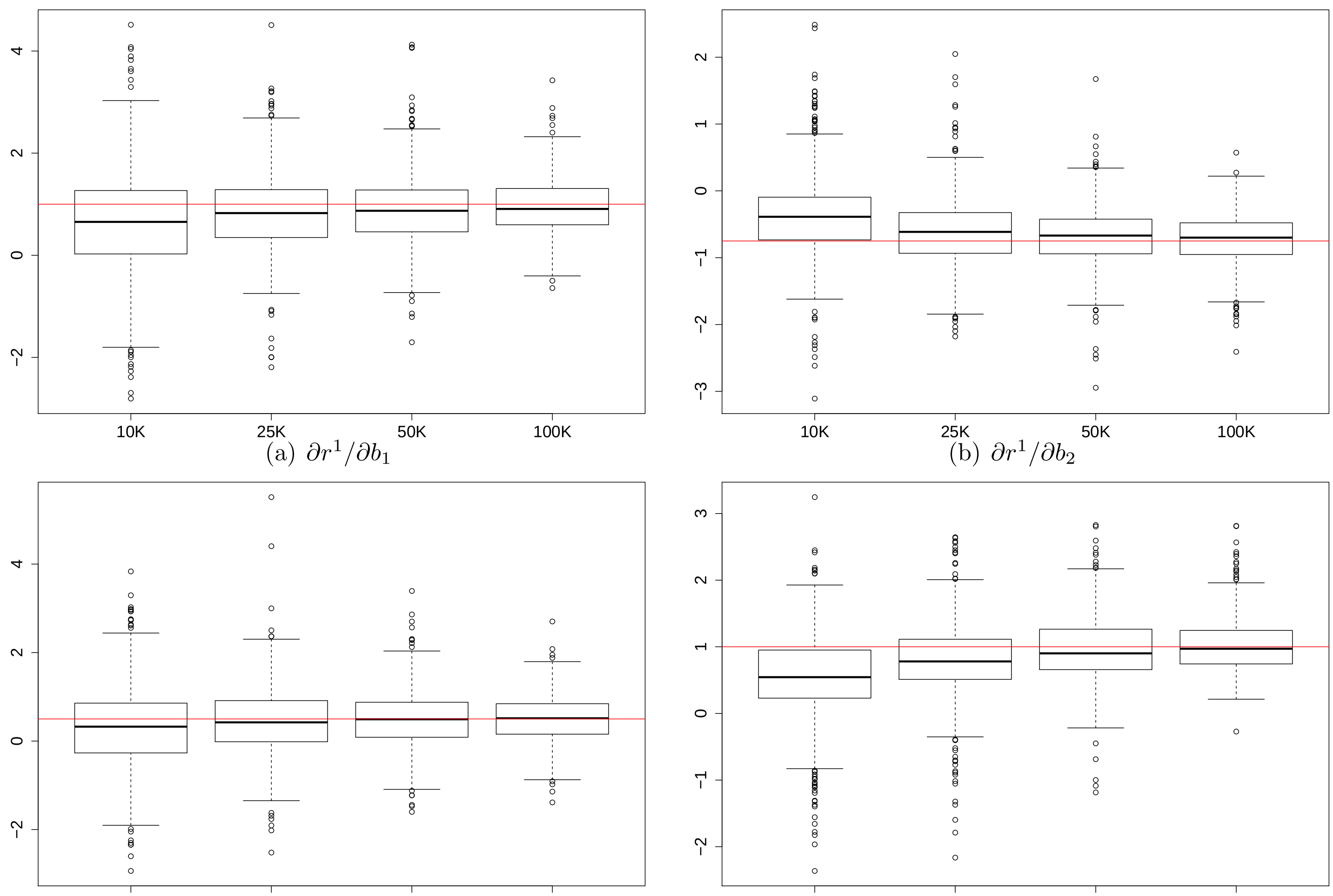
Table 1.3: Simulation Result: Nonlinear Model

$r^1_{b_1} = -0.2$										
$k = 6$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
100,000	-0.1552	-0.1332	1.4436	1.4429	6.3101	355.0971	1.00	-0.1705	0.0234	0.0396
250,000	-0.1099	-0.1083	1.0461	1.0490	3.5009	265.2227	1.00	-0.1596	0.0171	0.0269
500,000	-0.1496	-0.1156	0.9597	0.9601	1.9870	37.2885	0.99	-0.1530	0.0131	0.0201
1,000,000	-0.1413	-0.0964	0.8239	0.8252	1.3143	30.9672	0.97	-0.1507	0.0105	0.0152
$k = 8$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
100,000	-0.1202	-0.1689	0.9897	0.9919	3.5186	163.0602	1.00	-0.1343	0.0246	0.0396
250,000	-0.1292	-0.1330	0.9087	0.9106	1.7953	48.8344	0.99	-0.1247	0.0176	0.0270
500,000	-0.1463	-0.1600	0.8084	0.8094	1.1108	12.4573	0.98	-0.1187	0.0134	0.0202
1,000,000	-0.1983	-0.1776	0.6984	0.6978	0.7815	1.1685	0.96	-0.1167	0.0106	0.0152
$r^1_{b_2} = -1$										
$k = 6$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
100,000	-0.0893	-0.1243	0.7231	1.1624	4.1868	155.9989	0.98	-1.0808	0.0966	0.1020
250,000	-0.2222	-0.2417	0.6192	0.9938	2.3231	117.0002	0.96	-1.0722	0.0732	0.0692
500,000	-0.3185	-0.2935	0.5558	0.8790	1.3187	16.4386	0.91	-1.0616	0.0550	0.0519
1,000,000	-0.4207	-0.4350	0.5622	0.8069	0.8722	13.6235	0.88	-1.0574	0.0454	0.0390
$k = 8$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
100,000	-0.2098	-0.2393	0.6702	1.0357	2.3346	71.6969	0.96	-1.0588	0.1035	0.1021
250,000	-0.3997	-0.4572	0.7103	0.9295	1.1917	21.5846	0.91	-1.0542	0.0774	0.0695
500,000	-0.5900	-0.5591	0.6009	0.7270	0.7372	5.4788	0.88	-1.0458	0.0575	0.0521
1,000,000	-0.6884	-0.7016	0.5680	0.6474	0.5186	0.5023	0.89	-1.0443	0.0472	0.0392

Table 1.4: Simulation Result: Nonlinear Model (Continued)

$r^2_{b_1} = -0.1$										
$k = 6$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
100,000	-0.1004	-0.0309	1.7169	1.7151	7.4219	1020.8505	1.00	-0.1134	0.0264	0.0447
250,000	-0.1940	-0.1726	1.0857	1.0886	3.4340	335.8757	0.99	-0.1062	0.0167	0.0296
500,000	-0.0782	-0.0468	0.9728	0.9721	1.9691	36.1284	0.98	-0.1020	0.0133	0.0218
1,000,000	-0.0833	-0.0938	0.8310	0.8303	1.3844	29.5566	0.98	-0.0995	0.0103	0.0162
$k = 8$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
100,000	-0.0947	-0.0125	1.2861	1.2848	3.3784	106.1826	0.99	-0.1066	0.0288	0.0458
250,000	-0.1728	-0.0913	0.9952	0.9969	1.7948	40.5969	0.98	-0.0991	0.0179	0.0302
500,000	-0.0559	-0.0748	0.8522	0.8525	1.1082	13.3834	0.96	-0.0946	0.0140	0.0221
1,000,000	-0.1307	-0.0968	0.7357	0.7356	0.7749	0.7629	0.96	-0.0920	0.0107	0.0163
$r^2_{b_2} = 1$										
$k = 6$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
100,000	0.0780	0.0984	0.7798	1.2070	4.9250	451.4259	0.98	1.0883	0.1163	0.1153
250,000	0.1617	0.1784	0.5999	1.0305	2.2785	147.7338	0.95	1.0714	0.0798	0.0763
500,000	0.2862	0.2968	0.5968	0.9301	1.4428	47.8213	0.94	1.0598	0.0584	0.0563
1,000,000	0.3314	0.3843	0.5742	0.8810	0.9187	13.0154	0.87	1.0594	0.0475	0.0417
$k = 8$										
N	Mean	Med	STDEV	RMSE	$\widehat{SD}$	RMSE	95%	Mean	STDEV	$\widehat{SD}$
100,000	0.2298	0.2490	0.6779	1.0256	2.2417	46.7785	0.97	1.0659	0.1264	0.1180
250,000	0.3471	0.3773	0.6525	0.9226	1.1913	17.9479	0.92	1.0524	0.0852	0.0777
500,000	0.6158	0.6102	0.6254	0.7335	0.7355	5.8909	0.93	1.0431	0.0615	0.0570
1,000,000	0.6721	0.6758	0.5778	0.6639	0.5142	0.3285	0.86	1.0456	0.0494	0.0420

1.7.C Figures



(a) $\partial r^1/\partial b_1$

(b) $\partial r^1/\partial b_2$

(c) $\partial r^2/\partial b_1$

(d) $\partial r^2/\partial b_2$

Figure 1.1: Average Derivative Estimator: Linear case, 6th-order Gaussian Kernel

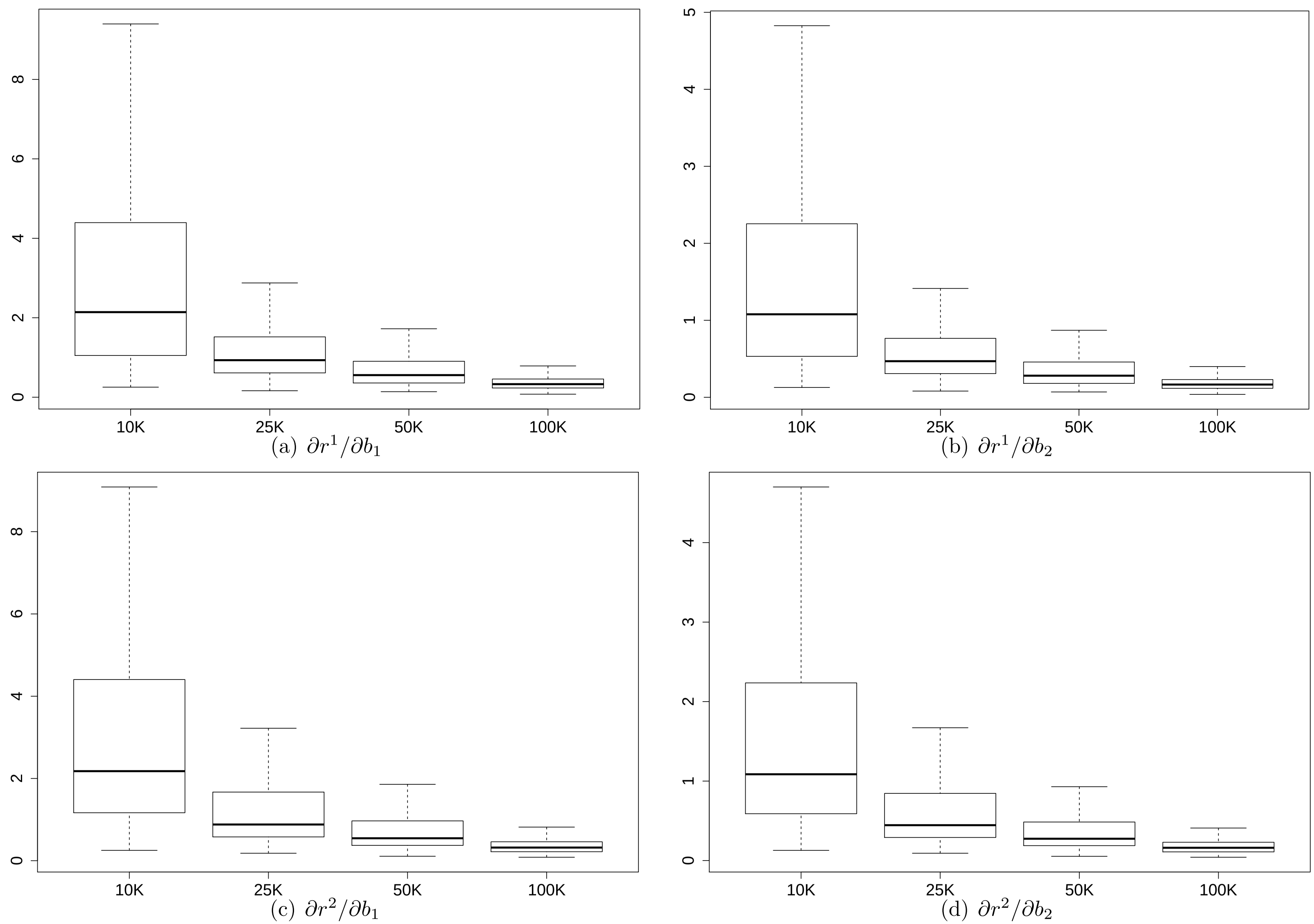


Figure 1.2: Asymptotic Variance Estimator: Linear case, 6th-order Gaussian Kernel

