ESSAYS IN MULTIPLE FRACTIONAL RESPONSES WITH ENDOGENOUS EXPLANATORY VARIABLES

By

Suhyeon Nam

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Economics - Doctor of Philosophy

2014

ABSTRACT

ESSAYS IN MULTIPLE FRACTIONAL RESPONSES WITH ENDOGENOUS EXPLANATORY VARIABLES

By

Suhyeon Nam

This dissertation consists of three chapters. The first and second chapters develop new estimation methods for multiple fractional response variables with endogeneity. Multiple fractional response variables have two features. Each response is between zero and one, and the sum of the responses for a unit is one. The first chapter proposes an estimation method accounting for these features when there is a continuous endogenous explanatory variable (EEV). It is a two step estimation method combining a control function approach. The first step generates a control function using a linear regression, and the second step maximizes a multinomial log likelihood function with a multinomial logit conditional mean which depends on the control function generated in the first step. Monte Carlo simulations examine the performance of the estimation method when the conditional mean in the second step is misspecified. The simulation results provide evidence that the method's average partial effect (APE) estimates approximate well true APEs as long as an instrument is not weak and that the method's approximation outperforms an alternative linear method's. We apply the proposed two step estimation method to Michigan's fourth grade math test data to estimate the average partial effects of spending on student performance.

The second chapter develops and evaluates an estimation method allowing for the discrete nature of an EEV. We modify the two step estimation method proposed in the first chapter by following Wooldridge (2014); instead of unstandardized residual, we use the generalized residuals as control functions The Monte Carlo simulation demonstrate that although the two step estimation method cannot provide consistent estimators for the mean parameters and average partial effects under the conditional mean misspecification,

it yields a decent approximation to average partial effects.

In the third chapter, we clarify some issues in computing average partial (or marginal) effects in models that have been estimated using control function or correlated random effects approaches (or some combination). In particular, we show that a common method of estimating average partial effects, where the averaging is done across all variables and across the entire sample, estimates an interesting parameter. Nevertheless, the method differs from averaging across the observed covariates the partial effects obtained via the average structural function. In the special case where unobservables are independent of the observed covariates the two methods are identical.

Copyright by SUHYEON NAM 2014 To my husband, Sungsam Chung, my family, and God.

ACKNOWLEDGMENTS

I would not have been able to accomplish my doctoral study without the help and support from many people.

First, I am deeply grateful to my advisor, Professor Jeffrey Wooldridge for his invaluable guidance and continuous support during these long years. I would also like to thank my dissertation committee members, Professor Leslie Papke, Professor Peter Schmidt, and Dr. Kenneth Frank for their insightful comments and encouragements.

I want to express my special thanks to Hajin Kim and Eunsil Lee, who have been my family in Michigan.

My special thanks also go to my family. Especially, I am truly grateful to my parents and parents-in-law for their unconditional love and support. And I thank my husband, Sungsam Chung, who shared with me every moment of this journey.

Finally, I thank God, who has always been with me.

TABLE OF CONTENTS

LIST O	PF TABLES	iii
LIST O	FFIGURES	x
CHAPT	FER 1 MULTIPLE FRACTIONAL RESPONSE VARIABLES WITH CON-	
	TINUOUS ENDOGENOUS EXPLANATORY VARIABLES	1
1.1	INTRODUCTION	1
1.2	THE MODEL AND ESTIMATION WITH ENDOGENEITY	4
1.3	MONTE CARLO SIMULATIONS	11
	1.3.1 The Quantities of Interest	11
	1.3.2 Data Generating Process	12
	1.3.3 Simulation Results	14
1.4	APPLICATION: MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM	
	MATH TEST	50
1.5	CONCLUSION	56
APF	PENDIX	57
REF	ERENCES	78
CHAPT	TER 2 MULTIPLE FRACTIONAL RESPONSE VARIABLES WITH A	
	BINARY ENDOGENOUS EXPLANATORY VARIABLE 8	31
2.1	INTRODUCTION	31
2.2	THE MODIFIED TWO STEP ESTIMATION	33
2.3	MONTE CARLO SIMULATION	39
	2.3.1 Data Generating Process	39
	2.3.2 Simulation Results	90
2.4	CONCLUSION	15
REF	ERENCES	16
CHAPT	FER 3 ON COMPUTING AVERAGE PARTIAL EFFECTS IN MODELS	
	WITH ENDOGENEITY OR HETEROGENEITY	18
3.1	INTRODUCTION	18
3.2	THE AVERAGE STRUCTURAL FUNCTION AND AVERAGE PARTIAL	
	EFFECTS	21
3.3	APES ACROSS THE ENTIRE POPULATION	28
3.4	EMPIRICAL EXAMPLE	34
3.5	CONCLUSION	36
REF	ERENCES	38

LIST OF TABLES

Table 1.1	Average APEs under Condition 1	16
Table 1.2	Percentile APEs under Condition 1 and Normal distribution	17
Table 1.3	Percentile APEs under Condition 1 and χ_3^2 distribution	19
Table 1.4	F statistics of the 1 st step ($H_0: \pi_2 = 0$)	33
Table 1.5	Average APEs under Condition 2 and Normal distribution	34
Table 1.6	Mean Squared Errors of Average APE estimates under Condition 2	44
Table 1.7	Percentile APEs under Condition 2, $\pi_2 = 0.1$, and Normal distribution	46
Table 1.8	Percentile APEs under Condition 2, $\pi_2 = 0.2$, and Normal distribution	48
Table 1.9	Percentile APEs under Condition 2, $\pi_2 = 0.5$, and Normal distribution	50
Table 1.10	Percentile APEs under Condition 2, $\pi_2 = 0.1$, and χ^2_3 distribution	52
Table 1.11	Percentile APEs under Condition 2, $\pi_2 = 0.2$, and χ^2_3 distribution	54
Table 1.12	Percentile APEs under Condition 2, $\pi_2 = 0.5$, and χ^2_3 distribution	56
Table 1.13	Mean Squared Errors of Percentile APE estimates under Condition 2 .	58
Table 1.14	Four levels of MEAP	60
Table 1.15	Summary statistics of the dependent variables	61
Table 1.16	Summary statistics of the data	62
Table 1.17	The first step estimation result	63
Table 1.18	Average APE estimates of spending on the fourth grade MEAP math test	64
Table 1.19	Percentile APE estimates of spending on the fourth grade MEAP math test	65

Table 1.20	(y_{i1}, y_{i2}, y_{i3}) generated across the simulations
Table 2.1	APEs with $r = \rho u + e$, $\pi_2 = 1$, and $e _v \sim Normal(0, 1)$ 92
Table 2.2	APEs with $r = \rho u + e$, $\pi_2 = 1$, and $e _v \sim \chi_3^2$
Table 2.3	MSEs of APEs with $r = \rho u + e$ and $\pi_2 = 1$
Table 2.4	F and Wald statistics of the first stage/step ($H_0: \pi_2 = 0$) 95
Table 2.5	APEs with $w = 1[z_2\pi_2 + u > 0]$, $\rho = 1$, and $e _v \sim Normal(0,1)$ 97
Table 2.6	APEs with $w = 1[z_2\pi_2 + u > 0]$, $\rho = 1$, and $e _v \sim \chi_3^2$
Table 2.7	MSEs of APE estimates with $w = 1[z_2\pi_2 + u > 0]$ and $\rho = 1 \dots 99$
Table 2.8	Rejection Frequencies for $\alpha = 0.05$ test in Case 1 with varying ρ 101
Table 2.9	Rejection Frequencies for $\alpha = 0.05$ test in Case 1 with varying π_2 104
Table 2.10	APEs with $r = \rho u + e$ and $\pi_2 = 1$ in Case 2
Table 2.11	MSEs of APE estimates with $r = \rho u + e$ and $\pi_2 = 1$ in Case 2 109
Table 2.12	APEs with $w = 1[z_2\pi_2 + u > 0]$ and $\rho = 1$ in Case 2
Table 2.13	MSEs of APE estimates with $w = 1[z_2\pi_2 + u > 0]$ and $\rho = 1$ in Case 2 111
Table 2.14	Rejection Frequencies for $\alpha = 0.05$ test in Case 2 with varying ρ 113
Table 2.15	Rejection Frequencies for $\alpha = 0.05$ test in Case 2 with varying π_2 113
Table 3.1	APE estimates in Papke and Wooldridge (2008)
Table 3.2	APE estimates in Chapter 1

LIST OF FIGURES

Figure 1.1	Empirical distributions of Average APE estimates under Condition 1 and Normal distribution	21
Figure 1.2	Empirical distributions of Average APE estimates under Condition 1 and Logistic distribution	22
Figure 1.3	Empirical distributions of Average APE estimates under Condition 1 and χ_3^2 distribution	23
Figure 1.4	Empirical distributions of Percentile APE estimates including w^2 under Condition 1 and Normal distribution	24
Figure 1.5	Empirical distributions of Percentile APE estimates including w^2 under Condition 1 and Logistic distribution	27
Figure 1.6	Empirical distribution of Percentile APEs under Condition 1 and χ_3^2 distribution	30
Figure 1.7	Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.1$, and Normal distribution	35
Figure 1.8	Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.2$, and Normal distribution	36
Figure 1.9	Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.5$, and Normal distribution	37
Figure 1.10	Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.1$, and Logistic distribution	38
Figure 1.11	Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.2$, and Logistic distribution	39
Figure 1.12	Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.5$, and Logistic distribution	40
Figure 1.13	³ Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.1$, and χ_3^2 distribution	41

Figure 1.14	Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.2$, and χ_3^2 distribution
Figure 1.15	Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.5$, and χ_3^2 distribution
Figure 2.1	Rejection Frequencies for Normal(0,1) with varying ρ
Figure 2.2	Rejection Frequencies for χ_3^2 with varying ρ
Figure 2.3	Rejection Frequencies for Normal(0,1) with varying π_2 , ($\pi_1 = \tau = 0$). 105
Figure 2.4	Rejection Frequencies for χ_3^2 with varying π_2 , ($\pi_1 = \tau = 0$) 105
Figure 2.5	Rejection Frequencies for Normal $\left(0, 1 + \frac{1}{2}v^2\right)$ with varying π_2 , $(\pi_1 = \tau = 0)$
Figure 2.6	Rejection Frequencies for Normal $\left(0, 1 + \frac{1}{2}v^2\right)$ with varying π_2 , $(\pi_1 = \tau = 0)$

CHAPTER 1

MULTIPLE FRACTIONAL RESPONSE VARIABLES WITH CONTINUOUS ENDOGENOUS EXPLANATORY VARIABLES

1.1 INTRODUCTION

Fractional responses have interesting functional form issues that have been attracting econometricians' attentions. The research began with a single fractional response, a fractional scalar y_i , which has a salient feature - the bounded nature: $0 \le y_i \le 1$. Then it has moved to two kinds of systems of fractional responses. One is panel data setting in which a cross sectional unit has relatively smaller time periods. The other is multiple responses in which a cross sectional unit has a set of several choices.

For a single fractional response, an OLS estimator or an IV estimator of a linear model are consistent for the parameters in the linear projection. They, however, do not guarantee that their fitted values lie within the unit interval nor that their partial effects estimates for regressors' extreme values are good.¹ The log-odds transformation, $\log \frac{y}{1-y}$, is a traditional solution to recognize the bounded nature. But it requires the response to be strictly inside the unit interval. Papke and Wooldridge (1996) introduce a quasi maximum likelihood estimation (QMLE), which is applicable even when the response takes the boundary values. Their nonlinear estimation method directly models the conditional mean of the response as an appropriate function.

¹ Theses are the same drawbacks as the linear probability model for a binary response has.

Papke and Wooldridge (2008) extend their single fractional response discussion to a panel data setting with allowing for endogeneity. They allow time invariant unobserved effect to be correlated with explanatory variables and develop another QMLE method employing a control function approach to account for endogeneity.

Multiple fractional responses have one additional feature as well as the bounded nature: an adding-up constraint. The sum of a cross sectional unit's multiple responses is one. For example, a researcher studies student performance for a test in a state. Suppose that she is interested in how the public school spending affects the test outcome and her response variable is a set of students' pass and fail shares in a district *i*, $(y_{i,pass}, y_{i,fail})$ where $y_{i,pass} + y_{i,fail} = 1$. This example can fall into the single fractional response category since there are only two shares. However, if the test outcomes are graded into the student's level of proficiency from Level 1 to Level 4 and district-level data she can access contain the four shares $(y_{i,level1}, y_{i,level2}, y_{i,level3}, y_{i,level4})$ where $\sum_{1}^{4} y_{i,levelg} = 1$ instead of the pass and fail rates, the types of behaviors described by the single fractional response analysis are limited. Hence, an alternative estimation method is required to exploit all of the available information when there are more than two shares.

Such an estimation method is proposed by Sivakumar and Bhat (2002). It is a method of QMLE with the multinomial distribution and the multinomial logit conditional mean specification. It is a multivariate generalization of the method proposed by Papke and Wooldridge (1996). Mullahy (2010) studies this method with more detail. Buis (2008) has written a STATA[®] module of this QMLE, and dubs it as "fractional multinomial logit (fmlogit)." In this chapter, we also refer this QMLE as fractional multinomial logit or fmlogit.

Although these studies develop a new estimation method, which can consistently estimate the parameters in the mean as long as the mean specification is correct, they do not address endogeneity. In empirical works, however, endogeneity often arises. In the student performance example, school spending could be endogenous because it is likely to be correlated with some unobserved district effects. This endogeneity issue may lead to inconsistency of the fractional multinomial logit estimation.

In this chapter, we develop an estimation method for multiple fractional responses with endogenous explanatory variables. In the model, we allow a continuous endogenous explanatory variable to be correlated with an unobserved omitted variable. Then we propose a two step estimation method employing a control function approach to deal with the endogeneity. The first step generates a control function and the second step applies fractional multinomial logit with including the control function as an extra regressor in the conditional mean. The method can provide a consistent estimator of the conditional mean parameters provided that the conditional mean specification in the second step is correct.

A distinct feature of this method is that although the multinomial logit specification in the second step is sensible as a multiple fractional responses' conditional mean, it is not underpinned by usual structural assumptions. The functional form of the conditional mean in the second step is determined by the two structural components. One is the functional form of the conditional mean depending on the unobserved omitted variable (structural conditional mean). The other is the distributional assumption of the error, which appears when the control function approach is combined. However, there are no closed forms for them to allow the second step conditional mean to be multinomial logit. Thus we suggest directly specifying the conditional mean of the second step as multinomial logit without usual assumptions about them.

But we would like to examine how the two step estimation method works if the multinomial logit specification is wrong by conducting Monte Carlo simulations. The simulations focus on whether or not the estimates by the proposed two step estimation method can approximate well the average partial effects of the endogenous explanatory variable, which is the partial effects of the endogenous explanatory variable on the conditional mean averaged across the population distribution of the endogenous explanatory vari-

3

able. Further, we compare the method's approximation ability with an alternative linear model's.

The simulation results provide evidence that even though the conditional mean is misspecified, the two step estimation method with a strong instrument yields a good approximation. Although a weak instrument deteriorates its approximation performance, it still outperforms an alternative linear approach.

The rest of the chapter is organized as follows. Section 1.2 describes the model and the two step estimation method. Section 1.3 presents a Monte Carlo simulation design and results where the conditional mean of the two step estimation method is misspecified. Section 1.4 includes an application of the method to examine the relationship between public school spending and the fourth grade math test outcome for Michigan. And Section 1.5 concludes the chapter.

1.2 THE MODEL AND ESTIMATION WITH ENDOGENE-ITY

We assume that a random sampling across the cross section is available, and each cross sectional unit i has G choices where the sum of i's responses is one. The dependent variable for i is expressed as

$$y_{i} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{ig} \\ \vdots \\ y_{iG} \end{pmatrix}_{G \times 1}$$
(1.1)

where

$$0 \le y_{ig} \le 1, \quad g = 1, \cdots, G,$$
 (1.2)

and

$$\sum_{g}^{G} y_{ig} = 1.$$
 (1.3)

(1.2) and (1.3) represent the bounded nature and the adding-up constraint, respectively.

To represent endogeneity in the model, we assume that there is a continuous endogenous explanatory variable w_{ig} , and that it is correlated with an unobserved omitted variable r_{ig} . To simplify the exposition, w_{ig} and r_{ig} are assumed to be invariant across choices: $\forall g, w_{ig} = w_i$ and $r_{ig} = r_i$. Then, for a set of explanatory variables in all choices, $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iG})$, we assume

$$E(y_{ig}|\mathbf{X}_{i},r_{i}) = E(y_{ig}|\mathbf{Z}_{i},w_{i},r_{i}) = G_{g}(\mathbf{Z}_{i1},w_{i},r_{i};\boldsymbol{\beta}), \quad g = 1, \cdots, G,$$
(1.4)

where

$$0 < G_g(\cdot) < 1 \tag{1.5}$$

and

$$\sum_{g}^{G} G_g(\cdot) = 1. \tag{1.6}$$

 $\mathbf{Z}_i \equiv (\mathbf{z}_{i1}, \cdots, \mathbf{z}_{iG})$ is a set of exogenous variables in all choices where $\mathbf{z}_{ig} = (\mathbf{z}_{i1g} \ \mathbf{z}_{i2g})$ indicates exogenous variables for choice g and $\mathbf{Z}_{i1} \equiv (\mathbf{z}_{i11}, \cdots, \mathbf{z}_{i1G})$ is a set of $\mathbf{z}_{i1g}, \forall g$. (1.5) ensures that the fitted value will lie between zero and one. The adding-up constraint (1.3) leads to (1.6). Any function satisfying both (1.5) and (1.6) can be specified for $G(\cdot)$.

 w_i can appear very flexibly in (1.4); for example, we can add w_i^2 to allow for the quadratic effect of w. If w_i and w_i^2 appear in the specification, plug-in methods are subject to the "forbidden regression" problem as Wooldridge (2010) discusses.

To deal with the endogeneity, we employ a control function approach. It includes extra regressors in the estimating equation so that the remaining variation in the endogenous explanatory variable would not be correlated with the unobservables. Since the approach requires an exclusion restriction, a part of Z_i appears in $G_g(\cdot)$. We further assume

$$w_i = f(\mathbf{Z}_i \; ; \; \boldsymbol{\pi}) + v_i \tag{1.7}$$

$$r_i = \rho v_i + e_i \tag{1.8}$$

and

$$(r_i, v_i)$$
 is independent of \mathbf{Z}_i . (1.9)

(1.7) models the endogenous variable w_i as a function of Z_i where π is the parameter vector. It includes the exogenous variables excluded from (1.4) so that the instruments could be allowed to be correlated with w. (1.8) models the omitted variable r_i as a linear function of the reduced form error v_i , which plays a role of the control function - the extra regressor - in this study, where e_i is independent of w_i . (1.8) is for simplicity; it can be allowed to be more flexible by modeling it with polynomial functions of v_i as well as v_i . (1.8) reveals that if there is any correlated with r_i , and consequently tells whether w_i is endogenous or not. Due to (1.9), w_i cannot have discreteness. The independence assumption implies

$$D(e_i | \mathbf{Z}_i, v_i) = D(e_i). \tag{1.10}$$

(1.8) and (1.9) ensure that a single control function, v_i , can correct the endogeneity of w_i even when flexible functional forms for w_i appear in (1.4).

If we assume a parametric model for the distribution in (1.10), then one could derive the mean function conditional on X_i as

$$E(y_{ig}|\mathbf{X}_i) = K_g(\mathbf{Z}_{i1}, w_i, v_i; \boldsymbol{\theta})$$
(1.11)

where

$$0 < K_g(\cdot) < 1 \tag{1.12}$$

and

$$\sum_{g}^{G} K_g(\cdot) = 1. \tag{1.13}$$

If we know the functional form of $K_g(\cdot)$ and v_i is observed, θ can be estimated by nonlinear least squares or a QMLE using multinomial distribution by specifying $K_g(\cdot)$ as a proper functional form satisfying (1.12) and (1.13). However, since v_i is unobserved here, a simple way to estimate θ is to replace v_i with a consistent estimator of v_i and apply one of those estimation methods.

In general, it is difficult to start with function $G_g(\cdot)$ and a distribution for e_i and obtain $K_g(\cdot)$ as a simple function. Instead, the proposal is to directly model $K_g(\cdot)$ parametrically. A natural choice for a proper functional form of $K_g(\cdot)$ is multinomial logit,

$$K_{g}(\mathbf{h}_{i};\boldsymbol{\theta}) = \frac{\exp\left(\mathbf{h}_{i}\boldsymbol{\theta}_{g}\right)}{\sum_{h}^{G}\exp\left(\mathbf{h}_{i}\boldsymbol{\theta}_{h}\right)}$$
(1.14)

where $\mathbf{h}_i = (\mathbf{x}_{i1} \ v_i) = (\mathbf{z}_{i1} \ w_i \ v_i)$ is a 1 × *p* vector, $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1 \ \dots \ \boldsymbol{\theta}'_G)'$ is a *pG* × 1 parameter vector, $\boldsymbol{\theta}_g$ is a *p* × 1 vector, $g = 2, \dots, G$, and $\boldsymbol{\theta}_1 = \mathbf{0}$.² In the basic multinomial logit model, a set of explanatory variables change by unit *i* but not by choice *g*. Its coefficient parameters change by choice *g*, instead.³ In accordance with this choice, we rewrite (1.7) as a linear function of $\mathbf{z}_i = (\mathbf{z}_{i1} \ \mathbf{z}_{i2})_{1 \times M}$:

$$w_i = \mathbf{z}_i \pi + v_i = \mathbf{z}_{i1} \pi_1 + \mathbf{z}_{i2} \pi_2 + v_i \tag{1.15}$$

where $\pi = (\pi'_1 \ \pi'_2)'$ is a $M \times 1$ parameter vector and the constant is subsumed in \mathbf{z}_{i1} . The transformation of w should be carefully chosen to yield (1.15) where v_i is arguably independent of z_i . Plus, we can add \mathbf{z}_i in a flexible way. (1.15) is to simplify the notation.

Then we propose the following procedure for θ :

² The first choice is a reference.

³ Hence this specification is appropriate for problems where the characteristics of choices are unimportant or are not of interest.

Procedure 1.2.1

Step 1. Obtain the OLS residual \hat{v}_i from the regression of w_i on z_i .

Step 2. Apply fractional multinomial logit of $(y_{i1}, y_{i2}, \dots, y_{iG})$ on \mathbf{z}_{i1}, w_i and \hat{v}_i to estimate $\boldsymbol{\theta}$. This is a QMLE with (1.14) and the following log likelihood for *i* replacing v_i with \hat{v}_i :

$$\ell_i(\boldsymbol{\theta}) = \sum_{g}^{G} y_{ig} \log K_g(\mathbf{h}_i; \boldsymbol{\theta}).$$
(1.16)

Procedure 1.2.1 yields a consistent estimator of θ under (1.14). Its consistency does not hinge on whether or not (1.16) is true. It is because the multinomial distribution is a member of the linear exponential family (LEF). Gourieroux et al. (1984) show that a QMLE with a distribution in the LEF provides a consistent estimate of the parameters in a correctly specified conditional mean even when the rest of distribution is misspecified.

Furthermore, Procedure 1.2.1 is able to provide a very useful estimator for the quantity regarding the structural conditional mean $G_g(\cdot)$. Dropping the cross-sectional unit *i*, the partial effect of interest for a continuous explanatory variable x_{1j} , the *j*th element of \mathbf{x}_1 is

$$\frac{\partial E(y_g|\mathbf{x}, r)}{\partial x_{1j}} = \frac{\partial G_g(\mathbf{x}_1, r; \boldsymbol{\beta})}{\partial x_{1j}}, \quad \forall g,$$
(1.17)

where $\mathbf{x} = (\mathbf{z}, w)$ and $\mathbf{x}_1 = (\mathbf{z}_1, w)$. However, (1.17) is not identified because *r* is unobserved. Thus the quantity of more interest is the average partial effect (APE), which can be identified by averaging (1.17) over the distribution of *r*:

$$E_r\left(\frac{\partial G_g(\mathbf{x}_1^0, r; \boldsymbol{\beta})}{\partial x_{1j}}\right), \quad \forall g$$
(1.18)

where the APEs are evaluated at \mathbf{x}_{1}^{0} , a set of fixed values of the covariates. From Wooldridge

(2010, section 2.2.5)

$$E_r\left(\frac{\partial G_g(\mathbf{x}_1^0, r; \boldsymbol{\beta})}{\partial x_{1j}}\right) = E_v\left(\frac{\partial K_g(\mathbf{x}_1^0, v; \boldsymbol{\theta})}{\partial x_{1j}}\right)$$
(1.19)

under (1.9) and (1.15). Hence Procedure 1.2.1 can estimate the APE on the structural conditional mean $G_g(\cdot)$ even though it does not estimate β , the structural parameters.

The asymptotic variances of $\hat{\theta}$ and the APE estimator need to consider the additional variation from the first step of the procedure. The appendix derives their valid asymptotic variances.

Notice that the two step estimation method does not assume anything about a functional form of $G_g(\cdot)$ and a distribution of D(e) although they determine the functional form of $K_g(\cdot)$. If a combination of their specific forms were able to obtain a certain explicit functional form for $K_g(\cdot)$ satisfying (1.12) and (1.13), we would assume those specific forms and maximize the multinomial distribution with a derived $K_g(\cdot)$ from those assumptions. However, there are no closed forms of $G_g(\cdot)$ and D(e) to generate an explicitly known form satisfying (1.12) and (1.13). Considering that $G_g(\cdot)$ should satisfy (1.5) and (1.6), a natural choice for $G_g(\cdot)$ is also multinomial logit. Yet, it cannot derive a closed form of $K_g(\cdot)$ whatever D(e) is. If $G_g(\cdot)$ is specified as (1.20), it can derive an explicit form by assuming that e is normally distributed:

$$G_g(\mathbf{X}_i; \boldsymbol{\beta}) = \Phi(\mathbf{x}_{ig}\boldsymbol{\beta}), \qquad g = 1, \cdots, G - 1,$$

$$G_G(\mathbf{X}_i; \boldsymbol{\beta}) = 1 - \sum_{g}^{G-1} \Phi(\mathbf{x}_{ig}\boldsymbol{\beta}) \qquad (1.20)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Based on the mixing property of the normal distribution, the derived function is similar as (1.20). However, $G_G(\cdot)$ and the derived function for choice *G* are not necessarily between zero and one, which violate (1.5) and (1.12). It is the same drawback that the linear models have. So (1.20) is not appropriate, either. Alternatively, the two step estimation method suggests directly specifying $K_g(\cdot)$ as multinomial logit. That is, instead of making usual assumptions about $G_g(\cdot)$ and D(e), it implicitly assumes that their combination derives the multinomial logit functional form of $K_g(\cdot)$.⁴ To be cautious about this, we conduct Monte Carlo simulations to investigate how it works as an approximation when the specification is wrong.

Some researchers may be inclined to use a linear model rather than a nonlinear model, in which case it would be natural to drop one of the *G* equations and apply a linear method, say a linear control function (LCF) approach to remaining equations. Then, for choice g, w's coefficient parameter estimate by the linear method is comparable to (1.23). Procedure 1.2.2 summarizes the LCF approach:

Procedure 1.2.2

- **Step 1.** Obtain the OLS residual \hat{v}_i from the regression of w_i on z_i . This is the same as Step 1 of Procedure 1.2.1.
- **Step 2.** For each $g = 2, \dots, G$,⁵ regress y_{ig} on \mathbf{z}_{i1} , w_i and \hat{v}_i to estimate γ_g , where $\gamma_g = (\gamma'_{\mathbf{z}g} \ \gamma_{wg} \ \gamma_{vg})'$ is a 4 × 1 coefficient parameter vector for choice g. Obtain $\hat{\gamma}_1$ based on $\gamma_1 = \mathbf{e}_1 \gamma_2 \gamma_3$, where \mathbf{e}_1 is a 4 × 1 unit vector.⁶

The asymptotic variance of $\hat{\gamma}_g$ also needs the adjustment taking the extra variation from the first step into account; see the appendix.

The simulations compare the approximation by the two step estimation method with the misspecified conditional mean and one by this LCF approach.

⁴ The approach reflects the manner in which Petrin and Train (2010) employ a control function approach when their dependent variable is a multinomial choice. They divide the structural error in their consumer utility into two parts to generate a mixed logit. Without a distributional assumption of the structural error, one divided part is assumed to be normal and the other is assumed to be type 1 extreme value.

⁵ The first choice is dropped as the reference choice.

⁶ The coefficients of a variable across choices sum to be 0 and those of the constant sum to be 1 because of (1.3).

1.3 MONTE CARLO SIMULATIONS

1.3.1 The Quantities of Interest

The quantity of interest in the simulations is the APE of the endogenous explanatory variable w,

$$E_r\left(\frac{\partial G_g(\mathbf{x}_1^0, r; \boldsymbol{\beta})}{\partial w}\right) = E_r\left(\frac{\partial G_g(\mathbf{z}_1^0, w^0, r; \boldsymbol{\beta})}{\partial w}\right), \quad \forall g.$$
(1.21)

Since (1.21) depends on where it is evaluated, the simulations use two approaches to obtain a single number. One is averaging (1.21) out across the sample again and the other is evaluating (1.21) at a certain set of values, (\bar{z}_1, w_p) where \bar{z}_1 is the mean of z_1 and w_p stands for the p^{th} percentile of w's distribution. We call the former "average APE" and the latter "percentile APE."

If the two step estimation method's mean specification (1.14) is correct, (1.21) is obtained by estimating

$$E_{v}\left[\frac{\partial K_{g}(\mathbf{x}_{1}^{0}, v; \boldsymbol{\theta})}{\partial w}\right] = E_{v}\left[K_{g}(\mathbf{x}_{1}^{\circ}, v; \boldsymbol{\theta}) \cdot \left(\theta_{wg} - \frac{\sum_{h}^{G} \theta_{wh} \exp\left(\mathbf{x}_{1}^{\circ} \boldsymbol{\theta}_{\mathbf{x}h} + \theta_{vh} v\right)}{\sum_{h}^{G} \exp\left(\mathbf{x}_{1}^{\circ} \boldsymbol{\theta}_{\mathbf{x}h} + \theta_{vh} v\right)}\right)\right]$$
(1.22)

where $\theta_{\mathbf{x}h} = (\theta'_{\mathbf{z}h} \ \theta_{wh})'$. Since the distribution of v is not assumed, (1.22) can be estimated by averaging out \hat{v}_i across the sample, instead:

$$\frac{1}{N}\sum_{i}^{N}\left[K_{g}(\mathbf{x}_{1}^{0},\widehat{v}_{i};\widehat{\boldsymbol{\theta}})\cdot\left(\widehat{\theta}_{wg}-\frac{\sum_{h}^{G}\widehat{\theta}_{wh}\exp\left(\mathbf{x}_{1}^{\circ}\widehat{\boldsymbol{\theta}}_{\mathbf{x}h}+\widehat{\theta}_{vh}\widehat{v}_{i}\right)}{\sum_{h}^{G}\exp\left(\mathbf{x}_{1}^{\circ}\widehat{\boldsymbol{\theta}}_{\mathbf{x}h}+\widehat{\theta}_{vh}\widehat{v}_{i}\right)}\right)\right]$$
(1.23)

where $\hat{\theta}$ is obtained from Procedure 1.2.1.

