Binary and Fractional Response Models with Continuous and Binary Endogenous Explanatory Variables

Wei Lin*

Jeffrey M. Wooldridge†

March 26, 2017

Abstract

This paper considers models for binary and fractional responses with a binary endogenous explanatory variable (EEV) and many potentially continuous endogenous explanatory variables. A two-step control function (CF) approach combined with a quasi-likelihood method is proposed to account for endogeneity. The approach is flexible, and we are able to uncover average partial effects of the continuous variables and the binary variable EEV. The inference for the partial effects can be easily obtained through bootstrapping because of the computational simplicity of the two-step CF approach. A basic probit model, an endogenous switching probit model, and a fractional probit model are covered. Variable addition tests on generalized residuals are used to detect additional endogeneity from the binary EEV, and a simulation study shows that the average partial effects are well estimated using generalized residuals. An empirical illustration of the determination of housing budget shares shows that, in a fractional response model, using generalized residuals leads to a close approximation to joint estimation.

*Center for Real Estate, Massachusetts Institute of Technology, Cambridge, MA 02139 , United States. linwei1@mit.edu.
†Department of Economics, Michigan State University, East Lansing, MI 48824, United States. wooldri1@msu.edu.

1 Introduction

Binary response models play a significant role in empirical work across many fields. Examples include migration decisions (Dong and Lewbel, 2015), college admissions decisions, (Conlin et al., 2013), and firm takeover activity (Edmans et al., 2012), to name just a few. As is well known – see, for example, (Wooldridge, 2010) – accounting for endogeneity in nonlinear binary response models, especially if an endogenous explanatory variable (EEV) is binary, generally requires strong assumptions. To avoid modeling decisions, many empirical researchers prefer using a linear model estimated by, say, two stage least squares. However, a linear model only approximates an average treatment effect, perhaps only for an unobserved subpopulation (Imbens and Angrist, 1994). In many cases we might be willing to make stronger assumptions in order to estimate structural effects or treatment effects for observed subpopulations.

Another knock on nonlinear binary response models is that they can be difficicult to estimate in the presence of multiple EEVs, as the joint log-likelihood function, which requires integrating over high-dimensional probability density functions, is not always well behaved. As an example, suppose that one is modeling a consumer decision to purchase a product, and the decision depends on own price and that of substitutes. In addition, some consumers have access to online shopping while others do not. If prices and online shopping availability are all endogenous (assuming we have suitable instruments), a full MLE could be very difficult to implement.

This paper considers the estimation of a class of binary response models when there are possibly many continuous EEVs and a single binary EEV. The binary EEV could be an indicator of self-selecting into a treatment, into a region or into a sample, depending on how it appears in the binary response equation. Some empirical examples for the binary EEV are whether to own a house, whether to submit SAT scores in college applications, whether to be in the labor force, or whether to use online shopping. The continuous EEVs could be years of education, family income, or prices of various goods or inputs. Admittedly, the restriction to a single binary EEV is more restrictive than ideal, but it is a leading case, and includes the endogenous switching regression model with additional EEVs. The approach we take can be extended to other situations with more than one discrete EEV, with significant computational advantages, as will be apparent in what follows.

Our approach is parametric, and is a hybrid of a control function (CF) approach for the continuous EEVs and a bivariate probit maximum likelihood estimation. With a single continuous EEV, our approach reduces to the Rivers and Vuong (1988) two-step estimator. If we have only a binary EEV, the resulting estimator is the binary response version of the dummy endogenous variable model studied by Heckman (1978), Amemiya (1978), and Rivers and Vuong (1988). To allow for both kinds of EEVs, our approach effectively combines these two approaches and produces computationally simple two-step MLEs (or two-step quasi-MLEs).