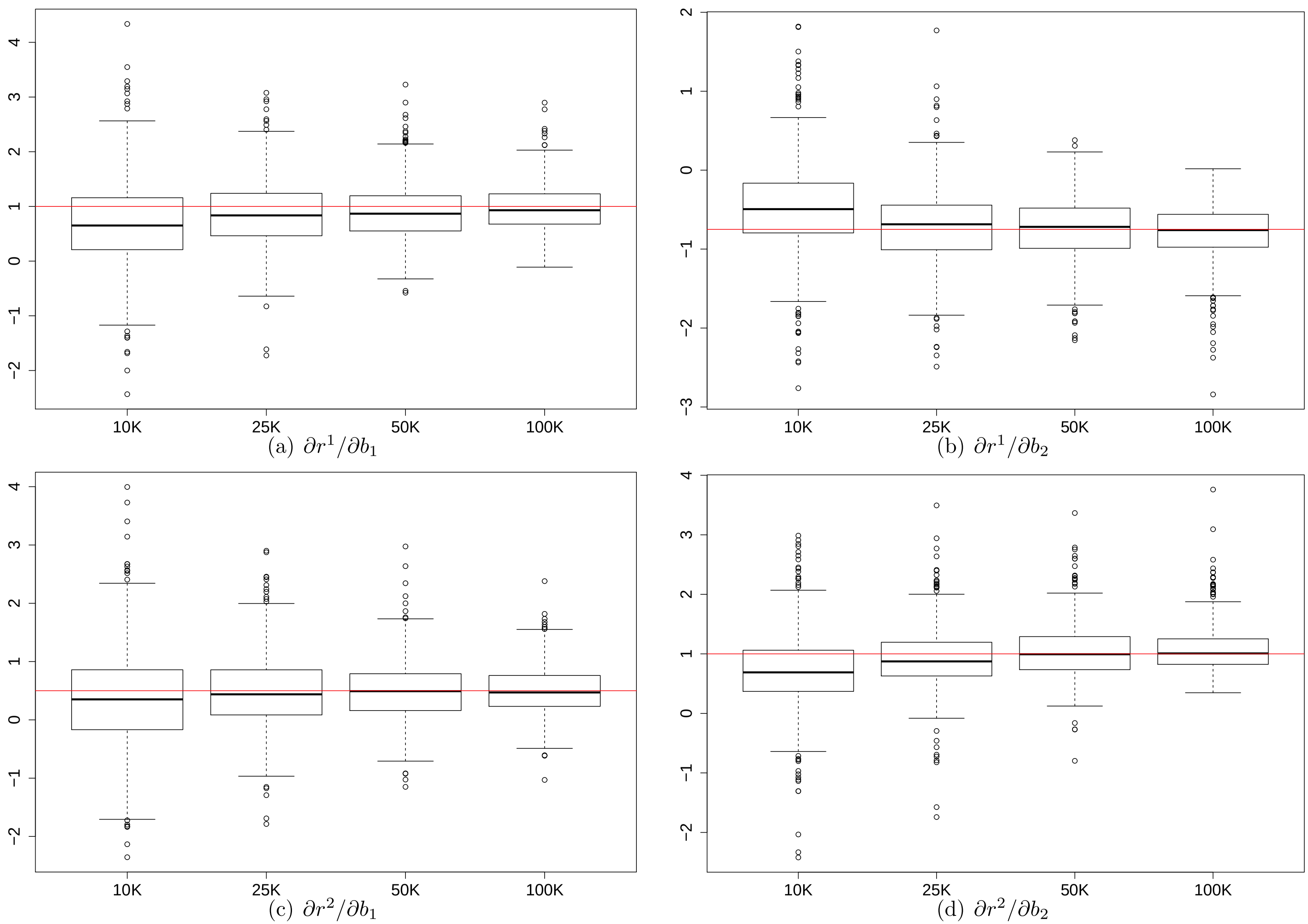


Figure 1.3: Average Derivative Estimator: Linear case, 8th-order Gaussian Kernel

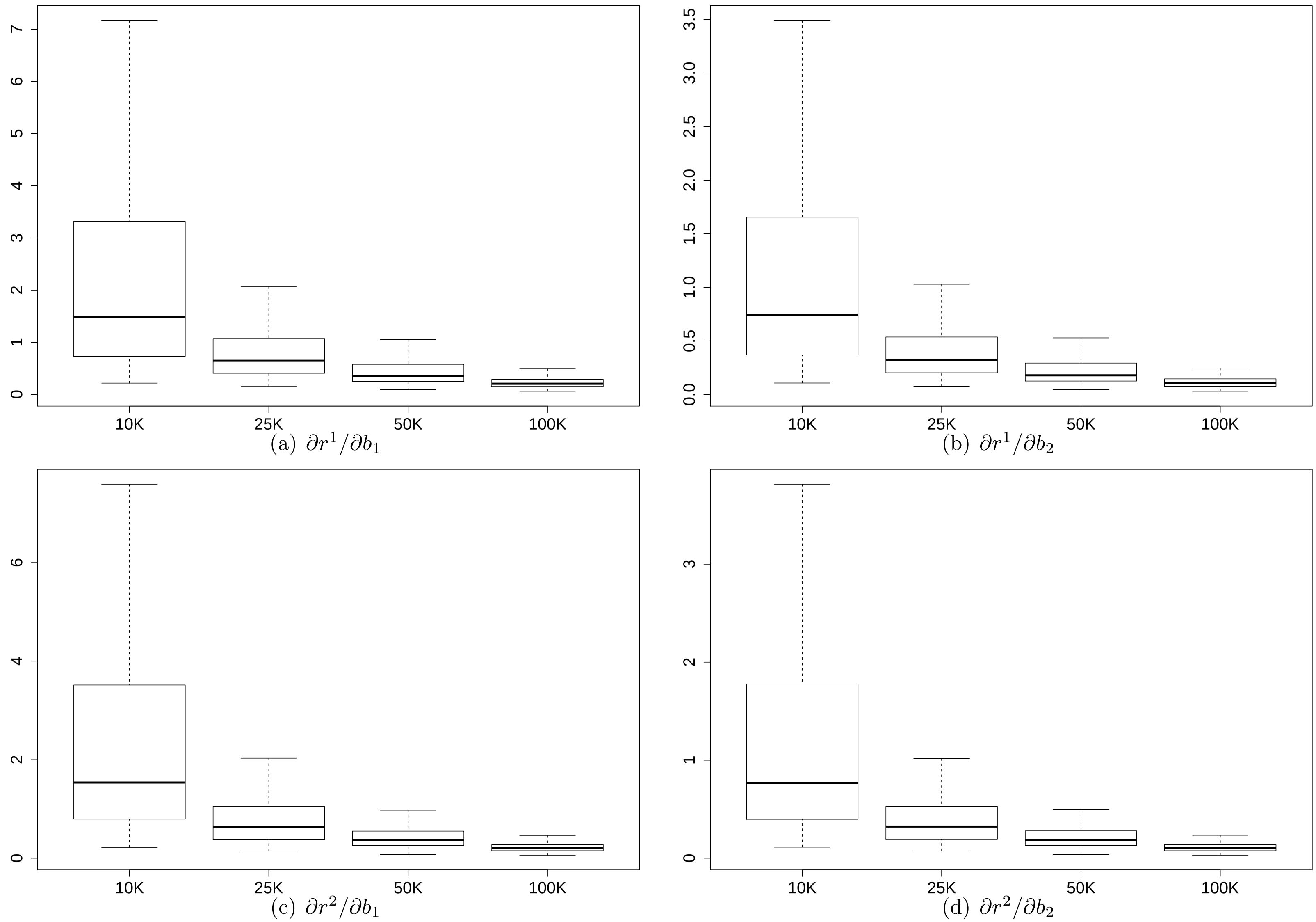
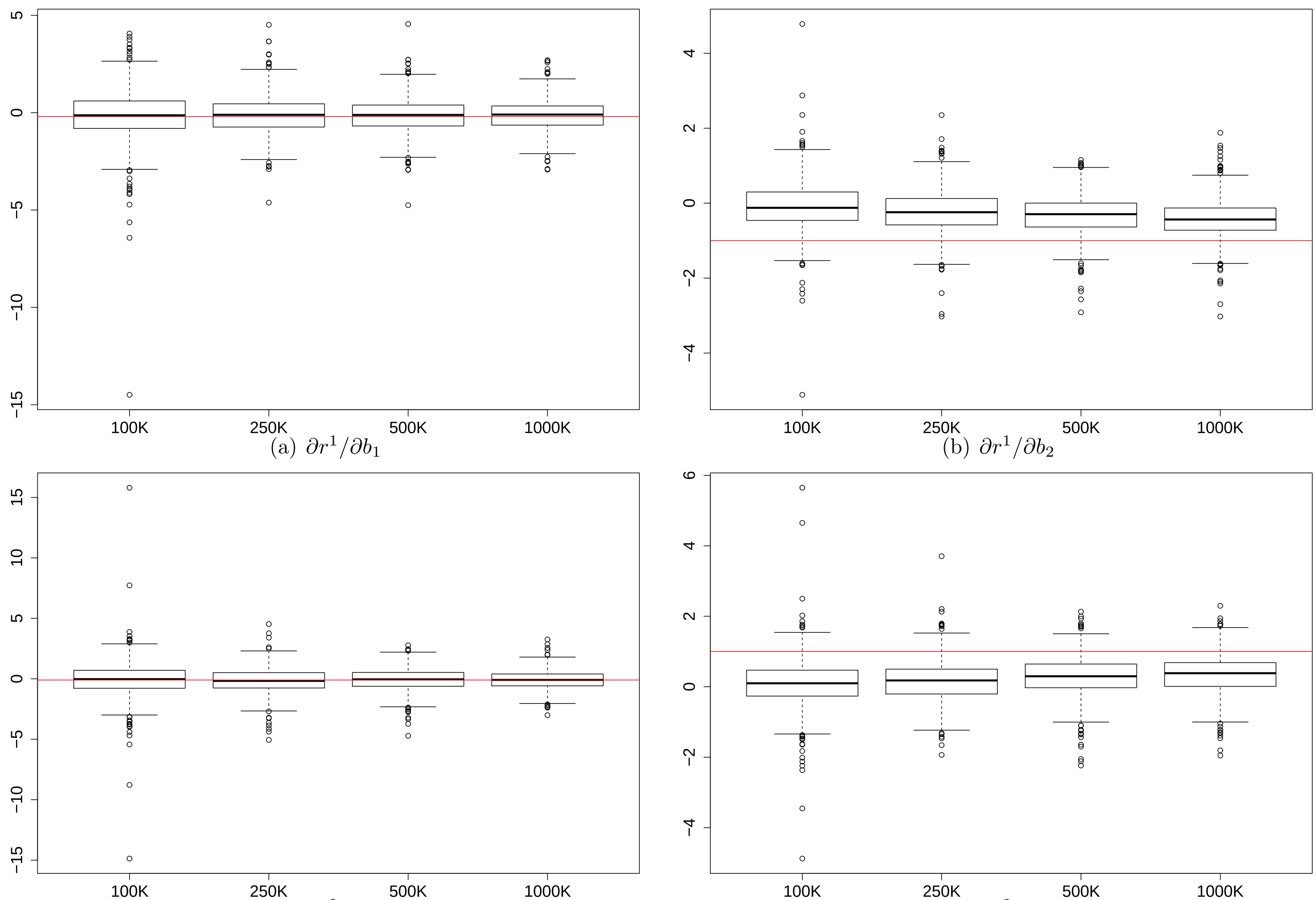


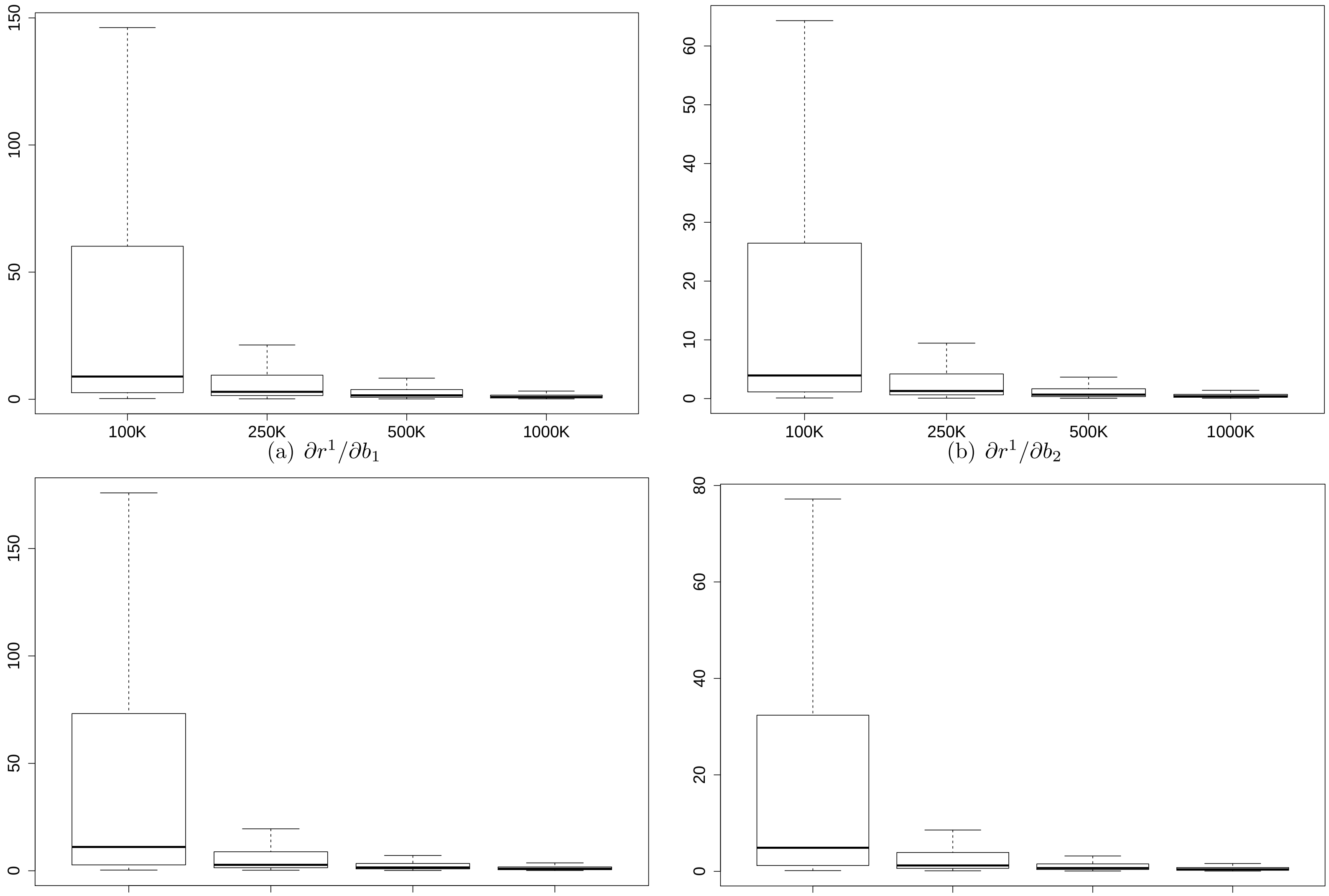
Figure 1.4: Asymptotic Variance Estimator: Linear case, 8th-order Gaussian Kernel