The simulations let (1.14) be misspecified, and so we examine how close (1.23) is to (1.21), if it is closer than the estimates by the LCF approach.

Some simulations allow for w's quadratic effect by including w^2 in the model. These simulations add \hat{v}_i^2 and \hat{v}_i^3 in the two procedures' second steps to see if it improves their

approximations.

1.3.2 Data Generating Process

For the simulations, the number of observation *N* and choice *G* are 500 and 3, respectively. We use 1000 replications.

The covariates For each replication, we generate 500 observations of \mathbf{z}_i , w_i , r_i , v_i and e_i as following.

• $\mathbf{z}_i = (\mathbf{z}_{i1} \ z_{i2}) = (1 \ z_{i1} \ z_{i2})_{1 \times 3}$ where $\mathbf{z}_{i1} = (1 \ z_{i1})$ and

$$\begin{pmatrix} z_{i1} \\ z_{i2} \end{pmatrix} \sim MV Normal \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix} \right], \quad \tau \in \{0, -0.5\}.$$

There are one included exogenous variable and one excluded exogenous variable where they are drawn from the multivariate normal distribution. A simulation allows them to be correlated: $\tau = -0.5$.

- *D*(*e*) is one of the three distributions:
 - (a) $e_i \sim Normal(0, 1)$
 - (b) $e_i \sim Logistic(0, 1)$
 - (c) $e_i \sim \chi_3^2$

To study various misspecifications, three distributions of *e* are in use: two symmetric distributions and one asymmetric distribution.

• $v_i \sim Normal(0, \sigma^2)^7$

 $^{7 \}sigma^2$ is adjusted for the variance of w_i to be invariant across the simulations.

• $w_i = \pi_1 z_{i1} + \pi_2 z_{i2} + v_i$

The endogenous variable is generated based on (1.15). The coefficient parameter for the constant is set to be zero.

• $r_i = \rho v_i + e_i$

The omitted variable is generated based on (1.8).

The structural conditional mean $G_{g}(\cdot)$ **specification:** We specify $G_{g}(\cdot)$ as multinomial logit because it satisfies (1.5) and (1.6):

$$E(y_{ig}|\mathbf{x}_i, r_i) = G_g(\mathbf{z}_{i1}, w_i, r_i; \boldsymbol{\beta}) = \frac{\exp\left(\mathbf{z}_{i1}\boldsymbol{\beta}_{zg} + w_i\boldsymbol{\beta}_{wg} + r_i\boldsymbol{\beta}_{rg}\right)}{\sum_h^3 \exp\left(\mathbf{z}_{i1}\boldsymbol{\beta}_{zh} + w_i\boldsymbol{\beta}_{wh} + r_i\boldsymbol{\beta}_{rh}\right)}$$
(1.24)

where $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1 \ \boldsymbol{\beta}'_2 \ \boldsymbol{\beta}'_3)'$ is a 12 × 1 parameter vector, $\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{zg} \ \boldsymbol{\beta}_{wg} \ \boldsymbol{\beta}_{rg})'$ is a 4 × 1 parameter vector for g = 2, 3, and $\boldsymbol{\beta}_1 = \mathbf{0}$ since the first choice is chosen as a reference. Other parameters are set to be 1: $\boldsymbol{\beta}_g = (1 \ 1 \ 1 \ 1)'$ for $g = 2, 3.^8$ Note that (1.14) is misspecified under (1.24) and any of the three distributions for *e*.

The multiple fractional dependent variables y: The multiple fractional dependent variables for each observation *i* are generated by the following process.

- (1) Calculate the response probabilities G_{i1} , G_{i2} , and G_{i3} by using (1.24) and the covariates generated above.⁹
- (2) Draw 100 multinomial outcomes among 1, 2, and 3 based on the calculated response probabilities.

$$G_g(\mathbf{z}_{i1}, w_i, r_i; \boldsymbol{\beta}) = \frac{\exp\left(\mathbf{z}_{i1}\boldsymbol{\beta}_{zg} + w_i\boldsymbol{\beta}_{wg} + w_i^2\boldsymbol{\beta}_{w^2g} + r_i\boldsymbol{\beta}_{rg}\right)}{\sum_h^3 \exp\left(\mathbf{z}_{i1}\boldsymbol{\beta}_{zh} + w_i\boldsymbol{\beta}_{wh} + w_i^2\boldsymbol{\beta}_{w^2h} + r_i\boldsymbol{\beta}_{rh}\right)}$$
(1.25)

where $\boldsymbol{\beta}$ is a 15 × 1 parameter vector, $\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{zg} \ \beta_{wg} \ \beta_{w^2g} \ \beta_{rg})' = (1 \ 1 \ 1 \ -0.1 \ 1)'$ for g = 2, 3. ⁹ (1.25) is used for the model including w^2 .

⁸ For the simulations including w^2 ,

(3) Count the frequencies and obtain the proportion for each outcome.

For instances, if 1 is drawn 50 times, 2 is drawn 30 times, and 3 is drawn 20 times for an observation *i*, then $y_{i1} = 0.5$, $y_{i2} = 0.3$, and $y_{i3} = 0.2$.¹⁰ Appendix includes a table showing the summary of (y_{i1}, y_{i2}, y_{i3}) generated by this process.

1.3.3 Simulation Results

The first columns of the simulation result tables show whether or not the quadratic effect of w is included in the model: w indicates that it is not, and w^2 indicates that it is. The tables report the mean of (1.21) over the 1000 replications (True) along with the results of the two step estimation method and the LCF approach - the means of the estimates (Mean), their standard deviations (SD), and the means of their adjusted standard errors (SE). For the model including the quadratic effect, there are two additional estimation results allowing for the control function in a flexible way: Two step (flexible) and LCF (flexible). They includ \hat{v}, \hat{v}^2 , and \hat{v}^3 in the second stage; We would like to examine if it helps the approximations.

Condition 1: $\pi_2 = 1, \rho = 1$

In generating the data for Condition 1, we allow the instrument and the endogeneity to be strong: $w_i = z_{i2} + v_i$ and $r_i = v_i + e_i$.^{11,12}

The simulation results under Condition 1 demonstrate that while both the two step estimation method and the LCF approach provide good approximations to average APEs, the two step estimation method provides better percentile APE estimates. In Table 1.1, the average APE estimates by the two methods are quite similar to true APEs. However,

¹⁰ Through this process, the upper corner 1 is generated only for the reference choice while the lower corner 0 is generated for all three choices because of the multinomial logit response probabilities structure.

¹¹ z_{i1} has no effect on w_i : $\pi_1 = 0$ ¹² The simulations allowing z_{i1} to affect w_i ($w_i = 0.5z_{i1} + z_{i2} + v_i$, $\tau = -0.5$, $\rho = 1$) provide similar results as those under Condition 1.

Table 1.2 and Table 1.3 illustrate that the percentile APE estimates by the two step estimation method have less biased without any sign distortions across the percentiles of w's distribution than those by the LCF approach.¹³ The percentile estimates by the LCF approach under the χ_3^2 distribution have the opposite directions to the true APEs at the 90th percentile.

The results also present that allowing for the flexible forms of \hat{v}_i does not help the approximations when w's quadratic effect is included in the models. In Table 1.1 through 1.3, the estimates with allowing for the flexible forms of \hat{v}_i are similar to those without it.

The empirical distributions of APE estimates in Figure 1.1 through 1.6, confirm these results.

¹³ The results of percentile APEs under the logistic distribution are similar to those under the normal distribution in general.

	D(e)			Normal			ogistic		χ_3^2			
	8		1	2	3	1	2	3	1	2	3	
w	True	Mean	-0.117	0.059	0.059	-0.108	0.054	0.054	-0.049	0.025	0.025	
	Two Step	Mean	-0.117	0.059	0.058	-0.108	0.054	0.054	-0.048	0.024	0.024	
		SD	0.008	0.005	0.005	0.011	0.006	0.006	0.007	0.004	0.004	
		SE	-	-	-	-	-	-	-	-	-	
	LCF	Mean	-0.109	0.055	0.054	-0.102	0.051	0.051	-0.052	0.026	0.026	
		SD	0.009	0.005	0.005	0.011	0.006	0.006	0.008	0.005	0.005	
		SE	0.009	0.005	0.005	0.011	0.006	0.006	0.008	0.005	0.005	
w^2	True	Mean	-0.127	0.063	0.063	-0.116	0.058	0.058	-0.061	0.031	0.031	
	Two Step	Mean	-0.127	0.063	0.063	-0.115	0.058	0.058	-0.060	0.030	0.030	
	-	SD	0.009	0.005	0.005	0.011	0.006	0.006	0.008	0.004	0.005	
		SE	-	-	-	-	-	-	-	-	-	
	Two Step	Mean	-0.126	0.063	0.063	-0.115	0.057	0.057	-0.058	0.029	0.029	
	(Flexible)	SD	0.009	0.005	0.005	0.011	0.006	0.006	0.008	0.004	0.004	
		SE	-	-	-	-	-	-	-	-	-	
	LCF	Mean	-0.118	0.059	0.059	-0.108	0.054	0.054	-0.064	0.032	0.032	
		SD	0.009	0.005	0.005	0.011	0.006	0.006	0.008	0.005	0.005	
		SE	0.008	0.005	0.005	0.011	0.006	0.006	0.008	0.004	0.004	
	LCF	Mean	-0.118	0.059	0.059	-0.109	0.054	0.054	-0.064	0.032	0.032	
	(Flexible)	SD	0.008	0.005	0.005	0.011	0.006	0.006	0.008	0.005	0.005	
		SE	0.008	0.005	0.005	0.011	0.006	0.006	0.008	0.004	0.004	

Table 1.1: Average APEs under Condition 1

^{1.} $\pi_1 = 0, \pi_2 = 1, \rho = 1$ ^{2.} We cannot obtain the standard errors of average APE estimates by the two step estimation method; the process in STATA to calculate them takes too much time to complete.

w_p		10^{th}		25^{th}			50^{th}			75 th				
	8		1	2	3	1	2	3	1	2	3	1	2	3
w	True	Mean	-0.181	0.090	0.090	-0.171	0.085	0.085	-0.132	0.066	0.066	-0.085	0.043	0.043
	Two Step	Mean	-0.184	0.092	0.092	-0.173	0.087	0.087	-0.131	0.066	0.066	-0.082	0.041	0.041
		SD	0.016	0.008	0.008	0.018	0.009	0.009	0.012	0.006	0.006	0.007	0.004	0.004
		SE	0.016	0.008	0.008	0.018	0.009	0.009	0.011	0.006	0.006	0.004	0.003	0.003
w^2	True	Mean	-0.241	0.121	0.121	-0.206	0.103	0.103	-0.133	0.066	0.066	-0.073	0.036	0.036
	Two Step	Mean	-0.240	0.120	0.120	-0.208	0.104	0.104	-0.133	0.067	0.066	-0.072	0.036	0.036
		SD	0.022	0.011	0.011	0.021	0.011	0.011	0.014	0.007	0.007	0.009	0.005	0.006
		SE	0.021	0.011	0.011	0.020	0.010	0.010	0.011	0.006	0.006	0.007	0.005	0.005
	Two Step	Mean	-0.240	0.120	0.120	-0.208	0.104	0.104	-0.132	0.066	0.066	-0.071	0.036	0.036
	(Flexible)	SD	0.022	0.011	0.011	0.021	0.011	0.011	0.014	0.008	0.007	0.010	0.006	0.006
		SE	0.021	0.011	0.011	0.020	0.010	0.010	0.011	0.006	0.006	0.008	0.005	0.005
	LCF	Mean	-0.206	0.103	0.103	-0.164	0.082	0.082	-0.118	0.059	0.059	-0.071	0.036	0.035
		SD	0.018	0.009	0.009	0.013	0.007	0.007	0.009	0.005	0.005	0.010	0.006	0.006
		SE	0.016	0.008	0.008	0.012	0.006	0.006	0.008	0.005	0.005	0.010	0.006	0.006
	LCF	Mean	-0.190	0.095	0.095	-0.156	0.078	0.078	-0.118	0.059	0.059	-0.080	0.040	0.040
	(Flexible)	SD	0.018	0.009	0.009	0.013	0.007	0.006	0.009	0.005	0.005	0.010	0.005	0.006
		SE	0.016	0.008	0.009	0.012	0.006	0.006	0.008	0.005	0.005	0.009	0.005	0.005

Table 1.2: Percentile APEs under Condition 1 and Normal distribution

^{1.} APEs at (\overline{z}_1, w_p) . ^{2.} $\pi_1 = 0, \pi_2 = 1, \rho = 1$.

	w_p		90^{th}					
	8		1	2	3			
w	True	Mean	-0.050	0.025	0.025			
	Two Step	Mean	-0.048	0.024	0.024			
		SD	0.006	0.004	0.004			
		SE	0.003	0.003	0.003			
w^2	True	Mean	-0.040	0.020	0.020			
	Two Step	Mean	-0.039	0.020	0.019			
		SD	0.007	0.005	0.006			
		SE	0.005	0.005	0.005			
	Two Step	Mean	-0.039	0.019	0.019			
	(Flexible)	SD	0.007	0.006	0.006			
		SE	0.006	0.005	0.005			
	LCF	Mean	-0.029	0.014	0.014			
		SD	0.016	0.008	0.009			
		SE	0.014	0.008	0.008			
	LCF	Mean	-0.045	0.023	0.023			
	(Flexible)	SD	0.014	0.008	0.008			
		SE	0.012	0.007	0.007			

Table 1.2: (cont'd)

w_p		10^{th}				25^{th}		50^{th}			75^{th}			
	8		1	2	3	1	2	3	1	2	3	1	2	3
w	True	mean	-0.100	0.050	0.050	-0.068	0.034	0.034	-0.038	0.019	0.019	-0.018	0.009	0.009
	Two Step	Mean	-0.096	0.048	0.048	-0.061	0.031	0.030	-0.033	0.017	0.017	-0.017	0.009	0.009
		SD	0.022	0.011	0.011	0.011	0.006	0.006	0.004	0.003	0.003	0.002	0.003	0.003
		SE	0.020	0.010	0.010	0.010	0.005	0.005	0.003	0.003	0.003	0.002	0.003	0.003
w^2	True	Mean	-0.153	0.076	0.076	-0.085	0.043	0.043	-0.038	0.019	0.019	-0.016	0.008	0.008
	Two Step	Mean	-0.134	0.067	0.067	-0.081	0.041	0.040	-0.040	0.020	0.020	-0.018	0.009	0.009
		SD	0.033	0.017	0.017	0.013	0.007	0.007	0.007	0.004	0.004	0.004	0.003	0.003
		SE	0.029	0.015	0.015	0.012	0.007	0.007	0.006	0.004	0.004	0.003	0.003	0.003
	Two Step	Mean	-0.133	0.067	0.066	-0.078	0.039	0.039	-0.039	0.019	0.019	-0.018	0.009	0.009
	(Flexible)	SD	0.032	0.016	0.016	0.013	0.007	0.007	0.007	0.004	0.004	0.004	0.003	0.004
		SE	0.028	0.015	0.015	0.012	0.007	0.007	0.006	0.005	0.005	0.004	0.005	0.005
	LCF	Mean	-0.162	0.081	0.081	-0.116	0.058	0.058	-0.064	0.032	0.032	-0.013	0.006	0.006
		SD	0.019	0.010	0.010	0.013	0.007	0.007	0.009	0.005	0.005	0.008	0.005	0.005
		SE	0.017	0.009	0.009	0.012	0.007	0.007	0.008	0.004	0.005	0.007	0.005	0.005
	LCF	Mean	-0.143	0.072	0.071	-0.106	0.053	0.053	-0.064	0.032	0.032	-0.023	0.011	0.011
	(Flexible)	SD	0.022	0.011	0.011	0.015	0.008	0.008	0.008	0.005	0.005	0.007	0.004	0.005
		SE	0.020	0.010	0.010	0.014	0.007	0.007	0.008	0.004	0.004	0.006	0.004	0.004

Table 1.3: Percentile APEs under Condition 1 and χ^2_3 distribution

^{1.} APEs at (\overline{z}_1, w_p) . ^{2.} $\pi_1 = 0, \pi_2 = 1, \rho = 1$. ^{3.} The grey colored cells indicate that at least one of the APE estimates for three choices has the opposite direction to its true APE.

	w_p		90^{th}				
	8		1	2	3		
w	True	Irue mean		0.004	0.004		
	Two Step	Mean	-0.009	0.005	0.005		
		SD	0.002	0.003	0.003		
		SE	0.001	0.003	0.003		
w^2	True	Mean	-0.007	0.004	0.004		
	Two Step	Mean	-0.008	0.004	0.004		
		SD	0.002	0.004	0.004		
		SE	0.002	0.004	0.004		
	Two Step	Mean	-0.008	0.004	0.004		
	(Flexible)	SD	0.002	0.004	0.005		
		SE	0.002	0.006	0.006		
	LCF	Mean	0.034	-0.017	-0.017		
		SD	0.012	0.007	0.007		
		SE	0.011	0.006	0.006		
	LCF	Mean	0.015	-0.007	-0.007		
	(Flexible)	SD	0.012	0.007	0.007		
		SE	0.010	0.006	0.006		

Table 1.3: (cont'd)



Figure 1.1: Empirical distributions of Average APE estimates under Condition 1 and Normal distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.2: Empirical distributions of Average APE estimates under Condition 1 and Logistic distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.3: Empirical distributions of Average APE estimates under Condition 1 and χ^2_3 distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.4: Empirical distributions of Percentile APE estimates including w^2 under Condition 1 and Normal distribution

Figure 1.4: (cont'd)


Figure 1.4: (cont'd)





Figure 1.5: Empirical distributions of Percentile APE estimates including w^2 under Condition 1 and Logistic distribution

Figure 1.5: (cont'd)



Figure 1.5: (cont'd)





Figure 1.6: Empirical distribution of Percentile APEs under Condition 1 and χ^2_3 distribution

Figure 1.6: (cont'd)



Figure 1.6: (cont'd)



Condition 2: $\pi_2 < 1, \rho = 1$

To see if the above results are dependent on the instrument's strong predictive power, we consider a condition where its predictive power is weaker than Condition 1 by generating the data with $w_i = \pi_2 z_{i2} + v_i$ where $\pi_2 \in \{0.1, 0.2, 0.5\}$ and $r_i = v_i + e_i$.

Staiger and Stock (1997) suggest a guideline for dividing weak and strong instruments by using the first step's F statistic, which tests the hypothesis that the instruments are not correlated with the endogenous regressor; a threshold that they suggest is a F statistic of 10. Table 1.4 shows that the mean of F statistics testing the null hypothesis of π_2 being 0 is not larger than 10 until $\pi_2 = 0.2$. It also shows that about a half of 1000 replications has the F statistic which is larger than 10 as $\pi_2 = 0.2$ and every replication has it as $\pi_2 = 0.5$. Thus, according to the Staiger and Stock's discussion, the instrument is weak as $\pi_2 = 0.1$ and strong as $\pi_2 = 0.5$. When $\pi_2 = 0.2$, it barely manages to have properties as a strong instrument.

Table 1.4: F statistics of the 1^{st} step ($H_0: \pi_2 = 0$)

π_2	2	0.1	0.2	0.5	1
F statistics	Mean	3.379	10.942	71.888	507.208
	SD	(3.436)	(6.665)	(19.440)	(80.357)
	(F > 10)	0.053	0.485	1.000	1.000

^{1.} (F > 10) stands for the proportion of the F statistics being greater than 10 among the 1000 replications.

Table 1.5 and Figure 1.7 to 1.9 illustrate that under the normal distribution, the average APE estimates by both the two methods are more biased and more volatile as π_2 decreases; the weak instrument makes their approximations worse.¹⁴ But the mean squared errors (MSEs) in Table 1.6 suggest that, for all three distributions, the LCF approach is worse than the two step estimation method as a weak instrument is in use.

¹⁴ Figure 1.10 through 1.15 show that the results under the other distributions are similar.

	$\frac{\pi_2}{g}$			0.1			0.2			0.5	
	8		1	2	3	1	2	3	1	2	3
w	True	Mean	-0.113	0.056	0.056	-0.113	0.057	0.057	-0.114	0.057	0.057
	Two Step	Mean	-0.087	0.045	0.042	-0.096	0.048	0.048	-0.111	0.055	0.055
		SD	0.110	0.059	0.067	0.061	0.032	0.033	0.020	0.011	0.011
		SE	-	-	-	-	-	-	-	-	-
	LCF	Mean	-0.008	0.033	-0.025	-0.085	0.043	0.042	-0.098	0.049	0.049
		SD	2.509	1.511	1.318	0.105	0.052	0.055	0.021	0.011	0.011
		SE	36.96	27.92	19.25	0.088	0.045	0.047	0.021	0.011	0.011
w^2	True	Mean	-0.121	0.061	0.061	-0.121	0.061	0.061	-0.122	0.061	0.061
	Two Step	Mean	-0.091	0.045	0.046	-0.102	0.051	0.051	-0.119	0.059	0.059
	_	SD	0.113	0.063	0.065	0.064	0.033	0.034	0.021	0.011	0.011
		SE	-	-	-	-	-	-	-	-	-
	Two Step	Mean	-0.085	0.042	0.043	-0.097	0.049	0.049	-0.117	0.058	0.058
	(Flexible)	SD	0.109	0.062	0.064	0.063	0.034	0.034	0.022	0.011	0.012
		SE	-	-	-	-	-	-	-	-	-
	LCF	Mean	0.035	-0.036	0.001	-0.091	0.045	0.046	-0.106	0.053	0.053
		SD	2.954	1.624	1.732	0.094	0.044	0.054	0.020	0.011	0.011
		SE	43.42	29.42	26.22	0.082	0.041	0.045	0.020	0.011	0.011
	LCF	Mean	0.032	-0.033	0.000	-0.092	0.046	0.046	-0.106	0.053	0.053
	(Flexible)	SD	3.301	1.482	2.076	0.099	0.046	0.056	0.020	0.010	0.011
		SE	47.57	22.03	36.71	0.082	0.041	0.045	0.019	0.010	0.010

Table 1.5: Average APEs under Condition 2 and Normal distribution

^{1.} $\pi_1 = 0, \rho = 1.$ ^{2.} We cannot obtain the standard errors of average APE estimates by the two step estimation method; the process in STATA to calculate them takes too much time to complete.

^{3.} The grey colored cells indicate that at least one of the APE estimates for three choices has the opposite direction to its true APE.



Figure 1.7: Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.1$, and Normal distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.8: Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.2$, and Normal distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.9: Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.5$, and Normal distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.10: Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.1$, and Logistic distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.11: Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.2$, and Logistic distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.12: Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.5$, and Logistic distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.13: Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.1$, and χ_3^2 distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.14: Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.2$, and χ_3^2 distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .



Figure 1.15: Empirical distributions of Average APE estimates under Condition 2, $\pi_2 = 0.5$, and χ_3^2 distribution

^{1.}The distributions in the first row are those of the estimates when the model includes only w and those in the second row are when the model includes w^2 .

	π_2		0.1			0.2			0.5			1	
	8	1	2	3	1	2	3	1	2	3	1	2	3
					Nc	rmal							
w	Two Step	0.013	0.004	0.005	0.004	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	LCF	6.305	2.285	1.743	0.012	0.003	0.003	0.001	0.000	0.000	0.000	0.000	0.000
w^2	Two Step	0.014	0.004	0.004	0.004	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	Two Step (Flexible)	0.013	0.004	0.004	0.005	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	LCF	8.752	2.647	3.003	0.010	0.002	0.003	0.001	0.000	0.000	0.000	0.000	0.000
	LCF (Flexible)	10.923	2.204	4.314	0.011	0.002	0.003	0.001	0.000	0.000	0.000	0.000	0.000
					Lo	gistic							
w	Two Step	0.015	0.004	0.005	0.005	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
	LCF	9.419	1.312	5.220	0.013	0.003	0.003	0.001	0.000	0.000	0.000	0.000	0.000
w^2	Two Step	0.016	0.005	0.005	0.006	0.001	0.002	0.001	0.000	0.000	0.000	0.000	0.000
	Two Step (Flexible)	0.015	0.005	0.004	0.006	0.001	0.002	0.001	0.000	0.000	0.000	0.000	0.000
	LCF	15.493	5.048	3.587	0.012	0.003	0.003	0.001	0.000	0.000	0.000	0.000	0.000
	LCF (Flexible)	12.334	3.486	3.412	0.012	0.003	0.004	0.001	0.000	0.000	0.000	0.000	0.000
						χ^2_3							
w	Two Step	0.011	0.003	0.004	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	LCF	2.386	0.855	0.538	0.007	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000
w^2	Two Step	0.010	0.004	0.003	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	Two Step (Flexible)	0.011	0.004	0.004	0.003	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	LCF	5.328	2.061	1.312	0.004	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	LCF (Flexible)	5.039	1.914	1.272	0.004	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000

Table 1.6: Mean Squared Errors of Average APE estimates under Condition 2

^{1.} The grey colored cells indicate that at least one of the APE estimates for three choices has the opposite direction to its true APE.

Table 1.7 through 1.12^{15} show that percentile APE estimates have similar patters with those in average APE estimates. Plus, the tables presents that the weak instrument causes huge standard errors at some replications.¹⁶ The standard errors significantly decrease as π_2 rises. As in the average APE estimates, Table 1.13 shows that the weak instrument causes the LCF approach to have much worse approximation than the two step estimation method.

Under Condition 2, the results does not provide enough evidence that including \hat{v}_i in a flexible way helps their approximations. Actually, with the weak instrument, it makes the two methods' approximations worse; it causes the two step estimation method to have enormous standard errors at several replications and the two methods to have more sign distortions.

Hence the results under Condition 2 demonstrate that the quality of an instrument affects the two methods' approximation performances, and that the two step estimation method is less sensitive to a weak instrument than the LCF approach. Plus, it is better not to include the additional terms of the control function for their approximations if a weak instrument is used.

¹⁵ The logistic distribution has similar results to the normal distribution.