Because we use a two-step method, we do not, strictly speaking, have to assume joint normality of all unobservables in the system. Nevertheless, the conditional normality assumptions almost get us to joint normality. One could argue, then, that one should just use full MLE. As mentioned earlier, full MLE can be very difficult to implement, and it can be difficult to see the source of any computational difficulties when multiple integrals need to be approximated numerically. Our two-step approach is transparent and forces a researcher to first study the strength of instruments before proceeding. It allows researchers to focus on the substantive problems in their applications rather than wrestling with knotty computational problems. Another competitor is substituting fitted values from various first-stage estimations for EEVs in the structural binary response equation, but it is well known that, especially with a binary EEV, this procedure yields inconsistent estimates of parameters and partial effects. In fact, Hausman (1975), coined the phrase ”forbidden regression” to describe this unfortunate practice.

The current paper can be viewed as an extension of Wooldridge (2014), who studied a two-step CF approach in a quasi-LIML framework. Wooldridge (2014) considers only models with either continuous EEVs or a binary EEV – but not both. We apply insights from Wooldridge (2014) to obtain an estimating equation that includes control functions to account for continuous EEVs. To carry out the two-step procedure, residuals from the first-stage estimation of reduced forms for the continuous EEVs are plugged into the second-stage joint estimation of the binary outcome and binary EEV. Routines in commonly used software can be exploited (or slightly modified) to carry out the procedure. Most importantly, due to the computational simplicity of the CF approach, bootstrapping can be easily applied to obtain inference for functions of the parameters, such as average partial effects (APEs), rather than using the complicated delta method.

In addition, as shown in Wooldridge (2014), simple variable addition tests (VATs) for endogeneity are obtained as by-products of the CF approach. This paper extends the VATs for a single EEV to VATs for mixed, multiple EEVs. The VATs are based on standard Wald tests of those plugged-in residuals obtained from the first stage estimations. We can test the null that a subset of the continuous EEVs is exogenous without taking a stand on the binary EEV. Or, by using generalized residuals, we can allow the continuous variables to be endogenous and test the null that the binary EEV is actually exogenous.Since EEVs are often correlated with each other, conditioning on residuals obtained from the continuous EEV reduced forms gives a cleaner test of whether the binary EEV is in fact exogenous.

Another feature of our two-step procedure is that it applies immediately to cases where the response variable in the structural equation is a fraction or proportion, as in Papke and Wooldridge (1996) for the case of exogenous explanatory variables. Wooldridge (2014) showed how the probit quasi-log likelihood is justified with either continuous EEVs or a single binary EEV. Our new approach greatly expands the scope of such models. In the example of a consumer purchase decision cited earlier, the response variable can instead be a budget share ratio. So long as the conditional mean of the fractional response is correctly specified, we consistently estimate parameters in the conditional mean even though other features of the distribution are misspecified.

Besides parametric approaches to estimating this triangular model for binary responses, semi-parametric and nonparametric approaches are also available. Nevertheless, while those approaches sensibly relax parametric assumptions in some directions, they inevitably impose restrictions in other directions. For example, the Blundell and Powell (2003) nonparametric CF approach to binary response models applies to continuous EEVs. Unfortunately, their assumption of additive, independent errors in the reduced forms rules out discrete EEVs. The special regressor method in Dong and Lewbel (2015) allows for both continuous and binary EEVs in semiparametric binary response models. However, their method requires a special regressor to be excluded from the reduced forms for the EEVs, and the special regressor cannot appear in the structural equation in a flexible way (such as just adding a quadratic, or including interactions). Further, as discussed in Lin and Wooldridge (2015), the average index functions (AIF), proposed as a basis for defining marginal effects for special regressor methods, lack a causal interpretation.

The rest of the paper is organized as follows. Section 2 starts with a basic model with one binary and many continuous EEVs. The same arguments are then extended to an endogenous switching model where there are two sources of unobservables in the structural equation and these can interact with the binary EEV. In other words, we consider a full binary response endogenous switching model with continous EEVs as well. Section 3 derives the VAT for endogeneity of the binary EEV given residuals from continuous EEVs. Section 4 confirms that the CF approach can be applied to fractional response models. Section 5 presents Monte Carlo simulation evidence on how differenet methods estimate APEs for the binary response model. Section 6 illustrates this approach by revisiting the study of the effects of price and total expenditure on housing budget share equation. Section 7 concludes.