(a) $\partial r^1/\partial b_1$ (b) $\partial r^1/\partial b_2$

(c) $\partial r^2/\partial b_1$ (d) $\partial r^2/\partial b_2$

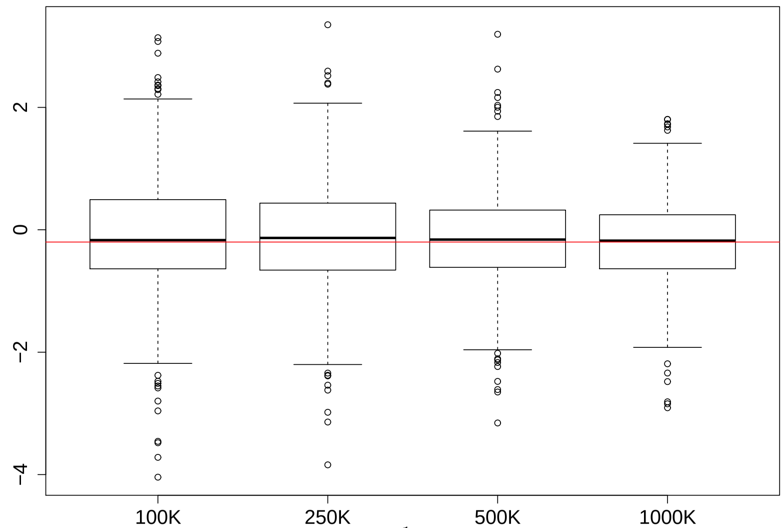
Figure 1.5: Average Derivative Estimator: Nonlinear case, 6th-order Gaussian Kernel



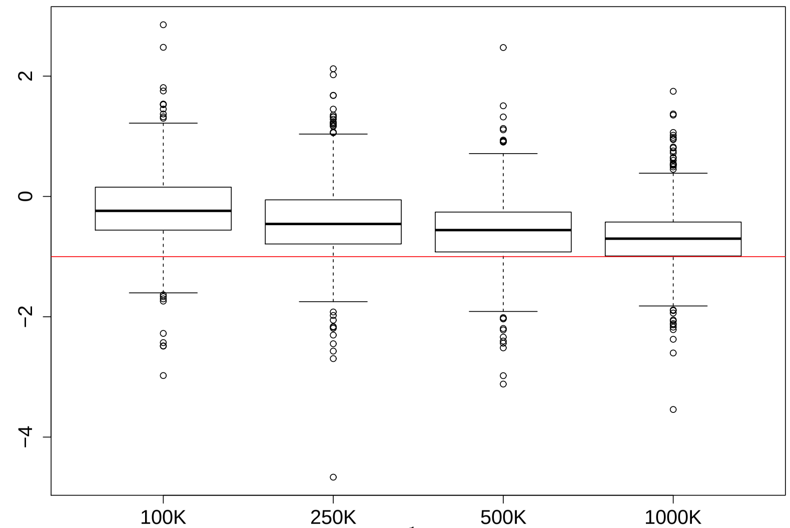
(a) $\partial r^1/\partial b_1$ (b) $\partial r^1/\partial b_2$

(c) $\partial r^2/\partial b_1$ (d) $\partial r^2/\partial b_2$

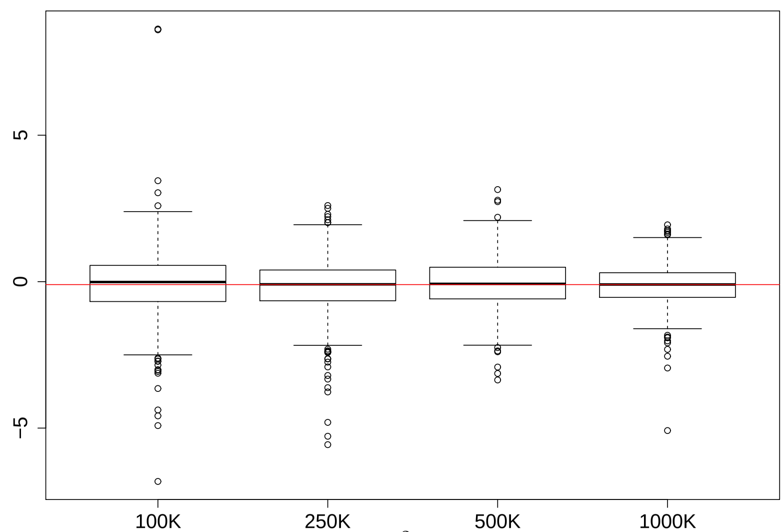
Figure 1.6: Asymptotic Variance Estimator: Nonlinear case, 6th-order Gaussian Kernel



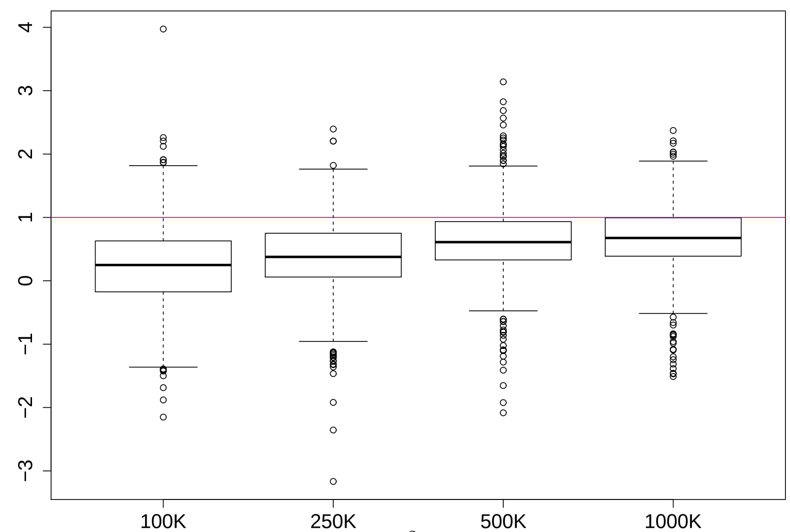
(a) $\partial r^1/\partial b_1$



(b) $\partial r^1/\partial b_2$

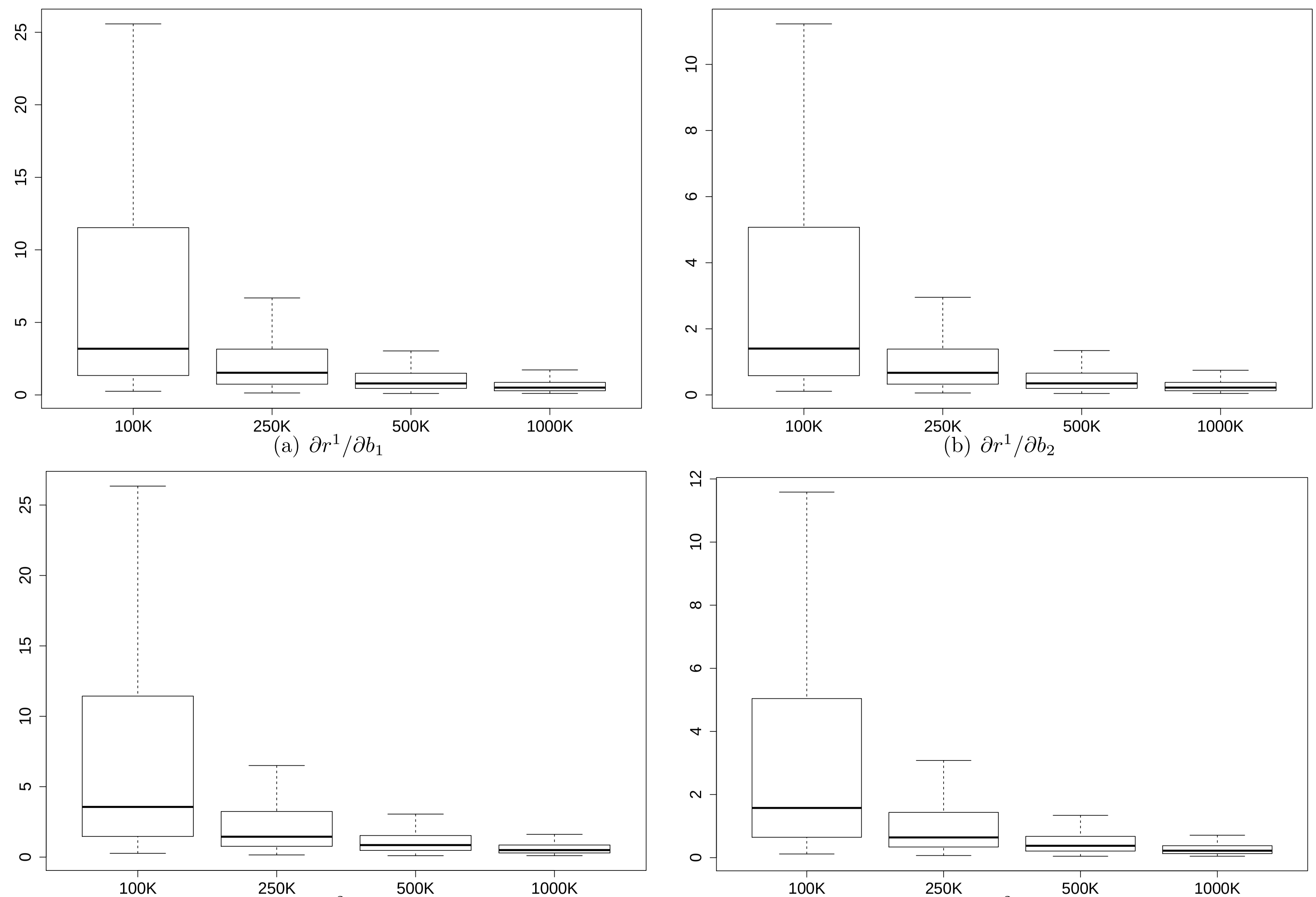


(c) $\partial r^2/\partial b_1$



(d) $\partial r^2/\partial b_2$

Figure 1.7: Average Derivative Estimator: Nonlinear case, 8th-order Gaussian Kernel



(a) $\partial r^1/\partial b_1$ (b) $\partial r^1/\partial b_2$

(c) $\partial r^2/\partial b_1$ (d) $\partial r^2/\partial b_2$

Figure 1.8: Asymptotic Variance Estimator: Nonlinear case, 8th-order Gaussian Kernel

CHAPTER 2

The Determinants of Smartphone Adoption among the Elderly in South Korea

2.1 Introduction

Smartphones and mobile devices have become indispensable in daily life. As of January 2015, the number of Korean smartphone subscribers reached 40.3 million[1]. The advent of smartphones has caused a dramatic change in access to information and media, leading to a super-connected world of real-time services customized for each individual, location-based services, and the Internet of Things (IoT). The constant dissemination of new technologies creates new areas of digital divide. Hence, the analysis of this division has become increasingly important for various stakeholders in or related to the industry.

For policymakers, digital divides serve to exacerbate the disparity of information, relations, social participation, creativity, productivity, and income among different population groups. For example, those alienated within a social system and who lack reliable ICT services are often excluded from the benefits that new technologies provide. Market-oriented planners, such as mobile app developers and start-up entrepreneurs, seek to understand the evolution of ICT markets and the extent to which product and service provisions can meet expected demand. Furthermore, they work to stimulate markets by providing platforms for newly developed services (Robertson et al., 2007).

The Korean government has taken a comprehensive approach to overcoming digital divide issues since the early stages of informatization. Various strategies have been implemented

[1]Source: IT Statistics of Korea (http://www.itstat.go.kr/en/)

throughout the country since the enactment of the the Act on Narrowing the Digital Divide in 2001 and the first and second Comprehensive Plan for Addressing the Digital Divide, established in 2001 and 2005, respectively. The First Comprehensive Plan sought to foster an information culture environment, whereas and the Second Comprehensive Plan focused on developing the necessary technologies to improve access to information and user environments for the underprivileged. The latter were classified into four categories: handicapped, economically unfortunate (recipients of national basic livelihood guarantees), the elderly (aged 50 and over), and rural (farmers and fishermen), with detailed implementation plans for each category.

During this time period, The Ministry of Information and Communication established the National Information Society Agency (NIA) in 2002 to support the development of policies related to the national informatization of national agencies, the creation of an information culture, and narrowing the digital divide. The NIA began at this time to disseminate surveys on the digital divide, leading to its annual publishing of *The Survey on the Digital Divide Index and Status*. Thanks to these efforts of the government, the digital divide related to traditional information devices and services, such as desktop PCs and wired internet services, all four categories of the underprivileged have improved over the past few decades. According to an NIA report in 2014, the overall informatization level of the underprivileged increased by 23.3 points, from 53.3 points in 2005 to 76.6 points in 2014. The informatization levels of the persons ages 50 and over has also improved, from 49.3 points in 2005 to 74.3 points in 2014 (National Information Society Agency (NIA), 2015).

Since new technological smart devices and mobile wireless internet have emerged, the digital divide has in turn become multi-layered. In order to examine this multifaceted divide, the NIA extended its survey to include new technologies created since 2013, as well as invented a new index called the "smart digital divide index." According to the NIA report in 2014, with the entire population represented by 100 points, the overall smart informatization level of the underprivileged was 57.4 points, and of persons ages 50 and over, 54.3 points. Among disadvantaged minority groups affected by the digital divide, there are several reasons to focus on information services in relation to the elderly in Korea. Older persons have little incentive to become familiar with new technologies due to the fact that many have no education concerning ICT, with the private costs of such education exceeding the benefits of obtaining new knowledge. This is unfortunate, as the elderly represent a demographic of individuals who can reap major benefits from informatization, especially social and medical benefits (Hagberg, 2004). Moreover, many older persons belong to low-income and rural demographics, both of which serve to make the digital divide more severe.

The goal of this paper is to analyze determinants affecting the adoption of smartphones among the elderly. More specifically, I examine individual, household, and regional factors that can influence the preferences of the elderly with regard to obtaining a new technological device. Using a random effect binary logit model with data from the Korean Media Panel Survey, I estimate the marginal effect of values related to the probability of owning a smartphone over time. To date, several previous studies have focused on the traditional digital divide caused by disparities in access to computer and internet service in the home, yet very little is known about the existence of smartphones among the elderly. The majority of studies on smartphones examine populations that include younger generations, arrive at results via the simple analysis of survey data, and fail to consider many socioeconomic variables. This paper contributes to knowledge of the technological environment following the introduction of smartphones and other smart mobile devices, with particular attention given to local and governmental informatization policies for the elderly. For firms in the ICT sector, the results of this paper can be used in marketing for the stimulation of demand for new ICT devices.

The two main conclusions obtained in this paper are as follows. First, smartphone ownership among the elderly is mainly determined by personal rather than family characteristics. With every year of age, the probability of an elderly persons switching to a smartphone decreases by 0.8%. Education and familiarity with traditional information technologies also affects choice probability. The choice probability of those who do not finish secondary education is 7.5% lower than that of high school graduates, while the regular and heavy internet users have a 10% and 20% higher probability of using smartphones, respectively. These figures can be contrasted with those related to computer and internet access, where household characteristics such as family structure and family income are more important factors (Fairlie, 2004; Chaudhuri et al., 2005). Indeed, a smartphone is an individual mobile device that is not shared by other household members, and therefore personal preferences are most integral to the decision.