¹⁶ There are big differences between the medians and the means of the standard errors except the two step estimation.

w _p				10^{th}			25^{th}			50^{th}	
	8		1	2	3	1	2	3	1	2	3
w	True	Mean	-0.164	0.082	0.082	-0.156	0.078	0.078	-0.127	0.064	0.064
	Two Step	Mean	-0.125	0.064	0.061	-0.160	0.082	0.078	-0.112	0.058	0.054
		SD	0.122	0.063	0.069	0.173	0.088	0.097	0.147	0.078	0.088
		SE	0.232	0.132	0.148	0.485	0.259	0.353	0.355	0.269	0.246
w^2	True	Mean	-0.219	0.110	0.110	-0.188	0.094	0.094	-0.127	0.064	0.064
	Two Step	Mean	-0.150	0.075	0.075	-0.182	0.090	0.092	-0.112	0.055	0.056
		SD	0.125	0.068	0.069	0.180	0.096	0.096	0.147	0.084	0.083
		SE	0.237	0.156	0.158	0.367	0.258	0.224	0.305	0.211	0.200
	Two Step	Mean	-0.146	0.074	0.073	-0.177	0.088	0.089	-0.110	0.054	0.056
	(Flexible)	SD	0.147	0.084	0.085	0.180	0.099	0.101	0.154	0.088	0.087
		SE	8E+04	2E+13	6E+11	5E+07	7E+12	3E+11	8E+09	3E+12	9E+10
		SE*	0.155	0.085	0.081	0.180	0.094	0.093	0.134	0.071	0.075
	LCF	Mean	-0.063	0.013	0.050	-0.016	-0.010	0.027	0.035	-0.036	0.001
		SD	2.953	1.624	1.732	2.954	1.624	1.732	2.954	1.624	1.732
		SE	43.42	29.42	26.22	43.42	29.41	26.22	43.42	29.41	26.22
		SE*	0.106	0.058	0.057	0.105	0.057	0.057	0.105	0.057	0.057
	LCF	Mean	-0.009	-0.012	0.021	0.011	-0.022	0.011	0.032	-0.032	0.001
	(Flexible)	SD	3.467	1.563	2.148	3.385	1.521	2.113	3.299	1.481	2.075
		SE	47.77	22.13	36.81	47.65	22.07	36.75	47.57	22.03	36.71
		SE*	0.168	0.088	0.090	0.130	0.068	0.069	0.099	0.055	0.055

Table 1.7: Percentile APEs under Condition 2, $\pi_2 = 0.1$, and Normal distribution

 APEs at (z
₁, w_p).
 π₁ = 0, ρ = 1.
 SE* is the median of the standard errors.
 The grey colored cells indicate that at least one of the APE estimates for three choices has the opposite direction to its true APE.

	w_p			75 th			90 th	
	8		1	2	3	1	2	3
w	True	Mean	-0.089	0.044	0.044	-0.057	0.028	0.028
	Two Step	Mean	-0.044	0.023	0.020	-0.014	0.008	0.006
		SD	0.092	0.056	0.063	0.058	0.043	0.045
		SE	0.456	0.489	0.438	0.213	0.235	0.193
w^2	True	Mean	-0.075	0.037	0.037	-0.044	0.022	0.022
	Two Step	Mean	-0.035	0.016	0.018	-0.003	0.001	0.003
		SD	0.094	0.058	0.062	0.062	0.044	0.046
		SE	0.294	0.284	0.273	0.251	0.178	0.190
	Two Step	Mean	-0.031	0.015	0.016	0.010	-0.005	-0.005
	(Flexible)	SD	0.116	0.070	0.069	0.098	0.065	0.065
		SE	2E+09	1E+12	5E+10	2E+14	5E+12	2E+11
		SE*	0.094	0.061	0.060	0.087	0.068	0.069
	LCF	Mean	0.087	-0.062	-0.025	0.134	-0.085	-0.049
		SD	2.955	1.624	1.732	2.955	1.625	1.732
		SE	43.42	29.41	26.22	43.42	29.42	26.22
		SE*	0.105	0.057	0.057	0.106	0.058	0.057
	LCF	Mean	0.054	-0.044	-0.010	0.074	-0.054	-0.020
	(Flexible)	SD	3.244	1.457	2.050	3.220	1.449	2.039
		SE	47.58	22.04	36.71	47.65	22.08	36.75
_		SE*	0.100	0.056	0.058	0.127	0.073	0.074

Table 1.7: (cont'd)

	w_p			10^{th}			25^{th}			50^{th}			75^{th}	
	8		1	2	3	1	2	3	1	2	3	1	2	3
w	True	Mean	-0.164	0.082	0.082	-0.157	0.078	0.078	-0.127	0.064	0.064	-0.089	0.044	0.044
	Two Step	Mean	-0.144	0.072	0.072	-0.147	0.074	0.073	-0.111	0.056	0.055	-0.065	0.033	0.032
		SD	0.089	0.045	0.046	0.101	0.051	0.052	0.074	0.039	0.040	0.045	0.026	0.028
		SE	0.088	0.045	0.045	0.102	0.052	0.052	0.072	0.039	0.039	0.039	0.027	0.027
w^2	True	Mean	-0.220	0.110	0.110	-0.188	0.094	0.094	-0.127	0.064	0.064	-0.075	0.037	0.037
	Two Step	Mean	-0.182	0.091	0.091	-0.174	0.087	0.087	-0.112	0.056	0.056	-0.055	0.027	0.028
		SD	0.092	0.046	0.047	0.108	0.055	0.055	0.075	0.039	0.041	0.048	0.027	0.029
		SE	0.090	0.046	0.046	0.108	0.055	0.055	0.073	0.039	0.039	0.043	0.028	0.028
	Two Step	Mean	-0.182	0.091	0.091	-0.174	0.087	0.087	-0.110	0.055	0.055	-0.050	0.025	0.025
	(Flexible)	SD	0.104	0.055	0.056	0.108	0.056	0.057	0.078	0.041	0.042	0.062	0.034	0.035
		SE	152.5	13.09	217.2	43.23	0.626	6.530	5.656	0.089	0.457	0.759	0.066	0.243
	LCF	Mean	-0.189	0.094	0.095	-0.143	0.071	0.072	-0.091	0.045	0.046	-0.040	0.020	0.020
		SD	0.097	0.046	0.055	0.095	0.045	0.054	0.094	0.044	0.054	0.094	0.044	0.054
		SE	0.085	0.042	0.046	0.083	0.041	0.045	0.082	0.041	0.045	0.082	0.041	0.045
	LCF	Mean	-0.152	0.076	0.076	-0.124	0.062	0.062	-0.092	0.046	0.046	-0.061	0.030	0.031
	(Flexible)	SD	0.159	0.078	0.083	0.123	0.059	0.067	0.098	0.046	0.056	0.103	0.050	0.059
		SE	0.120	0.060	0.063	0.099	0.049	0.052	0.082	0.041	0.045	0.078	0.041	0.045

Table 1.8: Percentile APEs under Condition 2, $\pi_2 = 0.2$, and Normal distribution

^{1.} APEs at (\bar{z}_1, w_p) . ^{2.} $\pi_1 = 0, \rho = 1$. ^{3.} The grey colored cells indicate that at least one of the APE estimates for three choices has the opposite direction to its true APE.

	w_p			90 th	
	8		1	2	3
w	True	Mean	-0.057	0.028	0.028
	Two Step	Mean	-0.037	0.019	0.018
		SD	0.033	0.022	0.023
		SE	0.033	0.024	0.025
w^2	True	Mean	-0.043	0.022	0.022
	Two Step	Mean	-0.024	0.012	0.012
		SD	0.038	0.024	0.025
		SE	0.037	0.026	0.025
	Two Step	Mean	-0.013	0.007	0.006
	(Flexible)	SD	0.063	0.037	0.037
		SE	8.476	0.256	0.690
	LCF	Mean	0.007	-0.004	-0.003
		SD	0.095	0.045	0.054
		SE	0.084	0.042	0.046
	LCF	Mean	-0.032	0.016	0.016
	(Flexible)	SD	0.130	0.065	0.072
		SE	0.088	0.048	0.052

Table 1.8: (cont'd)

	w_p			10^{th}			25^{th}			50^{th}			75^{th}	
	8		1	2	3	1	2	3	1	2	3	1	2	3
w	True	Mean	-0.168	0.084	0.084	-0.159	0.080	0.080	-0.129	0.064	0.064	-0.088	0.044	0.044
	Two Step	Mean	-0.164	0.082	0.082	-0.158	0.079	0.079	-0.125	0.062	0.062	-0.083	0.041	0.041
		SD	0.037	0.019	0.019	0.040	0.020	0.020	0.026	0.014	0.014	0.010	0.007	0.008
		SE	0.037	0.019	0.019	0.039	0.020	0.020	0.025	0.013	0.013	0.009	0.007	0.007
w^2	True	Mean	-0.224	0.112	0.112	-0.192	0.096	0.096	-0.129	0.064	0.064	-0.075	0.037	0.037
	Two Step	Mean	-0.213	0.106	0.107	-0.188	0.094	0.094	-0.126	0.063	0.063	-0.072	0.036	0.036
		SD	0.041	0.020	0.020	0.044	0.022	0.022	0.027	0.014	0.015	0.014	0.009	0.009
		SE	0.039	0.020	0.020	0.043	0.022	0.022	0.026	0.014	0.014	0.013	0.009	0.009
	Two Step	Mean	-0.214	0.107	0.107	-0.188	0.094	0.094	-0.125	0.062	0.062	-0.069	0.035	0.035
	(Flexible)	SD	0.044	0.022	0.022	0.044	0.023	0.023	0.028	0.015	0.015	0.018	0.011	0.011
		SE	0.042	0.022	0.022	0.042	0.022	0.022	0.026	0.014	0.014	0.017	0.010	0.010
	LCF	Mean	-0.202	0.101	0.101	-0.157	0.078	0.078	-0.106	0.053	0.053	-0.056	0.028	0.028
		SD	0.028	0.014	0.015	0.024	0.012	0.012	0.021	0.011	0.011	0.021	0.011	0.012
		SE	0.026	0.014	0.014	0.022	0.012	0.012	0.020	0.011	0.011	0.021	0.011	0.011
	LCF	Mean	-0.168	0.084	0.084	-0.139	0.069	0.070	-0.106	0.053	0.053	-0.074	0.037	0.037
	(Flexible)	SD	0.036	0.018	0.018	0.027	0.014	0.014	0.020	0.010	0.011	0.020	0.011	0.011
		SE	0.032	0.017	0.017	0.025	0.013	0.013	0.019	0.010	0.010	0.018	0.010	0.010

Table 1.9: Percentile APEs under Condition 2, $\pi_2 = 0.5$, and Normal distribution

^{1.} APEs at (\overline{z}_1, w_p) .

^{2.}
$$\pi_1 = 0, \rho = 1.$$

	w_p			90 th	
	8		1	2	3
w	True	Mean	-0.055	0.028	0.028
	Two Step	Mean	-0.052	0.026	0.026
		SD	0.007	0.007	0.007
		SE	0.006	0.007	0.007
w^2	True	Mean	-0.043	0.021	0.021
	Two Step	Mean	-0.041	0.021	0.020
		SD	0.010	0.008	0.008
		SE	0.010	0.008	0.008
	Two Step	Mean	-0.037	0.019	0.018
	(Flexible)	SD	0.018	0.012	0.012
		SE	0.016	0.012	0.011
	LCF	Mean	-0.010	0.005	0.005
		SD	0.026	0.013	0.014
		SE	0.024	0.013	0.013
	LCF	Mean	-0.044	0.022	0.022
	(Flexible)	SD	0.026	0.014	0.015
		SE	0.021	0.013	0.013

Table 1.9: (cont'd)

	w_p			10^{th}			25^{th}			50^{th}	
	8		1	2	3	1	2	3	1	2	3
w	True	Mean	-0.097	0.049	0.049	-0.070	0.035	0.035	-0.043	0.022	0.022
	Two Step	Mean	-0.133	0.069	0.064	-0.093	0.049	0.044	-0.030	0.017	0.013
		SD	0.160	0.083	0.089	0.150	0.078	0.089	0.109	0.065	0.071
		SE	0.237	0.886	0.877	0.268	0.685	0.612	0.306	0.573	0.674
w^2	True	Mean	-0.145	0.072	0.072	-0.087	0.044	0.044	-0.043	0.022	0.022
	Two Step	Mean	-0.146	0.073	0.072	-0.118	0.059	0.059	-0.044	0.023	0.021
		SD	0.164	0.087	0.087	0.160	0.089	0.086	0.109	0.069	0.065
		SE	0.271	3.391	3.362	0.458	3.292	3.084	0.434	1.627	1.667
	Two Step	Mean	-0.136	0.070	0.067	-0.100	0.051	0.049	-0.041	0.021	0.020
	(Flexible)	SD	0.192	0.103	0.107	0.155	0.090	0.087	0.140	0.081	0.078
		SE	4E+25	5E+21	1E+20	1E+29	2E+22	4E+21	1E+27	1E+31	6E+29
		SE*	0.209	0.117	0.116	0.136	0.078	0.079	0.126	0.072	0.074
	LCF	Mean	-0.058	0.028	0.030	0.002	-0.002	0.000	0.069	-0.035	-0.034
		SD	2.305	1.434	1.144	2.305	1.434	1.144	2.305	1.434	1.144
		SE	41.22	27.90	20.46	41.22	27.90	20.46	41.22	27.90	20.46
		SE*	0.090	0.051	0.051	0.089	0.050	0.051	0.089	0.050	0.051
	LCF	Mean	0.019	-0.008	-0.011	0.041	-0.019	-0.021	0.063	-0.031	-0.032
	(Flexible)	SD	2.401	1.420	1.238	2.302	1.391	1.175	2.241	1.383	1.124
		SE	37.35	26.65	19.09	37.13	26.55	18.98	36.96	26.47	18.91
		SE*	0.221	0.113	0.116	0.152	0.079	0.082	0.090	0.051	0.050

Table 1.10: Percentile APEs under Condition 2, $\pi_2 = 0.1$, and χ^2_3 distribution

 APEs at (z
₁, w_p).
 π₁ = 0, ρ = 1.
 SE* is the median of the standard errors.
 The grey colored cells indicate that at least one of the APE estimates for three choices has the opposite direction to its true APE.

	w_p			75^{th}			90 th	
	8		1	2	3	1	2	3
w	True	Mean	-0.023	0.012	0.012	-0.012	0.006	0.006
	Two Step	Mean	0.006	-0.001	-0.004	0.014	-0.007	-0.007
		SD	0.081	0.053	0.057	0.057	0.041	0.044
		SE	0.503	0.436	0.421	0.164	0.380	0.339
w^2	True	Mean	-0.020	0.010	0.010	-0.010	0.005	0.005
	Two Step	Mean	-0.003	0.002	0.001	0.007	-0.003	-0.004
		SD	0.078	0.056	0.052	0.051	0.042	0.041
		SE	0.525	1.332	0.998	0.561	0.537	0.947
	Two Step	Mean	0.020	-0.010	-0.010	0.048	-0.025	-0.023
	(Flexible)	SD	0.136	0.081	0.076	0.122	0.075	0.073
		SE	4E+24	5E+30	2E+29	2E+28	2E+30	4E+28
		SE*	0.150	0.108	0.113	0.189	0.162	0.169
	LCF	Mean	0.136	-0.069	-0.067	0.196	-0.099	-0.097
		SD	2.304	1.434	1.144	2.304	1.434	1.144
		SE	41.22	27.90	20.46	41.22	27.90	20.46
		SE*	0.089	0.050	0.050	0.089	0.051	0.051
	LCF	Mean	0.087	-0.044	-0.043	0.109	-0.056	-0.053
	(Flexible)	SD	2.248	1.400	1.111	2.308	1.435	1.131
		SE	37.01	26.50	18.95	37.22	26.59	19.05
		SE*	0.076	0.049	0.049	0.125	0.075	0.073

Table 1.10: (cont'd)

	w_p			10^{th}			25^{th}			50^{th}			75^{th}	
	8		1	2	3	1	2	3	1	2	3	1	2	3
w	True	Mean	-0.097	0.049	0.049	-0.070	0.035	0.035	-0.043	0.021	0.021	-0.023	0.011	0.011
	Two Step	Mean	-0.111	0.056	0.056	-0.066	0.033	0.033	-0.027	0.014	0.014	-0.009	0.005	0.004
		SD	0.113	0.057	0.057	0.069	0.037	0.037	0.038	0.024	0.023	0.032	0.022	0.020
		SE	0.107	0.055	0.055	0.066	0.036	0.036	0.031	0.022	0.022	0.024	0.021	0.021
w^2	True	Mean	-0.145	0.072	0.072	-0.087	0.044	0.044	-0.043	0.022	0.022	-0.020	0.010	0.010
	Two Step	Mean	-0.134	0.067	0.067	-0.091	0.045	0.045	-0.041	0.020	0.021	-0.016	0.008	0.008
		SD	0.124	0.063	0.063	0.080	0.042	0.042	0.036	0.022	0.023	0.027	0.019	0.020
		SE	0.120	0.061	0.062	0.077	0.041	0.041	0.029	0.021	0.022	0.023	0.021	0.020
	Two Step	Mean	-0.130	0.065	0.065	-0.080	0.040	0.040	-0.037	0.018	0.019	0.001	-0.001	0.000
	(Flexible)	SD	0.127	0.067	0.067	0.069	0.038	0.039	0.060	0.033	0.033	0.071	0.038	0.039
		SE	0.817	15.77	25.78	0.312	15.56	34.77	0.420	13.94	49.29	19.58	12.11	65.85
	LCF	Mean	-0.183	0.091	0.092	-0.123	0.062	0.062	-0.057	0.028	0.029	0.009	-0.005	-0.004
		SD	0.069	0.036	0.039	0.067	0.035	0.038	0.066	0.034	0.037	0.065	0.033	0.037
		SE	0.062	0.033	0.035	0.061	0.032	0.034	0.060	0.032	0.034	0.060	0.032	0.034
	LCF	Mean	-0.130	0.065	0.065	-0.096	0.048	0.048	-0.058	0.029	0.029	-0.019	0.009	0.010
	(Flexible)	SD	0.185	0.095	0.094	0.122	0.062	0.062	0.064	0.033	0.036	0.075	0.039	0.042
		SE	0.132	0.068	0.069	0.095	0.049	0.050	0.059	0.032	0.033	0.046	0.028	0.030

Table 1.11: Percentile APEs under Condition 2, $\pi_2 = 0.2$, and χ^2_3 distribution

 APEs at (z
₁, w_p).
 π₁ = 0, ρ = 1.
 SE* is the median of the standard errors.
 The grey colored cells indicate that at least one of the APE estimates for three choices has the opposite direction to its true APE.

	w_p		90^{th}					
	8		1	2	3			
w	True	Mean	-0.012	0.006	0.006			
	Two Step	Mean	-0.002	0.001	0.001			
		SD	0.029	0.021	0.020			
		SE	0.024	0.021	0.021			
w^2	True	Mean	-0.010	0.005	0.005			
	Two Step	Mean	-0.006	0.003	0.003			
		SD	0.023	0.018	0.018			
		SE	0.022	0.020	0.021			
	Two Step	Mean	0.025	-0.013	-0.012			
	(Flexible)	SD	0.085	0.047	0.046			
		SE	4143	490.5	400.8			
	LCF	Mean	0.069	-0.035	-0.034			
		SD	0.065	0.033	0.037			
		SE	0.061	0.033	0.035			
	LCF	Mean	0.016	-0.008	-0.008			
	(Flexible)	SD	0.129	0.068	0.069			
		SE	0.070	0.041	0.042			

Table 1.11: (cont'd)

w_p			10^{th}			25^{th}			50^{th}			75^{th}		
	8		1	2	3	1	2	3	1	2	3	1	2	3
w	True	Mean	-0.098	0.049	0.049	-0.070	0.035	0.035	-0.042	0.021	0.021	-0.022	0.011	0.011
	Two Step	Mean	-0.096	0.048	0.048	-0.063	0.031	0.031	-0.036	0.018	0.018	-0.019	0.010	0.010
		SD	0.045	0.023	0.023	0.023	0.012	0.012	0.008	0.006	0.006	0.004	0.005	0.005
		SE	0.042	0.021	0.021	0.021	0.012	0.012	0.007	0.006	0.006	0.003	0.005	0.005
w^2	True	Mean	-0.147	0.073	0.073	-0.087	0.044	0.044	-0.042	0.021	0.021	-0.019	0.010	0.010
	Two Step	Mean	-0.123	0.062	0.062	-0.085	0.043	0.043	-0.047	0.023	0.023	-0.022	0.011	0.011
		SD	0.056	0.028	0.028	0.028	0.014	0.015	0.010	0.007	0.007	0.006	0.006	0.006
		SE	0.052	0.027	0.027	0.025	0.014	0.014	0.009	0.007	0.007	0.005	0.006	0.006
	Two Step	Mean	-0.124	0.062	0.062	-0.079	0.039	0.040	-0.042	0.021	0.021	-0.019	0.010	0.010
	(Flexible)	SD	0.054	0.028	0.028	0.025	0.014	0.014	0.014	0.008	0.009	0.009	0.007	0.007
		SE	0.054	0.028	0.028	0.029	0.016	0.015	0.025	0.013	0.013	0.045	0.019	0.019
	LCF	Mean	-0.184	0.092	0.092	-0.127	0.063	0.064	-0.064	0.032	0.032	-0.001	0.000	0.000
		SD	0.026	0.014	0.014	0.021	0.011	0.011	0.018	0.010	0.010	0.017	0.010	0.010
		SE	0.023	0.012	0.012	0.019	0.010	0.010	0.016	0.009	0.009	0.017	0.009	0.009
	LCF	Mean	-0.138	0.069	0.069	-0.103	0.051	0.052	-0.064	0.032	0.032	-0.025	0.012	0.012
	(Flexible)	SD	0.046	0.024	0.024	0.032	0.016	0.016	0.017	0.010	0.010	0.013	0.009	0.009
		SE	0.040	0.021	0.021	0.028	0.015	0.015	0.016	0.009	0.009	0.010	0.007	0.007

Table 1.12: Percentile APEs under Condition 2, $\pi_2 = 0.5$, and χ^2_3 distribution

^{1.} APEs at (\overline{z}_1, w_p) .

^{2.} $\pi_1 = 0, \rho = 1.$ ^{3.} The grey colored cells indicate that at least one of the APE estimates for three choices has the opposite direction to its true APE.

	w_p		90 th					
	8		1	2	3			
w	True	Mean	-0.011	0.006	0.006			
	Two Step	Mean	-0.011	0.006	0.006			
		SD	0.003	0.005	0.005			
		SE	0.003	0.005	0.005			
w^2	True	Mean	-0.009	0.005	0.005			
	Two Step	Mean	-0.010	0.005	0.005			
		SD	0.004	0.006	0.006			
		SE	0.004	0.006	0.006			
	Two Step	Mean	-0.008	0.004	0.004			
	(Flexible)	SD	0.009	0.009	0.009			
		SE	0.123	0.040	0.040			
	LCF	Mean	0.056	-0.028	-0.028			
		SD	0.020	0.011	0.011			
		SE	0.019	0.011	0.011			
	LCF	Mean	0.011	-0.005	-0.005			
	(Flexible)	SD	0.023	0.014	0.014			
		SE	0.017	0.011	0.011			

Table 1.12: (cont'd)

			10^{th}		25 th			50 th			75 th			
	π_2	8	1	2	3	1	2	3	1	2	3	1	2	3
			·			Norr	nal							
w^2	0.1	Two step	0.020	0.006	0.006	0.032	0.009	0.009	0.022	0.007	0.007	0.010	0.004	0.004
		Two step (Flexible)	0.027	0.008	0.009	0.033	0.010	0.010	0.024	0.008	0.008	0.015	0.005	0.005
		LCF	8.747	2.646	3.002	8.753	2.648	3.003	8.754	2.648	3.003	8.757	2.649	3.004
		LCF (Flexible)	12.06	2.457	4.621	11.50	2.328	4.470	10.91	2.202	4.308	10.54	2.130	4.206
	0.2	Two step	0.010	0.002	0.003	0.012	0.003	0.003	0.006	0.002	0.002	0.003	0.001	0.001
		Two step (Flexible)	0.012	0.003	0.003	0.012	0.003	0.003	0.006	0.002	0.002	0.004	0.001	0.001
		LCF	0.010	0.002	0.003	0.011	0.003	0.003	0.010	0.002	0.003	0.010	0.002	0.003
		LCF (Flexible)	0.030	0.007	0.008	0.019	0.005	0.005	0.011	0.002	0.003	0.011	0.003	0.004
	0.5	Two step	0.002	0.000	0.000	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
		Two step (Flexible)	0.002	0.001	0.001	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
		LCF	0.001	0.000	0.000	0.002	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
		LCF (Flexible)	0.004	0.001	0.001	0.004	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
						χ^2_3								
w^2	0.1	Two step	0.027	0.008	0.008	0.027	0.008	0.008	0.012	0.005	0.004	0.006	0.003	0.003
		Two step (Flexible)	0.037	0.011	0.012	0.024	0.008	0.008	0.020	0.006	0.006	0.020	0.007	0.006
		LCF	5.321	2.060	1.310	5.320	2.059	1.310	5.324	2.060	1.311	5.335	2.063	1.314
		LCF (Flexible)	5.793	2.023	1.539	5.316	1.938	1.385	5.035	1.917	1.266	5.064	1.964	1.238
	0.2	Two step	0.015	0.004	0.004	0.006	0.002	0.002	0.001	0.000	0.001	0.001	0.000	0.000
		Two step (Flexible)	0.016	0.004	0.005	0.005	0.001	0.002	0.004	0.001	0.001	0.005	0.002	0.002
		LCF	0.006	0.002	0.002	0.006	0.002	0.002	0.005	0.001	0.001	0.005	0.001	0.002
		LCF (Flexible)	0.034	0.009	0.009	0.015	0.004	0.004	0.004	0.001	0.001	0.006	0.002	0.002
	0.5	Two step	0.004	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Two step (Flexible)	0.003	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		LCF	0.002	0.001	0.001	0.002	0.001	0.001	0.001	0.000	0.000	0.001	0.000	0.000
		LCF (Flexible)	0.002	0.001	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000

Table 1.13: Mean Squared Errors of Percentile APE estimates under Condition 2

^{1.} The Mean Squared Errors are calculated from Table 1.7 through 1.12. ^{2.} The grey colored cells indicate that at least one of the APE estimates for three choices has the opposite direction to its true APE.

		w _p	90 th				
	π_2	g	1	2	3		
		Normal					
w^2	0.1	Two step	0.005	0.002	0.002		
		Two step (Flexible)	0.013	0.005	0.005		
		LCF	8.765	2.651	3.006		
		LCF (Flexible)	10.38	2.104	4.158		
	0.2	Two step	0.002	0.001	0.001		
		Two step (Flexible)	0.005	0.002	0.002		
		LCF	0.012	0.003	0.004		
		LCF (Flexible)	0.017	0.004	0.005		
	0.5	Two step	0.000	0.000	0.000		
		Two step (Flexible)	0.000	0.000	0.000		
		LCF	0.002	0.000	0.000		
		LCF (Flexible)	0.001	0.000	0.000		
		χ^2_3					
w^2	0.1	Two step	0.003	0.002	0.002		
		Two step (Flexible)	0.006	0.006			
		LCF	5.353	2.067	1.318		
		LCF (Flexible)	5.340	2.063	1.282		
	0.2	Two step	0.001	0.000	0.000		
		Two step (Flexible)	0.008	0.003	0.002		
		LCF	0.010	0.003	0.003		
		LCF (Flexible)	0.017	0.005	0.005		
	0.5	Two step	0.000	0.000	0.000		
		Two step (Flexible)	0.000	0.000	0.000		
		LCF	0.005	0.001	0.001		
		LCF (Flexible)	0.001	0.000	0.000		

Table 1.13: (cont'd)

Condition 3: $\rho < 1$

We also generate the data by allowing the amount of endogeneity to be smaller than Conditions 1 and 2: $w_i = \pi_2 z_{i2} + v_i$ where $\pi_2 \in \{0.1, 0.2, 0.5, 1\}$ and $r_i = \rho v_i + e_i$ where $\rho \in \{0.1, 0.5\}$ Although a fewer sign distortions are observed than $\rho = 1$, the previous results remain almost the same, in general.

In summary, the simulations under Condition 1 through Condition 3 demonstrate that the two step estimation method with a strong instrument provides a good approximation even though its conditional mean is misspecified, and that it outperforms the LCF approach regardless of the instrument's quality and the amount of endogeneity. Furthermore, in the simulations, adding \hat{v}_i^2 and \hat{v}_i^3 in the estimation does not improve the two methods' approximations.¹⁷

1.4 APPLICATION: MICHIGAN EDUCATIONAL ASSESS-MENT PROGRAM MATH TEST

We apply the two step estimation method to Michigan Educational Assessment Program (MEAP) data of the school year 2004/2005 in order to estimate the effects of spending on the fourth grade math test outcome. The fourth grade MEAP math test is a statewide assessment test given by the State Board of Education in Michigan. It measures public school student achievement in relation to Michigan curriculum standards, which groups of educators, teachers and school administrators set. A student's outcome is rated at one of the four levels as described in Table 1.14 and public school districts' percentage shares of students for the four levels are available from the Michigan Department of Education (MDE) website.¹⁸

Table 1.14: Four levels of MEAP

Outcome	Description
Level 1	Exceeded Michigan Standards
Level 2	Met Michigan Standards
Level 3	demonstrated Basic knowledge and skills of Michigan Standards
Level 4	Apprentice level, showing little success in meeting Michigan standards

^{1.} The description is from Michgian Department of Education (2005).

Papke (2005, 2008), Papke and Wooldridge (2008), and Roy (2011) examine the relationship between spending and pass rates of this test by using panel data. During their data periods, the test had three performance levels (Satisfactory, Moderate, Low) and

¹⁷ That the simulations model r_i as a linear function of v_i could be one of the reasons.

¹⁸ http://www.michigan.gov/mde/

their pass rates indicate the percentage of students in the satisfactory level. They find that there exist significant positive causal effects of spending on the pass rates although their magnitudes are different. The application in this chapter can help understand how spending shifts students in the four different levels instead of between pass and fail.

We use district-level data including 512 districts.¹⁹ We turn these districts' percentage shares into proportions to obtain fractional dependent variables.²⁰ Table 1.15 illustrates the dependent variables' summary statistics. While the lower corner 0 appears for all of the four levels, the upper corner 1 appears only for level 1 like the dependent variables generated in the simulations. We choose the first level as the reference as in the simulations.

Variable	Mean	SD	Min	Max	Description
y_1	0.283	0.138	0	1.000	fraction of Level 1
<i>y</i> ₂	0.463	0.082	0	0.742	fraction of Level 2
<i>y</i> ₃	0.221	0.098	0	0.643	fraction of Level 3
y_4	0.033	0.035	0	0.278	fraction of Level 4
Total	1.000				

Table 1.15: Summary statistics of the dependent variables

The key explanatory variable, spending, is constructed as per pupil general fund expenditure in a logarithmic form, log(*per pupil GF expenditure*). Although we additionally control for the fraction of applications for the free and reduced-price lunch program as a measure of the poverty rate and log(*enrollment*) as a measure of school district size, we still suspect that spending is endogenous. There can be unobserved district effects such as parental involvement, which are correlated with spending and are able to affect student outcome as well. Thus we need an instrument to achieve more accurate estimation for the effects of spending on the student test outcome.

To find an instrument, we exploit Michigan's school funding system reform in 1994, which is called "Proposal A." The reform changed Michigan's school funding sources

¹⁹ In the year of 2004/2005, Michigan has 552 public school districts.

²⁰ The original percentage shares for some districts do not sum to 100 because of rounding errors. Thus We calculate the proportions not based on 100, but based on total percentage shares of the four levels.
and started to provide school districts with foundation allowances. It resulted in not only significantly raising districts spending, but also reducing the spending inequalities across districts by letting the low spending districts' foundation allowances increase faster than the others. The initial foundation allowance for a district, which was awarded in 1994/1995, was determined based on its per pupil spending in 1993/1994 and the dollar increases for the following years has been decided solely by the state legislature. Therefore per pupil foundation allowance in 2004/2005 meets the requirements of an instrument for spending if spending in 1993/1994 is controlled for.

Letting $\mathbf{z}_{i1} = (\log(enrollment_i) freelunch_i spending93_i)$, we apply Procedure 1.2.1 to the model

$$K_{g}(\mathbf{h}_{i};\boldsymbol{\theta}) = \frac{\exp\left(\mathbf{h}_{i}\boldsymbol{\theta}_{g}\right)}{\sum_{h}^{4}\exp\left(\mathbf{h}_{i}\boldsymbol{\theta}_{h}\right)}, \qquad \forall g = 1, \cdots, 4$$
(1.26)

where $\mathbf{h}_i = (\mathbf{z}_{i1} \text{ spending}_i v_i)$, freelunch_i represents the fraction of applications for the free and reduced-price lunch program, and spending93_i is spending in 1993/1994. The spending variable's reduced form is expressed as

$$spending_i = \mathbf{z}_{i1}\boldsymbol{\pi}_1 + \boldsymbol{\pi}_2 foundation_i + v_i \tag{1.27}$$

where $foundation_i = \log(per pupil foundation allowance_i)$. Table 1.16 includes summary statistics of the data.