2 Model Specification and Estimation for Binary Response

2.1 Probit Models with One Binary EEV and Many Continuous EEVs

We start with the model that includes only one unobservable in the structural equation, but we allow the EEVs – whether continous or binary – to appear very generally. We do assume that the continuous and binary EEVs have specific reduced forms. More specifically, consider a simple model for a binary response y_1 with continuous EEVs $\mathbf{y}_2$ and one binary EEV y_3. Write the model recursively in a triangular form as

$$y_1 = 1\left[\mathbf{x}_1\boldsymbol{\beta} + u_1 > 0\right], \tag{1a}$$

$$\mathbf{y}_2 = \mathbf{z}\Pi + \mathbf{v}_2, \tag{1b}$$

$$y_3 = 1\left[\mathbf{z}\boldsymbol{\delta} + u_3 > 0\right]. \tag{1c}$$

Equation (1a) is a structural equation that represents a causal relationship. Equations (1b) and (1c) are reduced forms for the continuous EEVs $\mathbf{y}_2$, of dimension $1 \times G$, and the scalar binary EEV y_3, respectively. The indicator function $1\left[\cdot\right]$ takes on a value of one when the statement in the bracket is true and zero otherwise. In addition, $\mathbf{x}_1$ is $1 \times K_1$ vector whose elements are general functions of $(\mathbf{z}_1, \mathbf{y}_2, y_3)$, such as polynomials, interactions, logarithms, and so on, with $\mathbf{x}_1 = (\mathbf{z}_1, \mathbf{y}_2, y_3)$ being the leading case. $\mathbf{z}_1$ is a $1 \times L_1$ vector and a strict subset of the entire $1\times\ L$ vector of exogenous variables $\mathbf{z} \equiv (\mathbf{z}_1, \mathbf{z}_2)$, with $L \equiv L_1 + L_2$ and $L_2 \geq G + 1$. Identifying parameters based on nonlinearity often times turns out poorly in practice, so we need at least one excluded instrument for the binary EEV. Because we are using a control function approach, we do not need separate instruments for all nonlinear functions of $(\mathbf{y}_2, y_3)$ in $\mathbf{x}_1$. We should point out that $\mathbf{z}$ can include nonlinear functions of underlying endogenous variables, as would be the case in semiparametric approaches to this problem. In almost all cases $\mathbf{z}_1$ would include unity as its first element, which means we can assume the error terms $(u_1, \mathbf{v}_2, v_3)$ have zero means. The matrix Π is an $L \times G$ matrix of reduced form parameters. In the simplest case, $G = 1$ and y_2 is a scalar.

The system above most directly applies to endogeneity due to omitted variables, rather than simultaneity. This is certainly the leading case. For example, a student decides to attend a certain kind of school, say a charter school (y_3) and then subsequently the score on a standardized test is revealed. A consumer is a price taker ($\mathbf{y}_2$ is a vector of prices) and then decides on purchasing decisions, and may be faced with the possibility of shopping online (y_3), which is determined by the grocery store chain not based on an individual consumer's purchases but on community characteristics.

Formally, endogeneity arises in the model when the structural error, u_1, is correlated with explanatory variables $\mathbf{y}_2$ and y_3, in that it contains an unobservable that also appears in error terms $\mathbf{v}_2$ and u_3. Write the linear projections of (u_1, u_3) on $\mathbf{v}_2$ in error forms:

$$u_1 = \mathbf{v}_2\boldsymbol{\theta} + v_1, \tag{2a}$$

$$u_3 = \mathbf{v}_2\boldsymbol{\eta} + v_3, \tag{2b}$$

where $\boldsymbol{\theta} \equiv E\left(\mathbf{v}_2'\mathbf{v}_2\right)^{-1} E\left(\mathbf{v}_2'u_1\right)$ and $\boldsymbol{\eta} \equiv E\left(\mathbf{v}_2'\mathbf{v}_2\right)^{-1} E\left(\mathbf{v}_2'u_3\right)$ are the $G \times 1$ vectors of the population regression coefficients.