Next, the area where a person lives has a significant effect on the probability of their owning a smartphone. This effect does not diminish even after controlling for macroeconomic variables. Living in Chungnam, Jeonnam, Gyeongbuk, or Gyeongnam, which are located in the Southern part of the Korean peninsula, significantly decreases the probability of ones having a smart phone by more than 5%. A possible reason for this is regional imbalance. The areas that show significantly lower probabilities are far from the Seoul Metropolitan Area (SMA). Over the last two decades, the development of economy, education, public welfare, culture, and communication in Korea has been concentrated in the SMA, which may play a role in usage of smartphones and other ICT, as well as informatization.

This paper is organized as follows. In the remainder of this section, I will briefly review the previous literature related to the subject. The mobile phone industry and current social issues related to ageing and informatization among the elderly in Korea will be discussed in Section 2. Section 3 presents the data used for analysis in this paper. Section 4 includes a description of a consumer preference model. Estimation results are presented and discussed in Section 5, and conclusions are presented in Section 6.

Related Literature

There are two streams of literature on the digital divide. One stems from the Technology Acceptance Model (TAM), which derives from Davis (1989). Davis suggested that the adoption of a technology is potentially influenced by consumers general perceptions of it as a useful communicating and interactive medium. The TAM literature suggests that when an ICT adoption choice is made, providing that it is not influenced unduly by peer pressure, perceptions as to a technologys usefulness and ease-of-use are key drivers. The original TAM model included a psychological framework that did not consider personal or household factors or product characteristics. Brown and Venkatesh (2005) point out that key demographics play significant roles throughout family lifecycles. They model computer adoption in terms of attitudinal beliefs, hedonic outcomes, word-of-mouth, media effects, and control beliefs, as well as household socio-demographics (e.g., age, income, marital status, child age).

Another stream of study concerning personal and household technology choices involves the application of econometric analyses. For example, Kridel et al. (1999) utilized a binary response model on high-speed internet choice using survey data captured from individuals that had residential internet access. Underlying this model is a concept of consumer utility that assumes consumers will always seek to maximize product-specific utility according to any constraints, such as product price and disposable income. Chaudhuri et al. (2005) utilize a direct residential internet model with a focus on the digital divide. They show how determinants of internet access may be assessed using a binary choice framework, expanding upon the work of Kridel et al. (1999) by assessing residential ICT choice across all households, not among those that already use the internet. Robertson et al. (2007) combine two approaches. They applied the basic findings of a technology adoption model into a discrete choice framework, heteroscedastically utilizing a heterogeneous probit model to study heterogeneity in household computer adoption.

The above-mentioned literature focuses on computer purchases and fixed-line Internet subscriptions as sources of digital divide. There are very few studies dealing with the dispersion of smartphones and other personalized digital devices as sources of digital divide. Researchers that have examined fixed-to-mobile substitutions (FMS) in telecommunications industry contexts have not considered substitutions between traditional desktop internet access and internet access on mobile devices. Grzybowski (2014) analyzes the substitutability between fixed-line and mobile telephony in 27 EU countries using cross-country panel data on household choices of telecommunications technologies. Rennhoff and Routon (2016) examine both wired and wireless US telephone services markets to understand consumer behaviors and welfare, estimating demand for both landline and wireless telephone services, and compute the consumer welfare due to the introduction of the smartphone.

2.2 Mobile Phone Industry and Informatization of the Elderly in Korea

2.2.1 Mobile Phone Industry

The first smartphone appeared in 2008 in Korea, but it was after the introduction of Apples iPhone 3G in 2009 that the smartphone came into wide use. The first model of Android smartphone, the Motorola Motoroi, came on the market in January of 2010. From there forward, the adoption of smartphones occurred rapidly. The proportion of smartphone users increased from 15.6% in 2010 to 70.9% in 2014. As expected, younger generations switched to the new service earlier than their elders. According to Gallup Korea Daily Opinion, more than 90% of 20- and 30-year-olds used smartphones by the end of 2012, with 90% of people in their 40s adopting smartphones by July 2014. The elderly have been reluctant to buy the new devices. By the second half of 2014, only 40% of people over 60 began to use a smartphone. Table 2.1 shows an overview of the smartphone distribution rate in Korea.

The rapid diffusion of the smartphone is due to the marketing strategies of mobile service

Table 2.1: Mobile Phone Service

Year	Mobile Phone Users	Portion of Smartphone Users	
		Total	Over 50
2010	50,767,241	15.6%	1.0%
2011	52,506,793	43.0%	6.2%
2012	53,624,427	61.1%	18.8%
2013	54,680,840	68.6%	41.5%
2014	57,207,957	70.9%	51.4%

Source: Ministry of Science, ICT and Future Planning providers and device manufacturers. In relation to these, three kinds of subsidies are granted to consumers. Many device manufacturers offer a subsidy or cash rebate to boost demand for their devices. Mobile service providers offer discounts in terms of one- or two-year phone contracts. Lastly, authorized retail stores offer discounts on phone charges by utilizing commission. As a result, consumers can purchase new devices much lower than factory price. The size of a subsidy is typically larger for a new device, and therefore consumers are more likely to adopt smartphones.

As the Korean mobile phone market reached saturation in 2011, subsidy competition became severe. Three mobile network operators (MNO), SK Telecom (SKT), Korea Telecom (KT) and LG U+, differentiated consumers to entice brand switching. To safeguard consumers welfare, the Korean government regulated subsidy competition, setting a maximum subsidy amount of KRW 270,000, and charging a fine or imposing suspensions of business if they were caught offering more than the limit. Furthermore, to revitalize the telecommunication market, the government promoted the entry of mobile virtual network operators (MVNO).[2]

Table 2.2: Market Shares of Mobile Service Operator

Year	SKT			KT			LG U+		
	MNO	MVNO	Total	MNO	MVNO	Total	MNO	MVNO	Total
2011	50.46%	0.11%	50.57%	30.95%	0.60%	31.54%	17.82%	0.06%	17.89%
2012	49.52%	0.76%	50.28%	29.55%	1.22%	30.77%	18.55%	0.40%	18.95%
2013	48.07%	1.95%	50.02%	27.95%	2.14%	30.09%	19.43%	0.46%	19.89%
2014	46.27%	3.74%	50.02%	26.65%	3.63%	30.29%	19.06%	0.64%	19.69%

Source: Ministry of Science, ICT and Future Planning

Table 2.2 shows the market shares of three Korean mobile network operators and MVNOs using their network infrastructure. SK Telecom is the largest wireless carrier in Korea, leading the market with a 50% share of the total market. Korea Telecom is the second

[2] A mobile virtual network operator, or MVNO, is a wireless communications services provider that does not own the network infrastructure, especially the frequency and base transceiver station, and that obtains bulk access to network services from a mobile network operator at wholesale rates, then sets retail prices independently.

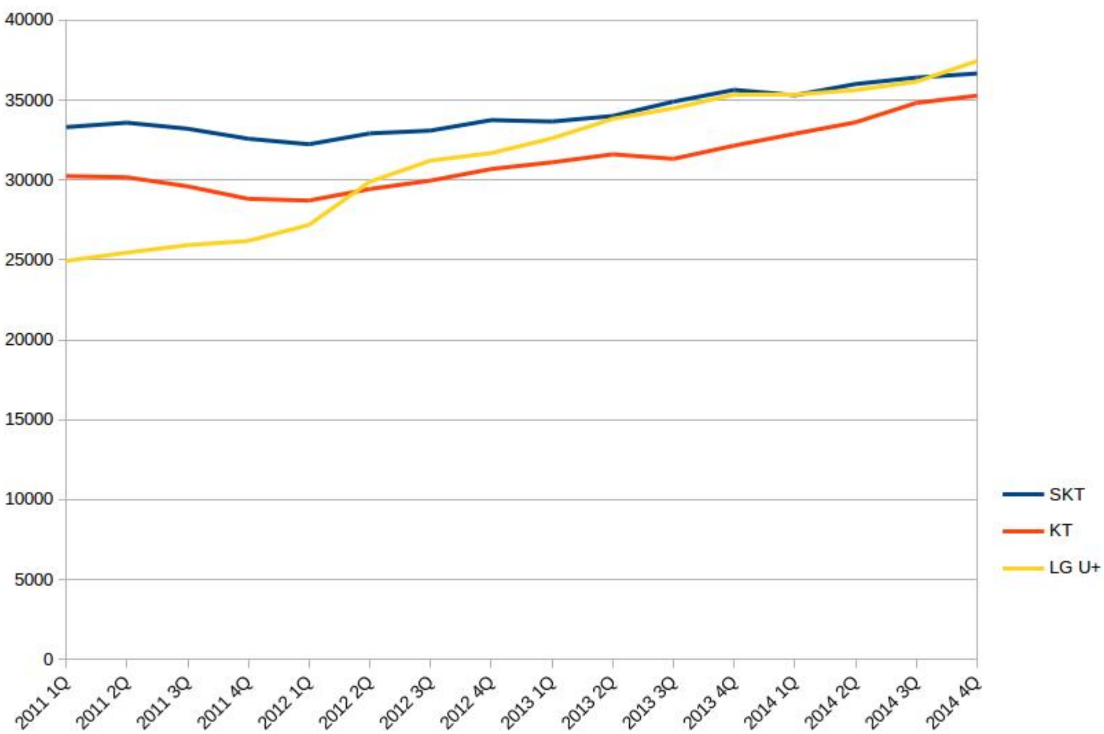


Figure 2.1: Billing ARPU by Operator (Unit: KRW)

largest provider, with a market share of 30%. Lastly LG U+ holds 20% of the market share. Following the governments revitalization policy, the market shares of MVNOs increased from 0.77% in 2011 to 8.01% in 2014. The government also tried to implement a fourth kind of mobile network operator into the market, but it failed in 2016.

Figure 2.1 shows the billing average revenue per user (ARPU) across different operators.[3] Interestingly, the ranking of operators by the ARPU turns out much differently than when done by market share. LG U+ earned the smallest in 2011, but caught up with Korea Telecom and SK Telecom in 2012 and 2013, respectively. Due to the governments policy on 3G mobile services, LG U+ was forced to apply the EV-DO technology rather than WCDMA, based on GSM. The former included many problems, such as slow download speeds and few device models. To recover its deficit from the failure of the 3G mobile market,

[3] ARPU is a measure used primarily by consumer communications and networking companies. ARPU is the total revenue divided by the number of subscribers. Billing ARPU excludes the sign-up fee from the total revenue.

LG U+ invested aggressively in a 4G LTE network infrastructure, launching a nationwide LTE network in March 2012. KoreaTelecom failed to retain frequencies for LTE service, and therefore could not provide an LTE network until the company discontinued its 2G service. Around roughly half a million KT subscribers 4 ARPU is a measure used primarily by consumer communications and networking companies. ARPU is calculated in terms of total revenue, excluding sign-up feeds, divided by number of subscribers. moved to LG U+ in 2012 for the fast data service. SK Telecom was also reluctant to expand its LTE network infrastructure, as the companys service focused more on voice service. These factors explain why, despite its low market share, LG U+ was able to earn the highest ARPU among its competitors following its adoption of smartphones and 4G LTE service.

2.2.2 Population Ageing and Informatization of the Elderly

Population ageing in Korea has become accelerated by an extremely low fertility rate and increased longevity. The fertility rate of South Korea fell as low as 1.30 in 2012, even lower than Japans 1.41. Furthermore, the population of the elderly aged over 60 has increased over time. The ageing index[4] reached 51.0 in 2006 and 77.7 in 2012. This index is anticipated to be over 100 in 2017. The speed of ageing is very fast in Korea compared to other developed countries. Table 2.3 shows a comparison in terms of speed of ageing. The ratio of the elderly was 7.2% in 2000, when Korea entered the ageing society, and is expected to reach 14.3% and 20.8% in 2018 and 2026, respectively.

Along with the increasing elderly demographic in the country, fast informatization in South Korea has made the digital divide problem severe. Older private and public services in the country have been converted into information-based services, with e-commerce, online banking, and online trading becoming standard. Some people even order their groceries over the internet in the country instead of going to a supermarket. With the informatization of South Korean society and the development of the ICT industry, however, the elderly have

[4]The ageing index is calculated as the number of persons 60 years old and over per hundred persons under age 15.

Table 2.3: The Speed of Ageing by Countries

Country	Year Reached			Year needed to reach	
	7% (Ageing)	14% (Aged)	20% (Hyper-aged)	7% → 14%	14% → 20%
Japan	1970	1994	2006	24	12
France	1864	1979	2018	115	39
United Kingdom	1929	1976	2026	47	50
United States	1942	2015	2036	73	21
South Korea	2000	2018	2026	18	8

Source: Korean Statistical Information Service

become more alienated. The elderly are reluctant to adapt the information technology for many reasons, including their lack of desire to obtain a personal digital device, and health problems, such as presbyopia, which prevent them from using small devices. Furthermore, the elderly are most familiar with the delivery of information via centralized media, such as TV and radio, and thus have difficulty searching for information they need on distributed information systems, such as the internet.

Table 2.4 shows the informatization level of the elderly compared to that of the overall population. The access level measures the ownership of traditional/smart digital devices,

Table 2.4: Informatization Level

Category		2008	2010	2011	2012	2013	2014
PC-based	Overall	64.2	67.5	69.2	71.2	72.6	74.3
	Access	92.5	93.8	94.3	94.9	94.3	94.9
	Ability	34.5	39.4	42.8	48.0	53.9	59.0
	Application	44.9	49.5	51.9	54.4	57.5	59.3
Smart	Overall				22.2	38.8	54.3
	Access				32.9	55.8	79.2
	Ability				16.1	30.7	35.5
	Application				22.9	32.1	53.0

Source: NIA (2012, 2013, 2014, 2015)

such as PC, smartphones, and other mobile devices, as well as subscriptions to fixed/mobile internet services. The ability level measures ability to use traditional/smart digital devices. People were asked the degree to which they can use computer programs and internet services. The application level indicates actual use of information services in terms of daily hours of use and the degree of use for certain purposes. The overall level of informatization among the elderly in terms of traditional ICT service use has improved over time, reaching nearly the same level as that of the entire population. However, ability and actual use of information services still remains at a low level. If we consider those in their 50s and 60s separately, the difference is much larger. In 2014, the ability and application level of persons aged 60 and over was 36.6 and 34.6, respectively, with the ability to use information services below 40% of the total population.