Variable	Mean (SD)
enrollment	3127.186 (6972.843)
fraction of applications for free and reduced lunch	0.354 (0.176)
per pupil expenditure in 2004/2005	8086.428 (1090.264)
per pupil expenditure in 1994/1995	4901.967 (943.182)
per pupil foundation allowance	6979.738 (655.772)
# of districts	512

Table 1.16: Summary statistics of the data

Table 1.17 contains the first step estimation result. The foundation variable's t statistic shows that spending is correlated with the foundation allowance, netting out the other

explanatory variables. In addition, the F statistic suggests that it can be declared as a strong instrument according to Staiger and Stock (1997).

	coefficient	SE	t	p-values
log(enrollment)	0.004	0.005	0.860	0.391
freelunch	0.323	0.024	13.230	0.000
spending93	0.180	0.053	3.390	0.001
foundation	0.738	0.121	6.110	0.000
constant	0.787	0.706	1.110	0.266
R^2	0.625			
$F(H_0: foundation = 0)$	37.28			

Table 1.17: The first step estimation result

Table 1.18 reports the average APE estimates by the two methods. The two step estimation method provides statistically significant effects on level 1 and level 3. When the quadratic effect of spending is not allowed for, a 10% increase in per pupil expenditure - an increase in spending by 0.1 - causes the fractions of student at level 1 and level 3 to raise 6.8 percentage points and decrease 6.2 percentage points, respectively.²¹ With the quadratic effect, the magnitudes of the effects are slightly larger. The LCF approach yields similar result except that the fraction at level 4 is also statistically significant. The magnitudes of the effects are bigger than those by the two step estimation method. Without the quadratic effect, its estimated effect of a 10% increase in per pupil expenditure on the fraction at level 1 is 7.7 percentage points and those on the fractions at level 3 and 4 are -6.3 and -1.0 percentage points, respectively. With the quadratic effect, those effects become larger. Including the flexible forms of \hat{v}_i causes the estimated effects by both the methods to be smaller except that on level 4 by the LCF approach.

Table 1.19 presents that the two methods' percentile APE estimates for level 1 and level 3 by the two methods and those for level 4 by the LCF approach at 75th and 90th percentiles are statistically significant. All of the estimates for level 1 are positive and those for level

²¹ Considering that 119,687 students took the MEAP math test in 2004/2005, one percentage point increase(decrease) in the number of students at a certain level translates into about 1200 student increase(decrease).

	level	1	2	3	4
w	Two Step	0.681*	0.017	-0.619*	-0.079
	_	(0.158)	(0.158)	(0.048)	(0.048)
	LCF	0.768*	-0.039	-0.625*	-0.104*
		(0.213)	(0.131)	(0.165)	(0.049)
w^2	Two Step	0.779*	-0.044	-0.667*	-0.068
		(0.067)	(0.126)	(0.095)	(0.035)
	Two Step	0.691*	0.014	-0.602*	-0.103
	(Flexible)	(0.038)	(0.158)	(0.174)	(0.054)
	LCF	0.856*	-0.051	-0.693*	-0.112*
		(0.218)	(0.143)	(0.182)	(0.056)
	LCF	0.771*	0.019	-0.626*	-0.163*
	(Flexible)	(0.227)	(0.146)	(0.171)	(0.068)

Table 1.18: Average APE estimates of spending on the fourth grade MEAP math test

^{1.} $w = spending = \log(per pupil GF expenditure).$

^{2.} Standard errors are in parentheses.

^{3.} The two step estimation method's standard errors are calculated using 1000 bootstrap replications.
^{4.} * is significant at, or below, 5 percent.

3 and 4 are negative. As the percentile of spending increases, the magnitudes of the estimated effects for level 1 and 3 by the LCF approach decrease, illustrating that given the same percentage increase in per pupil expenditure, low spending districts are affected more than high spending districts. On the other hand, the two step estimation method has different patterns. Without the quadratic effect, the higher the percentile of spending is, the bigger its estimated effect for level 1 is and the smaller that for level 3 is. But its level 1 estimate with the quadratic effect increases from the 10^{th} to the 75^{th} percentile and then drops at the 90^{th} percentile. Allowing for \hat{v}_i^2 and \hat{v}_i^3 in the estimation has the same result as the average APE estimates.

The effects on level 2 in Table 1.18 and Table 1.19 are not significant. It might be because the fractions of students who used to be at lower levels move to level 2 and who used to be at level 2 to level 1 are similar.

In summary, the two methods show that spending affects mainly on level 1 and 3. Plus, considering the magnitudes of their estimates along with the directions, we can

w_p		10) th		25 th					
evel	1	2	3	4	1	2	3	4		
Two Step	0.601*	0.213	-0.729*	-0.084	0.633*	0.157	-0.706*	-0.083		
	(0.135)	(0.156)	(0.209)	(0.074)	(0.156)	(0.142)	(0.200)	(0.067)		
Two Step	0.720*	0.163	-0.832*	-0.051	0.755*	0.095	-0.792*	-0.059		
-	(0.146)	(0.194)	(0.246)	(0.078)	(0.171)	(0.169)	(0.227)	(0.071)		
Two Step	0.628*	0.202	-0.710*	-0.120	0.658*	0.145	-0.688*	-0.115		
(Flexible)	(0.163)	(0.178)	(0.215)	(0.102)	(0.184)	(0.159)	(0.202)	(0.088)		
LCF	0.922*	-0.060	-0.745*	-0.117	0.903*	-0.057	-0.730*	-0.115		
	(0.231)	(0.163)	(0.208)	(0.066)	(0.227)	(0.156)	(0.200)	(0.063)		
LCF	0.791*	0.040	-0.646*	-0.185*	0.785*	0.034	-0.641*	-0.179*		
(Flexible)	(0.242)	(0.169)	(0.192)	(0.081)	(0.237)	(0.162)	(0.185)	(0.077)		
w_p		50	$)^{th}$			75	5^{th}			
evel	1	2	3	4	1	2	3	4		
Two Step	0.679*	0.062	-0.660*	-0.080	0.728*	-0.070	-0.584*	-0.074		
	(0.188)	(0.125)	(0.179)	(0.056)	(0.222)	(0.132)	(0.140)	(0.041)		
Two Step	0.798*	-0.012	-0.718*	-0.067	0.819*	-0.141	-0.607*	-0.072		
-										
	(0.206)	(0.141)	(0.190)	(0.059)	(0.236)	(0.143)	(0.138)	(0.042)		
Two Step	(0.206) 0.697*	(0.141) 0.052	(0.190) -0.643*	(0.059) -0.106	(0.236) 0.731*	(0.143) -0.072	(0.138) -0.568*	(0.042) -0.091*		
Two Step (Flexible)	(0.206) 0.697* (0.213)	$ \begin{array}{r} (0.141) \\ 0.052 \\ (0.140) \end{array} $	(0.190) -0.643* (0.175)	(0.059) -0.106 (0.067)	(0.236) 0.731* (0.239)	(0.143) -0.072 (0.147)	(0.138) -0.568* (0.134)	(0.042) -0.091* (0.043)		
Two Step (Flexible) LCF	(0.206) 0.697* (0.213) 0.872*	(0.141) 0.052 (0.140) -0.053	(0.190) -0.643* (0.175) -0.706*	(0.059) -0.106 (0.067) -0.113	(0.236) 0.731* (0.239) 0.828*	(0.143) -0.072 (0.147) -0.047	(0.138) -0.568* (0.134) -0.672*	(0.042) -0.091* (0.043) -0.109*		
Two Step (Flexible) LCF	(0.206) 0.697* (0.213) 0.872* (0.221)	(0.141) 0.052 (0.140) -0.053 (0.147)	(0.190) -0.643* (0.175) -0.706* (0.188)	(0.059) -0.106 (0.067) -0.113 (0.058)	(0.236) 0.731* (0.239) 0.828* (0.215)	(0.143) -0.072 (0.147) -0.047 (0.137)	(0.138) -0.568* (0.134) -0.672* (0.174)	(0.042) -0.091* (0.043) -0.109* (0.053)		
Two Step (Flexible) LCF LCF	(0.206) 0.697* (0.213) 0.872* (0.221) 0.776*	(0.141) 0.052 (0.140) -0.053 (0.147) 0.024	(0.190) -0.643* (0.175) -0.706* (0.188) -0.631*	(0.059) -0.106 (0.067) -0.113 (0.058) -0.169*	(0.236) 0.731* (0.239) 0.828* (0.215) 0.762*	(0.143) -0.072 (0.147) -0.047 (0.137) 0.010	(0.138) -0.568* (0.134) -0.672* (0.174) -0.618*	(0.042) -0.091* (0.043) -0.109* (0.053) -0.154*		
	w_p evelEwo StepEwo StepEwo StepFlexible)LCFLCFFlexible)w_pevelEwo StepEwo StepEwo Step	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	w_p 10 evel 1 2 Two Step 0.601* 0.213 (0.135) (0.156) Two Step 0.720* 0.163 (0.146) (0.194) Two Step 0.628* 0.202 Flexible) (0.163) (0.178) LCF 0.922* -0.060 (0.231) (0.163) LCF 0.791* 0.040 Flexible) (0.242) (0.169) w_p 50 evel 1 2 Two Step 0.679* 0.062 (0.188) (0.125) 5 Two Step 0.798* -0.012	w_p 10^{th} evel123Ewo Step 0.601^* 0.213 -0.729^* (0.135) (0.156) (0.209) Ewo Step 0.720^* 0.163 -0.832^* (0.146) (0.194) (0.246) Ewo Step 0.628^* 0.202 -0.710^* Flexible) (0.163) (0.178) (0.215) LCF 0.922^* -0.060 -0.745^* (0.231) (0.163) (0.208) LCF 0.791^* 0.040 -0.646^* Flexible) (0.242) (0.169) (0.192) wp 50^{th} evel123Ewo Step 0.679^* 0.062 -0.660^* (0.188) (0.125) (0.179) Ewo Step 0.798^* -0.012 -0.718^*	w_p 10^{th} evel1234Evo Step 0.601^* 0.213 -0.729^* -0.084 (0.135) (0.156) (0.209) (0.074) Evo Step 0.720^* 0.163 -0.832^* -0.051 (0.146) (0.194) (0.246) (0.078) Evo Step 0.628^* 0.202 -0.710^* -0.120 Flexible) (0.163) (0.178) (0.215) (0.102) LCF 0.922^* -0.060 -0.745^* -0.117 (0.231) (0.163) (0.208) (0.066) LCF 0.791^* 0.040 -0.646^* -0.185^* Flexible) (0.242) (0.169) (0.192) (0.081) w_p 50^{th} -0.080 (0.188) (0.125) (0.179) Evol1234Evol1234Evol0.679^* 0.062 -0.660^* -0.080 (0.188) (0.125) (0.179) (0.056) Evol0.798^* -0.012 -0.718^* -0.067	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$\begin{array}{c c c c c c c c c c c c c c c c c c c $		

Table 1.19: Percentile APE estimates of spending on the fourth grade MEAP math test

	w_p	90^{th}								
	level	1	2	3	4					
w	Two Step	0.764*	-0.252	-0.452*	-0.060*					
		(0.242)	(0.179)	(0.072)	(0.019)					
w^2	Two Step	0.775*	-0.273	-0.436*	-0.065*					
		(0.243)	(0.181)	(0.070)	(0.022)					
	Two Step	0.733*	-0.230	-0.437*	-0.066*					
	(Flexible)	(0.245)	(0.185)	(0.074)	(0.018)					
	LCF	0.761*	-0.038	-0.619*	-0.104*					
		(0.214)	(0.132)	(0.162)	(0.049)					
	LCF	0.741*	-0.011	-0.598*	-0.132*					
	(Flexible)	(0.216)	(0.132)	(0.159)	(0.058)					

^{1.} $w = spending = \log(per pupil GF expenditure).$ ^{2.} Standard errors are in parentheses. ^{3.} APEs at $(\overline{z}_1, spending_p).$

^{4.} * is significant at, or below, 5 percent.

conclude that an increase in spending tends to shift students who are rated at a lower level to an upper level, in general. It is consistent with the results of the four research that mentioned above.

1.5 CONCLUSION

This chapter develops a two step estimation method for multiple fractional dependent variables especially when endogenous explanatory variables are continuous. The method directly specifies the estimating conditional mean rather than the structural one and deals with the endogeneity by combining a control function approach. Although the method is not applicable when the characteristics of choices are of interest, it can provide not only a consistent estimator of the parameters in the estimating conditional mean as long as the conditional mean specification is correct, but also a useful estimator for the quantities of the structural conditional mean without estimating the structural mean parameters.

Monte Carlo simulations demonstrate that the method even with a misspecified conditional mean works well as an approximation to true APEs if a strong instrument is available. The simulations also provide evidence that the two step estimation is preferable to an alternative linear method - a LCF approach; the linear method's approximation is more sensitive to the quality of an instrument.

The application to the fourth grade MEAP math test of the year 2004/2005 illustrates that the two step estimation method and the LCF approach provide similar results in general: the more a school district spends, the more students are in the exceeded Michigan Standards level and the fewer students are in the demonstrated basic knowledge and skills of Michigan Standards level. That is, an increase in spending tends to shift students from a lower level to a upper level.

APPENDIX

Appendix for Chapter 1

In this appendix, we show how to obtain the estimators' standard errors taking the extra variations from the first steps of Procedure 1.2.1 and Procedure 1.2.2 into account especially when the quadratic effect is not allowed for. The model including w^2 and using the flexible forms of \hat{v} require modifications. This appendix also includes Table 1.20 showing how the response variables for three choices are generated across the simulations.

The standard errors of $\hat{\theta}$ First, we obtain the standard errors of $\hat{\theta}$ from Procedure 1.2.1. The first step of Procedure 1.2.1 applies OLS to (1.15) and so, under the standard regularity conditions,

$$\sqrt{N}(\hat{\pi} - \pi) = \frac{1}{\sqrt{N}} \sum_{i}^{N} E(\mathbf{z}'\mathbf{z})^{-1} \mathbf{z}'_{i} v_{i} + o_{p}(1) = \frac{1}{\sqrt{N}} \sum_{i}^{N} \mathbf{q}_{i} + o_{p}(1)$$
(1.28)

where $\mathbf{q}_i \equiv E(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'_i v_i$. The conditional mean of the second step obtaining $\hat{\boldsymbol{\theta}}$ is

$$E(y_{ig}|\mathbf{z}_i, w_i) = K_g(\widehat{\mathbf{h}}_i; \boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\theta}_g \widehat{\mathbf{h}}_i)}{1 + \sum_{h=2}^G \exp(\boldsymbol{\theta}_h \widehat{\mathbf{h}}_i)}, \quad g = 2, \cdots, G, \quad (1.29)$$

$$=\frac{1}{1+\sum_{h=2}^{G}\exp(\boldsymbol{\theta}_{h}\widehat{\mathbf{h}}_{i})}, \quad g=1, \quad (1.30)$$

where $\hat{\mathbf{h}}_i = (\mathbf{x}_{i1} \ \hat{v}_i)$ is a 1 × *p* vector, $v_i = w_i - \mathbf{z}_i \pi$, and we redefine $\boldsymbol{\theta}$ for the appendix by dropping $\boldsymbol{\theta}_1$ from it:

$$\boldsymbol{\theta} = \left(\boldsymbol{\theta}_2' \quad \cdots \quad \boldsymbol{\theta}_g' \quad \cdots \quad \boldsymbol{\theta}_G' \right)_{p(G-1) \times 1}'$$

where $\theta_g = (\theta'_z \ \theta_w \ \theta_v)'_{p \times 1}$, for $g = 2, \dots G$. Then the first order condition is

$$\sum_{i}^{N} \mathbf{s}_{i}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\pi}}) = \mathbf{0}$$
(1.31)

where

$$\mathbf{s}_{i}(\boldsymbol{\theta},\boldsymbol{\pi}) \equiv (\nabla_{\boldsymbol{\theta}}\ell_{i})' = \begin{pmatrix} (\frac{\partial\ell_{i}}{\partial\boldsymbol{\theta}_{2}})'\\ \vdots\\ (\frac{\partial\ell_{i}}{\partial\boldsymbol{\theta}_{g}})'\\ \vdots\\ (\frac{\partial\ell_{i}}{\partial\boldsymbol{\theta}_{G}})' \end{pmatrix} = \begin{pmatrix} \mathbf{s}_{i2}\\ \vdots\\ \mathbf{s}_{ig}\\ \vdots\\ \mathbf{s}_{iG} \end{pmatrix}$$

and

$$\mathbf{s}_{ig} = \left(\frac{\partial \ell_i}{\partial \boldsymbol{\theta}_g}\right)' = \mathbf{h}'_i(y_{ig} - K_g(\mathbf{h}_i; \boldsymbol{\theta})), \, p \times 1 \text{ vector.}$$

A mean value expansion (MVE) around θ gives

$$\sum_{i}^{N} \mathbf{s}_{i}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\pi}}) = \sum_{i} \mathbf{s}_{i}(\boldsymbol{\theta}, \widehat{\boldsymbol{\pi}}) + \left[\nabla_{\boldsymbol{\theta}} \sum_{i} \mathbf{s}_{i}(\ddot{\boldsymbol{\theta}}) \right] (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$
(1.32)

where $\ddot{\theta}$ is on the line segment between $\hat{\theta}$ and θ . By multiplying through by $\frac{1}{\sqrt{N}}$ and using (1.31) and the weak law of large numbers (WLLN), we rearrange (1.32):

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = -\frac{1}{\sqrt{N}} \sum_{i}^{N} \left(\mathbf{A}^{-1} \mathbf{s}_{i}(\boldsymbol{\theta}, \widehat{\boldsymbol{\pi}}) \right) + o_{p}(1)$$
(1.33)

where $\mathbf{A} = -E(\mathbf{H}_i(\boldsymbol{\theta})) = -E(\nabla_{\boldsymbol{\theta}}^2 \ell_i(\boldsymbol{\theta}))_{p(G-1) \times p(G-1)}$

$$\mathbf{H}_{i} = \begin{pmatrix} -\mathbf{h}_{i}'\mathbf{h}_{i}K_{i2}(1-K_{i2}) & \mathbf{h}_{i}'\mathbf{h}_{i}K_{i2}K_{i3} & \cdots & \cdots & \mathbf{h}_{i}'\mathbf{h}_{i}K_{i2}K_{iG} \\ \mathbf{h}_{i}'\mathbf{h}_{i}K_{i3}K_{i2} & -\mathbf{h}_{i}'\mathbf{h}_{i}K_{i3}(1-K_{i3}) & \cdots & \cdots & \mathbf{h}_{i}'\mathbf{h}_{i}K_{i2}K_{iG} \\ \vdots & \vdots & & \vdots \\ \mathbf{h}_{i}'\mathbf{h}_{i}K_{iG}K_{i2} & \cdots & \cdots & \mathbf{h}_{i}'\mathbf{h}_{i}K_{iG}K_{iG} & -\mathbf{h}_{i}'\mathbf{h}_{i}K_{iG}(1-K_{iG}) \end{pmatrix},$$

and $K_{ig} = K_g(\mathbf{h}_i; \boldsymbol{\theta})$.

Since $\sum_i \mathbf{s}_i(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}})$ still depends on the sample, we can not apply the central limit the-

orem (CLT) yet. Using a MVE around π , multiplying through by $\frac{1}{\sqrt{N}}$ and using (1.28) gives

$$\frac{1}{\sqrt{N}}\sum_{i}^{N}\mathbf{s}_{i}(\boldsymbol{\theta},\widehat{\boldsymbol{\pi}}) = \frac{1}{\sqrt{N}}\sum_{i}^{N}\mathbf{s}_{i}(\boldsymbol{\theta},\boldsymbol{\pi}) + E[\nabla_{\boldsymbol{\pi}}\mathbf{s}_{i}(\boldsymbol{\theta},\boldsymbol{\pi})]\sqrt{N}(\widehat{\boldsymbol{\pi}}-\boldsymbol{\pi}) + o_{p}(1).$$
$$= \frac{1}{\sqrt{N}}\sum_{i}^{N}(\mathbf{s}_{i}+\mathbf{F}\mathbf{q}_{i}) + o_{p}(1)$$
(1.34)

where

$$\mathbf{F} = E[\nabla_{\pi}\mathbf{s}_{i}(\boldsymbol{\theta},\boldsymbol{\pi})]_{p(G-1)\times M} = E\left[\left(\frac{\partial\mathbf{s}_{i2}}{\partial\boldsymbol{\pi}}\right)' \cdots \left(\frac{\partial\mathbf{s}_{ig}}{\partial\boldsymbol{\pi}}\right)' \cdots \left(\frac{\partial\mathbf{s}_{iG}}{\partial\boldsymbol{\pi}}\right)'\right]'$$

and

$$\frac{\partial \mathbf{s}_{ig}}{\partial \pi} = \frac{\partial \mathbf{h}'_i}{\partial \pi} (y_{ig} - K_{ig}) + \mathbf{h}'_i \mathbf{z}_i K_{ig} \left(\theta_{vg} - \frac{\sum_{h=2} \theta_{vh} \exp(\mathbf{h}_i \theta_h)}{1 + \sum_{h=2} \exp(\mathbf{h}_i \theta_h)} \right)$$

By plugging (1.34) into (1.33),

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{A}^{-1} \left(-\frac{1}{\sqrt{N}} \sum_{i}^{N} \mathbf{d}_{i}(\boldsymbol{\theta}, \boldsymbol{\pi}) \right) + o_{p}(1)$$
(1.35)

where $\mathbf{d}_i \equiv \mathbf{s}_i + \mathbf{F} \mathbf{q}_i$. Therefore,

Avar
$$\left(\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right) = \mathbf{A}^{-1}\mathbf{D}\mathbf{A}^{-1}$$
 (1.36)

where $\mathbf{D} \equiv \text{Var}(\mathbf{d}_i) = \text{Var}(\mathbf{s}_i + \mathbf{F}\mathbf{q}_i)$, and so a valid estimator of $\text{Avar}(\widehat{\boldsymbol{\theta}})$ is

$$\frac{1}{N}\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{D}}\widehat{\mathbf{A}}^{-1} \tag{1.37}$$

where

$$\widehat{\mathbf{D}} \equiv \frac{1}{N} \sum_{i}^{N} \widehat{\mathbf{d}}_{i} \widehat{\mathbf{d}}_{i}' = \frac{1}{N} \sum_{i} (\widehat{\mathbf{s}}_{i} + \widehat{\mathbf{F}} \widehat{\mathbf{q}}_{i}) (\widehat{\mathbf{s}}_{i} + \widehat{\mathbf{F}} \widehat{\mathbf{q}}_{i})', \qquad (1.38)$$

$$\widehat{\mathbf{s}}_i = \mathbf{s}_i(\widehat{\mathbf{h}}_i; \widehat{\boldsymbol{\theta}}), \qquad (1.39)$$

$$\widehat{\mathbf{F}} = \mathbf{F}_i(\widehat{\mathbf{h}}_i; \widehat{\boldsymbol{\theta}}), \qquad (1.40)$$

$$\widehat{\mathbf{q}} = \left(\frac{1}{N} \sum_{i}^{N} \mathbf{z}_{i}' \mathbf{z}_{i}\right)^{-1} \mathbf{z}_{i}' \widehat{v}_{i}, \qquad (1.41)$$

and

$$\widehat{\mathbf{A}} = -\frac{1}{N} \sum_{i}^{N} \mathbf{H}_{i}(\widehat{\mathbf{h}}_{i}; \widehat{\boldsymbol{\theta}}).$$
(1.42)

The square roots of (1.37)'s diagonal elements are the standard errors.

The standard errors of the two step APE estimator Next, we derive the standard errors of (1.23). Let's define (1.22) as $\delta_g(\mathbf{x}_1; \boldsymbol{\eta})$:

$$\delta_{g}(\mathbf{x}_{1};\boldsymbol{\eta}) \equiv E_{v} \left[K_{g}(\mathbf{x}_{1},v;\boldsymbol{\theta}) \cdot \left(\theta_{wg} - \frac{\sum_{h}^{G} \theta_{wh} \exp\left(\mathbf{x}_{1} \boldsymbol{\theta}_{\mathbf{x}h} + \theta_{vh} v\right)}{\sum_{h}^{G} \exp\left(\mathbf{x}_{1} \boldsymbol{\theta}_{\mathbf{x}h} + \theta_{vh} v\right)} \right) \right]$$
(1.43)

where $\eta = (\theta' \ \pi')'_{(p(G-1)+M)\times 1}$. Since it depends on the value of \mathbf{x}_1 , we use two approaches to obtain a single number. One is the average APEs,

$$\delta_g^{AVG} = E_{\mathbf{x}_1} \left(\delta_g(\mathbf{x}_1; \boldsymbol{\eta}) \right) \tag{1.44}$$

and the other is the percentile APEs,

$$\delta_g^{PCT} = \delta_g(\mathbf{x}_1^\circ; \boldsymbol{\eta}) \tag{1.45}$$

where $\mathbf{x}_1^\circ = (\overline{\mathbf{z}}_1, w_p)$.

(1.44) is estimated as

$$\widehat{\delta}_{g}^{AVG} = \frac{1}{N} \sum_{j}^{N} \widehat{\delta}_{g}(\mathbf{x}_{j1}; \widehat{\boldsymbol{\eta}})$$
(1.46)

where
$$\widehat{\delta}_{g}(\mathbf{x}_{j1};\widehat{\boldsymbol{\eta}}) = \frac{1}{N}\sum_{i}^{N}K_{g}(\mathbf{x}_{j1},\widehat{v}_{i};\widehat{\boldsymbol{\theta}}) \cdot \left(\widehat{\theta}_{wg} - \frac{\sum_{h}^{G}\widehat{\theta}_{wh}\exp\left(\mathbf{x}_{j1}\widehat{\theta}_{\mathbf{x}h} + \widehat{\theta}_{vh}\widehat{v}_{i}\right)}{\sum_{h}^{G}\exp\left(\mathbf{x}_{j1}\widehat{\theta}_{\mathbf{x}h} + \widehat{\theta}_{vh}\widehat{v}_{i}\right)}\right)$$
. Based on

(1.28) and (1.35), we can write

$$\sqrt{N}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = \sqrt{N} \begin{pmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \widehat{\boldsymbol{\pi}} - \boldsymbol{\pi} \end{pmatrix} = \frac{1}{\sqrt{N}} \sum_{i}^{N} \begin{pmatrix} \mathbf{A}^{-1} \mathbf{d}_{i} \\ \mathbf{q}_{i} \end{pmatrix} + o_{p}(1) = \frac{1}{\sqrt{N}} \sum_{i}^{N} \mathbf{k}_{i} + o_{p}(1). \quad (1.47)$$

Applying a MVE to (1.46) around η , multiplying through \sqrt{N} , using (1.47) and WLLN gives

$$\sqrt{N}\widehat{\delta}_{g}^{AVG} = \frac{1}{\sqrt{N}}\sum_{j}^{N} \left[\delta_{g}(\mathbf{x}_{j1};\boldsymbol{\eta}) + E\left(\nabla_{\boldsymbol{\eta}}\delta_{g}(\mathbf{x}_{j1};\boldsymbol{\eta})\right)\mathbf{k}_{j}\right] + o_{p}(1).$$
(1.48)

We subtract $\sqrt{N}\delta_g^{AVG}$ from both sides of (1.48):

$$\sqrt{N}(\widehat{\delta}_{g}^{AVG} - \delta_{g}^{AVG}) = \frac{1}{\sqrt{N}} \sum_{j}^{N} \left[\delta_{g}(\mathbf{x}_{j1}; \boldsymbol{\eta}) - \delta_{g}^{AVG} + E\left(\nabla_{\boldsymbol{\eta}} \delta_{g}(\mathbf{x}_{j1}; \boldsymbol{\eta})\right) \mathbf{k}_{j} \right] + o_{p}(1) \quad (1.49)$$

where $E\left[\delta_g(\mathbf{x}_{j1}; \boldsymbol{\eta}) - \delta_g^{AVG} + E\left(\nabla_{\boldsymbol{\eta}}\delta_g(\mathbf{x}_{j1}; \boldsymbol{\eta})\right)\mathbf{k}\right] = 0$. Therefore, based on the CLT,

Avar
$$\left(\sqrt{N}(\widehat{\delta}_{g}^{AVG} - \delta_{g}^{AVG})\right) = \operatorname{Var}\left(\delta_{g}(\mathbf{x}_{j1}; \boldsymbol{\eta}) - \delta_{g}^{AVG} + \boldsymbol{\Delta}_{g}(\boldsymbol{\eta})\mathbf{k}\right)$$
 (1.50)

where $\Delta_g(\eta) = E\left[\nabla_{\eta}\delta_g(\mathbf{x}_{j1};\eta)\right]$, the $1 \times \{p(G-1) + M\}$ Jacobian of $\delta_g(\mathbf{x}_{j1};\eta)$, and a valid estimator of (1.50) is

$$\frac{1}{N}\sum_{j}^{N}\left(\widehat{\delta}_{g}(\mathbf{x}_{j1};\widehat{\boldsymbol{\eta}}) - \widehat{\delta}_{g}^{AVG} + \widehat{\boldsymbol{\Delta}}_{g}(\widehat{\boldsymbol{\eta}})\widehat{\mathbf{k}}_{j}\right) \left(\widehat{\delta}_{g}(\mathbf{x}_{j1};\widehat{\boldsymbol{\eta}}) - \widehat{\delta}_{g}^{AVG} + \widehat{\boldsymbol{\Delta}}_{g}(\widehat{\boldsymbol{\eta}})\widehat{\mathbf{k}}_{j}\right)'$$
(1.51)

where

$$\widehat{\Delta}_{g}(\widehat{\eta}) = \frac{1}{N} \sum_{j}^{N} \left[\nabla_{\eta} \widehat{\delta}_{g}(\mathbf{x}_{j1}; \widehat{\eta}) \right], \qquad (1.52)$$

and

$$\widehat{\mathbf{k}}_{j} = \begin{pmatrix} \widehat{\mathbf{A}}^{-1} \widehat{\mathbf{d}}_{j} \\ \widehat{\mathbf{q}}_{j} \end{pmatrix}.$$
(1.53)

The percentile APE (1.45) is estimated as

$$\widehat{\delta}_{g}^{PCT} = \widehat{\delta}_{g}(\mathbf{x}_{1}^{\circ}; \widehat{\boldsymbol{\eta}}) = \frac{1}{N} \sum_{i}^{N} j_{g}(\mathbf{x}_{1}^{\circ}, v_{i}; \widehat{\boldsymbol{\eta}})$$
(1.54)

where $j_g(\mathbf{x}_1^\circ, v_i; \widehat{\boldsymbol{\eta}}) = K_g(\mathbf{x}_1^\circ, \widehat{v}_i; \widehat{\boldsymbol{\theta}}) \cdot \left(\widehat{\theta}_{wg} - \frac{\sum_h^G \widehat{\theta}_{wh} \exp\left(\mathbf{x}_1^\circ \widehat{\boldsymbol{\theta}}_{xh} + \widehat{\theta}_{vh} \widehat{v}_i\right)}{\sum_h^G \exp\left(\mathbf{x}_1^\circ \widehat{\boldsymbol{\theta}}_{xh} + \widehat{\theta}_{vh} \widehat{v}_i\right)}\right)$. Through the same process as in the average APE estimate,

$$\sqrt{N}(\hat{\delta}_{g}^{PCT} - \delta_{g}^{PCT}) = \frac{1}{\sqrt{N}} \sum_{i}^{N} \left[j_{g}(\mathbf{x}_{1}^{\circ}, v_{i}; \boldsymbol{\eta}) - \delta_{g}^{PCT} + \mathbf{J}_{g}(\boldsymbol{\eta})\mathbf{k}_{i} \right] + o_{p}(1)$$
(1.55)

where $\mathbf{J}_{g}(\boldsymbol{\eta}) = E\left[\nabla_{\boldsymbol{\eta}} j_{g}(\mathbf{x}_{1}^{\circ}, v; \boldsymbol{\eta})\right]$, the $1 \times \{p(G-1) + M\}$ Jacobian of $j_{g}(\mathbf{x}_{1}^{\circ}, v; \boldsymbol{\eta})$. Since $E\left[j_{g}(\mathbf{x}_{1}^{\circ}, v; \boldsymbol{\eta}) - \delta_{g}^{PCT} + \mathbf{J}_{g}(\boldsymbol{\eta})\mathbf{k}\right] = 0$,

$$Avar\left(\sqrt{N}(\hat{\delta}_{g}^{PCT} - \delta_{g}^{PCT})\right) = Var\left(j_{g}(\mathbf{x}_{1}^{\circ}, v; \boldsymbol{\eta}) - \delta_{g}^{PCT} + \mathbf{J}_{g}(\boldsymbol{\eta})\mathbf{k}\right).$$
(1.56)

Thus, a valid estimator of (1.56) is

$$\frac{1}{N}\sum_{j}^{N}\left(j_{g}(\mathbf{x}_{1}^{\circ},v_{i};\widehat{\boldsymbol{\eta}})-\widehat{\delta}_{g}^{PCT}+\widehat{\mathbf{J}}_{g}(\widehat{\boldsymbol{\eta}})\widehat{\mathbf{k}}_{i}\right)\left(j_{g}(\mathbf{x}_{1}^{\circ},v_{i};\widehat{\boldsymbol{\eta}})-\widehat{\delta}_{g}^{PCT}+\widehat{\mathbf{J}}_{g}(\widehat{\boldsymbol{\eta}})\widehat{\mathbf{k}}_{i}\right)^{\prime}$$
(1.57)

where

$$\widehat{\mathbf{J}}_{g}(\widehat{\boldsymbol{\eta}}) = \frac{1}{N} \sum_{i}^{N} \left[\nabla_{\boldsymbol{\eta}} j_{g}(\mathbf{x}_{1}^{\circ}, v_{i}; \widehat{\boldsymbol{\eta}}) \right].$$
(1.58)

Hence, the the asymptotic standard errors of $\hat{\delta}_g^{AVG}$ and $\hat{\delta}_g^{PCT}$ are obtained as the square roots of (1.51) and (1.57), divided by \sqrt{N} , respectively.