A convenient joint normality assumption among $(u_1, \mathbf{v}_2, u_3)$ (Heckman, 1978; Amemiya, 1978; Rivers and Vuong, 1988) leaves us with a bivariate normally distributed vector of errors (v_1, v_3) that is independent of $\mathbf{v}_2$ (by definition of a linear projection and by a property of multivariate normality):

$$D\begin{pmatrix} v_1 \\ v_3 \end{pmatrix} = D\left(\begin{matrix} v_1 \\ v_3 \end{matrix}\middle| \mathbf{v}_2\right) \sim \text{Normal}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right], \tag{3}$$

where $D(\cdot)$ denotes the distribution, the variances of v_1 and v_3 are normalized to one (without loss of generality for estimating partial effects), and $\rho \equiv Cov\left(v_1, v_3\right)$ is the correlation.

In fact, if we are willing to assume a strong enough exogeneity condition for the instruments $\mathbf{z}$, the bivariate distribution (v_1, v_3) becomes independent not only of $\mathbf{v}_2$ but also of $\mathbf{z}$ and thus of $\mathbf{y}_2$ (because $\mathbf{y}_2$ is a deterministic function of $\mathbf{z}, \mathbf{v}_2$):

$$D\begin{pmatrix} v_1 \\ v_3 \end{pmatrix} = D\left(\begin{matrix} v_1 \\ v_3 \end{matrix}\middle| \mathbf{z}, \mathbf{v}_2\right) = D\left(\begin{matrix} v_1 \\ v_3 \end{matrix}\middle| \mathbf{y}_2, \mathbf{z}, \mathbf{v}_2\right). \tag{4}$$

Given the distributional assumptions, we arrive at a bivariate probit model that accounts for endogeneity issues in (1) by adding $\mathbf{v}_2$ as additional explanatory variables:

$$y_1 = 1\left[\mathbf{x}_1\boldsymbol{\beta} + \mathbf{v}_2\boldsymbol{\theta} + v_1 \geq 0\right], \tag{5a}$$

$$y_3 = 1\left[\mathbf{z}\boldsymbol{\delta} + \mathbf{v}_2\boldsymbol{\eta} + v_3 \geq 0\right]. \tag{5b}$$

Adding reduced form errors to control for endogeneity is the essence of a control function approach. However, since we cannot observe $\mathbf{v}_2$, to operationalize it, a simple two-step procedure proceeds as follows:

1. Estimate (1b), the reduced forms for $\mathbf{y}_2$, by ordinary least squares (OLS), equation by equation, to obtain the residuals $\widehat{\mathbf{v}}_2 = \mathbf{y}_2 - \mathbf{z}\widehat{\Pi}$.

2. Estimate (5), the bivariate probit model for y_1 and y_3, jointly by maximum likelihood estimation (MLE), replacing $\mathbf{v}_2$ with $\widehat{\mathbf{v}}_2$.

As there is no one-to-one mapping between the reduced form error v_3 and the binary EEV y_3, we cannot obtain a proxy for v_3 and hence have to rely on a joint estimation in the second step. By the usual consistency argument of two-step M-estimations (see, for example, Wooldridge, 2010, section 12.4.1), the resulting control function estimator $\left(\widehat{\Pi}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\eta}}\right)$ is consistent for parameters identified by the following population problems. Formally,

$$\Pi = E\left(\mathbf{z}'\mathbf{z}\right)^{-1} E\left(\mathbf{z}'\mathbf{y}_2\right), \tag{6}$$

and $(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\theta}, \boldsymbol{\eta})$ is the unique solution to