When it comes to the smart digital divide, the problem becomes more severe. Smart devices have been rapidly distributed among the elderly, yet they do not know how to utilize them. If owning a smartphone, the elderly commonly only use basic features, such as using the feature phone. They do not use a smart device to communicate using a mobile messaging app, to search for information over mobile Internet, or watch live streaming video. These services may not be indispensable to them, but a survey on smartphone usage has shown that the elderly do not know how to use them. Furthermore, for old persons using information services, there is a large gap in terms of SNS usage, network formation, community activity, information processing, and social activity between groups using both wired and mobile Internet or wired internet only. Persons using both services actively to participate in social activities, such as joining an internet community and writing replies to online articles, make up 47.5%, whereas only 13.4% of persons using fixed-line internet services do so. This difference results in an information differential that in turn leads to income and social differentials.

2.3 Data

I explore smartphone adoption among the elderly by examining the KISDI's (Korea Information Society Development Institute) Korean Media Panel Survey dataset. Since 2010, the KISDI annually conducts a survey on a sample of approximately 5,000 households and 10,000 members across them ages six and over using stratified sampling based on the distribution of Korean demographics. The 2010 dataset includes only six metropolitan areas, but the number respondents has been expanded nationwide since 2011, including into rural areas. Hence, I used the datasets from 2011 to 2014 in this analysis.[5]

The survey asks respondents if they subscribe to a wireless phone service. Respondents indicating wireless phone ownership are also asked to indicate which service provider they are connected to, if their mobile phone is a smartphone, how much they pay for the service on average, and who pays the bill. In addition to this usage information, the survey collects individual and household information on consumers, including extensive information on media device ownership, media device connectivity, and the media use patterns of each household and individual. A feature that is lacking from this survey is a range detailed phone characteristics. While the survey contains information on the manufacturer and general characteristics of a phone, such as digital camera and camcorder features, Wi-Fi connectivity, and Wibro connectivity, this information is not examined across different phone models. For example, nearly every new feature phone has a camera and camcorder function, yet none utilizes Wi-Fi, all smartphones have camera and camcorder functions with Wi-Fi connectivity. Therefore, we cannot identify the marginal effects of each product characteristic based on preference. Furthermore, the model number, as well as specific information, such as operating system (e.g., iOS, Android), processor, size of memory, and resolution, which have been used to analyze consumer choice in the literature, are not collected. Hence, this analysis was undertaken without this product information, which makes up part of the

[5] The first model of iPhone was introduced in 2009 in South Korea, and the first model of Android smartphone in 2010. The data says that only 1% of consumers over 50 adopted a smartphone in 2010, so the exclusion of year 2010 would not affect the analysis.

socioeconomic variables that determine smartphone preference.

The objective of this study is to determine the factors responsible for smartphone adoption among the elderly. The range of age referred to as elderly is 65 and over in traditional economics literature; however, in the context of the digital industry, people aged 50 and over have been considered an older, information alienated group. I follow this convention and focus on individuals aged 50 and over in 2014. [6] I examine those who used a mobile phone at least once and had no missing data for all variables used in the analysis. For individuals using more than one mobile phone, the primary phone information reported in the survey is used. The dataset has a balanced panel data structure, containing four periods (years 2011–2014) and 2,375 individuals from 1,752 households.

Table 2.5 presents the sample means of the individual and household demographic variables in the analysis. These variables include a gender (female) dummy, age, an indicator variable for those living in rural areas, an indicator for whether or not phone bills are paid out of ones own pocket, two dummies for the degree of internet use, and 16 area dummies. The dataset also contains categorical variables pertaining to educational attainment and household income. From the educational attainment variable, two indicator variables were created, one for lower education and the other for higher education. Lower education refers to those who only finished middle school, including high school dropouts, whereas higher education refers to those who hold a post-secondary school degree, including associate, bachelor, masters, and Ph.D. degree holders. For household income, I generated two indicator variables, one for low income and one for high income. In the original dataset, there are over 20 categories for household income, but for the households of the elderly, most samples are concentrated within the lower income categories. I therefore set upper and lower cutoff points from the middle household income to be a monthly income from KRW 500,000 to KRW 3,500,000, respectively. As a result, low family income refers to households earning less than KRW 500,000 per month, and high family income refers to households earning

[6] I performed the analysis including the individuals who reach their 50s in 2011, and found no significant changes in the result. So, I include those at the boundary for efficiency.

Table 2.5: Means of demographic variables by the user

Variable	Full sample	Feature phone	Smartphone
Female	0.527	0.543	0.462
Age	63.556	64.802	58.432
Rural area	0.228	0.256	0.111
Higher education	0.093	0.063	0.216
Lower education	0.579	0.663	0.235
Regular Internet use	0.223	0.169	0.444
Heavy Internet use	0.090	0.043	0.281
Payer	0.711	0.678	0.849
High family income	0.100	0.072	0.217
Low family income	0.096	0.115	0.017
N	9500	7641	1859

more than KRW 3.5 million per month. This classification of the income category is similar to that found in Rennhoff and Routon (2016).

The means for each type of the mobile phone user are presented in Table 2.5 alongside the full sample means to show how socioeconomic characteristics vary across stated choice alternatives. From these simple descriptive statistics, it appears that males are more likely to adopt a smartphone. The same can be said for youth, the residents of a city, the educated, internet users, and those who pay a phone bill out of their own pockets. Members of high income households are more concentrated within the smartphone group, while those of low income households are more concentrated within the feature phone group.

2.4 Empirical Model

As in Fairlie (2004), a simple linear random utility model of the decision to purchase a smartphone is used in this study, which explores the underlying causes of racial differences in rates of computer and internet access. The difference from the approach in Fairlie is that

I allowed for a panel data structure in the model,[7] assuming that the utility associated with having (purchasing) a smartphone or a feature phone is a function of an individuals characteristics, x, and an additive error term, ϵ. Let u_{ift} and U_{ist} stand for the ith persons indirect utilities associated with having a feature phone and a smartphone at year t, respectively. These indirect utilities can be expressed as

$$U_{ijt} = \delta_{jt} + \nu_{ij} + x'_{it}\beta_j + \epsilon_{ijt} \qquad j \in \{f, s\} \tag{2.1}$$

The ith person at year t has a smartphone if $U_{is} > U_{if}$. Let $y_{it} = 1$ if the ith person at year t owns a smartphone. Then:

$$\begin{aligned} P(y_i = 1) &= P(U_{is} > U_{if}) = F[(\delta_{st} - \delta_{ft}) + (\nu_{is} - \nu_{if}) + x'_i(\beta_s - \beta_f)], \\ &= F(x'_i\beta + \delta_t + \nu_i) \end{aligned} \tag{2.2}$$

where F is the cumulative distribution function of $\epsilon_{is} - \epsilon_{if}$. The model can be estimated using standard random effect logit regression assuming that $\epsilon_{is} - \epsilon_{if}$ has a logistic distribution and ν_i has a normal distribution with some variance.

Indirect utilities are functions of several measurable individual characteristics. Income is likely to be a key factor, as it has an effect on budget constraints underlying (2.1) and (2.2), and may also relate to preferences for owning a smartphone and the affordability of charges, as the charges for a smartphone are higher than that of a feature phone due to data plans. I consider household income rather than personal income for two reasons. First, though the mandatory retirement age is 65 in Korea, many workers retire in their 50s. Second, many females over 50 in Korea have never worked, even during their working age, and share their husbands income or pension for a living. I do include, however, the variable of who pays the phone bill to control for another aspect of ones decision to purchase a new smartphone.

Preferences for using a smartphone are likely to vary across individuals and may depend on exposure to and familiarity with information technology. Furthermore, these preferences

[7] An alternative approach is to estimate the choice decision among three alternatives, a smartphone, a feature phone, and no use. The model cannot be estimated, however, because of the lack of variation in measurable characteristics of the alternative choices.

may be related to a persons age, gender, education level, region of the country, and ability to use the internet. While the price of a mobile device and a mobile phone plan also affect the decision to use a smartphone, I have not included these two variables as independent variables, as these data are not available in the dataset.[8] As an alternative, I include brand as a variable in the model. There are three major mobile service providers in Korea: SK Telecom (SKT), Korea Telecom (KT), and LG U+. According to the market shares of each provider and government regulations, each company charges different prices for a similar service, which can serve as proxy variables for the price of a mobile service. In addition, network effect can be captured by including provider dummy variables. To control for the dispersion of a smartphone over time and by region, I include year and area dummy variables in certain specifications.

2.5 Result

Table 2.6 shows the estimates of logit regression concerning the probability of owning a smartphone. Both the marginal effects and their standard errors are reported.[9] Specification 1 includes age and a dummy variable for gender. The coefficient estimates capture the age effects, showing that elderly females tend to use smartphones less than older males. Specification 2 controls for personal characteristics, such as education, use of the internet, rural areas, and the payer of a phone bill, in addition to age and gender, though no household characteristics. Here, the high education dummy indicates whether or not a person received post-secondary education, and low education indicates whether or not a person finished high school, including dropping out. The use of the internet variable is a level variable that either has a value of one if a person can use basic applications, such as web-browsing and e-mail, or a value of two if a person participates in more advanced activities, such as social media or

[8] Vogelsang (2010) identifies the lack of price variation and information as one of the primary difficulties in studying the wireless industry.

[9] The reported marginal effect provides an estimate of the effect of a 1-unit increase in the independent variable on the smartphone probability. In regards to the dummy variable, the marginal effect provides the change of the probability by deviating from the baseline group. The standard error is computed by the cluster-robust standard error estimator.

online polls, etc. The payer dummy has a value of one if a phone bill is paid in full by a user without financial help from others or other benefits. The coefficient of the female dummy becomes statistically insignificant after controlling for the individual characteristics. As the descriptive statistics suggests, females in this age group are likely to be less educated and to use the internet less, so there may be an omitted variable bias in the coefficient of the female dummy in specification 1.

As expected, education is an important determinant of owning a smartphone, with persons who did not graduate high school having a roughly 10% lower probability than those who graduated high school. Individuals holding a college degree have a 3.6% higher probability of using a smartphone than high school graduates. Education may be a proxy for familiarity with new technologies, or may have an effect on preferences for smartphones in terms of taste, exposure, perceived usefulness, or conspicuous consumption. The use of the internet also plays a major role in determining who uses a smartphone, with the relationship between smartphone probability and internet utilization monotonically increasing. People who know how to use the basic internet services have a 10% higher probability to use smartphones, with those who actively participate in network environments have a probability of 21%. These percentages are not surprising, as people who use the internet on their home computers are already familiar with various information and communication technologies, and therefore are likely not afraid of adopting new devices, as well as desire to search and gather information from all locations. The inclusion of education, internet utilization, and other controls was found to have a notable effect on the age coefficients, with the marginal effect of age becoming cut in half in Specification 1, from -0.016 to -0.008. However, it still accounts for a large portion of the differences in smartphone penetration rate.

Household characteristics may also affect the probability of owning a smartphone. While there are many variables that can be used for controlling household effects, none of them except family income were statistically significant. I categorize family income into three categories: low, middle, and high. The low family income group earns less than KRW 500,000 a month, whereas the high family income groups earn more than KRW 5 million a month. All

Table 2.6: Logit Regressions for Probability of Having a Smartphone

Explanatory Variables	(1)	(2)	(3)	(4)	(5)
Age	−0.0162***	−0.0081***	−0.0077***	−0.0080***	−0.0080***
	(0.0007)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Female	−0.0578***	−0.0072	−0.0065	−0.0139	−0.0145
	(0.0096)	(0.0091)	(0.0091)	(0.0093)	(0.0093)
Higher Education		0.0362**	0.0323**	0.0293*	0.0291*
		(0.0155)	(0.0153)	(0.0152)	(0.0152)
Lower Education		−0.0974***	−0.0934***	−0.0752***	−0.0752***
		(0.0117)	(0.0117)	(0.0119)	(0.0119)
Regular Internet use		0.1061***	0.1051***	0.1044***	0.1046***
		(0.0091)	(0.0090)	(0.0090)	(0.0090)
Heavy Internet use		0.2104***	0.2075***	0.2016***	0.2010***
		(0.0155)	(0.0154)	(0.0153)	(0.0152)
Rural area		−0.0321***	−0.0290***	−0.0189*	−0.0191*
		(0.0103)	(0.0102)	(0.0115)	(0.0115)
Phone bill payer		0.0165**	0.0166**	0.0150**	0.0152**
		(0.0076)	(0.0076)	(0.0076)	(0.0076)
High family income			0.0353***	0.0297**	0.0298***
			(0.0118)	(0.0115)	(0.0115)
Low family income			−0.0305*	−0.0263	−0.0256
			(0.0173)	(0.0171)	(0.0172)
Area: Busan				−0.0520***	−0.0308
				(0.0200)	(0.0649)
Area: Daegu				−0.0679***	−0.0431
				(0.0227)	(0.0765)
Area: Gwangju				−0.0566***	−0.0337
				(0.0207)	(0.0697)
Area: Chungnam				−0.0901***	−0.1084**
				(0.0225)	(0.0524)
Area: Jeonnam				−0.0661***	−0.0703***
				(0.0212)	(0.0245)
Area: Gyeongbuk				−0.0791***	−0.0801***
				(0.0206)	(0.0203)
Area: Gyeongnam				−0.0842***	−0.0802***
				(0.0199)	(0.0226)

Notes: (1) Marginal effects and their standard errors are reported.

(2) *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

other households belong to the middle family income group. The third column in Table 2.6 reports the marginal effects after controlling for household characteristics. The relationship between smartphone probability and income can be seen to be monotonically increasing. However, the effect is rather small. In fact, when I use a different specification with finer categorical family income variables, most of the coefficients are not statistically significant, save for the high family income group. Even if a husband earns most of a households income and a wife does not work in a typical household in this group, family factors seem not to affect the purchasing probability of a smartphone. This suggests that individual characteristics are more important determinants of owning a smartphone than household variables. Indeed, smartphones are personal belongings, and their personal utilization is more important to users in terms of their preferences. Furthermore, consumers can receive discounts on devices and phone bills in one- or two-year contracts, so family and individual income may not serve as constraints.