The standard errors of $\hat{\gamma}$ For $g = 2, \dots, G$, the LCF approach models

$$y_{ig} = \mathbf{z}_{i1}\gamma_{zg} + \gamma_{wg}w_i + u_{ig} \tag{1.59}$$

and

$$u_{ig} = \rho_g v_i + e_{ig} \tag{1.60}$$

with the reduced form of the endogenous variable w, (1.15) and $y_{i1} = \sum_{g=2}^{G} y_{ig}$. Plugging (1.60) into (1.59),

$$y_{ig} = \mathbf{z}_{i1}\gamma_{zg} + \gamma_{wg}w_i + \rho_g v_i + e_{ig} = \mathbf{h}_i\gamma_g + e_{ig}, \quad g = 2, \cdots, G$$
(1.61)

where $\gamma_g = (\gamma'_{zg} \ \gamma_{wg} \ \rho_g)' = (\gamma'_{zg} \ \gamma_{wg} \ \gamma_{vg})'_{p\times 1}$. Considering that the second step of Procedure 1.2.2 replaces v with \hat{v} , the estimating equation of the LCF approach for $g = 2, \dots, G$ is

$$y_{ig} = \mathbf{z}_{i1}\gamma_{zg} + \gamma_{wg}w_i + \rho_g\widehat{v}_i + (\rho_g(v_i - \widehat{v}_i) + e_{ig})$$

$$= \widehat{\mathbf{h}}_i\gamma_g + (\rho_g(v_i - \widehat{v}_i) + e_{ig})$$

$$= \widehat{\mathbf{h}}_i\gamma_g + (\mathbf{h}_i - \widehat{\mathbf{h}}_i)\gamma_g + e_{ig}$$
(1.62)

and the LCF estimator is expressed as

$$\widehat{\boldsymbol{\gamma}}_{g} = \boldsymbol{\gamma}_{g} + \left(\sum_{i} \widehat{\mathbf{h}}_{i}^{\prime} \widehat{\mathbf{h}}_{i}\right)^{-1} \sum_{i} \widehat{\mathbf{h}}_{i}^{\prime} \left[(\mathbf{h}_{i} - \widehat{\mathbf{h}}_{i}) \boldsymbol{\gamma}_{g} + e_{ig} \right], \quad g = 2, \cdots, G \quad (1.63)$$

and

$$\widehat{\gamma}_1 = \mathbf{e}_1 - \sum_{g=2}^G \widehat{\gamma}_g. \tag{1.64}$$

Therefore,

$$\sqrt{N}(\widehat{\gamma}_{g} - \gamma_{g}) = \left(\frac{1}{N}\sum_{i}\widehat{\mathbf{h}}_{i}'\widehat{\mathbf{h}}_{i}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i}\widehat{\mathbf{h}}_{i}'\left[(\mathbf{h}_{i} - \widehat{\mathbf{h}}_{i})\gamma_{g} + e_{ig}\right], \quad g = 2, \cdots, G, \quad (1.65)$$

$$= \left(\frac{1}{N}\sum_{i}\widehat{\mathbf{h}}_{i}'\widehat{\mathbf{h}}_{i}\right)^{-1}\sum_{g=2}^{G}\left(-\frac{1}{\sqrt{N}}\sum_{i}\widehat{\mathbf{h}}_{i}'[(\mathbf{h}_{i} - \widehat{\mathbf{h}}_{i})\gamma_{g} + e_{ig}]\right), \quad g = 1. \quad (1.66)$$

A similar reasoning in Wooldridge (2010, Appendix 6A) rewrites (1.65) and (1.66) as

$$\sqrt{N}\left(\widehat{\boldsymbol{\gamma}}_{g}-\boldsymbol{\gamma}_{g}\right) = \mathbf{C}^{-1}\left\{\frac{1}{\sqrt{N}}\sum_{i}^{N}\left(\mathbf{h}_{i}^{\prime}\boldsymbol{e}_{ig}-\mathbf{R}_{g}\mathbf{B}^{-1}\mathbf{z}_{i}^{\prime}\boldsymbol{v}_{i}\right)\right\} + o_{p}(1), \quad g = 2, \cdots, G, \quad (1.67)$$

$$= \mathbf{C}^{-1} \left\{ -\frac{1}{\sqrt{N}} \sum_{i}^{N} \sum_{g=2}^{G} \left(\mathbf{h}_{i}^{\prime} e_{ig} - \mathbf{R}_{g} \mathbf{B}^{-1} \mathbf{z}_{i}^{\prime} v_{i} \right) \right\} + o_{p}(1), \quad g = 1.$$
(1.68)

where

$$\mathbf{C} = E(\mathbf{h}'\mathbf{h}),$$
$$\mathbf{R}_g = E\left[(\boldsymbol{\gamma}_g \otimes \mathbf{h})' \nabla_{\boldsymbol{\pi}} \mathbf{h}\right],$$

and

$$\mathbf{B} = E(\mathbf{z}'\mathbf{z}). \tag{1.69}$$

Since $E\left[\mathbf{h}' e_g - \mathbf{R}_g \mathbf{B}^{-1} \mathbf{z}' v\right] = 0 \ \forall g$,

$$Avar\left(\sqrt{N}(\widehat{\gamma}_g - \gamma_g)\right) = \mathbf{C}^{-1}\mathbf{M}_g\mathbf{C}^{-1}$$
(1.70)

where

$$\mathbf{M}_{g} = \operatorname{Var}\left(\mathbf{h}' e_{g} - \mathbf{R}_{g} \mathbf{B}^{-1} \mathbf{z}' v\right), \quad g = 2, \cdots, G.$$
(1.71)

and

$$\mathbf{M}_{1} = \operatorname{Var}\left(\sum_{g=2}^{G} \left(\mathbf{h}' e_{g} - \mathbf{R}_{g} \mathbf{B}^{-1} \mathbf{z}' v\right)\right)$$
(1.72)

Therefore, $\operatorname{Avar}(\widehat{\boldsymbol{\gamma}}_g)$ is estimated as

$$\frac{1}{N} \left(\widehat{\mathbf{C}}^{-1} \widehat{\mathbf{M}}_{g} \widehat{\mathbf{C}}^{-1} \right)$$
(1.73)

where

$$\widehat{\mathbf{C}} = \frac{1}{N} \sum_{i}^{N} \widehat{\mathbf{h}}_{i}^{\prime} \widehat{\mathbf{h}}_{i}, \qquad (1.74)$$

$$\widehat{\mathbf{M}}_{g} = \frac{1}{N} \sum_{i}^{N} \left(\widehat{\mathbf{h}}_{i}^{\prime} \widehat{e}_{ig} - \widehat{\mathbf{R}}_{g} \widehat{\mathbf{B}}^{-1} \mathbf{z}_{i}^{\prime} \widehat{v}_{i} \right) \left(\widehat{\mathbf{h}}_{i}^{\prime} \widehat{e}_{ig} - \widehat{\mathbf{R}}_{g} \widehat{\mathbf{B}}^{-1} \mathbf{z}_{i}^{\prime} \widehat{v}_{i} \right)^{\prime}, \quad g = 2, \cdots, G, \quad (1.75)$$

$$\widehat{\mathbf{M}}_{1} = \sum_{g=2}^{G} \widehat{\mathbf{M}}_{g} + \frac{1}{N} \sum_{g \neq k} \left[\left(\widehat{\mathbf{h}}_{i}' \widehat{e}_{ig} - \widehat{\mathbf{R}}_{g} \widehat{\mathbf{B}}^{-1} \mathbf{z}_{i}' v_{i} \right) \left(\widehat{\mathbf{h}}_{i}' \widehat{e}_{ik} - \widehat{\mathbf{R}}_{k} \widehat{\mathbf{B}}^{-1} \mathbf{z}_{i}' v_{i} \right)' \right], \qquad (1.76)$$

$$\widehat{\mathbf{R}}_{g} = \frac{1}{N} \sum_{i}^{N} \left(\widehat{\gamma}_{g} \otimes \widehat{\mathbf{h}}_{i} \right)^{\prime} \nabla_{\pi} \mathbf{x}_{i}(\widehat{\boldsymbol{\pi}}), \qquad (1.77)$$

$$\widehat{\mathbf{B}} = \frac{1}{N} \sum_{i}^{N} \mathbf{z}_{i}' \mathbf{z}_{i}, \qquad (1.78)$$

and

$$\widehat{e}_{ig} = y_{ig} - \widehat{\mathbf{h}}_i \widehat{\boldsymbol{\gamma}}_g. \tag{1.79}$$

The standard error of $\hat{\gamma}_g$ is obtained as the square root of (1.73).

(y_1 , y_2 , y_3) generated across the simulations In Table 1.20, we report the average outcome in each of the three choices across the simulations, and the fraction of times at least one choice being below 0.05. The fractions are about 0.25 ~ 0.45 with the two symmetric distributions and the χ^2 distribution has higher fractions, about 0.70. They suggest that the dependent variable generating process covers the cases where y_i has a set of extreme values such as (0.95, 0.05,0).

D	(<i>e</i>)		No	rmal		Logistic χ_3^2							
π_2	ρ		mean		fraction		mean		fraction		mean		fraction
		y_1	y_2	<i>y</i> 3		y_1	y_2	<i>y</i> 3		y_1	y_2	<i>y</i> 3	
1	1	0.296	0.352	0.352	0.39	0.313	0.344	0.344	0.44	0.103	0.449	0.449	0.68
		(0.014)	(0.007)	(0.007)	(0.02)	(0.015)	(0.008)	(0.008)	(0.02)	(0.009)	(0.005)	(0.005)	(0.02)
0.5	1	0.289	0.356	0.356	0.37	0.307	0.346	0.346	0.42	0.095	0.452	0.452	0.68
		(0.014)	(0.007)	(0.007)	(0.02)	(0.015)	(0.008)	(0.008)	(0.02)	(0.008)	(0.005)	(0.005)	(0.02)
0.2	1	0.286	0.357	0.357	0.36	0.305	0.347	0.347	0.42	0.093	0.453	0.453	0.68
		(0.013)	(0.007)	(0.007)	(0.02)	(0.015)	(0.008)	(0.007)	(0.02)	(0.008)	(0.005)	(0.004)	(0.02)
0.1	1	0.286	0.357	0.357	0.36	0.305	0.347	0.347	0.42	0.093	0.453	0.453	0.68
		(0.013)	(0.007)	(0.007)	(0.02)	(0.015)	(0.008)	(0.007)	(0.02)	(0.008)	(0.005)	(0.004)	(0.02)
1	0.5	0.277	0.361	0.361	0.34	0.299	0.351	0.351	0.40	0.085	0.458	0.457	0.69
		(0.013)	(0.007)	(0.007)	(0.02)	(0.014)	(0.007)	(0.007)	(0.02)	(0.008)	(0.004)	(0.004)	(0.02)
0.5	0.5	0.268	0.366	0.366	0.31	0.291	0.354	0.354	0.38	0.077	0.462	0.462	0.69
		(0.012)	(0.006)	(0.006)	(0.02)	(0.014)	(0.007)	(0.007)	(0.02)	(0.007)	(0.004)	(0.004)	(0.02)
0.2	0.5	0.265	0.368	0.368	0.30	0.289	0.355	0.355	0.37	0.074	0.463	0.463	0.69
		(0.012)	(0.006)	(0.006)	(0.02)	(0.014)	(0.007)	(0.007)	(0.02)	(0.007)	(0.004)	(0.004)	(0.02)
0.1	0.5	0.264	0.368	0.368	0.30	0.289	0.356	0.355	0.37	0.074	0.463	0.463	0.69
		(0.012)	(0.006)	(0.006)	(0.02)	(0.014)	(0.007)	(0.007)	(0.02)	(0.006)	(0.004)	(0.004)	(0.02)
1	0.1	0.263	0.368	0.368	0.30	0.289	0.356	0.356	0.37	0.073	0.463	0.463	0.69
		(0.012)	(0.006)	(0.006)	(0.02)	(0.014)	(0.007)	(0.007)	(0.02)	(0.007)	(0.004)	(0.004)	(0.02)
0.5	0.1	0.251	0.374	0.374	0.27	0.280	0.360	0.360	0.35	0.065	0.468	0.468	0.70
		(0.011)	(0.006)	(0.006)	(0.02)	(0.013)	(0.007)	(0.007)	(0.02)	(0.006)	(0.004)	(0.004)	(0.02)
0.2	0.1	0.248	0.376	0.376	0.26	0.277	0.361	0.361	0.34	0.062	0.469	0.469	0.70
		(0.011)	(0.006)	(0.006)	(0.02)	(0.013)	(0.007)	(0.007)	(0.02)	(0.005)	(0.003)	(0.003)	(0.02)
0.1	0.1	0.247	0.376	0.376	0.25	0.277	0.361	0.361	0.34	0.062	0.469	0.469	0.70
		(0.011)	(0.006)	(0.006)	(0.02)	(0.013)	(0.007)	(0.007)	(0.02)	(0.005)	(0.003)	(0.003)	(0.02)

Table 1.20: (y_{i1}, y_{i2}, y_{i3}) generated across the simulations

 $^{1.} \tau = 0.$

^{2.} Standard deviations are in parentheses.
^{3.} These are the results when the quadratic effect of *w* is not allowed.

REFERENCES

REFERENCES

Arsen, D., and D. N. Plank., *Michigan School Finance Under Proposal A: State Control, Local Consequences.*

Buis, M. L. 2008. "FMLOGIT: Stata module fitting a fractional multinomial logit model by quasi maximum likelihood." Statistical Software Components, Boston College Department of Economics, June.

Gourieroux, C., A. Monfort, and A. Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica*, 52(3): 681–700.

Greene, W. H. 2008. Econometric Analysis.: Prentice Hall.

Lockwood, A. 2002. *School finance reform in Michigan, Proposal A: Retrospective.*: Office of Revenue and Tax Analysis, Michigan Department of Treasury.

Michgian Department of Education. 2005. "Michigan Educational Assessment Program (MEAP) Frequently Asked Questions – Winter 2005." State of Michigan, Department of Education, May.

Mullahy, J. 2010. "Multivariate Fractional Regression Estimation of Econometric Share Models." NBER Working Papers 16354, National Bureau of Economic Research, Inc.

Papke, L. E. 2005. "The effects of spending on test pass rates: evidence from Michigan." *Journal of Public Economics*, 89(5-6): 821–839.

———, and J. M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates." *Journal of Applied Econometrics*, 11(6): 619–32.

_____, and **_____**. 2008. "Panel data methods for fractional response variables with an application to test pass rates." *Journal of Econometrics*, 145(1-2): 121–133.

Petrin, A., and K. Train. 2010. "A Control Function Approach to Endogeneity in Consumer Choice Models." *Journal of Marketing Research*, 47(1): 3–13.

Roy, J. 2011. "Impact of School Finance Reform on Resource Equalization and Academic Performance: Evidence from Michigan." *Education Finance and Policy*, 6(2): 137–167.

Sivakumar, A., and C. Bhat. 2002. "Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis." *Transportation Research Record*, 1790(1): 80–88.

Staiger, D., and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65(3): 557–586.

Wicksall, B., and M. A. Cleary. 2009. "The Basics of the Foundation Allowance – FY 2006-07." House Fiscal Agency, January, memorandum.

Wooldridge, J. M. 2003. Solutions Manual and Supplementary Materials for Econometric Analysis of Cross Section and Panel Data.: MIT Press.

------ . 2010. Econometric Analysis of Cross Section and Panel Data.: MIT Press.

CHAPTER 2

MULTIPLE FRACTIONAL RESPONSE VARIABLES WITH A BINARY ENDOGENOUS EXPLANATORY VARIABLE

2.1 INTRODUCTION

In this chapter, we extend Chapter 1 by allowing endogenous explanatory variables (EEVs) to be discrete. In Chapter 1, we developed a two step estimation method for the response variable. Unlike in a linear model, the response variable's two features are inherent in the nonlinear model for the two step estimation method; each response lies in the unit interval¹ and the sum of responses for a cross sectional unit is one. It maximizes a multinomial distribution with specifying the conditional mean, which includes the first step residuals, as multinomial logit. That is, the method combines fractional multinomial logit developed by Sivakumar and Bhat (2002) and Mullahy (2010) with a control function (CF) approach solving the endogeneity problem by including extra regressors, which is called a control function, in the equation so that the correlation between EEVs and the unobservables would not exist.

The method, however, restricts the probabilistic nature of EEVs. The EEVs should be continuous since it relies on the assumption that the exogenous variables are independent of the EEVs' reduced form errors. So it is not applicable for a research which has discrete EEVs such as "how much people have their pension funds invested in stocks, bonds, and

¹ It can take a corner value, zero or one.

other" where only the shares for these three assets are available and the key explanatory variable is whether or not taking a class of financial education, which could be correlated with some factors that a researcher cannot observe.

Thus, in this chapter we modify the method in order to take the discrete nature of EEVs into account. Especially we consider a binary EEV in the model. A control function approach handling discrete EEVs in nonlinear models was discussed by Terza et al. (2008).² They suggested "two-stage residual inclusion" (2SRI), which includes the unstandardized residuals – an estimator of the difference between the EEV and its conditional mean – as additional regressors in the second step. Wooldridge (2014) extended two-stage residual inclusion and proposed another control function approach. Motivated by variable addition tests, he suggests using the standardized or generalized residuals as control functions instead of the unstandardized residuals. We employ the approach in Wooldridge (2014) to modify the two step estimation method proposed in Chapter 1.

The modified two step estimation method generates the generalized residuals from a probit regression at the first step. Then, it applies the fractional multinomial logit with including the generalized residuals in the conditional mean at the second step.

The second step is a quasi maximum likelihood estimation (QMLE) using a distribution belonging to the linear exponential family. Therefore, it provides a consistent estimator of the mean parameter if the multinomial logit conditional mean is correctly specified, which Gourieroux et al. (1984) describe. In other words, the consistency does not require other distributional assumptions than the conditional mean specification.

Notice that it is the parameter in the "estimating" conditional mean that the method consistently estimates; it is not the structural mean parameter. However, without estimating the structural mean parameter, the method can provide a consistent estimator of average partial effect (APE) on the "structural" conditional mean, which is often more interesting than the mean parameter itself. For the APE estimator to be consistent, the

² Actually, they did not restrict the EEVs' nature.

multinomial logit specification needs to be true.

To see how the method works as an approximation when the specification is wrong, we conduct Monte Carlo simulations. The quantities that we are interested in are the true APEs of the binary EEV. We evaluate the method's approximation to these APEs, compared to several alternative estimation methods' including two stage least squares (*2SLS*), linear control function (*LCF*) approaches, and forbidden regressions. In addition, we compare the performance of its test for endogeneity with one from a linear control function using the same control function – the generalized residuals.

The simulations provide evidence that although the two step estimation method's approximation to APEs depends on how strong the instrument is, it is generally as good as or often better than the alternative methods'. In addition, the two step estimation method's test for endogeneity have not only about the correct size but also better power.

The remainder of this chapter is structured as follows. In the next section, we describe the set of assumptions and the modified two step estimation method. Section 2.3 contains the Monte Carlo simulation design and presents the simulation results where the conditional mean of the method is misspecified. And we conclude the chapter in Section 2.4.

2.2 THE MODIFIED TWO STEP ESTIMATION

Let's consider a random sample in the cross section where each cross sectional observation has *G* choices or shares. The dependent response of interest – multiple fractional response – for observation *i* is written as,

$$\mathbf{y}_{i} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{ig} \\ \vdots \\ y_{iG} \end{pmatrix}_{G \times 1}$$
(2.1)

where $0 \le y_{ig} \le 1$ and $\sum_{g}^{G} y_{ig} = 1$.

Dropping the cross sectional observation index *i*, write the structural conditional mean of the response for choice *g*:

$$E(y_g | \mathbf{x}, r) = E(y_g | \mathbf{z}, w, r) = G_g(\mathbf{z}_1, w, r; \boldsymbol{\beta}), \qquad g = 1, 2, \cdots, G, \qquad (2.2)$$

where $0 < G_g(\cdot) < 1$ and $\sum_g^G G_g = 1$. These two conditions for $G_g(\cdot)$ are required because of the response variable's two features: the bounded nature and the adding-up constraint. \mathbf{x} is the set of explanatory variables including a binary endogenous explanatory variable w and the set of exogenous variables $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2)$ where \mathbf{z}_1 includes an intercept. r is an unobserved omitted variable. Note that the covariates do not have are the g subscript in (2.2). That is, each choice has the same covariates; choice specific covariates are not allowed for. The modified two step estimation method specifies an estimating conditional mean, derived from (2.2), as multinomial logit.³ The multinomial logit model allows covariates to contain characteristics varying across cross sectional observations, not choices. Yet, $E(y_g | \mathbf{x}, r) \neq E(y_h | \mathbf{x}, r)$ for $g \neq h$ – is still possible since the model allows the parameters to vary across g. The second equality of (2.2) shows that \mathbf{z}_2 is redundant in the structural conditional mean, indicating that there is an exclusion restriction.

To set up endogeneity into the model, we use an omitted variable problem by assuming w is correlated with r in the following fashion;

$$w = 1 \left[\mathbf{z} \boldsymbol{\pi} + \boldsymbol{u} > 0 \right], \tag{2.3}$$

$$u \sim Normal(0,1), \tag{2.4}$$

and we are allowing *r* to be correlated with *u*.

³ In (2.2), we are not specifying $G_g(\cdot)$ as multinomial logit. The assumptions regarding $G_g(\cdot)$ are just the two conditions and the restriction for the covariates. Due to the restriction, this approach is more appropriate for the cases where the characteristics for choices are not important.

Additionally, add the following conditional independence assumption:

$$D(r|\mathbf{z}, w) = D(r|v), \tag{2.5}$$

where *v* is the generalized error ⁴ of *w*, which playing a role of a sufficient statistic to control for *w*'s endogeneity. Since $u \sim Normal(0, 1)$,

$$v \equiv E(u|\mathbf{z}, w) = w\lambda(\mathbf{z}\boldsymbol{\pi}) - (1-w)\lambda(-\mathbf{z}\boldsymbol{\pi})$$
(2.6)

where $\lambda(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$ is the Inverse Mill's Ratio.

Based on the assumptions above, the estimating conditional mean

$$E(y_g|\mathbf{x}) = K_g(\mathbf{z}_1, w, v; \boldsymbol{\theta}).$$
(2.7)

Its functional form is determined by the functional form of $G_g(\cdot)$ and the distribution of e. The estimating conditional mean, however, can be specified as any function satisfying $0 < K_g(\cdot) < 1$ and $\sum_g^G K_g = 1$, which are from the two conditions of $G_g(\cdot)$, because we have not yet made any assumptions about those. Instead, we assume that their combination leads to the multinomial logit form for $K_g(\cdot)$:

$$K_{g}(\mathbf{h};\boldsymbol{\theta}) = \frac{\exp\left(\mathbf{h}\boldsymbol{\theta}_{g}\right)}{\sum_{h}^{G}\exp\left(\mathbf{h}\boldsymbol{\theta}_{h}\right)}$$
(2.8)

where $\mathbf{h} = (\mathbf{z}_1 \ w \ v)$ is a $1 \times p$ vector, $\boldsymbol{\theta}_g$ is a $p \times 1$ parameter vector for choice g. $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1 \ \dots \ \boldsymbol{\theta}'_G)'$ is a $pG \times 1$ vector where $\boldsymbol{\theta}_1 = \mathbf{0}$ as a reference.⁵

We can start with specifying $G_g(\cdot)$ as a multinomial logit in (2.2), and assume joint normality of (r, u); then, in principle, we can find $E(y_g | \mathbf{x})$. But it would not be in closed form. So, instead, we use (2.5) as an approximation, and the model $E(y_g | \mathbf{x}) = E(y_g | \mathbf{x}, v)$

⁴ Gourieroux et al. (1987)

⁵ (2.8) satisfies the two conditions for $K_g(\cdot)$

directly as a multinomial logit. Neither approach contains the other, as Wooldridge (2014) discussed; they use different assumptions.

In order to estimate θ , a consistent estimator of generalized error v_i should be obtained first since it is not observed. Therefore, the modified two step estimation method is summarized as follows:

Procedure 2.2

Step 1. Obtain $\hat{\pi}$ from the probit regression of w_i on \mathbf{z}_i and compute the generalized residual \hat{v}_i ,

$$\widehat{v}_i = w\lambda(\mathbf{z}\widehat{\boldsymbol{\pi}}) - (1 - w)\lambda(-\mathbf{z}\widehat{\boldsymbol{\pi}}) = \frac{\phi(\mathbf{z}_i\widehat{\boldsymbol{\pi}})(w_i - \Phi(\mathbf{z}_i\widehat{\boldsymbol{\pi}}))}{\Phi(\mathbf{z}_i\widehat{\boldsymbol{\pi}})(1 - \Phi(\mathbf{z}_i\widehat{\boldsymbol{\pi}}))}$$
(2.9)

Step 2. Run fractional multinomial logit (fmlogit) of (y_{i1}, \dots, y_{iG}) on \mathbf{z}_{i1}, w_i and \hat{v}_i , which maximizes a multinomial log-likelihood:

$$\sum_{i}^{N} \ell_{i}(\boldsymbol{\theta}) = \sum_{i}^{N} \left(\sum_{g}^{G} y_{ig} \log K_{g}(\widehat{\mathbf{h}}_{i}; \boldsymbol{\theta}) \right).$$
(2.10)

where $\widehat{\mathbf{h}}_i = (\mathbf{z}_{i1} \ w_i \ \widehat{v}_i)$.

The method provides a convenient test for endogeneity with the null hypothesis that w is exogenous by obtaining an asymptotic robust t statistic on \hat{v}_i . This is a variable addition test discussed in Wooldridge (2014).

Furthermore, it is able to estimate average partial effect (APE), which is often the quantity of more interest, without estimating the structural mean parameter β . It is because the assumptions above ensure the following equivalence:

$$\delta_g(\mathbf{z}^\circ) = APE_g(\mathbf{z}^\circ) \equiv E_r[E(y_g|\mathbf{z}^\circ, w=1, r) - E(y_g|\mathbf{z}^\circ, w=0, r)]$$
(2.11)

$$= E_r[G_g(\mathbf{z}_1^{\circ}, w = 1, r; \boldsymbol{\beta}) - G_g(\mathbf{z}_1^{\circ}, w = 0, r; \boldsymbol{\beta})]$$
(2.12)

$$= E_v[K_g(\mathbf{z}_1^\circ, w = 1, v; \boldsymbol{\theta}) - K_g(\mathbf{z}_1^\circ, w = 0, v; \boldsymbol{\theta})], \qquad (2.13)$$

where \mathbf{z}° denotes a fixed value of $\mathbf{z}^{.6}$ As shown in Wooldridge (2010, section 2.2.5), (2.13) holds under (2.5) and $E(y_g | \mathbf{x}, r, v) = E(y_g | \mathbf{x}, r)$, which can be derived under the assumptions that have been made so far.

In order to have a representative single number summarizing $\delta_g(\mathbf{z}^\circ)$, we average it out across the sample for \mathbf{z} :

$$\delta_g = E_{\mathbf{z}} \left[\delta_g(\mathbf{z}) \right] \tag{2.14}$$

Thus, δ_g can be estimated by obtaining

$$\widehat{\delta}_{g}^{2step} = \frac{1}{N} \sum_{j}^{N} \left[\frac{1}{N} \sum_{i}^{N} \left[K_{g}(\mathbf{z}_{j1}, w = 1, v_{i}; \widehat{\boldsymbol{\theta}}) - K_{g}(\mathbf{z}_{j1}, w = 0, v_{i}; \widehat{\boldsymbol{\theta}}) \right] \right], \quad (2.15)$$

where $\hat{\theta}$ is the two step estimator of θ .