$$\begin{aligned} & \max_{\mathbf{b}\in\mathbb{R}^{K_1}, \mathbf{d}\in\mathbb{R}^{L}, \mathbf{r}\in\mathbb{R}^{G}, \mathbf{g}\in\mathbb{R}^{G}, \rho\in\mathbb{R}} E\left[\log P\left(y_1, y_3 | \mathbf{y}_2, \mathbf{z}, \mathbf{v}_2\right)\right] \\ = \; & E\left[y_1 y_3 \log \int_{-q_3}^{\infty} \Phi(d)\,\phi(v_3)\,dv_3 \right. \\ & + (1-y_1)\,y_3 \log \int_{-q_3}^{\infty} \left[1-\Phi(d)\right]\phi(v_3)\,dv_3 \\ & + y_1 (1-y_3) \log \int_{-\infty}^{-q_3} \Phi(d)\,\phi(v_3)\,dv_3 \\ & \left. + (1-y_1)(1-y_3) \log \int_{-\infty}^{-q_3} \left[1-\Phi(d)\right]\phi(v_3)\,dv_3 \right], \end{aligned} \tag{7}$$

where

$$d \equiv \frac{\mathbf{x_1 b} + \mathbf{v}_2\mathbf{r} + \rho v_3}{\sqrt{1-\rho^2}}, \tag{8}$$

$$q_3 \equiv \mathbf{zd} + \mathbf{v}_2\mathbf{g}. \tag{9}$$

As the magnitude of $\boldsymbol{\beta}$ depends on the normalization of the error terms and thus is identified only up to scale, interpreting $\boldsymbol{\beta}$ is not especially meaningful. Instead, the primary goal in empirical studies is to explain marginal effects of a variable of interest on response probabilities. In the presence of EEVs, $P(y_1 = 1|\mathbf{x}_1)$, the conditional response probability is hardly of any interest: it is affected by y_2 and y_3

having correlations with the omitted variable in the unobservables u_1. We must use care in constructing an interesting response function for deriving partial effects. Fortunately, Blundell and Powell (2003, 2004) have proposed the average structural function (ASF), which is intuitively appealing and can be obtained via counterfactual reasoning. In defining the ASF for the structural equation (1a), we break the correlations by holding the observables $\mathbf{x}_1$ as fixed arguments and averaging out the unobservable u_{i1} without conditioning on $\mathbf{x}_1$:

$$\text{ASF}(\mathbf{x}_1) = E_{u_{i1}}\{1[\mathbf{x}_1\boldsymbol{\beta} + u_{i1} > 0]\}, \tag{10}$$

where the subscript i on u_{i1} emphasizes that it is a random variable, and $E_{u_{i1}}\{\cdot\}$ is the expected value with respect to u_{i1}.

In the two-step CF procedure above, we identify parameters that correspond to the conditional normality of u_1 given $\mathbf{v}_2$, namely,

$$u_1|\mathbf{v}_2 \sim \text{Normal}(\mathbf{v}_2\boldsymbol{\theta}, 1). \tag{11}$$

Thus, by the usual law of iterated expectations, the ASF defined in (10) can be obtained in two steps. First, we treat $\mathbf{v}_{i2}$ as fixed, and then we average them out as random variables:

$$\begin{aligned}\text{ASF}(\mathbf{x}_1) &= E_{\mathbf{v}_{i2}}\left\{E_{u_{i1}|\mathbf{v}_{i2}}\{1[\mathbf{x}_1\boldsymbol{\beta} + \mathbf{v}_{i2}\boldsymbol{\theta} + v_{i1} > 0]\,|\mathbf{v}_{i2}\}\right\} \\ &= E_{\mathbf{v}_{i2}}\{\Phi(\mathbf{x}_1\boldsymbol{\beta} + \mathbf{v}_{i2}\boldsymbol{\theta})\} \\ &= \int_{-\infty}^{\infty}\Phi(\mathbf{x}_1\boldsymbol{\beta} + \boldsymbol{v}_{i2}\boldsymbol{\theta})\,\phi(\boldsymbol{v}_{i2})\,d\boldsymbol{v}_{i2},\end{aligned} \tag{12}$$

where $\phi(\cdot)$ is the density function for the random variables $\mathbf{v}_{i2}$.