Specification 4 (shown in Table 2.6) includes dummy variables for 17 areas in Korea. After adding a regional control, an interesting result was found. Adding the area dummy variables did not change the marginal effects of personal and household variables, but revealed that elderly living in Busan, Daegu, and Gwangju, which are metropolitan cities in Korea, have far less probability of owning a smartphone than those living in Seoul. Compared to the elderly living in Seoul, those living in Busan have a 5.2% lower probability, those in Daegu have a 6.8% lower probability, and those in Gwangju have a 5.7% lower probability of using a smartphone. Furthermore, living in Gangwon province, which has the lowest GRDP per capita and most mountainous areas, the latter of which may potentially damage the quality of mobile service, was found to have no significant effect on probability. One plausible explanation for this is regional imbalance. The areas that show significantly lower probabilities of smartphone usage are far from the Seoul Metropolitan Area (SMA). Over the past two decades, the development of the economy, education, public welfare, culture, and communications in Korea has been concentrated in the SMA, which may influence the dissemination of smartphones, the utilization of other ICTs, and informatization. The

Table 2.7: Basic Information of Selected Areas

Area	Population		Land area	GRDP per capita
	Total	Over 50	(km^2)	(in KRW thousands)
Busan	3,411,930	1,306,223	770	21,614
Daegu	2,459,901	836,417	884	18,941
Gwangju	1,515,992	444,403	501	20,448
Gangwon	1,501,041	590,911	16,826	24,574
Chungnam	2,078,771	741,908	8,214	47,024
Jeonnam	1,758,340	764,327	12,309	35,883
Gyeongbuk	2,640,441	1,058,604	19,029	34,711
Gyeongnam	3,273,079	1,139,034	10,538	31,311

Source: Korean Statistical Information Service

marketing strategies of mobile service providers are also affected by regional imbalances, as marketers have little incentive to sell new devices to areas where there is little demand. To control for the effect, I added the GRDP of each area in specification 5. The majority of the coefficients of the regional dummy variables were insignificant, save for those of Chungnam, Jeonnam, Gyeongbuk, and Gyeongnam province. Controlling for regional economic situation did not change other marginal effects. I will leave the detailed analysis of this issue for future research.

2.6 Conclusion

This paper discussed the effects of individual, household, and regional factors that may influence preferences among the elderly with regard to obtaining a smartphone. Ownership of a smartphone was found to predominantly be affected by personal characteristics. Age, education and familiarity with traditional information technologies have a significant effect on the probability of owning a smartphone, while family characteristics, such as income and structure, were found not to have an effect. These results are in contrast with the determinants affecting the purchase of a personal computer and traditional internet service subscription, which were found to be shared by family members. Moreover, the area where a person lives can have an impact on his or her adoption of a smartphone. Individuals living in regions far from the Seoul Metropolitan Area were found to be far less likely to use a smartphone. Considering the regional income and purchasing power of these areas, lack of smartphone usage may reflect the imbalance of development between regions. Policymakers should thus consider these effects when drafting informatization policy for future ICTs.

My analysis includes several limitations that speak to various avenues for future research. Product characteristics such as the price of a device and mobile phone subscription plan were not considered in the analysis due to lack of information or little variation of characteristics. Using a smartphone requires a mobile data subscription, which increases ones monthly phone charge in addition to a voice service charge. Furthermore, the higher price of a smartphone compared to that of a traditional feature phone may prevent the elderly from purchasing a new device. Many mobile virtual network operators have entered the market since 2014 with low phone charges that attract consumers. Changes in the mobile phone market such as this one can alter incentives to use a smartphone among the elderly. One can obtain a more accurate decomposition if detailed product characteristics are available. Moreover, the regional effect should be investigated in detail.

2.7 Appendix

2.7.A Tables

Table 2.8: Full Results of Logit Regressions for Probability of Having a Smartphone

Explanatory Variables	(1)	(2)	(3)	(4)	(5)
Age	-0.0162^{***}	-0.0081^{***}	-0.0077^{***}	-0.0080^{***}	-0.0080^{***}
	(0.0007)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Female	-0.0578^{***}	-0.0072	-0.0065	-0.0139	-0.0145
	(0.0096)	(0.0091)	(0.0091)	(0.0093)	(0.0093)
Higher Education		0.0362^{**}	0.0323^{**}	0.0293^{*}	0.0291^{*}
		(0.0155)	(0.0153)	(0.0152)	(0.0152)
Lower Education		-0.0974^{***}	-0.0934^{***}	-0.0752^{***}	-0.0752^{***}
		(0.0117)	(0.0117)	(0.0119)	(0.0119)
Regular Internet use		0.1061^{***}	0.1051^{***}	0.1044^{***}	0.1046^{***}
		(0.0091)	(0.0090)	(0.0090)	(0.0090)
Heavy Internet use		0.2104^{***}	0.2075^{***}	0.2016^{***}	0.2010^{***}
		(0.0155)	(0.0154)	(0.0153)	(0.0152)
Rural area		-0.0321^{***}	-0.0290^{***}	-0.0189^{*}	-0.0191^{*}
		(0.0103)	(0.0102)	(0.0115)	(0.0115)
Phone bill payer		0.0165^{**}	0.0166^{**}	0.0150^{**}	0.0152^{**}
		(0.0076)	(0.0076)	(0.0076)	(0.0076)
High family income			0.0353^{***}	0.0297^{**}	0.0298^{***}
			(0.0118)	(0.0115)	(0.0115)
Low family income			-0.0305^{*}	-0.0263	-0.0256
			(0.0173)	(0.0171)	(0.0172)
Area: Busan^{a}				-0.0520^{***}	-0.0308
				(0.0200)	(0.0649)
Area: Daegu^{a}				-0.0679^{***}	-0.0431
				(0.0227)	(0.0765)
Area: Incheon^{a}				-0.0090	0.0091
				(0.0242)	(0.0573)
Area: Gwangju^{a}				-0.0566^{***}	-0.0337
				(0.0207)	(0.0697)

Table 2.9: Full Results of Logit Regressions (Continued)

Explanatory Variables	(1)	(2)	(3)	(4)	(5)
Area: Daejeon[a]				-0.0270	-0.0043
				(0.0264)	(0.0715)
Area: Ulsan[a]				-0.0162	-0.0684
				(0.0316)	(0.1339)
Area: Gyeonggi				-0.0207	-0.0078
				(0.0179)	(0.0409)
Area: Gangwon				-0.0338	-0.0172
				(0.0279)	(0.0550)
Area: Chungbuk				0.0168	0.0210
				(0.0216)	(0.0252)
Area: Chungnam				-0.0901^{***}	-0.1084^{**}
				(0.0225)	(0.0524)
Area: Jeonbuk				-0.0758^{***}	-0.0610
				(0.0207)	(0.0482)
Area: Jeonnam				-0.0661^{***}	-0.0703^{***}
				(0.0212)	(0.0245)
Area: Gyeongbuk				-0.0791^{***}	-0.0801^{***}
				(0.0206)	(0.0203)
Area: Gyeongnam				-0.0842^{***}	-0.0802^{***}
				(0.0199)	(0.0226)
Area: Jeju				-0.0794^{**}	-0.0636
				(0.0311)	(0.0552)

Notes: (1) Marginal effects and their standard errors are reported.

(2) ^{a}: Metropolitan city

(3) *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

CHAPTER 3

The Evolution of Preferences for Brands in Digital Camera Market

3.1 Introduction

An economic agent considers the brand and the characteristics of products to decide which product to buy. Some features of a product are observed both by consumers and by econometricians, while others affecting the purchase decision of a consumer are observed only by consumers, not by econometricians. One way to capture this effect is to use brand-specific dummy variables. Nevo (2000, 2001) expanded the random coefficient logit model in Berry et al. (1995, BLP henceforth) by using dummy variables in order to capture brand-specific fixed effects.

When purchasing a digital camera, consumers examine attributes of a camera: the resolution of an image sensor, zoom function, the size of LCD, and the price. Also, they consider the brand of a camera, since each manufacturer has some intrinsic characteristics; for example, Nikon and Canon have been two major manufacturers in film camera industry, so their products have better optical features. Fuji and Kodak have produced film for a long time so that their cameras are said to have good chromatic features. Sony is a well-known manufacturer of home appliance, and its products are easy to use.

During the period from April 1997 through May 1999, the data on unit sales of digital cameras show that the unit sales of Sony has increased steadily, while those of Casio has decreased gradually, as shown in Figure 3.1. However, as other consumer electronics, the technological development of a camera does not vary across manufacturers. For instance, the maximum resolution of the newest camera is not quite different between each other. It suggests that the difference of technological features across brands is not sufficient to fully explain the trend of unit sales in the digital camera industry. Even including brand-specific dummy variables does not account for the variation of the unit sales. In this sense, we may assume that the preference for digital camera brands changes over time.

This paper suggests a method to capture the time-varying brand preference under the specification of BLP model. In particular, I assume that the brand preference is affected by the advertising expenditure of each brand as well as the reputation among consumers. I assume that the brand preference at time period t is a linear function of a preference at time $t-1$ and advertisement expenditure of the brand at time $t-1$. Then, this is a state equation related to the brand preference in two consequential time periods. Furthermore, to allow for flexible price elasticities among products, I admit the heterogeneity of a household with an aggregate data. Hence, the estimation strategy can be considered as an extension of the random coefficient logit model in BLP. For estimation, I use data for the U.S. digital camera market over 26 months from April 1997 through May 1999.

My work is based on Sriram et al. (2006), in which they used a nested logit model. However, when fitting the nested model, I faced a problem that estimated market shares do not converge to a certain value since some of them diverged to infinity. For this technical reason, I changed my model to "standard" logit model. Except that, I used exactly the same dataset as of Sriram.

3.2 The Empirical Framework

3.2.1 Model

During each period t, a household h faces a decision problem of choosing a digital camera j offered by brand b. Specifically, a consumer chooses to buy a model from the set of $M_{bt} = \{1, 2, \cdots, J_{bt}\}$ models offered by brand b where $b = 1, 2, \cdots, B$ and J_{bt} is the number of models offered by brand b at time t. Under the discrete choice model framework, the indirect utility function of a household h from choosing model j offered by brand b at time t is given by

$$U_{hjbt} \quad = \quad \alpha_t + \beta_{0hbt} + \theta H_{bt} + \beta_h X_{jbt} + \eta_{jbt} + \epsilon_{hjbt}, \tag{3.1}$$

where β_{0hbt} is the household h's intrinsic preference for the brand b at time t, H_{bt} is a set of dummy variables of holiday seasons, X_{jbt} is the vector of attributes of model j offered by brand b at time t such as the resolution of the image sensor, the maximum number of images that can be stored in the internal memory, the type of external storage media and the price, and β_h is the vector of coefficients corresponding to the product attribues, varying across households. In addition, X_{jbt} might include other factors such as the age of a model, which may have an effect on the consumer's preference for the model. To allow for the possibility of the nonlinear effect of the age, we include the quadratic term of an age in addition to the linear term. The term α_t is a time-specific dummy variable, common to all brands and models, which captures the utility of owning a digital camera at time t. As in Nevo (2000), η_{jbt} is the utility derived from unobserved product characteristics, and ϵ_{hjbt} is a mean-zero stochastic error term which is distributed according to a Type-I extreme-value distribution. Also, we may normalize the utility from choosing the outside option as $U_{h00t} = \epsilon_{h00t}$ for identification.

Under the assumptions of the standard random coefficient multinomial logit model, we can express the probability of household h purchasing model j offered by brand b at time t, p_{hjbt}, as

$$p_{hjbt} \quad = \quad Pr[y_{hjbt} = 1] = \frac{\exp(\delta_{jbt} + \mu_{hjbt})}{1 + \sum_{b'=1}^{B} \sum_{j \in M_{b'}} \exp\left(\delta_{jb't} + \mu_{hjb't}\right)}, \tag{3.2}$$

where δ_{jbt} is the mean (across households) utility of model j offered by brand b at time t, and μ_{hjbt} is the deviation in the utility of household h from this mean, which captures household heterogeneity. Specifically,

$$\delta_{jbt} \quad = \quad \alpha_t + \beta_{0bt} + \theta H_{bt} + \beta X_{jbt} + \eta_{jbt}, \tag{3.3}$$

$$\mu_{hjbt} \;\; = \;\; (\beta_{0hbt} - \beta_{0bt}) + (\beta_h - \beta) X_{jbt} = \Delta\beta_{0hbt} + \Delta\beta_h X_{jbt}. \tag{3.4}$$

The parameters β_{0bt} and β are the mean level of the brand preference and average marginal effects of product characteristics, respectively. $\Delta\beta_{0hbt}$ and $\Delta\beta_h$ are the household-specific deviation from β_{0bt} and β, respectively.

If there is no household heterogeneity, that is, $\mu_{hjbt} \equiv 0$ for all h, j, b, t, we obtain a standard multinomial logit model. However, this model suffers from IIA property without allowing for the heterogeneity. To overcome this limitation, we need to account for unobserved heterogeneity in the model by allowing μ_{hjbt} to be different from zero. For simplicity, I assume that the vector $\nu = (\Delta\beta_{0hb}, \Delta\beta_h,\ b = 1, 2, \cdots, B)$ follows a multivariate normal distribution $\nu \sim N(0, \Sigma)$, independent from all other variables and error terms. Given the distributional assumption on ν, the market share of model j offered by brand b at time t can be obtained by integrating over ν, that is,

$$s_{jbt} = \int_A Pr[y_{hjbt=1}]\phi(v)dv = \int_A \frac{\exp(\delta_{jbt} + \mu_{hjbt})}{1 + \sum_{b'=1}^{B} \sum_{j \in M_{b'}} \exp\left(\delta_{jb't} + \mu_{hjb't}\right)} \phi(v)dv, \tag{3.5}$$

where $\phi(\cdot)$ denotes the density of a multivariate normal distribution and A is the region of integration which results in the choice of model j of brand b. Hence, our model is a random coefficient multinomial logit model, as in Berry et al. (1995) and Nevo (2001).

As mentioned in Equation (3), I allow the parameter β_{0bt}, the intrinsic preference for brand b, to vary over time. With the intuition that advertising activity of a brand has an effect on the evolution of the intrinsic brand preference, we model the dynamics of the mean brand preferences as

$$\beta_{0bt} = \bar{\beta}_b + \lambda\beta_{0bt-1} + \omega_b AD_{bt} + \psi_{bt}, \qquad \psi_{bt} \sim N(0, \sigma^2_{\psi b}), \tag{3.6}$$

where β_{0bt} is the mean preference for brand b at time t, $\bar{\beta}_b$ is the time-invariant part of the mean preference for brand b, and AD_{bt} is the advertising cost of brand b at time t. The parameters ω_b capture the effects of advertisement on brand b's intrinsic preference. The parameter λ captures the extent to which the intrinsic brand preference carries over from period to period.