One thing to be aware of is that the standard errors of $\hat{\theta}$ and $\hat{\delta}_g^{2step}$ need to take the additional variation caused by the firs step into consideration by using the delta method or bootstrapping the two steps.

 θ is a QMLE estimator using a distribution, which is one of linear exponential family. Therefore, based on the discussion of Gourieroux, Monfort, and Trognon (1984), its consistency is ensured only by (2.8). In other words, it is still consistent even though the distribution specification is completely wrong except the conditional mean.

If (2.8) is wrong, $\hat{\delta}_g^{2step}$ is inconsistent because $\hat{\theta}$ is inconsistent and the wrong functional form is used in (2.15). So we conduct Monte Carlo simulations in the next section

⁶ (2.11) can be written as $ASF(\mathbf{z}^{\circ}, w = 1) - ASF(\mathbf{z}^{\circ}, w = 0)$, $ASF(\cdot)$ denotes "average structure function" defined in Blundell and Powell (2003) and Wooldridge (2005).

in order to examine how well the two step estimation method approximate to δ_g when the multinomial logit conditional mean is misspecified. The simulations compare its approximation with several alternative methods that researchers would use. These methods are 1) Two stage least squares (2*SLS*),⁷ 2) Linear control function approach using the generalized residual (*LCF*), 3) Linear IV using the fitted probability $\widehat{\Phi}(\cdot)$ as an instrument (*LIV*), 4) Linear plug-in method, and 5) Two step plug-in method.⁸

The adding-up constraint is not inherent in the alternative linear models. So they should apply their estimation methods to G - 1 equations by dropping one of G choices. Then, the dropped one's parameters are obtained based on the constraint and the estimates for the other choices.⁹ The coefficient estimates of these linear models are comparable to $\hat{\delta}_{g}$.

The alternative methods include two forbidden regressions – the two plug-in methods. They substitute w_i with its fitted probability $\widehat{\Phi}(\cdot)$ at their second steps. Researchers often attempt to use this kind of approach, believing that it is legitimate because it emulates the 2SLS procedure. However, they are inappropriate especially for nonlinear models.

In addition to 1) through 5), we include the generalized residual – the control function – in a flexible way¹⁰ for the two step estimation method and *LCF* in order to examine if it helps their approximations. We call them 6) *Two step flexible* and 7) *LCF flexible*, respectively.

Furthermore, we compare the performance of tests for endogeneity done by the two step method and *LCF* with significance level $\alpha = 0.05$ by varying degree of endogeneity and the instrument's predictive power; the null hypotheses of the tests are that there is no endogeneity.¹¹

⁷ This is the same as a linear control function approach using the residual from the linear regression w on **z**.

⁸ Terza et al. (2008) call this approach "two-stage predictor substitution (2SPS)."

⁹ In the simulations, we drop the first choice (g = 1), which is the reference choice for the two step estimation method, and obtain its estimates using $\gamma_1 = \mathbf{e}_1 - \sum_{g=2}^G \gamma_g$ where γ_g is the coefficient parameter for choice g in the linear models and \mathbf{e}_1 is a unit vector.

 $^{^{10} \, \}hat{v}^2$ and \hat{v}^3 are additionally included in the second step.

¹¹ $H_0: \theta_v = 0$ for the two step estimation method and $H_0: \gamma_v = 0$ for LCF.

2.3 MONTE CARLO SIMULATION

2.3.1 Data Generating Process

We use 1000 replications where each replication has 500 cross sectional observations (N = 500) and 3 choices (g = 3).

The data generating process for each replication is as follows:

The covariates

•
$$\mathbf{z}_i = \begin{pmatrix} \mathbf{z}_{i1} & z_{i2} \end{pmatrix} = \begin{pmatrix} 1 & z_{i1} & z_{i2} \end{pmatrix}_{1 \times 3}$$

where $\begin{pmatrix} z_{i1} \\ z_{i2} \end{pmatrix} \sim MV Normal \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \tau \\ \tau & 1 \end{pmatrix} \end{bmatrix}$ and $\tau \in \{0, -0.5\}$.

•
$$u_i \sim Normal(0, \sigma_u^2)^{12}$$

• $w_i = 1[w_i^* > 0] = 1[\mathbf{z}_i \pi + u_i > 0] = 1[z_{i1}\pi_1 + z_{i2}\pi_2 + u_i > 0]$ where $\pi_1 \in \{0, 0.5\}$ and $\pi_2 \in \{0.1, 0.2, 0.5, 1\}$.

•
$$v_i = \sigma_u \left[w_i \lambda \left(\frac{\mathbf{z}_i \pi}{\sigma_u} \right) - (1 - w_i) \lambda \left(- \frac{\mathbf{z}_i \pi}{\sigma_u} \right) \right]$$

• $r_i = \rho u_i + e_i$ where $\rho \in \{0.1, 0.25, 0.5, 0.75, 0.9, 1\}$ and $D(e_i | v_i)$ is one of the three:

(a)
$$e_i | v_i \sim Normal(0, 1)$$

(b)
$$e_i | v_i \sim \chi_3^2$$

(c)
$$e_i | v_i \sim Normal\left(0, 1 + \frac{1}{2}v_i^2\right)$$
.

Two symmetric distributions and an asymmetric distribution are in use for the distribution of $e_i | v_i$ With (a) or (b), *r* is generated to be uncorrelated with *w* if $\rho = 0$.

¹² The values of π_1 and π_2 affect $Var(w_i)$. As (π_1, π_2) varies, we adjust σ_u for $Var(w_i^*)$ to be invariant instead of $Var(w_i)$: $Var(w_i^*) = 2$.

However, with (c), where there is heteroskedasticity, *r* still depends on *w* even though $\rho = 0$.

The structural conditional mean $G_g(\cdot)$ specification: We specify the structural conditional mean as multinomial logit as well since it satisfies the two conditions for $G_g(\cdot)$:

$$E(y_{ig}|\mathbf{x}_i, r_i) = G_g(\mathbf{z}_{i1}, w_i, r_i; \boldsymbol{\beta}) = \frac{\exp\left(\mathbf{z}_{i1}\boldsymbol{\beta}_{zg} + w_i\boldsymbol{\beta}_{wg} + r_i\boldsymbol{\beta}_{rg}\right)}{\sum_h^3 \exp\left(\mathbf{z}_{i1}\boldsymbol{\beta}_{zh} + w_i\boldsymbol{\beta}_{wh} + r_i\boldsymbol{\beta}_{rh}\right)}$$
(2.16)

where $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1 \ \boldsymbol{\beta}'_2 \ \boldsymbol{\beta}'_3)'$ is a 12 × 1 vector, $\boldsymbol{\beta}_g = (\boldsymbol{\beta}'_{zg} \ \boldsymbol{\beta}_{wg} \ \boldsymbol{\beta}_{rg}) = (1 \ 1 \ 1 \ 1)'$ for g = 2, 3, and $\boldsymbol{\beta}_1 = \mathbf{0}$. Under (2.16), neither of the three distributions of *e* can derives (2.8).

The multiple fractional dependent variables y: The dependent variable generating process is the same as that in Chapter 1; we first draw 100 multinomial outcomes among 1, 2, and 3 based on (2.16), and then calculate the proportions for the three outcomes.

2.3.2 Simulation Results

Case 1: Endogeneity comes only through *u*.

Let's consider the two settings where the data are generated so that w becomes exogenous when $\rho = 0$.

Table 2.1 through 2.3 contain simulation results as $\pi_2 = 1$.¹³ In Table 2.1, with the standard normal distribution, $\hat{\delta}_g^{2step}$ is pretty similar to δ_g regardless of the degree of endogeneity, and the alternative methods also provide good approximations. Even the forbidden regressions' approximations are good. Including the flexible forms of \hat{v} does not improve the approximations for both the two step estimation method and *LCF*. In addition, allowing for the correlation between \mathbf{z}_1 and \mathbf{z}_2 does not change the story.¹⁴ Table 2.2 shows that, under the asymmetric distribution, the biases are slightly larger than those in

 $^{^{13}}$ The results with other values of ρ are available upon request.

¹⁴ $\pi_1 = 0.5$ and $\tau = -0.5$

Table 2.1. Yet, in considering the efficiency as well as the bias, all of the methods under the asymmetric distribution also approximate well, as shown in Table 2.3.

ρ			1			1	
π_1			0.5			0	
g		1	2	3	1	2	3
δ_g	Mean	-0.107	0.054	0.054	-0.107	0.054	0.054
Two Step	Mean	-0.107	0.054	0.053	-0.108	0.054	0.054
	SD	(0.035)	(0.020)	(0.019)	(0.029)	(0.016)	(0.016)
Two Step	Mean	-0.105	0.053	0.052	-0.106	0.053	0.053
Flexible	SD	(0.046)	(0.026)	(0.026)	(0.035)	(0.020)	(0.020)
Two Step	Mean	-0.106	0.053	0.053	-0.107	0.054	0.053
Plug-in	SD	(0.037)	(0.021)	(0.020)	(0.030)	(0.017)	(0.016)
LCF	Mean	-0.107	0.054	0.053	-0.107	0.054	0.053
	SD	(0.037)	(0.020)	(0.020)	(0.030)	(0.017)	(0.016)
LCF	Mean	-0.107	0.054	0.053	-0.107	0.054	0.053
Flexible	SD	(0.047)	(0.026)	(0.026)	(0.036)	(0.020)	(0.020)
2SLS	Mean	-0.106	0.053	0.053	-0.107	0.054	0.053
	SD	(0.037)	(0.021)	(0.020)	(0.031)	(0.017)	(0.017)
LIV	Mean	-0.106	0.053	0.053	-0.107	0.054	0.053
$(\text{IV}=\widehat{\Phi}(\cdot))$	SD	(0.037)	(0.021)	(0.020)	(0.030)	(0.017)	(0.016)
Linear	Mean	-0.106	0.053	0.053	-0.107	0.054	0.053
Plug-in	SD	(0.037)	(0.021)	(0.020)	(0.030)	(0.017)	(0.016)
ρ			0.5			0.1	
π_1				()		
8		1	2	3	1	2	3
δ_g	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
Two Step	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
		(0,000)	(0, 0, 1, 1)	(0, 0, 1, 0)	(0.001)	(0, 0, 1, 0)	(0, 01, 0)

Table 2.1: APEs with $r = \rho u + e$, $\pi_2 = 1$, and $e|_v \sim Normal(0, 1)$

Ρ			0.0			0.1	
π_1				()		
8		1	2	3	1	2	3
δ_g	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
Two Step	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
	SD	(0.023)	(0.014)	(0.013)	(0.021)	(0.013)	(0.012)
Two Step	Mean	-0.108	0.054	0.054	-0.109	0.055	0.054
Flexible	SD	(0.028)	(0.017)	(0.017)	(0.025)	(0.016)	(0.015)
Two Step	Mean	-0.107	0.054	0.054	-0.108	0.054	0.054
Plug-in	SD	(0.024)	(0.014)	(0.014)	(0.021)	(0.013)	(0.013)
LCF	Mean	-0.107	0.054	0.054	-0.108	0.054	0.054
	SD	(0.024)	(0.014)	(0.014)	(0.022)	(0.013)	(0.013)
LCF	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
Flexible	SD	(0.030)	(0.017)	(0.017)	(0.027)	(0.016)	(0.016)
2SLS	Mean	-0.107	0.054	0.053	-0.108	0.054	0.054
	SD	(0.025)	(0.015)	(0.014)	(0.022)	(0.014)	(0.013)
LIV	Mean	-0.107	0.054	0.054	-0.108	0.054	0.054
$(\text{IV}=\widehat{\Phi}(\cdot))$	SD	(0.024)	(0.014)	(0.014)	(0.022)	(0.013)	(0.013)
Linear	Mean	-0.107	0.054	0.053	-0.108	0.054	0.054
Plug-in	SD	(0.024)	(0.014)	(0.014)	(0.022)	(0.013)	(0.013)

^{1.} As $\pi_1 = 0.5$, $\tau = -0.5$; otherwise, $\tau = 0$.

0			1		1				
<u> </u>			0.5			 			
<u></u>		1	<u>0.5</u>	2	1	2	2		
<u> </u>	Moon	1	<u> </u>	0.017	1	<u> </u>	0.016		
Ug Two Stop	Moon	-0.033	0.017	0.017	-0.032	0.010	0.010		
Two Step	SD	(0.039)	(0.020)	(0.019)	(0.030)	(0.019)	(0.010)		
Two Stop	Moon	0.015)	0.012)		(0.012)		0.009)		
Floviblo	SD	(0.033)	(0.016)	(0.017)	(0.033)	(0.017)	(0.010)		
	Maan	(0.020)	(0.010)	(0.013)	(0.014)	(0.012)	(0.012)		
Two Step	Niean	-0.029	(0.013)	0.014	-0.029	0.015	0.014		
Plug-in	50	(0.019)	(0.013)	(0.013)	(0.015)	(0.011)	(0.010)		
LCF	Mean	-0.028	0.015	0.014	-0.028	0.015	0.014		
	SD	(0.020)	(0.013)	(0.013)	(0.015)	(0.011)	(0.010)		
LCF	Mean	-0.034	0.018	0.016	-0.032	0.017	0.016		
Flexible	SD	(0.025)	(0.017)	(0.017)	(0.018)	(0.013)	(0.013)		
2SLS	Mean	-0.029	0.015	0.014	-0.029	0.015	0.014		
	SD	(0.019)	(0.013)	(0.013)	(0.015)	(0.011)	(0.011)		
LIV	Mean	-0.029	0.015	0.014	-0.029	0.015	0.014		
$(\text{IV}=\widehat{\Phi}(\cdot))$	SD	(0.019)	(0.013)	(0.013)	(0.015)	(0.011)	(0.010)		
Linear	Mean	-0.029	0.015	0.014	-0.029	0.015	0.014		
Plug-in	SD	(0.019)	(0.013)	(0.013)	(0.015)	(0.011)	(0.010)		
ρ			0.5			0.1			
π_1				()				
8		1	2	3	1	2	3		
δ_g	Mean	-0.028	0.014	0.014	-0.027	0.013	0.013		
Two Step	Mean	-0.029	0.015	0.014	-0.027	0.014	0.013		
-	SD	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)		
Two Step	Mean	-0.028	0.015	0.014	-0.027	0.014	0.013		
Flexible	SD	(0.011)	(0.011)	(0.011)	(0.010)	(0.011)	(0.011)		
Two Step	Mean	-0.028	0.014	0.013	-0.027	0.014	0.013		
Plug-in	SD	(0.011)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)		
LCF	Mean	-0.027	0.014	0.013	-0.027	0.014	0.013		
	SD	(0.011)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)		

Table 2.2: APEs with $r = \rho u + e$, $\pi_2 = 1$, and $e|_v \sim \chi_3^2$

LCF Mean -0.028 0.015 0.013 -0.027 0.014 0.013 Flexible (0.013)(0.011)(0.011)(0.011)(0.011) SD (0.011)2SLS 0.014 0.013 Mean -0.027 0.014 0.013 -0.027 (0.009)(0.009)(0.009)SD (0.011)(0.009)(0.009)LIV Mean -0.027 0.014 0.013 -0.027 0.014 0.013 $(IV = \widehat{\Phi}(\cdot))$ SD(0.011)(0.009)(0.009)(0.009)(0.009)(0.008)Linear Mean -0.027 -0.027 0.014 0.013 0.014 0.013 Plug-in (0.009)SD (0.011)(0.009)(0.009)(0.009)(0.008)

^{1.} As $\pi_1 = 0.5$, $\tau = -0.5$; otherwise, $\tau = 0$.

	ρ	1				1			0.5		0.1		
D(e v)	π_1		0.5		0								
	8	1	2	3	1	2	3	1	2	3	1	2	3
Normal	Two Step	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
(0,1)	Two Step Flexible	0.002	0.001	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
	Two Step Plug-in	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	LCF	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	LCF Flexible	0.002	0.001	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
	2SLS	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	LIV (IV= $\widehat{\Phi}(\cdot)$)	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	Linear Plug-in	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
χ^2_3	Two Step	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0	Two Step Flexible	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Two Step Plug-in	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	LCF	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	LCF Flexible	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2SLS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	LIV (IV= $\widehat{\Phi}(\cdot)$)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Linear Plug-in	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2.3: MSEs of APEs with $r = \rho u + e$ and $\pi_2 = 1$

^{1.} As $\pi_1 = 0.5$, $\tau = -0.5$; otherwise, $\tau = 0$. ^{2.} MSEs are calculated from Table 2.1 and 2.2.

We weaken the predictive power $(\pi_2 < 1)^{15}$ while keeping the strongest degree of endogeneity ($\rho = 1$) in order to examine if these good approximations hinges on the instrument's strong predictive power.

To determine if z_2 is a strong instrument, we use the rule of thumb suggested by Staiger and Stock (1997); the first stage F statistic, testing the null hypothesis that instruments are uncorrelated with EEVs, should be larger than 10 for the instruments to have properties as strong instruments. Since the two step estimation method's first step is a probit regression, we apply their rule to the first step Wald statistics testing the same null hypothesis. Table 2.4 gives the summary of the Wald and the F statistics.¹⁶

		τ			0		-0.5
D(e v)	1 st step	π_1			0		0.5
		π_2	0.1	0.2	0.5	1	1
Normal	OLS	F statistic	2.661	7.923	53.215	443.384	261.906
(0,1)			(2.914)	(5.733)	(19.551)	(84.613)	(60.292)
		[F>10]	2.6%	28.9%	100.0%	100.0%	100.0%
	Probit	Wald statistic	2.582	7.410	39.589	133.797	107.881
			(2.746)	(5.025)	(10.962)	(8.861)	(11.291)
		[Wald>10]	2.5%	26.2%	99.8%	100.0%	100.0%
	Re	plication	995	999	1000	1000	1000
χ^2_3	OLS	F statistic	2.861	7.992	54.203	443.938	263.869
0			(3.215)	(5.905)	(19.308)	(85.330)	(59.632)
		[F>10]	4.0%	32.0%	99.8%	100.0%	100.0%
	Probit	Wald statistic	2.763	7.464	40.174	134.474	108.501
			(2.988)	(5.204)	(10.957)	(9.186)	(11.390)
		[Wald>10]	3.0%	29.4%	99.8%	100.0%	100.0%
	Re	plication	987	999	1000	1000	1000

Table 2.4: F and Wald statistics of the first stage/step (H_0 : $\pi_2 = 0$)

^{1.} Standard deviations are in parentheses

^{2.} [F>10] and [Wald>10] indicate the proportions of the F and Wald statistics being greater than 10 among the replications, respectively. ^{3.} The tests are robust to heteroskedasticity.

They are not greater than 10 until $\pi_2 = 0.5$, on average. When $\pi_2 = 0.1$, only 25

¹⁵ $\pi_1 = \tau = 0.$

¹⁶ The first stages of 2SLS and LIV are linear regressions. We report only the former because the latter are quite similar. (The summary of *LIV*'s first stage F statistics are available upon request.)

~ 40 replications have the statistics being larger than 10. The proportions increase as π_2 increases, and then they are larger than 10 in almost every replication as $\pi_2 = 0.5$. Therefore, we consider z_2 as a weak instrument if $\pi_2 \leq 0.2$ and as a strong one if $\pi_2 \geq 0.5$.

Table 2.5, containing the results with the standard normal distribution, show that, in general, the higher the predictive power is, the less biased and the more precise the estimates are, and that as π_2 increases, the nonlinear models more quickly recover their approximation abilities. When $\pi_2 = 0.1$, the linear models yield worse approximations than the three nonlinear models, providing much larger biases and huge standard deviations. When $\pi_2 = 0.2$, the nonlinear models' estimates become closer to δ_g and more precise whereas the linear models still provide poor approximations. As $\pi_2 = 0.5$, which makes z_2 a strong instrument, the approximations by all the methods are quite good.

Table 2.7 clearly shows that, among the nonlinear models, the two step estimation method yields a better approximation in both bias and efficiency criteria, and that it does not improve the approximations to include \hat{v} in a flexible way. The flexible way does not help the estimation to obtain smaller biases but causes bigger standard deviations. Table 2.6 and 2.7 also present that, in general, the asymmetric distribution have a similar results as the standard normal distribution.

π_2			0.1			0.2			0.5	
8		1	2	3	1	2	3	1	2	3
δ_g	Mean	-0.106	0.053	0.053	-0.106	0.053	0.053	-0.106	0.053	0.053
Two Step	Mean	-0.171	0.092	0.079	-0.082	0.042	0.039	-0.102	0.051	0.051
	SD	(0.300)	(0.186)	(0.170)	(0.189)	(0.110)	(0.103)	(0.071)	(0.040)	(0.038)
Two Step	Mean	-0.184	0.097	0.088	-0.127	0.067	0.060	-0.100	0.050	0.050
Flexible	SD	(0.427)	(0.283)	(0.290)	(0.350)	(0.210)	(0.223)	(0.123)	(0.070)	(0.069)
Two Step	Mean	-0.147	0.082	0.064	-0.053	0.028	0.025	-0.101	0.050	0.050
Plug-in	SD	(0.440)	(0.270)	(0.249)	(0.273)	(0.151)	(0.147)	(0.075)	(0.041)	(0.039)
LCF	Mean	-0.061	0.061	0.000	-0.017	0.010	0.007	-0.101	0.051	0.050
	SD	(6.387)	(3.274)	(3.165)	(0.752)	(0.370)	(0.397)	(0.075)	(0.041)	(0.039)
LCF	Mean	1.427	-0.861	-0.566	0.108	-0.035	-0.073	-0.099	0.050	0.049
Flexible	SD	(15.611)	(9.053)	(7.174)	(2.567)	(1.639)	(1.504)	(0.132)	(0.073)	(0.073)
2SLS	Mean	2.545	-1.303	-1.242	-0.023	0.014	0.009	-0.101	0.050	0.050
	SD	(91.074)	(47.430)	(43.651)	(0.524)	(0.260)	(0.292)	(0.075)	(0.041)	(0.040)
LIV	Mean	0.112	-0.068	-0.045	-0.027	0.015	0.013	-0.101	0.050	0.050
$(\text{IV}=\widehat{\Phi}(\cdot))$	SD	(7.392)	(4.376)	(3.079)	(0.496)	(0.259)	(0.257)	(0.075)	(0.041)	(0.040)
Linear	Mean	-0.218	0.150	0.067	-0.019	0.012	0.007	-0.101	0.050	0.050
Plug-in	SD	(5.287)	(2.830)	(2.590)	(0.597)	(0.284)	(0.337)	(0.075)	(0.041)	(0.040)
Replicat	ions		995/1000			999/1000)	1	1000/100	0

Table 2.5: APEs with $w = 1[z_2\pi_2 + u > 0]$, $\rho = 1$, and $e|_v \sim Normal(0, 1)$

^{1.} $\pi_1 = \tau = 0.$
π_2			0.1			0.2			0.5	
8		1	2	3	1	2	3	1	2	3
δ_g	Mean	-0.036	0.018	0.018	-0.036	0.018	0.018	-0.035	0.018	0.018
Two Step	Mean	-0.064	0.035	0.028	-0.029	0.018	0.012	-0.044	0.023	0.021
	SD	(0.275)	(0.168)	(0.166)	(0.159)	(0.093)	(0.093)	(0.031)	(0.022)	(0.022)
Two Step	Mean	0.011	-0.005	-0.006	-0.002	0.000	0.002	-0.030	0.016	0.013
Flexible	SD	(0.431)	(0.279)	(0.283)	(0.345)	(0.215)	(0.209)	(0.082)	(0.053)	(0.051)
Two Step	Mean	-0.036	0.018	0.018	-0.012	0.008	0.003	-0.027	0.014	0.013
Plug-in	SD	(0.428)	(0.252)	(0.253)	(0.218)	(0.119)	(0.126)	(0.042)	(0.026)	(0.026)
LCF	Mean	-0.046	0.021	0.025	0.042	-0.013	-0.029	-0.026	0.014	0.012
	SD	(3.994)	(2.047)	(2.060)	(1.484)	(0.654)	(0.838)	(0.041)	(0.026)	(0.025)
LCF	Mean	-1.639	0.772	0.868	-0.464	0.227	0.237	-0.039	0.021	0.018
Flexible	SD	(9.191)	(5.188)	(4.902)	(4.056)	(2.409)	(1.797)	(0.081)	(0.054)	(0.051)
2SLS	Mean	0.057	-0.044	-0.013	-0.029	0.018	0.012	-0.026	0.014	0.012
	SD	(2.241)	(1.396)	(1.118)	(0.645)	(0.275)	(0.392)	(0.040)	(0.025)	(0.025)
LIV	Mean	-0.074	0.194	-0.120	-0.007	0.007	0.000	-0.026	0.014	0.012
$(IV=\widehat{\Phi}(\cdot))$	SD	(1.891)	(4.725)	(3.054)	(0.411)	(0.195)	(0.250)	(0.040)	(0.025)	(0.025)
Linear	Mean	0.077	-0.055	-0.022	-0.043	0.024	0.019	-0.026	0.014	0.012
Plug-in	SD	(2.632)	(1.632)	(1.229)	(1.044)	(0.443)	(0.614)	(0.040)	(0.025)	(0.025)
Replicat	ions		987/1000)		999/1000		1	1000/100	0

Table 2.6: APEs with $w = 1[z_2\pi_2 + u > 0]$, $\rho = 1$, and $e|_v \sim \chi_3^2$

^{1.} $\pi_1 = \tau = 0.$

	π_2		0.1			0.2			0.5	
D(e v)	8	1	2	3	1	2	3	1	2	3
Normal	Two Step	0.094	0.036	0.030	0.036	0.012	0.011	0.005	0.002	0.001
(0,1)	Two Step Flexible	0.189	0.082	0.085	0.123	0.044	0.050	0.015	0.005	0.005
	Two Step Plug-in	0.196	0.074	0.062	0.078	0.024	0.022	0.006	0.002	0.002
	LCF	40.802	10.722	10.023	0.574	0.139	0.160	0.006	0.002	0.002
	LCF Flexible	246.059	82.795	51.851	6.636	2.693	2.279	0.017	0.005	0.005
	2SLS	8301.533	2251.416	1907.082	0.281	0.069	0.087	0.006	0.002	0.002
	LIV (IV= $\widehat{\Phi}(\cdot)$)	54.686	19.166	9.490	0.252	0.069	0.068	0.006	0.002	0.002
	Linear Plug-in	27.963	8.016	6.706	0.364	0.082	0.116	0.006	0.002	0.002
χ^2_3	Two Step	0.076	0.029	0.028	0.025	0.009	0.009	0.001	0.001	0.000
· ·	Two Step Flexible	0.188	0.078	0.081	0.120	0.047	0.044	0.007	0.003	0.003
	Two Step Plug-in	0.183	0.063	0.064	0.048	0.014	0.016	0.002	0.001	0.001
	LCF	15.953	4.189	4.244	2.208	0.428	0.705	0.002	0.001	0.001
	LCF Flexible	87.046	27.488	24.755	16.638	5.848	3.277	0.007	0.003	0.003
	2SLS	5.032	1.952	1.251	0.416	0.075	0.153	0.002	0.001	0.001
	LIV (IV= $\widehat{\Phi}(\cdot)$)	3.579	22.353	9.343	0.170	0.038	0.063	0.002	0.001	0.001
	Linear Plug-in	6.941	2.670	1.512	1.091	0.196	0.377	0.002	0.001	0.001

Table 2.7: MSEs of APE estimates with $w = 1[z_2\pi_2 + u > 0]$ and $\rho = 1$

^{1.} $\pi_1 = \tau = 0$. ^{2.} MSEs are calculated from Table 2.5 and 2.6.

Table 2.8 compares the rejection frequencies for the different degrees of endogeneity when the strongest instrument is in use. The tests by both the two step estimation method and *LCF* have good size properties; as $\rho = 0$, their rejection frequencies are quite close to the nominal value. Allowing for the additional terms of \hat{v} makes them over-reject. The two step estimation method, however, has better power than *LCF* for both the two distributions. Figure 2.1 and 2.2¹⁷ illustrate a pattern that the two step estimation method's rejection frequencies increase faster than *LCF*'s as $\rho > 0$. With χ_3^2 distribution, the pattern is more evident and the flexible way helps the two step estimation method have slightly better power.

¹⁷ The graphs show average rejection frequencies for the two choices at each value of ρ in Table 2.8.

	ρ	0		0.1		0.25		0.5		0.75		0.9		1	
D(e v)	8	2	3	2	3	2	3	2	3	2	3	2	3	2	3
Normal	Two Step	0.05	0.06	0.11	0.12	0.46	0.46	0.92	0.92	1.00	1.00	1.00	1.00	1.00	1.00
(0,1)	LCF	0.05	0.06	0.09	0.09	0.30	0.32	0.79	0.79	0.97	0.97	0.99	0.99	1.00	1.00
	Two Step flexible	0.11	0.11	0.15	0.16	0.40	0.41	0.89	0.89	0.99	0.99	1.00	1.00	1.00	1.00
	LCF flexible	0.10	0.10	0.13	0.12	0.26	0.25	0.66	0.68	0.93	0.94	0.98	0.98	0.99	0.99
χ^2_3	Two Step	0.06	0.07	0.09	0.08	0.22	0.22	0.54	0.55	0.78	0.79	0.88	0.88	0.91	0.90
	LCF	0.06	0.04	0.06	0.05	0.10	0.07	0.20	0.18	0.37	0.38	0.48	0.51	0.56	0.60
	Two Step flexible	0.17	0.17	0.19	0.18	0.28	0.29	0.58	0.58	0.85	0.85	0.94	0.93	0.96	0.96
	LCF flexible	0.11	0.11	0.12	0.11	0.13	0.13	0.18	0.21	0.32	0.31	0.41	0.41	0.48	0.49

Table 2.8: Rejection Frequencies for $\alpha = 0.05$ test in Case 1 with varying ρ

^{1.} $\pi_2 = 1$, $\pi_1 = \tau = 0$. ^{2.} The null hypotheses for the flexible models are that the coefficients of \hat{v} , \hat{v}^2 , and \hat{v}^3 are zeroes.



Figure 2.1: Rejection Frequencies for Normal(0,1) with varying ρ

Figure 2.2: Rejection Frequencies for χ^2_3 with varying ρ



Furthermore, the rejection frequencies in Table 2.9 suggest that the stronger the instru-

ment is, the higher the power is, except *LCF flexible*. Figure 2.3 and 2.4¹⁸ show that the two step estimation method has slightly better power with the standard normal distribution and much better power with the χ_3^2 distribution than *LCF*. Again, adding additional forms of \hat{v} in the estimation have the two step estimation method have slightly higher rejection frequencies for χ_3^2 distribution. *LCF flexible* has a different pattern, however; it has a U-shape graph whereas the other three are monotonically increasing, suggesting *LCF* could have higher power by including \hat{v} in a flexible way if a weak instrument is in use.