The average partial effects (APEs) for a given $\mathbf{x}_1$ are then obtained by taking derivatives or differences of (12)

$$\text{APE}_{y_2}(\mathbf{x}_1) = \beta_{y_2}\int_{-\infty}^{\infty}\phi(\mathbf{x}_1\boldsymbol{\beta} + \boldsymbol{v}_{i2}\boldsymbol{\theta})\,\phi(\boldsymbol{v}_{i2})\,d\boldsymbol{v}_{i2}, \tag{13a}$$

$$\text{APE}_{y_3}(\mathbf{x}_1) = \int_{-\infty}^{\infty}\left[\Phi\left(\mathbf{x}_1^{(1)}\boldsymbol{\beta} + \mathbf{v}_{i2}\boldsymbol{\theta}\right) - \Phi\left(\mathbf{x}_1^{(0)}\boldsymbol{\beta} + \mathbf{v}_{i2}\boldsymbol{\theta}\right)\right]\phi(\boldsymbol{v}_{i2})\,d\boldsymbol{v}_{i2}, \tag{13b}$$

where β_{y_2} is the coefficient on y_2 and $\mathbf{x}_1^{(1)}$ denotes explanatory variables at a particular fixed value with $y_3 = 1$ and $\mathbf{x}_1^{(0)}$ denotes the same fixed value of the explanatory variables except that $y_3 = 0$. Those APEs can be consistently estimated by using sample analogues and inserting consistent estimators of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ from the two-step CF approach:

$$\widehat{\mathrm{APE}}_{y_2}(\mathbf{x}_1) = \hat{\beta}_{y_2}\left[N^{-1}\sum_{i=1}^{N}\phi\left(\mathbf{x}_1\widehat{\boldsymbol{\beta}}+\widehat{\mathbf{v}}_{i2}\widehat{\boldsymbol{\theta}}\right)\right], \tag{14a}$$

$$\widehat{\mathrm{APE}}_{y_3}(\mathbf{x}_1) = N^{-1}\sum_{i=1}^{N}\left[\Phi\left(\mathbf{x}_1^{(1)}\widehat{\boldsymbol{\beta}}+\widehat{\mathbf{v}}_{i2}\widehat{\boldsymbol{\theta}}\right)-\Phi\left(\mathbf{x}_1^{(0)}\widehat{\boldsymbol{\beta}}+\widehat{\mathbf{v}}_{i2}\widehat{\boldsymbol{\theta}}\right)\right]. \tag{14b}$$

To obtain inferences for the estimators of APEs as in (14a) and (14b), analytical standard errors can be derived by the delta method and by setting the two-step control function problem as a one-step method-of-moments problem. However, because all the procedures involved in the estimations are standard routines, bootstrap standard errors can be easily obtained to account for the sampling errors.

As shown in (13a) and (13b), APEs for the binary response model have the attractive feature of built-in heterogeneity—-they deliver varying partial effects when evaluated at different values of $\mathbf{x}_1$. However, if one is interested in using a single summary statistic for marginal effects, further averaging across $\mathbf{x}_1$ should be applied. A joint averaging across $\mathbf{x}_1, \widehat{\mathbf{v}}_2$ (as "margins" command does in STATA) is computationally easier but bears a different causal interpretation from sequentially averaging out $\widehat{\mathbf{v}}_2$ and $\mathbf{x}_1$ (Nam and Wooldridge, 2014).