3.2.2 Estimation

The parameters which needs to be estimated are categorized into three sets: $\Theta_1 = \{\alpha_t, \theta, \bar{\beta}_b, \lambda, \omega\}$, $\Theta_2 = \{\beta\}$, and $\Theta_3 = \{\sigma_{\Delta h \beta}\}$. The first set Θ_1 is the collection of parameters that correspond to the mean preferences and other response parameters that influence the utility of all the models offered by a brand, not varying across each model offered by the same brand. The second set Θ_2 captures the effects of consumers' mean valuations of product characteristics. And the last set, Θ_3, corresponds to the Cholesky decomposition of the matrix Σ, the covariance matrix corresponding to the heterogeneity distribution in Equation (5). As suggested in BLP, for a given set of the heterogeneity parameters Θ_3, we can uniquely obtain the mean utility level δ_{jbt}. After recovering the mean utilities δ_{jbt}, we estimate the parameter Θ_2 by an instrumental variables regression, and then estimate the parameter Θ_1 by the Kalman filter algorithm.

Recall that η_{jbt} captures the effects of omitted characteristics of a specific model j as well as the brand-specific effect of brand b, both of which are observed by consumers but not by researchers. We can express this as

$$\eta_{jbt} = \eta_{bt} + \Delta\eta_{jbt} \tag{3.7}$$

where η_{bt} is the brand-specific unobserved factors common to all the models offed by brand b at time t, and $\Delta\eta_{jbt}$ captures the model-specific unobserved heterogeneity. To identify these two effects, I make a strong assumption that the model-specific deviation $\Delta\eta_{jbt}$ is equal to zero for one of the models of each brand for all time periods, which is called as a baseline model. Without loss of generality, I set $\Delta\eta_{1bt} = 0$ for all b and t.

Using the contraction mapping of BLP, we could obtain the estimates of $\{\delta_{jbt}\}$. Under the assumption that $\Delta\eta_{1bt} = 0$, denote the mean utility of baseline models for each b and t as δ_{1bt}. Therefore, the deviation of mean utility of each model from each baseline model can be represented as

$$\delta_{jbt} - \delta_{1bt} = \delta'_{jbt} = \beta\Delta X_{jbt} + \Delta\eta_{jbt}, \tag{3.8}$$

where $\Delta X_{jbt} = X_{jbt} - X_{1bt}$. Observe that the brand-specific unobserved effect η_{bt} is cancelled out in equation (8). Since the left term of Equation (8) is uniquely determined, the parameters $\beta(=\Theta_2)$ can be estimated by an instrumental variables regression that accounts for potential correlation between $\Delta\eta_{jbt}$ and the prices embedded in ΔX_{jbt}.

Given the heterogeneity parameter Θ_3, we have obtained δ_{jbt} and β so far through the above estimation procedure. Now, we need to estimate the parameters that influence choices at the brand level, Θ_1. For doing this estimation, we define the term R_{bt}, the total value of brand b at time t as follows:

$$R_{bt} = \ln\left\{\Sigma_{j\in M_b}\exp\left(\delta_{jbt}\right)\right\}. \tag{3.9}$$

If we plug δ_{jbt} from (3) into (9), we have

$$R_{bt} = \ln\left\{\Sigma_{j\in M_b}\exp\left(\beta X_{1bt} + \delta'_{jbt}\right)\right\} + (\alpha_t + \beta_{0bt} + \theta H_{bt} + \eta_{bt}). \tag{3.10}$$

Therefore, we have

$$Q_{bt} = \alpha_t + \beta_{0bt} + \theta H_{bt} + \eta_{bt}, \text{ where} \tag{3.11}$$

$$Q_{bt} = R_{bt} - \ln\left\{\Sigma_{j\in M_b}\exp\left(\beta X_{1bt} + \delta'_{jbt}\right)\right\}. \tag{3.12}$$

The second term in the right-hand side of Equation (12) is similar to the inclusive value of a nest (or a brand in our analysis) in the nested logit model, and so it can be regarded as a measure of the effect of a brand's product line on its performance in the market. Similarly, Q_{bt} can be interpreted as the intrinsic value of a brand. Given Q_{bt}, the parameters in Equation (11) can be estimated by a simple linear regression. However, it is impossible to directly apply the linear regression, since we cannot observe the values of brand preferences, β_{0bt}, at each time period t. Thus, regarding Equation (11) as an Observation Equation and Equation (6) as a System Equation, we can apply the Kalman filter algorithm to our model.

The method how we estimate Θ_1 and Θ_2 given the heterogeneity parameters Θ_3 are explained above. Through these two estimation procedures, we have the system of error terms $\Delta\eta_{jbt}$ and η_{bt}. Then, the heterogeneity parameters Θ_3 can be achieved by minimizing a quadratic function of these error terms. The quadratic object can be entertained by a generalized method of moments (GMM) procedure. In short, similar to BLP method, $\{\Theta_1, \Theta_2\}$ are computed in an inner loop for a fixed $\{\Theta_3\}$, and $\{\Theta_3\}$ are the arguments which minimize the GMM object in an outer loop.

3.3 Data

3.3.1 The Data

The data set used in estimation is collected from two different sources. One is aggregate monthly observations on unit sales of compact digital cameras in United States for a period of 26 months from April 1997 through May 1999. These data include information on the features of each model, such as the price, maximum resolution of the image sensor, maximum number of photos, the availablity of internal and external memory, type of storage media, and the presence of self-timer capabilities. The other is monthly advertising expenditures by each of the brands during the corresponding period. Sales, price and attribute data are at the model level, while advertising data are at the brand level.

We perform the analysis on the four leading brands in this category: Casio, Kodak, Olympus, and Sony. These brands account for more than 93% of the sales in this category, and the four brands are present during all the 26 months of the data. We report the descriptive statistics for the four brands in Table 3.1. From Table 3.1, we can see Sony has the highest market share, which is almost twice that of the nearest competitor, Kodak. Note that Sony has the highest market share in spite of the highest price. It may be attributed to the attractiveness of models in its product line (inclusive value) and/or to a high intrinsic preference for the brand.

Figure 3.1 shows the time trend in monthly sales of each brand over 26 months. As said before, Figure 3.1 reveals that while Casio possessed the largest market share at the beginning of the time periods, its unit sales steadily decreased over time and it ended up as the lowest selling brand. In contrast, although Sony has the lowest proportion at the

Table 3.1: Descriptive Statistics for the Digital Camera Brands

Brand	Average price ($)	Total unit sales	Market share (%)	Total advertising($ 1,000)	Average age (months)	Average number of model
Casio	320	176,049	10.69	848.3	13.79	10.50
Kodak	485	360,778	21.90	9,223.4	14.18	8.00
Olympus	606	305.385	18.54	4,432.0	9.92	6.31
Sony	675	691,457	41.98	11,890.6	8.43	4.96

beginning, it soon overtook all the other brands. It also shows the seasonality of compact digital camera sales, which reached its peak every December. It suggests the necessity of dummy variables to control this seasonal effect, which are the holiday dummy variables in the model.

3.3.2 Description of Variables

3.3.2.1 Product Characteristics

We estimate the consumer valuation of five features, maximum resolution of the image sensor, maximum number of photos, size of external memory, type of storage media, and the presence of self-timer capabilities. The price variable was operationalized as the logarithm of the price of the model. In addition, we constructed the age and square of age for each model. The age is measured by the number of month since it was first introduced to the market.

3.3.2.2 Market Size and Outside Alternative

In order to compute share of the model, we need to define the potential size of the market and the outside or no-purchase alternative. Song and Chintagunta (2003) used the number of households that used computers at home as the potential size of the market, since using digital cameras requires access to a computer. Following this approach, we set the total potential market size as 10 million (U.S. Census Bureau 1997).

3.3.2.3 Instrument Variables

Due to the endogeneity of the price, we need to find instruments for the price for IV regression step in 2.2.1. As in Berry et al. (1995), we use three sets of instruments variables: the observed product characteristics, the sums of the values of the same characteristics of other products offered by the manufacturer, and the sums of the values of the same characteristics of products offered by other firms. In addition, we also use producer price index for computer peripheral equipment (SIC code 3577) from the U.S. Bureau of Labor Statistics.

3.4 Result

3.4.1 Result from the model

Through the estimation procedure discussed in Section 2, the results for parameters are being reported in Table 3.2. As expected, the effect of a price is negative on a model choice. The presence of a self-timer and resolution has a significant positive effect on a model's share. The results of age and age squared parameters present that, when the age of a camera becomes 4 months, the effect of the age on a choice probability is maximum. This can be explained by the claim that some period after the introduction of a new model is necessary for a consumer to believe it as a flawless model. Next, in the case of parameters for brand choice part, The parameter λ that captures the carry-over of brand preferences from period to period is 0.947. This means that the intrinsic brand preferences are highly persistent. In addition, the parameters for advertising effects on brand preferences show positive estimates. The positive estimate for holiday seasonal dummy is consistent with the high unit sales in December.

3.4.2 Intrinsic Brand Preferences and Inclusive Values over Time

I present the intrinsic brand preferences and the inclusive values of the brands over time in Figure 2 and 3, respectively. The time trend in the intrinsic brand shows that the brand

Table 3.2: Model Results

Parameter	Estimates	Heterogeneity
Model Choice		
Price	-0.229	0.022
	(0.124)	
Resolution	0.296	0.015
	(0.149)	
No. of Images	-0.006	0.008
	(0.003)	
External Memory	-0.170	0.047
	(0.047)	
Self-timer	0.703	0.022
	(0.261)	
Age	0.252	0.010
	(0.036)	
Agesq	-0.030	0.020
	(0.001)	
Brand Choice		
Carry-over(λ)	0.947	
Constant (Casio)	-0.069	
Constant (Kodak)	0.080	
Constant (Olympus)	0.023	
Constant (Sony)	0.020	
Advertising (Casio)	0.516	
Advertising (Kodak)	0.090	
Advertising (Olympus)	0.304	
Advertising (Sony)	0.409	
Holiday	1.019	

preference for Casio has a declining trend. This is consistent with the facts that Casio's advertising expenditure is the smallest among four brands and the brand preference evolves by an advertisement in Equation (6). On the other hand, the brand preference for Sony shows the highest brand preference among four brands, though it starts with the lowest brand preference at the beginning period. The brand preferences for Kodak and Olympus have slightly increasing trends. In the case of the inclusive values, Kodak and Olympus show smoother trends than Casio and Sony. Kodak and Olympus have an almost flat trend, while the inclusive values of Casio and Sony have lots of fluctuation. However, The inclusive values of both companies coincide between period 10 and period 18. which is difficult to explain the overturn of market shares. Hence, The increasing and decreasing market shares of Sony and Casio, respectively, are better explained by the intrinsic brand preferences than by the inclusive values. Roughly speaking, it seems that the effects of the brand preferences for unit sales are more significant than the effects of the product lines of the brands.

3.5 Conclusion

This paper extends the work of Berry et al. (1995) and Nevo (2000, 2001) to allow for the time-varying brand preference. In particular, the model considered in this paper answers two questions: (a) What are the relative importances of intrinsic brand preferences, prices and product attributes in driving the performance of a brand, or the market share of a manufacturer? (b) Does advertising have any short-term and/or long-term effect in driving preferences? Even though set in the context of high-technology device markets, my model can be applied to the analysis of consumer packaged goods markets.

We find that intrinsic brand preferences have a much bigger effect on the performance of the brand than the inclusive values, which reflects model-level prices and product characteristics. Furthermore, we find that some brands can increase their advertising expenditures and increase their profitability. Casio, which has dominance in the earlier age of compact digital camera market but which has a relatively small advertising budget, lost their advantage and fell to the smallest competitor. On the other hand, Sony succeeded in obtaining the dominance by aggressive advertisement. Furthermore, very high value of carry-over parameter λ, which was close to 1, may imply that the reputation of a brand last for a long time. Also it implies that the empirical analysis considering only the product characteristics, not the marketing factors such as advertisement, promotion or discount, could fail to recover the true demand parameters.

Our approach is subject to several caveats and limitations. First of all, we tried to build a "nested" logit model, since each digital camera has its own brand and it seems to exist a correlation in choices within a brand, and people tend to choose the brand of a camera first and compare a few models within that brand when purchasing a camera. However, in the contraction mapping of the nested logit BLP model, we faced on the problem that the estimated market share, δ_{bjt}, does not converge to a certain value. It may be because the method proposed by BLP is very sensitive to the initial conditions and others. Recently, Dubé et al. (2012) suggest the nest fixed point algorithm to improve the numerical performance of BLP model. This paper could be a starting point to solve the problem. Additionally, our framework does not account for the dynamics in the consumer valuation of individual attributes in any general way. This issue also should be solve in a further study.