¹⁸ The graphs show average rejection frequencies for the two choices at each value of π_2 where $\pi_1 = \tau = 0$ in Table 2.9.

	τ	-0	.5	0									
	π_1	0.5					()					
	π_2		L	1		0.5		0.2		0.	.1		
D(e v)	8	2	3	2	3	2	3	2	3	2	3		
Normal	Two Step	1.00	1.00	1.00	1.00	0.97	0.97	0.43	0.43	0.15	0.14		
(0,1)	LCF	1.00	1.00	1.00	1.00	0.92	0.92	0.32	0.33	0.14	0.12		
	Two Step flexible	1.00	1.00	1.00	1.00	0.98	0.97	0.40	0.39	0.16	0.15		
	LCF flexible	0.98	0.99	0.99	0.99	0.86	0.87	0.60	0.58	0.69	0.68		
	Replication	1000		10	00	10	00	99	99	99	95		
χ_3^2	Two Step	0.88	0.89	0.91	0.90	0.61	0.61	0.17	0.16	0.09	0.09		
-	LCF	0.58	0.60	0.56	0.60	0.35	0.36	0.10	0.10	0.08	0.07		
	Two Step flexible	0.95	0.95	0.96	0.96	0.70	0.70	0.23	0.24	0.16	0.16		
	LCF flexible	0.48	0.48	0.48	0.49	0.31	0.28	0.35	0.33	0.53	0.52		
	Replication	10	00	10	00	10	00	99	99	98	37		

Table 2.9: Rejection Frequencies for $\alpha = 0.05$ test in Case 1 with varying π_2

^{1.} $\rho = 1$ ^{2.} The null hypotheses for the flexible models are that the coefficients of \hat{v} , \hat{v}^2 , and \hat{v}^3 are zeroes.



Figure 2.3: Rejection Frequencies for Normal(0,1) with varying π_2 , ($\pi_1 = \tau = 0$)

Figure 2.4: Rejection Frequencies for χ^2_3 with varying π_2 , ($\pi_1 = \tau = 0$)



Therefore, the simulations with the data generated with the two distributions demonstrate that when the instrument is strong, the two step estimation's approximation to APEs is as good as the alternative methods. Plus, although a weak instrument deteriorates its approximation, it makes much worse the approximations by the alternative linear models. In addition, the two step estimation's test for endogeneity show good size and better power than *LCF*'s.

Case 2: Endogeneity comes through *e* as well as *u*.

Now we consider a setting where D(e|v) is a heteroskedastic normal, in which w is not exogenous even though $\rho = 0$. For this case, we include two additional estimation methods treating that all covariates are exogenous: *Fmlogit* and *OLS*. We would like to examine how they work as $\rho = 0$.

When z_2 is a strong instrument,¹⁹ Table 2.10 shows that the estimates have bigger biases than those with the standard normal distribution, and that the larger ρ is, the larger their biases are.²⁰ Among the estimation methods, *LCF flexible* is the best with regard to bias. Yet, in considering the mean squared errors (MSEs) in Table 2.11, the two step estimation method and the other alternative methods also provide good approximations.²¹

Interestingly, as $\rho = 0$, the estimates of *Fmlogit* and *OLS* are quite similar to δ_g . So these estimates suggest that the dependence w and r due to the heteroskedasticity of e does not cause any distortion in estimating APEs.

With the strong degree of endogeneity ($\rho = 1$), although the biases in Table 2.12 are not monotonically decreasing as π_2 is larger, the MSEs in Table 2.13 are. Note that when a weak instrument is used, *LCF Flexible* provides more biased estimates than *Fmlogit* and *OLS*, and the MSEs of the methods taking the endogeneity into account are not smaller than them except the two step estimation method with $\pi_2 = 0.2$. The results also suggest that the flexible methods are not helpful at all.

¹⁹ The first stage F statistics and the first step Wald statistics in this case have similar summary statistics as in Table 2.4.

²⁰ The results with other values of ρ are available upon request.

²¹ When $\tau = -0.5$ and $(\pi_1, \pi_2) = (0.5, 1)$, the results are not much different from those as $\pi_2 = 1$ in Table 2.10 and 2.11.

ρ			1			0.5	
8		1	2	3	1	2	3
δ_g	Mean	-0.106	0.053	0.053	-0.107	0.053	0.053
Two Step	Mean	-0.115	0.058	0.057	-0.112	0.056	0.056
-	SD	(0.031)	(0.018)	(0.017)	(0.028)	(0.016)	(0.015)
Two Step	Mean	-0.112	0.056	0.056	-0.115	0.058	0.057
Flexible	SD	(0.038)	(0.021)	(0.021)	(0.033)	(0.019)	(0.019)
Two Step	Mean	-0.114	0.057	0.057	-0.111	0.055	0.055
Plug-in	SD	(0.032)	(0.018)	(0.017)	(0.027)	(0.016)	(0.015)
Fmlogit	Mean	-0.224	0.112	0.112	-0.167	0.083	0.083
	SD	(0.018)	(0.010)	(0.010)	(0.016)	(0.009)	(0.009)
LCF	Mean	-0.115	0.058	0.058	-0.111	0.056	0.055
	SD	(0.032)	(0.018)	(0.018)	(0.029)	(0.016)	(0.016)
LCF	Mean	-0.106	0.053	0.053	-0.107	0.053	0.053
Flexible	SD	(0.039)	(0.021)	(0.021)	(0.034)	(0.019)	(0.019)
2SLS	Mean	-0.113	0.057	0.056	-0.110	0.055	0.055
	SD	(0.032)	(0.018)	(0.018)	(0.028)	(0.016)	(0.016)
LIV	Mean	-0.114	0.057	0.057	-0.110	0.055	0.0554
$(\text{IV}=\widehat{\Phi}(\cdot))$	SD	(0.032)	(0.018)	(0.017)	(0.028)	(0.016)	(0.015)
Linear	Mean	-0.114	0.057	0.057	-0.110	0.055	0.055
Plug-in	SD	(0.032)	(0.018)	(0.017)	(0.028)	(0.016)	(0.015)
OLS	Mean	-0.224	0.112	0.112	-0.167	0.083	0.083
	SD	(0.018)	(0.010)	(0.010)	(0.016)	(0.009)	(0.009)
		1 -					

Table 2.10: APEs with $r = \rho u + e$ and $\pi_2 = 1$ in Case 2

$${}^{1.} e|_v \sim Normal\left(0, 1 + rac{1}{2}v^2
ight)$$

 ${}^{2.} \pi_1 = au = 0.$

^{2.}
$$\pi_1 = \tau = 0$$
.

Table 2.10: (cont'd)

ρ			0.1			0	
8		1	2	3	1	2	3
δ_g	Mean	-0.107	0.054	0.054	-0.107	0.054	0.054
Two Step	Mean	-0.109	0.055	0.054	-0.108	0.054	0.054
	SD	(0.026)	(0.015)	(0.015)	(0.026)	(0.015)	(0.015)
Two Step	Mean	-0.115	0.058	0.057	-0.115	0.057	0.057
Flexible	SD	(0.031)	(0.018)	(0.018)	(0.031)	(0.017)	(0.018)
Two Step	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
Plug-in	SD	(0.025)	(0.015)	(0.015)	(0.025)	(0.015)	(0.014)
Fmlogit	Mean	-0.119	0.059	0.059	-0.107	0.053	0.053
	SD	(0.015)	(0.009)	(0.009)	(0.015)	(0.009)	(0.009)
LCF	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
	SD	(0.027)	(0.015)	(0.015)	(0.027)	(0.015)	(0.015)
LCF	Mean	-0.107	0.054	0.053	-0.107	0.054	0.054
Flexible	SD	(0.031)	(0.018)	(0.018)	(0.031)	(0.018)	(0.018)
2SLS	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
	SD	(0.026)	(0.015)	(0.015)	(0.026)	(0.015)	(0.015)
LIV	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
$(IV=\widehat{\Phi}(\cdot))$	SD	(0.026)	(0.015)	(0.015)	(0.025)	(0.015)	(0.015)
Linear	Mean	-0.108	0.054	0.054	-0.108	0.054	0.054
Plug-in	SD	(0.026)	(0.015)	(0.015)	(0.026)	(0.015)	(0.015)
OLS	Mean	-0.119	0.059	0.059	-0.107	0.053	0.053
	SD	(0.016)	(0.009)	(0.009)	(0.016)	(0.009)	(0.009)

ρ		1			0.5			0.1			0	
8	1	2	3	1	2	3	1	2	3	1	2	3
Two Step	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
Two Step Flexible	0.002	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
Two Step Plug-in	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
Fmlogit	0.014	0.004	0.004	0.004	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
LCF	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
LCF Flexible	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
2SLS	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
LIV (IV= $\widehat{\Phi}(\cdot)$)	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
Linear Plug-in	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
OLS	0.014	0.004	0.004	0.004	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000

Table 2.11: MSEs of APE estimates with $r = \rho u + e$ and $\pi_2 = 1$ in Case 2

^{1.} $e|_v \sim Normal\left(0, 1 + \frac{1}{2}v^2\right)$ ^{2.} MSEs are calculated from Table 2.10.

π_2			0.1			0.2			0.5	
8		1	2	3	1	2	3	1	2	3
δ_g	Mean	-0.102	0.051	0.051	-0.102	0.051	0.051	-0.103	0.051	0.051
Two Step	Mean	-0.175	0.093	0.082	-0.109	0.057	0.052	-0.125	0.063	0.062
_	SD	(0.324)	(0.192)	(0.192)	(0.212)	(0.116)	(0.117)	(0.081)	(0.044)	(0.043)
Two Step	Mean	-0.198	0.097	0.101	-0.160	0.081	0.079	-0.120	0.060	0.060
Flexible	SD	(0.440)	(0.286)	(0.293)	(0.378)	(0.223)	(0.227)	(0.149)	(0.079)	(0.081)
Two Step	Mean	-0.138	0.073	0.064	-0.090	0.047	0.042	-0.126	0.063	0.063
Plug-in	SD	(0.483)	(0.280)	(0.280)	(0.271)	(0.146)	(0.148)	(0.080)	(0.043)	(0.042)
Fmlogit	Mean	-0.333	0.167	0.167	-0.330	0.165	0.165	-0.306	0.153	0.153
	SD	(0.020)	(0.010)	(0.011)	(0.020)	(0.010)	(0.011)	(0.020)	(0.010)	(0.011)
LCF	Mean	-0.205	0.112	0.093	-0.103	0.056	0.047	-0.128	0.064	0.064
	SD	(2.644)	(1.284)	(1.509)	(0.782)	(0.469)	(0.336)	(0.082)	(0.044)	(0.043)
LCF	Mean	0.884	-0.446	-0.438	0.111	-0.060	-0.051	-0.094	0.047	0.047
Flexible	SD	(11.653)	(7.151)	(5.858)	(3.230)	(2.032)	(1.919)	(0.156)	(0.082)	(0.085)
2SLS	Mean	-0.194	0.108	0.085	-0.089	0.051	0.038	-0.126	0.063	0.063
	SD	(2.241)	(1.314)	(1.189)	(0.575)	(0.398)	(0.232)	(0.081)	(0.044)	(0.043)
LIV	Mean	-0.221	0.142	0.079	-0.079	0.043	0.037	-0.127	0.064	0.063
$(\text{IV}=\widehat{\Phi}(\cdot))$	SD	(2.618)	(1.842)	(1.314)	(0.397)	(0.207)	(0.219)	(0.081)	(0.044)	(0.043)
Linear	Mean	-0.197	0.112	0.086	-0.091	0.052	0.039	-0.127	0.064	0.063
Plug-in	SD	(2.213)	(1.268)	(1.203)	(0.628)	(0.430)	(0.246)	(0.081)	(0.044)	(0.043)
OLS	Mean	-0.333	0.167	0.167	-0.330	0.165	0.165	-0.306	0.153	0.153
	SD	(0.020)	(0.010)	(0.011)	(0.020)	(0.010)	(0.011)	(0.020)	(0.010)	(0.011)
Replicat	ions	ç	997/1000		1	1000/100	0	1	1000/100)

Table 2.12: APEs with $w = 1[z_2\pi_2 + u > 0]$ and $\rho = 1$ in Case 2

^{1.} $e|_{v} \sim Normal\left(0, 1 + \frac{1}{2}v^{2}\right)$ ^{2.} $\pi_{1} = \tau = 0.$

².
$$\pi_1 = \tau = 0$$
.

π_2		0.1			0.2			0.5	
8	1	2	3	1	2	3	1	2	3
Two Step	0.110	0.039	0.038	0.045	0.014	0.014	0.007	0.002	0.002
Two Step Flexible	0.203	0.084	0.088	0.146	0.051	0.052	0.023	0.006	0.007
Two Step Plug-in	0.234	0.079	0.079	0.073	0.021	0.022	0.007	0.002	0.002
Fmlogit	0.054	0.013	0.014	0.052	0.013	0.013	0.042	0.010	0.010
LCF	7.004	1.652	2.278	0.611	0.220	0.113	0.007	0.002	0.002
LCF Flexible	136.772	51.391	34.559	10.480	4.141	3.695	0.024	0.007	0.007
2SLS	5.031	1.730	1.415	0.330	0.159	0.054	0.007	0.002	0.002
LIV (IV= $\widehat{\Phi}(\cdot)$)	6.867	3.403	1.727	0.158	0.043	0.048	0.007	0.002	0.002
Linear Plug-in	4.908	1.611	1.449	0.395	0.185	0.060	0.007	0.002	0.002
OLS	0.054	0.013	0.014	0.053	0.013	0.013	0.042	0.010	0.010

Table 2.13: MSEs of APE estimates with $w = 1[z_2\pi_2 + u > 0]$ and $\rho = 1$ in Case 2

 $\frac{1}{2}e|_{v} \sim Normal\left(0, 1 + \frac{1}{2}v^{2}\right)$

^{2.}
$$\pi_1 = \tau = 0$$

^{3.} MSEs are calculated from Table 2.12.

The rejection frequencies in Table 2.14 and 2.15 tell that the tests by the two step estimation method and *LCF* cannot detect the endogeneity due to the heteroskedasticity of e; as $\rho = 0$, their rejection frequencies are still close to 0.05. Yet, those by their flexible methods can detect it. The form of heteroskedasticity and the flexible methods contain v^2 and \hat{v}^2 , respectively. We suspect that it might be related with that the flexible methods yield higher rejection frequencies.

As $\rho > 0$, the patters in the rejection frequencies are similar as those with the standard normal distribution although the rejection frequencies themselves are generally smaller. It is more obviously shown in Figure 2.5 and 2.6.²²

Overall, we find that the results with the heteroskedastic normal distribution are pretty similar to those of the standard normal distribution regarding approximations to APEs and a test of endogeneity. The two additional estimation methods provide evidence that the endogeneity caused only by heteroskedasticity does not matter to obtain an approximation to APEs. Plus, although including the higher polynomial \hat{v} ensure the two step estimation method and *LCF* to have tests detecting this kind of endogeneity, it does not helps their approximations to APEs.

In summary, the simulations demonstrate that the two step estimation method with a misspecified conditional mean works well as an approximation if the instrument is strong. Even with a weak instrument, it yields a better approximation than the alternative methods. In addition, its test for endogeneity has about the correct size and better power than *LCF*'s.

²² The graphs are average rejection frequencies for the two choices at each value of ρ and π_2 in Table 2.14 and 2.15, respectively.

Table 2.14: Rejection Frequencies for $\alpha = 0.05$ test in Case 2 with varying ρ

	ρ	(0		0.1		0.25		0.5		75	0.9		1	
D(e v)	8	2	3	2	3	2	3	2	3	2	3	2	3	2	3
Normal	Two Step	0.08	0.07	0.08	0.08	0.20	0.20	0.53	0.53	0.84	0.85	0.94	0.94	0.96	0.96
$(0, 1 + \frac{1}{2}v^2)$	LCF	0.07	0.07	0.07	0.07	0.16	0.17	0.46	0.47	0.77	0.77	0.90	0.90	0.94	0.94
	Two Step flexible	0.43	0.43	0.45	0.45	0.50	0.50	0.70	0.70	0.88	0.89	0.95	0.94	0.97	0.97
	LCF flexible	0.31	0.30	0.31	0.32	0.39	0.37	0.58	0.59	0.80	0.81	0.89	0.89	0.93	0.92

^{1.} $\pi_2 = 1$, $\pi_1 = \tau = 0$. ^{2.} The null hypotheses for the flexible models are that the coefficients of \hat{v} , \hat{v}^2 , and \hat{v}^3 are zeroes.

	τ	-0	-0.5		0								
	π_1	0	.5	0									
	π_2		1	1		0.5		0.2		0.	.1		
D(e v)	8	2	3	2	3	2	3	2	3	2	3		
Normal	Two Step	0.94	0.94	0.96	0.96	0.69	0.70	0.21	0.21	0.09	0.10		
$(0, 1 + \frac{1}{2}v^2)$	LCF	0.91	0.91	0.94	0.94	0.62	0.65	0.18	0.18	0.09	0.09		
, ,	Two Step flexible	0.94	0.94	0.97	0.97	0.66	0.66	0.18	0.18	0.10	0.10		
	LCF flexible	0.90	0.92	0.93	0.92	0.63	0.62	0.39	0.39	0.51	0.52		
	Replication	10	00	10	00	10	00	10	00	99	97		

Table 2.15: Rejection Frequencies for $\alpha = 0.05$ test in Case 2 with varying π_2

^{1.} $\rho = 1$ ^{2.} The null hypotheses for the flexible models are that the coefficients of \hat{v} , \hat{v}^2 , and \hat{v}^3 are zeroes.



Figure 2.5: Rejection Frequencies for Normal $\left(0, 1 + \frac{1}{2}v^2\right)$ with varying π_2 , $(\pi_1 = \tau = 0)$

Figure 2.6: Rejection Frequencies for Normal $\left(0, 1 + \frac{1}{2}v^2\right)$ with varying π_2 , $(\pi_1 = \tau = 0)$



2.4 CONCLUSION

This chapter studies a two step estimation method for multiple fractional dependent variables especially when there is a binary endogenous explanatory variable. By employing a control function approach suggested by Wooldridge (2014), this method is an extension of a two step estimation method developed in Chapter 1 where the endogenous explanatory variable is continuous.

With the assumption that the conditional mean, conditional on only observed variables and the generalized residual, is multinomial logit, it applies a QMLE to obtain consistent estimator of the parameters in the conditional mean. Although it is not able to estimate the mean parameter in the structural conditional mean, it is able to provide a consistent estimator of APE, which is often more interesting.

By conducting Monte Carlo simulations, this chapter provides evidence that even when the conditional mean is misspecified, the two step estimation method yields a decent approximation to average partial effect. Its approximation is as good as or often better than the alternative methods including a linear control function approach, the standard two stage least squares, and plug-in methods. This results tell a consistent story with Chapter 1. In addition, the simulations demonstrate that the two step estimation method's test for endogeneity outperforms the linear control function approach in power although both of the two methods have approximately the correct size.

REFERENCES

REFERENCES

Blundell, R., and J. L. Powell. 2003. "Endogeneity in Nonparametric and Semiparametric Regression Models." In *Advances in Economics and Econometrics*. eds. by L. P. H. Dewatripont, Mathias, and S. J. Turnovsky, 2: Cambridge: Cambridge University Press, , 1st edition 312–357.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon. 1987. "Generalised residuals." *Journal of Econometrics*, 34(1-2): 5–32.

——, and A. Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica*, 52(3): 681–700.

Mullahy, J. 2010. "Multivariate Fractional Regression Estimation of Econometric Share Models." NBER Working Papers 16354, National Bureau of Economic Research, Inc.

Nam, S. 2014. "Multiple fractional response variables with continuous endogenous explanatory variables." June.

Sivakumar, A., and C. Bhat. 2002. "Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis." *Transportation Research Record*, 1790(1): 80–88.

Staiger, D., and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65(3): 557–586.

Terza, J. V., A. Basu, and P. J. Rathouz. 2008. "Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling." *Journal of Health Economics*, 27(3): 531–543.

Wooldridge, J. M. 2005. "Unobserved Heterogeneity and Estimation of Average Partial Effects." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. eds. by D. W. K. Andrews, and J. H. Stock: Cambridge University Press, , Chap. 3.

------ . 2010. Econometric Analysis of Cross Section and Panel Data.: MIT Press.

Wooldridge, J. M. 2014. "Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables." *Journal of Econometrics*.

CHAPTER 3

ON COMPUTING AVERAGE PARTIAL EFFECTS IN MODELS WITH ENDOGENEITY OR HETEROGENEITY

3.1 INTRODUCTION

It is now widely recognized that magnitudes of partial effects are important for determining the importance of policy interventions and for understanding the strengths of relationships posited by economic theory. In models where unobservables are assumed to be independent of observed covariates (or, in some cases, a weaker assumption, such as mean independence), there is little controversy about how one should compute partial effects – at least once one has decided whether the conditional mean or some other feature of a conditional distribution is the focus. Assuming, as we do in this chapter, that the conditional mean is the focus, one typically computes partial derivatives or discrete changes of a conditional mean function with respect to the explanatory variables of interest. To obtain a single number summarizing the effect of a covariate, one often averages the partial effects across the distribution of the covariates. This average leads to the notion of an "average partial effect" (or APE), and APEs are typically straightforward to estimate in parametric, semiparametric, and even nonparametric models with exogenous explanatory variables.¹ Other possibilities for summarizing the effect of a covariate is to insert specific values of the covariates, such as means or medians, into the estimated

¹We prefer the name "partial effect" to "marginal effect" because "marginal" connotes a small change. Average partial effects can be computed using derivatives or discrete changes, and the discrete changes can be of any magnitude.

partial effects, but these seems less desirable and is less popular.

When unobservables are correlated with one or more covariates – so we have some form of "endogeneity" – it is not clear how one should summarize partial effects. Blundell and Powell (2003) propose the notion of an "average structural function" (ASF). The ASF is defined as a function of covariates after the unobservables have been averaged out. More precisely, suppose a response *Y* is determined as $Y = G(\mathbf{X}, \mathbf{U})$ for observed covariates \mathbf{X} and unobservables \mathbf{U} . We obtain the ASF as a function of \mathbf{x} by inserting \mathbf{x} into $G(\cdot, \cdot)$ and then averaging across the unobservables. More precisely, $ASF(\mathbf{x}) \equiv E_{\mathbf{U}}[G(\mathbf{x}, \mathbf{U})]$. Partial effects are then obtained by taking partial derivatives or differences of $ASF(\mathbf{x})$. In general, the partial effects defined in this way depend on \mathbf{x} . Blundell and Powell (2003) showed how to estimate the ASF very generally in models with endogenous explanatory variables, provided, of course, one has sufficient instrumental variables. Wooldridge (2005b) focused on the partial effects, which in the continuous case are defined as $\partial ASF(\mathbf{x})/\partial x_i$, in a similar setting, but with parametric models.²

Part of the appeal of the ASF is that its definition is the same regardless of whether **U** and **X** are dependent. Because the ASF is a function of **x**, one can see how partial effects on the ASF change as elements of **x** change, and the differences across different values of **x** can be of significant interest. Even so, one often wants to compare estimates from nonlinear models with estimates from simple linear models – which are often estimated by two stage least squares or, in the case of panel data, standard fixed effects methods. The question is: How should one summarize the partial effects of observed covariates in nonlinear models to make them comparable to linear estimates? There are two possibilities. First, we can obtain the partial effects from the ASF and then average across the distribution of the observed covariates, **X**. A second possibility is to compute partial effects with respect to **X** from $Y = G(\mathbf{X}, \mathbf{U})$ and then average across (**X**, **U**). As we will see in Section 3.3, these two methods are not generally the same, and they can actually be

²In Wooldridge (2010)and the first edition, the notion of an APE is used regularly for models with endogeneity or correlated random effects.

very different.

Wooldridge (2010) discusses APEs based on the average structural function and those based directly on a conditional mean specification such as $E(Y|\mathbf{X}, \mathbf{U}) = G(\mathbf{X}, \mathbf{U})$. But he makes no systematic attempt to compare these APEs, and does not even mention that they can be different. This chapter has two goals. The first is to clarify the relationship between the two kinds of partial effects and demonstrate that they generally differ. We will not resolve which partial effect is "better" because it is mainly a matter of test. Second, we show that the kind of control function/correlated random effects approaches discussed in Blundell and Powell (2003), Altonji and Matzkin (2005), and Wooldridge (2005b, 2010) can be used to consistently estimate both types of partial effects.

The rest of the chapter is organized as follows. Section 3.2 defines the average structural function, slightly extending the definition in Blundell and Powell (2003). We then show that the key Blundell and Powell result about identifying the ASF when suitable "proxies" are available carries through with the more general definition. We also note that, assuming derivatives can be passed under integrals, the ASF identifies average partial effects.

Section 3.3 offers a different way to define a single average partial effect and shows that it is generally different from a definition based on averaging the observed covariates out of the average structural function. We also show that if the heterogeneity and covariates are assumed to be independent, the two ways of computing APEs coincide.

Section 3.4 illustrates how to compute the two types of partial effects in two empirical examples using Michigan Educational Assessment Program data, and Section 3.5 summarizes and concludes.

3.2 THE AVERAGE STRUCTURAL FUNCTION AND AV-ERAGE PARTIAL EFFECTS

In what follows, it is helpful to use a notation that clearly distinguishes random vectors from particular outcomes of those vectors. We use the traditional convention from probability that upper case letters are random variables or vectors and the lower case counterparts are specific possible values.

We are interested in the conditional mean of a response variable, *Y*, conditional on a vector of observed covariates, **X**, and a vector of unosbervables, **U**:

$$E(Y|\mathbf{X}, \mathbf{U}) = G(\mathbf{X}, \mathbf{U}), \tag{3.1}$$

where $(Y, \mathbf{X}, \mathbf{U})$ has a joint distribution in a population. We can also write

$$E(Y|\mathbf{X} = \mathbf{x}, \mathbf{U} = \mathbf{u}) = G(\mathbf{x}, \mathbf{u}).$$
(3.2)

This setup subsumes that in Blundell and Powell (2003), who assume that Y is a deterministic function of (**X**, **U**). Because conditional probabilities can be written as conditional means we cover partial effects for probabilities as a special case.

In many applications, $G(\cdot, \mathbf{u})$ is continuously differentiable on the the support of **X**, \mathcal{X} , which we would then assume to be an open set. But we are also interested in cases where $G(\cdot, \cdot)$ is not differentiable in either argument. For example, if Y is binary and we assume

$$Y = 1[\alpha + \mathbf{X}\boldsymbol{\beta} + U > 0]$$

for a $K \times 1$ vector β , $G(\cdot, \cdot)$ is not differentiable. In this case we define partial effects as changes. For example, in moving the K^{th} variable from x_K^0 to x_K^1 , holding the other

elements of \mathbf{x} , say $\mathbf{x}_{(K)}$ fixed, the partial effect is

$$1[\alpha + \mathbf{x}_{(K)}\boldsymbol{\beta}_{(K)} + \beta_{K}x_{K}^{1} + u > 0] - 1[\alpha + \mathbf{x}_{(K)}\boldsymbol{\beta}_{(K)} + \beta_{K}x_{K}^{0} + u > 0]$$

More generally, the effect of changing **x** from \mathbf{x}_0 to \mathbf{x}_1 is

$$G(\mathbf{x}_1, \mathbf{u}) - G(\mathbf{x}_0, \mathbf{u}).$$

When $G(\cdot, \mathbf{u})$ is continuously differentiable in x_j we can define the partial effect with respect to x_j as

$$g_j(\mathbf{x}, \mathbf{u}) = \frac{\partial G(\mathbf{x}, \mathbf{u})}{\partial x_j}.$$
(3.3)

Wooldridge (2005b, 2010) defines the average partial effect as

$$APE_{j}(\mathbf{x}) = E_{\mathbf{U}}[g_{j}(\mathbf{x}, \mathbf{U})], \qquad (3.4)$$

so that the unobservables are averaged out. It is important to see that the APE is obtained by inserting a fixed value for **x** and then averaging across the *unconditional* distribution of **U**. We are not using the conditional distribution, $D(\mathbf{U}|\mathbf{X})$, in the averaging, and we are not restricting this conditional distribution. In the special case that **U** and **X** are independent, $D(\mathbf{U}|\mathbf{X}) = D(\mathbf{U})$ and then the distinction is irrelevant.

As discussed in Wooldridge (2005b, 2010, Section 2.2.5), $APE_j(\mathbf{x})$ is closely related to the notion of an average structural function (Blundell and Powell, 2003):

$$ASF(\mathbf{x}) = E_{\mathbf{U}}[G(\mathbf{x}, \mathbf{U})]. \tag{3.5}$$

Actually, this definition is somewhat more general than that used by Blundell and Powell, who effectively write $Y = G(\mathbf{X}, \mathbf{U})$. The more general definition is important in situations where Y is discrete and we model endogeneity as correlation between one or more omitted variables, in U, and one or more observed covariates, in X.

The ASF is a very useful function because one can take derivatives or changes with respect to elements of **x**. Assuming that the partial derivative can be taken through the sum or integral defining $E_{\mathbf{U}}[G(\mathbf{x}, \mathbf{U})]$ – see, for example, Bartle (1966) for general conditions – we have

$$APE_j(\mathbf{x}) = \frac{\partial ASF(\mathbf{x})}{\partial x_j}.$$
 (3.6)

As mentioned earlier, the definition of the ASF takes no stand on $D(\mathbf{U}|\mathbf{X})$. However, if **U** and **X** are independent then the ASF is the same as the conditional expectation. This follows immediately from the the law of iterated expectations and the general Lebesgue representation of conditional expectation. Let

$$E(Y|\mathbf{X} = \mathbf{x}) = E[E(Y|\mathbf{X}, \mathbf{U})|\mathbf{X} = \mathbf{x}] = \int G(\mathbf{x}, \mathbf{u})Q(d\mathbf{u}|\mathbf{x})$$
$$= \int G(\mathbf{x}, \mathbf{u})Q(d\mathbf{u}) = E_{\mathbf{U}}[G(\mathbf{x}, \mathbf{U})] \equiv ASF(\mathbf{x}), \qquad (3.7)$$

where $Q(\cdot)$ is an appropriate σ -finite measure.