Although serving as a starting point, the modelling strategy in (1) for the CF approach is limited in several ways. One restrictive feature is that the reduced form error $\mathbf{v}_2$ needs to be independent of the exogenous variables $\mathbf{z}$. Thus, the linear function form for the conditional mean of $\mathbf{y}_2$ is unrealistic and can be relaxed to be any generic function $\boldsymbol{\pi}(\cdot)$ for $\mathbf{z}$ as in Blundell and Powell (2003, 2004). More importantly, $\mathbf{v}_2$ here acts as a sufficient statistic to control for any endogeneity from $\mathbf{y}_2$ in the structural error u_1: that is, $\mathbf{y}_2$ is correlated with u_1 only through $\mathbf{v}_2$ on its level form. However, as shown in Murtazashvili and Wooldridge (2015), in case of more heterogeneity such as random coefficients, the unobservable u_1 can contain full interactions between $\mathbf{v}_2$ and $\mathbf{z}, \mathbf{x}_1$.

2.2 Probit Endogenous Switching Models with Many Continuous EEVs

As we are interested in modeling some heterogeneity besides EEVs, we turn to a probit switching regression with EEVs. The binary EEV y_3 can be viewed as a switching indicator. In addition to shifting intercepts when y_3 appears by itself in the linear index, the switching can be made more general. Interacting y_3 with all the observables allows us to switch into regimes of differing slopes. The interaction between y_3 and unobservables indicates that the two regimes have differing unobservables. The switching is endogenous because y_3 is correlated with the unobservables. In the treatment effect framework, y_3 is the treatment indicator and the treatment effect is heterogenous. To see this, first write the model as follows:

$$y_1 = 1\left[(1-y_3)\,\mathbf{x}_1\boldsymbol{\beta}_0 + y_3\mathbf{x}_1\boldsymbol{\beta}_1 + (1-y_3)\,u_0 + y_3 u_1 > 0\right] \tag{15a}$$

$$\mathbf{y}_2 = \mathbf{z}\Pi + \mathbf{v}_2 \tag{15b}$$

$$y_3 = 1\left[\mathbf{z}\boldsymbol{\delta} + u_3 > 0\right], \tag{15c}$$

Under a similar set of notations and assumptions as in (1), write the linear projection of u_1, u_0 and u_3 onto the reduced form error $\mathbf{v}_2$ in error forms:

$$u_0 = \mathbf{v}_2\boldsymbol{\theta}_0 + v_0 \tag{16a}$$

$$u_1 = \mathbf{v}_2\boldsymbol{\theta}_1 + v_1 \tag{16b}$$

$$u_3 = \mathbf{v}_2\boldsymbol{\eta} + v_3, \tag{16c}$$

where $\boldsymbol{\theta}_0 \equiv E\left(\mathbf{v}_2'\mathbf{v}_2\right)^{-1} E\left(\mathbf{v}_2'u_0\right)$, $\boldsymbol{\theta}_1 \equiv E\left(\mathbf{v}_2'\mathbf{v}_2\right)^{-1} E\left(\mathbf{v}_2'u_1\right)$ and $\boldsymbol{\eta} \equiv E\left(\mathbf{v}_2'\mathbf{v}_2\right)^{-1} E\left(\mathbf{v}_2'u_3\right)$ are the $G \times 1$ vectors of the population regresion coefficients. Then, we maintain a strong exogeneity assumption that the remaining error terms v_0 and v_1 are independent of $\mathbf{v}_2$ and a parametric assumption that they have a bivariate normal distribution with the remaining error term v_3 with covariance ρ_0 and ρ_1, respectively:

$$D\left(\begin{array}{c} v_0 \\ v_3 \end{array}\middle| \mathbf{v}_2\right) \sim \text{Normal}\left[\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & \rho_0 \\ \rho_0 & 1 \end{array}\right)\right], \tag{17a}$$

$$D\left(\begin{array}{c} v_1 \\ v_3 \end{array}\middle| \mathbf{v}_2\right) \sim \text{Normal}\left[\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & \rho_1 \\ \rho_1 & 1 \end{array}\right)\right]. \tag{17b}$$