3.6 Appendix

3.6.A Figures

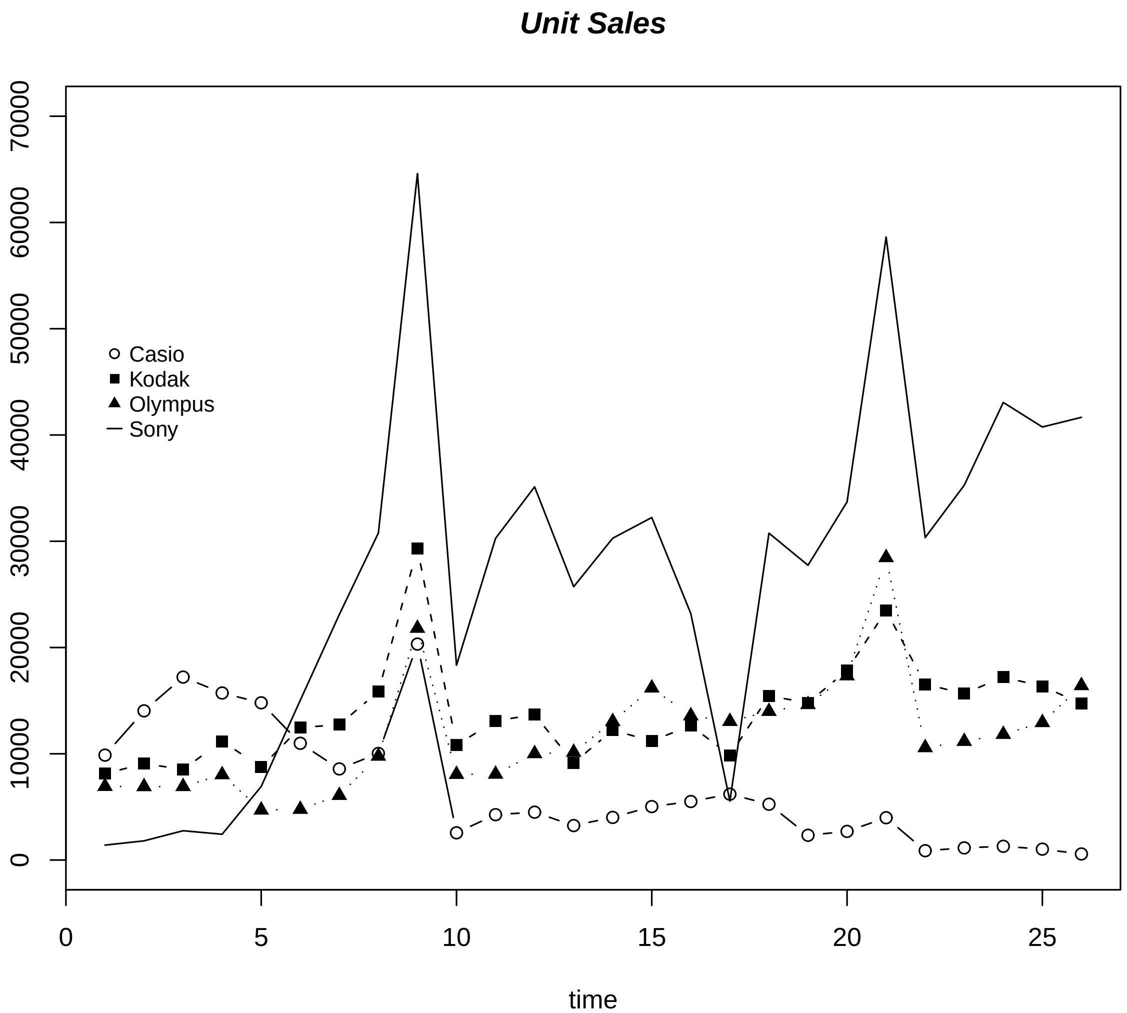


Figure 3.1: Unit Sales of Digital Camera Brands

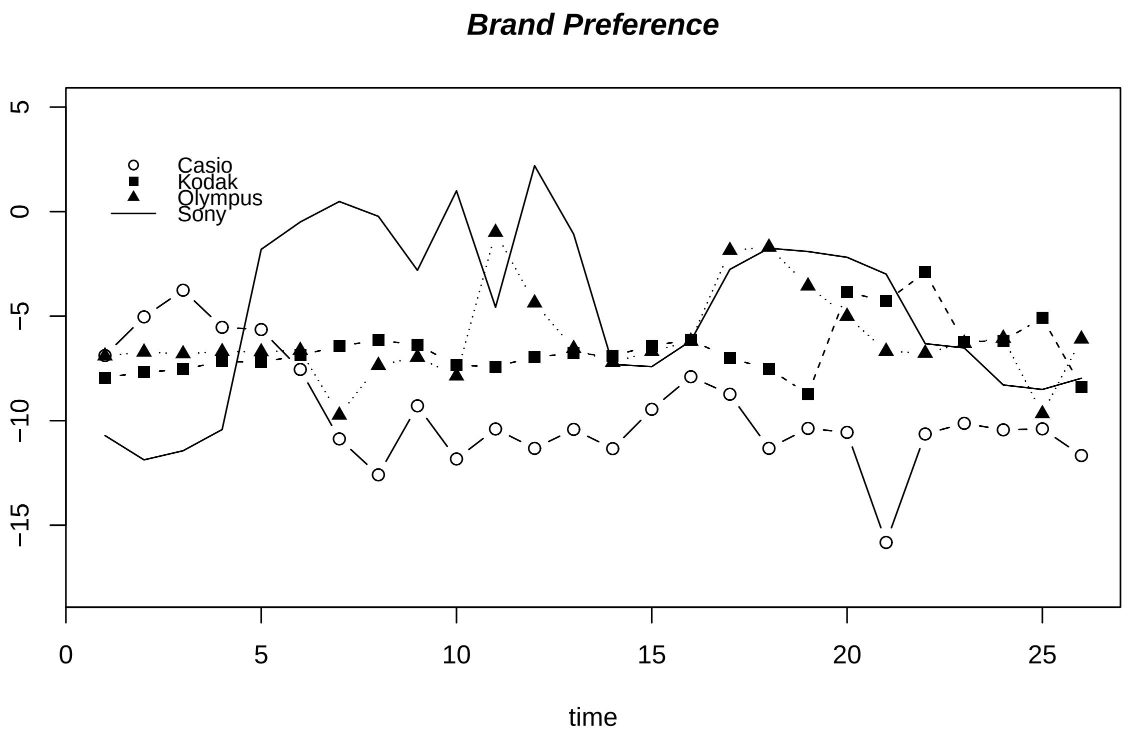


Figure 3.2: Brand Preferences of Digital Camera Brands

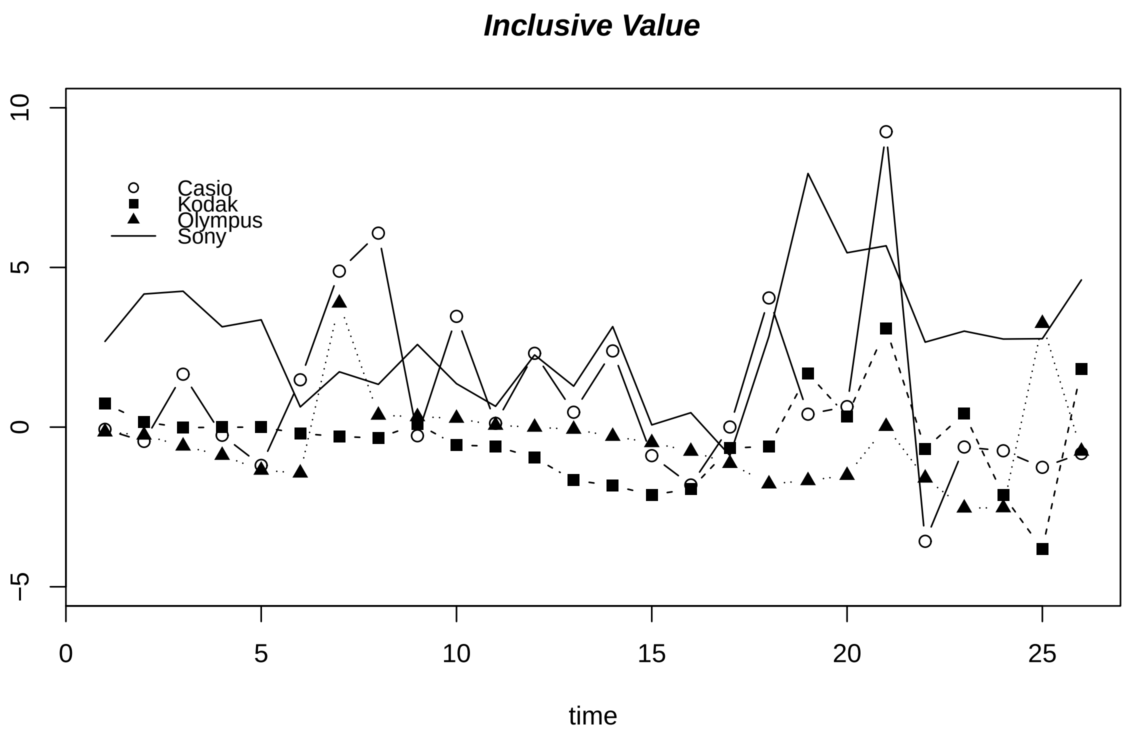


Figure 3.3: Inclusive Values of Digital Camera Brands

Bibliography

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Berry, S. T. and P. A. Haile (2010). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Cowles Foundation Discussion Papers 1718, Cowles Foundation for Research in Economics, Yale University.

Berry, S. T. and P. A. Haile (2014). Identification in differentiated products markets using market level data. *Econometrica 82*(5), 1749–1797.

Blau, D. and J. Currie (2006). *Pre-school, day care, and after-school care: who's minding the kids?*, Volume 2 of *Handbook of the Economics of Education*, Chapter 20, pp. 1163–1278. Elsevier.

Blau, D. M. and P. K. Robins (1988). Child-care costs and family labor supply. *The Review of Economics and Statistics 70*(3), 374–381.

Blundell, R. and R. L. Matzkin (2014). Control functions in nonseparable simultaneous equations models. *Quantitative Economics 5*(2), 271–295.

Blundell, R. W. and J. L. Powell (2004). Endogeneity in semiparametric binary response models. *The Review of Economic Studies 71*(3), 655–679.

Brown, B. W. (1983). The identification problem in systems nonlinear in the variables. *Econometrica 51*(1), 175–196.

Brown, S. A. and V. Venkatesh (2005). Model of adoption and technology in households: a baseline model test and extension incorporating household life cycle. *MIS Quarterly 29*(3), 399–436.

Buchmueller, T. and J. Dinardo (2002). Did community rating induce an adverse selection death spiral? Evidence from New York, Pennsylvania, and Connecticut. *The American Economic Review 92*(1), 280–294.

Chaudhuri, A., K. S. Flamm, and J. Horrigan (2005). An analysis of the determinants of internet access. *Telecommunications Policy 29*(9), 731–755.

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014). Local identification of nonparametric and semiparametric models. *Econometrica 82*(2), 785–809.

Chen, X. and D. Pouzo (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica 80*(1), 277–321.

Chernozhukov, V. and C. Hansen (2005). An iv model of quantile treatment effects. *Econometrica 73*(1), 245–261.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007). Instrumental variable estimation of nonseparable models. *Journal of Econometrics 139*(1), 4–14.

Chiappori, P. and I. Komunjer (2009). On the nonparametric identification of multiple choice models.

Clemens, J. et al. (2015). Regulatory redistribution in the market for health insurance. *American Economic Journal: Applied Economics 7*(2), 109–34.

Connelly, R. (1992). The effect of child care costs on married women's labor force participation. *The review of Economics and Statistics 74*(1), 83–90.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly 13*(3), 319–340.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of economic literature 48*(2), 424–455.

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica 80*(5), 2231–2267.

Fairlie, R. W. (2004). Race and the digital divide. *Contributions in Economic Analysis & Policy 3*(1).

Florens, J.-P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica 76*(5), 1191–1206.

Grzybowski, L. (2014). Fixed-to-mobile substitution in the european union. *Telecommunications Policy 38*(7), 601–612.

Hagberg, J.-E. (2004). Old people, new and old artefacts- technology for later life. In *Changing worlds and the ageing subject: dimensions in the study of ageing and later life*, New perspectives on ageing and later life, pp. 161–185. Ashgate.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica 42*(4), 679–694.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system1. *Econometrica 46*(4), 931–959.

Heckman, J. J. and S. Urzúa (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics 1*(156), 27–37.

Honoré, B. E. and A. Lewbel (2002). Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica 70*(5), 2053–2063.

Horowitz, J. L. and W. Härdle (1996). Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association 91*(436), 1632–1640.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics 58*(1), 71–120.

Imbens, G. W. (2010). Better LATE than nothing: some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature 48*(2), 399–423.

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512.

Klein, R. W., R. H. Spady, et al. (1993). An efficient semiparametric estimator for binary response models. *Econometrica 61*(2), 387–421.

Kridel, D. J., P. N. Rappoport, and L. D. Taylor (1999). An econometric study of the demand for access to the internet. In *The future of the telecommunications industry: Forecasting and demand analysis*, pp. 21–41. Springer.

Lee, M.-J. (2012). Semiparametric estimators for limited dependent variable (LDV) models with endogenous regressors. *Econometric Reviews 31*(2), 171–214.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics 97*(1), 145–177.

Lewbel, A. (2010). Identification of limited dependent variable models using a special regressor with limited support.

Lewbel, A., D. McFadden, and O. Linton (2011). Estimating features of a distribution from binomial data. *Journal of Econometrics 162*(2), 170–188.

Li, Q. and J. Racine (2003). Nonparametric estimation of distributions with categorical and continuous data. *journal of multivariate analysis 86*(2), 266–292.

Matzkin, R. L. (2007, January). Nonparametric identification. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6 of *Handbook of Econometrics*, Chapter 73, pp. 5307–5368. Elsevier.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. *Econometrica 76*(5), 945–978.

Matzkin, R. L. (2012). Identification in nonparametric limited dependent variable models with simultaneity and unobserved heterogeneity. *Journal of Econometrics 166*(1), 106–115.

Matzkin, R. L. (2015). Estimation of nonparametric models with simultaneity. *Econometrica 83*(1), 1–66.

National Information Society Agency (NIA) (2015). *The Survey on the Digital Divide Index and Status in 2014*.

Nelson, F., L. Olson, et al. (1978). Specification and estimation of a simultaneous-equation model with limited dependent variables. *International Economic Review 19*(3), 695–709.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy 9*(4), 513–548.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Nevo, A. and M. D. Whinston (2010). Taking the dogma out of econometrics: Structural modeling and credible inference. *The Journal of Economic Perspectives 24*(2), 69–81.

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics 36*(3), 231–250.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory 10*(02), 1–21.

Powell, J. L., J. Stock, and T. M. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica 57*(6), 1403–1430.

Rennhoff, A. D. and P. W. Routon (2016). Can you hear me now? The rise of smartphones and their welfare effects. *Telecommunications Policy 40*(1), 39 – 51.

Rivers, D. and Q. H. Vuong (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of econometrics 39*(3), 347–366.

Robertson, A., D. Soopramanien, and R. Fildes (2007). Segmental new-product diffusion of residential broadband services. *Telecommunications Policy 31*(5), 265–275.

Roehrig, C. S. (1988). Conditions for identification in nonparametric and parametic models. *Econometrica 56*(2), 433–47.

Rothe, C. (2009). Semiparametric estimation of binary response models with endogenous regressors. *Journal of Econometrics 153*(1), 51–64.

Rothschild, M. and J. Stiglitz (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics 90*(4), 630–49.

Smith, R. and R. Blundell (1986). An exogeneity test for a simultaneous equation tobit-model with an application to labor supply. *ECONOMETRICA 54*(3), 679–685.

Song, I. and P. K. Chintagunta (2003). A micromodel of new product adoption with heterogeneous and forward-looking consumers: Application to the digital camera category. *Quantitative Marketing and Economics 1*(4), 371–407.

Sriram, S., P. K. Chintagunta, and R. Neelamegham (2006). Effects of brand preference, product attributes, and marketing mix variables in technology product markets. *Marketing Science 25*(5), 440–456.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica 54*(6), 1461–81.

Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica 83*(3), 1185–1197.

Van de Ven, W. P. and B. M. Van Praag (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of econometrics 17*(2), 229–252.

Vogelsang, I. (2010). The relationship between mobile and fixed-line communications: A survey. *Information Economics and Policy 22*(1), 4–17.