When **U** and **X** are dependent, we generally cannot obtain interesting partial effects by estimating $E(Y|\mathbf{X})$. There are exceptions, of course. For example, if $Y = \alpha + \mathbf{X}\mathbf{fi} + U$ with E(U) = 0, $E(\mathbf{X}'U) = \mathbf{0}$, but $E(U|\mathbf{X}) \neq E(U)$, OLS using a random sample is generally consistent for (α, β) , which indexes the ASF: $ASF(\mathbf{x}) = \alpha + \mathbf{x}\mathbf{fi}$. However, simple extensions with correlated random slopes do permit OLS to consistently estimate the ASF. And in nonlinear models, directly estimating $E(Y|\mathbf{X})$ rarely leads to interesting quantities unless **U** and **X** are independent.

Blundell and Powell (2003) and Wooldridge (2005b) show how to identify the ASF when "proxy" variables for **U**, say **V**, are available. Sometimes we assume that we observe suitable proxies, such as standardized test scores to proxy for cognitive ability. In the context of BP (2003), **V** is a vector of reduced form errors for endogenous elements of **X**. When **V** is a vector of reduced form errors we need exogenous variables from outside

the equation to serve as instruments. Suitable proxies are also available in a panel data context, where **V** can be a vector of functions of a time series of covariates { $X_1, X_2, ..., X_T$ }. A leading case is $\mathbf{V} = T^{-1} \sum_{t=1}^{T} \mathbf{X}_t$. See, for example, Altonji and Matzkin (2005) and Wooldridge (2005a, 2010).

Two key restrictions on **V** suffice to identify average partial effects. The first is that **V** is redundant in the "structural" conditional expectation. The second is that **V** is a good enough proxy for **U** so that **U** and **X** are independent conditional on **V**. Formally, we state these assumptions as

$$E(Y|\mathbf{X}, \mathbf{U}, \mathbf{V}) = E(Y|\mathbf{X}, \mathbf{U}) \text{ (redundancy of } \mathbf{V})$$
(3.8)

$$D(\mathbf{U}|\mathbf{X}, \mathbf{V}) = D(\mathbf{U}|\mathbf{V})$$
 (conditional independence) (3.9)

Sometimes assumption (3.8) is called a "conditional mean independence" assumption, because the mean of Y is independent of **V** once we also condition on (**X**, **U**).

Under assumptions (3.8) and (3.9), we have an important identification result, which was discovered independently in different settings by Blundell and Powell (2003), Altonji and Matzkin (2005), and Wooldridge (2002, 2005b). In what follows we assume that conditional means exist, along with standard moment conditions.

Proposition 1: Let (Y, X, U, V) be a random vector such that (3.8) and (3.9) hold. Define

$$H(\mathbf{x}, \mathbf{v}) \equiv E(Y | \mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}).$$
(3.10)

Then

$$ASF(\mathbf{x}) = E_{\mathbf{V}}[H(\mathbf{x}, \mathbf{V})]$$
(3.11)

Proof: By the law of iterated expectations and the redundancy condition (3.8),

$$E(Y|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}) = E[E(Y|\mathbf{X} = \mathbf{x}, \mathbf{U}, \mathbf{V} = \mathbf{v})|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}]$$
$$= E[E(Y|\mathbf{X} = \mathbf{x}, \mathbf{U})|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}]$$
$$= E[G(\mathbf{x}, \mathbf{U})|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}]$$

Next, by conditional independence (3.9),

$$E[G(\mathbf{x}, \mathbf{U})|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}] = E[G(\mathbf{x}, \mathbf{U})|\mathbf{V} = \mathbf{v}],$$

and so we have established the key relationship

$$H(\mathbf{x}, \mathbf{v}) = E[G(\mathbf{x}, \mathbf{U}) | \mathbf{V} = \mathbf{v}].$$
(3.12)

Now integrate (in the measure theoretic sense) both sides with respect to the distribution of **V** and use iterated expectations on the right:

$$E_{\mathbf{V}}[H(\mathbf{x},\mathbf{V})] = E_{\mathbf{V}} \{ E[G(\mathbf{x},\mathbf{U})|\mathbf{V}] \} = E_{\mathbf{U}}[G(\mathbf{x},\mathbf{U})] = ASF(\mathbf{x}). \square$$

We can use Proposition 1 to obtain useful formulas for partial effects based on discrete changes. The "structural" quantity of interest is

$$ASF(\mathbf{x}_1) - ASF(\mathbf{x}_0) = E_{\mathbf{U}}[G(\mathbf{x}_1, \mathbf{U}) - G(\mathbf{x}_0, \mathbf{U})]$$

and Proposition 1 shows that this is the same as

$$E_{\mathbf{V}}[H(\mathbf{x}_1,\mathbf{V})-H(\mathbf{x}_0,\mathbf{V})].$$

An important special case, suppose X_K is a binary treatment indicator – perhaps indicat-

ing participation in a program. Generally, we can estimate the average treatment effect of the program with the observables set at particular values. In particular, under the conditions of Proposition 1,

$$E_{\mathbf{V}}[H(\mathbf{x}_{(K)}, 1, \mathbf{V}) - H(\mathbf{x}_{(K)}, 0, \mathbf{V})] = APE_{K}(\mathbf{x}_{(K)}) = E_{\mathbf{U}}[G(\mathbf{x}_{(K)}, 1, \mathbf{U}) - G(\mathbf{x}_{(K)}, 0, \mathbf{U})],$$
(3.13)

where $\mathbf{x}_{(K)}$ denotes \mathbf{x} without x_K .

For partial effects defined as derivatives, we have the following.

Proposition 2: Under the same assumptions in Proposition 1, assume in addition that $G(\cdot, \mathbf{u})$ is continuously differentiable and the partial derivative with to x_j can be passed through the integrals defining $E_{\mathbf{V}}[H(\mathbf{x}, \mathbf{V})]$ and $E_{\mathbf{U}}[G(\mathbf{x}, \mathbf{U})]$ then

$$E_{\mathbf{V}}\left[\frac{\partial H(\mathbf{x},\mathbf{V})}{\partial x_j}\right] = E_{\mathbf{U}}\left[\frac{\partial G(\mathbf{x},\mathbf{U})}{\partial x_j}\right]$$
(3.14)

and so the APEs can be gotten from $E(Y|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v})$ by taking derivatives (or changes) with respect to x_i and then averaging out \mathbf{V} . \Box

The conclusion in equation (3.11) is very powerful, especially considering that there are general conditions under which $H(\mathbf{x}, \mathbf{v})$ is nonparametrically identified. Blundell and Powell (2003) study the case where some elements of **X** are correlated with **U**, but exogenous variables **Z** are available such that endogenous variables **X**₂ can be represented as

$$\mathbf{X}_2 = \mathbf{F}(\mathbf{Z}) + \mathbf{V} \tag{3.15}$$

where (\mathbf{U}, \mathbf{V}) is independent of \mathbf{Z} . Provided there are enough elements \mathbf{Z}_2 , where $\mathbf{Z} = (\mathbf{X}_1, \mathbf{Z}_2)$, the function $H(\mathbf{x}, \mathbf{v})$ is nonparametrically identified. Of course, one can use parametric models or semiparametric models to estimate the functions $\mathbf{F}(\mathbf{z})$ and $H(\mathbf{x}, \mathbf{v})$ – either as the true models or as convenient approximations.

We can turn the population formulas into estimators very generally. Given a random

sample { $(Y_i, \mathbf{X}_i, \mathbf{V}_i) : i = 1, ..., N$ } and a consistent estimator $\hat{H}(\mathbf{x}, \mathbf{v})$ of $E(Y|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v})$, a generally consistent estimator of $ASF(\mathbf{x})$ is

$$\widehat{ASF}(\mathbf{x}) = N^{-1} \sum_{i=1}^{N} \widehat{H}(\mathbf{x}, \mathbf{V}_i), \qquad (3.16)$$

and the previous analysis shows that we can take derivatives or changes with respect to elements in **x** to estimate average partial effects. In some cases, including in the Blundell and Powell (2003) setup, V_i must be replaced with \hat{V}_i , which depends on parameters or functions that are consistently estimated in a first stage.

In many studies one wants to report a single number that measures the effect of, say, x_j on $ASF(\mathbf{x})$. We might do this by evaluating $APE_j(\mathbf{x})$ at a central values of the elements of \mathbf{x} , such as the means or medians. In the spirit of the average treatment effect literature, it is probably more appealing to average $APE_j(\mathbf{X})$ across the distribution of \mathbf{X} :

$$\delta_j = E_{\mathbf{X}}[APE_j(\mathbf{X})]. \tag{3.17}$$

A consistent estimate of δ_j is immediate under standard regularity conditions, provided $\widehat{APE}_j(\mathbf{x})$ is consistent for each **x**:

$$\hat{\delta}_j = N^{-1} \sum_{i=1}^N \widehat{APE}_j(\mathbf{X}_i).$$
(3.18)

It turns out that the quantity in (3.18) is not the most commonly reported in empirical studies, and it is a bit cumbersome to compute. Plus, obtaining a standard error via analytical methods is a bit tricky. In the next section we discuss APEs where we average jointly across the distribution of (\mathbf{X}, \mathbf{U}) .

3.3 APES ACROSS THE ENTIRE POPULATION

As in the previous section, we assume interest centers on $E(Y|\mathbf{X} = \mathbf{x}, \mathbf{U} = \mathbf{u}) = G(\mathbf{x}, \mathbf{u})$, but now we wish to average the partial effects across the joint distribution of the observables and unobservables. For example, in moving x_K from x_K^0 to x_K^1 , the average partial effect across the entire population is

$$E_{[\mathbf{X}_{(K)},\mathbf{U}]}\left[G(\mathbf{X}_{(K)},x_{K}^{1},\mathbf{U})-G(\mathbf{X}_{(K)},x_{K}^{0},\mathbf{U})\right].$$

For a continuous variable X_i we may define the partial effect as

$$g_j(\mathbf{x}, \mathbf{u}) = \frac{\partial G(\mathbf{x}, \mathbf{u})}{\partial x_j}$$
(3.19)

and then the parameter of interest is

$$\eta_j = E_{(\mathbf{X},\mathbf{U})} \left[g_j(\mathbf{X},\mathbf{U}) \right].$$
(3.20)

Notice that **X** and **U** are treated symmetrically in equation (3.20), as in the treatment effects literature [for example, Imbens and Wooldridge (2009)]. In studying identification of (3.20) it makes no sense to start with the ASF because $E_{(\mathbf{X},\mathbf{U})} [G(\mathbf{X},\mathbf{U})] = E(Y)$ by iterated expectations, and so the expected value of the ASF with respect (\mathbf{X},\mathbf{U}) conveys no useful information about how X_i affects Y.

Focusing for now on the the case where X_j is continuous, the partial effect defined by (3.20) generally differs from δ_j in (3.17). In other words,

$$E_{\mathbf{X}}\left[\frac{\partial ASF(\mathbf{X})}{\partial x_{j}}\right] \neq E_{(\mathbf{X},\mathbf{U})}\left[g_{j}(\mathbf{X},\mathbf{U})\right].$$
(3.21)

To see why, it is helpful to consider a simple example, where *X* and *U* are both scalars.

Assume the conditional mean function is

$$E(Y|X,U) = \beta_0 + \beta_1 X + \beta_2 U + \beta_3 X^2 U, \qquad (3.22)$$

where E(U) = 0. Then

$$ASF(x) = E_U(\beta_0 + \beta_1 x + \beta_2 U + \beta_3 x^2 U) = \beta_0 + \beta_1 x$$

and so

$$\frac{\partial ASF(x)}{\partial x} = \beta_1. \tag{3.23}$$

No further averaging is needed to obtain a single effect because APE(x) is constant.

By contrast, the partial effect of *x* on E(Y|X = x, U = u) is

$$PE(x,u) = \frac{\partial E(Y|X=x, U=u)}{\partial x} = \beta_1 + 2\beta_3 x u$$
(3.24)

and so

$$\eta \equiv E_{(X,U)}[PE(X,U)] = \beta_1 + 2\beta_3 Cov(X,U).$$
(3.25)

Only if *U* and *X* are uncorrelated does (3.25) equal the partial derivative of the ASF. With substantial correlation between *X* and *U* or if β_3 is large in magnitude, the difference between (3.25) and (3.23) can be substantial.

When one wants to study how the partial effects differ across a range of values of the explanatory variables, the ASF seems to be natural quantity of interest. But it is less clear that (3.17) is the best definition of the average effect across the population. If we follow the approach from the average treatment effects literature then (3.20) is preferred.

As in the previous section, we first state a result that can be applied to partial effects defined by differences.

Proposition 3: Under the same assumptions as Proposition 1, let x_K be a fixed value.

Then

$$E_{[\mathbf{X}_{(K)},\mathbf{V}]}\left[H(\mathbf{X}_{(K)},x_K,\mathbf{V})\right]=E_{[\mathbf{X}_{(K)},\mathbf{U}]}\left[G(\mathbf{X}_{(K)},x_K,\mathbf{U})\right].$$

Proof: From equation (3.12) we can write $H(\mathbf{x}_{(K)}, x_K, \mathbf{v}) = E[G(\mathbf{x}_{(K)}, x_K, \mathbf{U}) | \mathbf{V} = \mathbf{v}]$. Now use conditional independence again:

$$E[G(\mathbf{x}_{(K)}, x_K, \mathbf{U}) | \mathbf{V} = \mathbf{v}] = E[G(\mathbf{x}_{(K)}, x_K, \mathbf{U}) | \mathbf{X}_{(K)} = \mathbf{x}_{(K)}, \mathbf{V} = \mathbf{v}]$$

= $E[G(\mathbf{X}_{(K)}, x_K, \mathbf{U}) | \mathbf{X}_{(K)} = \mathbf{x}_{(K)}, \mathbf{V} = \mathbf{v}],$

which we can write in terms of random variables as

$$H(\mathbf{X}_{(K)}, x_K, \mathbf{V}) = E[G(\mathbf{X}_{(K)}, x_K, \mathbf{U}) | \mathbf{X}_{(K)}, \mathbf{V}].$$

The proof is finished by taking the expected value with respect to $(\mathbf{X}_{(K)}, \mathbf{V})$ and using iterated expectations.

For the continuous case, we have the following. Define

$$h_j(\mathbf{x}, \mathbf{v}) = \frac{\partial H(\mathbf{x}, \mathbf{v})}{\partial x_j},$$
(3.26)

the partial effect of $E(Y|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v})$ with respect to x_j .

Proposition 4: Define η_j as in equation (3.20). Under the same assumptions as Proposition 2,

$$\eta_j = E_{(\mathbf{X}, \mathbf{V})} \left[h_j(\mathbf{X}, \mathbf{V}) \right]. \tag{3.27}$$

Proof: From equation (3.12), $H(\mathbf{x}, \mathbf{v}) = E[G(\mathbf{x}, \mathbf{U})|\mathbf{V} = \mathbf{v}]$, and assuming the partial derivative with respect to x_i can be passed through the integrals,

$$h_j(\mathbf{x}, \mathbf{v}) = E[g_j(\mathbf{x}, \mathbf{U}) | \mathbf{V} = \mathbf{v}].$$

Now use the conditional independence assumption again:

$$h_j(\mathbf{x}, \mathbf{v}) = E[g_j(\mathbf{x}, \mathbf{U}) | \mathbf{V} = \mathbf{v}] = E[g_j(\mathbf{x}, \mathbf{U}) | \mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}]$$
$$= E[g_j(\mathbf{X}, \mathbf{U}) | \mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}],$$

which we can write in terms of random variables as

$$h_i(\mathbf{X}, \mathbf{V}) = E[g_i(\mathbf{X}, \mathbf{U}) | \mathbf{X}, \mathbf{V}].$$

Now use iterated expectations to get

$$E_{(\mathbf{X},\mathbf{V})}\left[h_j(\mathbf{X},\mathbf{V})\right] = E_{(\mathbf{X},\mathbf{U})}\left[g_j(\mathbf{X},\mathbf{U})\right]. \Box$$
(3.28)

The four propositions stated in this and the previous section show that, under the same set of "control function" assumptions, average partial effects obtained by averaging **X** out of the ASF and those obtained by averaging the partial effects of $E(Y|\mathbf{X}, \mathbf{U})$ are generally identified. For better or worse, these partial effects are not generally the same, and there is unlikely to be concensus on which measure is "best." The differences can be economically important. For example, in equation (3.22), the partial effect of the ASF, averaged across *X*, is simply β_1 . The average partial effect across (*X*, *U*) is $\beta_1 + 2\beta_3\sigma_{XU}$. In principal, these need not even have the same sign, let alone similar magnitudes. (In practice, β_3 might be small because it is the coefficient on the interaction X^2U .)

The example in equation (3.22) can also be used to illustrate the main result of this section, namely, that the η_j in (3.20) are identified if we have access to a suitable proxy variable, *V*. Assume *V* – which is observed or depends on parameters that we can consistently estimate – has a zero mean. Make the redundancy assumption along with a

linearity assumption on E(U|V):

$$E(U|X,V) = E(U|V) = \theta V$$
(3.29)

Then

$$E(Y|X,V) = \beta_0 + \beta_1 X + \beta_2 E(U|X,V) + \beta_3 X^2 E(U|X,V)$$
$$= \beta_0 + \beta_1 X + \beta_2 \theta V + \beta_3 X^2 \theta V$$
$$\equiv h(X,V)$$

The partial derivative of h(x, v) with respect to x is

$$\frac{\partial h(x,v)}{\partial x} = \beta_1 + 2\gamma_3 xv$$

where $\gamma_3 = \beta_3 \theta$. Now

$$E_{(X,V)}(\beta_1 + 2\gamma_3 XV) = \beta_1 + 2\gamma_3 E(XV)$$
$$= \beta_1 + 2\beta_3 E(X\theta V)$$

and $E(X\theta V) = E[XE(U|V)] = E[E(XU|X, V)] = E(XU)$ by iterated expectations. This shows that estimating E(Y|X, V), taking the partial effect with respect to X, and then averaging across (X, V) is the same as starting with E(Y|X, U) and performing the same operations. This same analysis shows that $\beta_1 = \partial ASF(x)/\partial x$ is also identified by E(Y|X, V).

As mentioned earlier, there is one important case where the different definitions of the average partial effects are the same. We already showed that when **U** and **X** are independent then $ASF(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$. We now can say more. Namely, basing the average partial effect on the ASF, or first taking the partial derivative of the original structural function, give the same answer after all random variables are averaged out.

Proposition 5: Under the assumptions of Proposition 4, assume that **U** and **X** are independent. Then

$$E_{(\mathbf{X},\mathbf{U})}\left[g_j(\mathbf{X},\mathbf{U})\right] = E_{\mathbf{X}}\left[\frac{\partial ASF(\mathbf{X})}{\partial x_j}\right].$$

Proof: By the law of iterated expectations,

$$E_{(\mathbf{X},\mathbf{U})}\left[g_j(\mathbf{X},\mathbf{U})\right] = E_{\mathbf{X}}\left\{E[g_j(\mathbf{X},\mathbf{U})|\mathbf{X}]\right\}.$$

Now, we can write

$$E[g_j(\mathbf{X}, \mathbf{U}) | \mathbf{X} = \mathbf{x}] = \int g_j(\mathbf{x}, \mathbf{U}) Q(d\mathbf{u} | \mathbf{x}) = \int \frac{\partial g}{\partial x_j}(\mathbf{x}, \mathbf{u}) Q(d\mathbf{u}) = \frac{\partial \int g(\mathbf{x}, \mathbf{u}) Q(d\mathbf{u})}{\partial x_j} = \frac{\partial ASF(\mathbf{x})}{\partial x_j},$$

where the second equality uses independence and the third uses the technical assumption of interchanging the derivative and the integral. It follows that

$$E_{\mathbf{X}}\{E[g_{j}(\mathbf{X},\mathbf{U})|\mathbf{X}]\} = E_{\mathbf{X}}\left[\frac{\partial ASF(\mathbf{X})}{\partial x_{j}}\right]$$

and this completes the proof. \Box
3.4 EMPIRICAL EXAMPLE

In this section, we illustrate how to obtain the estimates of (3.17) and (3.20) with two empirical examples using Michigan Educational Assessment Program (MEAP) math test outcome for the fourth graders.

Papke and Wooldridge (2008) apply a control function approach to estimate the effects of spending on student's performance for this test using 501 school districts from 1992 through 2001. Their dependent variable is a pass rate³ and they use *foundation allowance* as the instrument variable for *spending*.

For this example, (3.17) and (3.20) are estimated as

$$\widehat{\boldsymbol{\delta}} = \widehat{\boldsymbol{\beta}} \cdot \frac{1}{NT} \sum_{j}^{N} \sum_{s}^{T} \left[\frac{1}{NT} \sum_{i}^{N} \sum_{t}^{T} \left\{ \boldsymbol{\phi}(\widehat{\boldsymbol{\psi}} + \mathbf{x}_{js}\widehat{\boldsymbol{\beta}} + \overline{\mathbf{h}}_{i}\widehat{\boldsymbol{\xi}} + \widehat{\boldsymbol{\rho}}\widehat{\boldsymbol{v}}_{it}) \right\} \right],$$
(3.30)

and

$$\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\beta}} \cdot \left[\frac{1}{NT} \sum_{i}^{N} \sum_{t}^{T} \left\{ \phi(\widehat{\psi} + \mathbf{x}_{it} \widehat{\boldsymbol{\beta}} + \overline{\mathbf{h}}_{i} \widehat{\boldsymbol{\xi}} + \widehat{\rho} \widehat{v}_{it}) \right\} \right], \qquad (3.31)$$

where \mathbf{x}_{it} includes *spending*_{it}, *freelunch*_{it}, $\log(enrollment)_{it}$, *spending*_{i,1994}, time dummies and the interactions between *spending*_{i,1994} and time dummies, and $\mathbf{\overline{h}}_i$ contains time averages of *freelunch*_{it} and $\log(enrollment)_{it}$.^{4,5} \hat{v}_{it} is the residual obtained by the first step of, and $\hat{\boldsymbol{\beta}}$, $\hat{\psi}$, $\hat{\boldsymbol{\xi}}$ and $\hat{\rho}$ are the estimates obtained by the second step of **Procedure 4.1** in Papke and Wooldridge (2008), respectively.

Table 3.1 presents $\hat{\delta}$ and $\hat{\eta}$ for *spending*_{*it*}, *freelunch*_{*it*}, and log(*enrollment*)_{*it*}. The estimates $\hat{\eta}$ are the same as those reported in Papke and Wooldridge (2008).⁶ We see that $\hat{\delta}$ is more

³In their data period, the test outcome is graded as one of Satisfactory, Moderate, or Low. The pass rate is a fraction of students rated at the Satisfactory level.

⁴In Papke and Wooldridge (2008), *spending_{it}* is constructed as the average of real expenditures per pupil for the recent four years in logarithmic form, *freelunch_{it}* is the fraction of students who are eligible for free and reduced-price lunch program.

⁵They allow the correlation between the district individual heterogeneity and the time averages of *freelunch_{it}* and $log(enrollment)_{it}$.

⁶Table 5 in Papke and Wooldridge (2008).

precisely estimated than $\hat{\eta}$ but they are similar.⁷

	APE		
variables	$\widehat{\eta}$	$\widehat{\delta}$	
spending	0.583	0.558	
	(0.256)	(0.210)	
freelunch	-0.100	-0.096	
-	(0.069)	(0.065)	
log(lenrollment)	0.096	0.092	
	(0.073)	(0.061)	
Scale factor	0.337	0.323	

Table 3.1: APE estimates in Papke and Wooldridge (2008)

^{1.} The standard errors, in parentheses, are obtained by using 1000 bootstrap replications.

^{2.} Scale factor indicates the average sum of $\phi(\cdot)$ in (3.30) and (3.31).

We now calculate the two APE estimates of the empirical application in Chapter 1, which also examines how spending affects the MEAP test outcomes using the 2004 school year data with multiple fractional response variables. In this application, the dependent variable is a set of student fractions of the four levels for a district where the sum of a district's four fractions becomes one.⁸

The two APE of spending on these four levels are estimated by

$$\widehat{\delta}_{g} = \frac{1}{N} \sum_{j}^{N} \left[\frac{1}{N} \sum_{i}^{N} \left\{ \frac{\exp\left(\mathbf{x}_{j} \widehat{\boldsymbol{\theta}}_{\mathbf{x}g} + \widehat{\theta}_{vg} \widehat{v}_{i}\right)}{\sum_{h}^{4} \exp\left(\mathbf{x}_{j} \widehat{\boldsymbol{\theta}}_{\mathbf{x}h} + \widehat{\theta}_{vh} \widehat{v}_{i}\right)} \cdot \left(\widehat{\theta}_{wg} - \frac{\sum_{h}^{4} \widehat{\theta}_{wh} \exp\left(\mathbf{x}_{j} \widehat{\boldsymbol{\theta}}_{\mathbf{x}h} + \widehat{\theta}_{vh} \widehat{v}_{i}\right)}{\sum_{h}^{4} \exp\left(\mathbf{x}_{j} \widehat{\boldsymbol{\theta}}_{\mathbf{x}h} + \widehat{\theta}_{vh} \widehat{v}_{i}\right)} \right) \right\} \right]$$
(3.32)

⁷Their difference comes from the difference between their scale factors. Considering that the scale factors are normal density functions raging from zero to 0.4, the difference could not be large unless the coefficient estimate $\hat{\beta}$ is huge.

⁸In 2004, there were four categories.

, and

$$\widehat{\eta}_{g} = \frac{1}{N} \sum_{i}^{N} \left[\frac{\exp\left(\mathbf{x}_{i}\widehat{\boldsymbol{\theta}}_{\mathbf{x}g} + \widehat{\theta}_{vg}\widehat{v}_{i}\right)}{\sum_{h}^{4}\exp\left(\mathbf{x}_{i}\widehat{\boldsymbol{\theta}}_{\mathbf{x}h} + \widehat{\theta}_{vh}\widehat{v}_{i}\right)} \cdot \left(\widehat{\theta}_{wg} - \frac{\sum_{h}^{4}\widehat{\theta}_{wh}\exp\left(\mathbf{x}_{i}\widehat{\boldsymbol{\theta}}_{\mathbf{x}h} + \widehat{\theta}_{vh}\widehat{v}_{i}\right)}{\sum_{h}^{4}\exp\left(\mathbf{x}_{i}\widehat{\boldsymbol{\theta}}_{\mathbf{x}h} + \widehat{\theta}_{vh}\widehat{v}_{i}\right)} \right) \right], \quad (3.33)$$

where \mathbf{x}_i contains *spending*_{*i*}, *freelunch*_{*i*}, log(*enrollment*_{*i*}), and *spending*93_{*i*},⁹, \hat{v}_i is the OLS residual in the first step, and $\hat{\theta}$ is the fmlogit estimates in the second step.

Table 3.2 reports the two estimates. As in the above illustration, they are similar in general.

			level 1	level 2	level 3	level 4
spending	\widehat{v}	$\widehat{\eta}$	0.699	0.016	-0.634	-0.081
			(0.223)	(0.139)	(0.190)	(0.054)
		$\widehat{\delta}$	0.681	0.017	-0.619	-0.079
			(0.205)	(0.135)	(0.172)	(0.057)
spending ²	\widehat{v} only	$\widehat{\eta}$	0.806	-0.053	-0.679	-0.074
			(0.239)	(0.153)	(0.199)	(0.055)
		$\widehat{\delta}$	0.779	-0.044	-0.667	-0.068
			(0.216)	(0.150)	(0.183)	(0.056)
	$\widehat{v}, \widehat{v}^3, \widehat{v}^3$	$\widehat{\eta}$	0.709	0.012	-0.618	-0.103
			(0.248)	(0.158)	(0.188)	(0.061)
		$\widehat{\delta}$	0.691	0.014	-0.602	-0.103
			(0.229)	(0.154)	(0.170)	(0.068)

Table 3.2: APE estimates in Chapter 1

^{1.} The standard errors, in parentheses, are obtained by using 1000 bootstrap replications.

^{2.} *spending*² indicates, the model includes *spending*² as well as *spending*.

². \hat{v} , \hat{v}^3 , \hat{v}^3 indicates that the estimation includes them.

3.5 CONCLUSION

In this chapter, we have examined two types of a single APE that summarizes the APEs of observed covariates. One is obtained by calculating ASF – averaging the conditional

⁹*spending*_{*i*} = log(per pupil GF expenditure), and *spending*93_{*i*} is spending in 1993/1994.

mean over the distribution of unobservables – and then averaging its partial derivatives or discrete changes across the observed covariates. The other, which is more commonly used in empirical studies, is obtained by averaging the partial effects of conditional mean over the joint distribution of observed covariates and unobservables.

Through the propositions, we have shown that the two APEs are identified in general as long as there are suitable proxy variables satisfying the redundancy and the conditional independence assumptions. Furthermore, they are not generally the same unless the unobservables and observed covariates are independent.

We have also illustrated how the two APEs are estimated in the empirical examples using MEAP math test outcomes. In these examples, the two types of APE estimates are similar.

REFERENCES

REFERENCES

Altonji, J. G., and R. L. Matzkin. 2005. "Cross section and panel data estimators for nonseparable models with endogenous regressors." *Econometrica*, 73(4): 1053–1102.

Bartle, R. G. 1966. The Elements of Integration .: New York: Wiley.

Blundell, R., and J. L. Powell. 2003. "Endogeneity in Nonparametric and Semiparametric Regression Models." In *Advances in Economics and Econometrics*. eds. by L. P. H. Dewatripont, Mathias, and S. J. Turnovsky, 2: Cambridge: Cambridge University Press, , 1st edition 312–357.

Imbens, G. W., and J. M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47(1): 5–86.

Nam, S. 2014. "Multiple fractional response variables with continuous endogenous explanatory variables." June.

Papke, L. E., and J. M. Wooldridge. 2008. "Panel data methods for fractional response variables with an application to test pass rates." *Journal of Econometrics*, 145(1-2): 121–133.

Wooldridge, J. M. 2002. Econometric analysis of cross section and panel data.: MIT press.

Wooldridge, J. M. 2005b. "Unobserved Heterogeneity and Estimation of Average Partial Effects." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. eds. by D. W. K. Andrews, and J. H. Stock: Cambridge University Press, , Chap. 3.

——— . 2010. *Econometric Analysis of Cross Section and Panel Data,*. MIT Press Books: MIT Press, , 2nd edition.