Again, assuming (v_1, v_3) and (v_0, v_3) are independent of $\mathbf{z}$ leads to an independence between $\mathbf{y}_2$ and the joint distribution of (v_0, v_3) and (v_1, v_3)

$$D\left(\begin{array}{c} v_0 \\ v_3 \end{array}\right) = D\left(\begin{array}{c} v_0 \\ v_3 \end{array}\middle| \mathbf{z}, \mathbf{v}_2\right) = D\left(\begin{array}{c} v_0 \\ v_3 \end{array}\middle| \mathbf{y}_2, \mathbf{z}, \mathbf{v}_2\right), \tag{18a}$$

$$D\left(\begin{array}{c} v_1 \\ v_3 \end{array}\right) = D\left(\begin{array}{c} v_1 \\ v_3 \end{array}\middle| \mathbf{z}, \mathbf{v}_2\right) = D\left(\begin{array}{c} v_1 \\ v_3 \end{array}\middle| \mathbf{y}_2, \mathbf{z}, \mathbf{v}_2\right). \tag{18b}$$

Then, rewrite model (15) in the treatment effect framework

$$y_1 = (1-y_3)\, y_1^{(0)} + y_3 y_1^{(1)} \tag{19}$$

$$y_1^{(0)} = 1\left[\mathbf{x_1}\boldsymbol{\beta}_0 + \mathbf{v}_2\boldsymbol{\theta}_0 + v_0 > 0\right], \tag{20}$$

$$y_1^{(1)} = 1\left[\mathbf{x_1}\boldsymbol{\beta}_1 + \mathbf{v}_2\boldsymbol{\theta}_1 + v_1 > 0\right], \tag{21}$$

$$y_3 = 1\left[\mathbf{z}\boldsymbol{\delta} + \mathbf{v}_2\boldsymbol{\eta} + v_3 > 0\right], \tag{22}$$

where $y_1^{(0)}$ is the potential outcome when the treatment y_3 equals zero and $y_1^{(1)}$ is the potential outcome when the treatment is one. The self-selection problem is represented by the non-zero correlation between the treatment indicator y_3 and the unobservables v_0 and v_1 in the potential outcomes. Those who self-select into treatment inherently have a different distribution of unobservables from those who do not.

To consistently estimate the parameters in this model, a simple three-step control function approach splits the above model into two Heckman sample selection models with sub-samples defined by the treatment status:

1. Using all observation, estimate (15b), the reduced forms for $\mathbf{y}_2$, by ordinary least squares (OLS), equation by equation, to obtain the residuals $\widehat{\mathbf{v}}_2 = \mathbf{y}_2 - \mathbf{z}\widehat{\Pi}$.

2. Since $y_1^{(1)}$ is observed only when $y_3 = 1$, jointly estimate (21) and (22), the binary outcome equation for $y_1^{(1)}$ and sample selection equation for indicator y_3, by MLE, replacing $\mathbf{v}_2$ with $\widehat{\mathbf{v}}_2$, to obtain $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\theta}}_1$.

3. Since $y_1^{(0)}$ is observed only when $y_3 = 0$, jointly estimate (20) and (22), the binary response model for $y_1^{(0)}$ and sample selection equation for indicator $1 - y_3$, by MLE, replacing $\mathbf{v}_2$ with $\widehat{\mathbf{v}}_2$, to obtain $\widehat{\boldsymbol{\beta}}_0$ and $\widehat{\boldsymbol{\theta}}_0$.

The above procedure is justified by splitting the objective function for the second-step estimation into two parts. Namely, solving

$$\begin{aligned}
&\max_{\mathbf{b}_0\in\mathbb{R}^{K_1},\mathbf{b}_1\in\mathbb{R}^{K_1},\mathbf{d}\in\mathbb{R}^{L},\mathbf{r}_0\in\mathbb{R}^{G},\mathbf{r}_1\in\mathbb{R}^{G},\mathbf{g}\in\mathbb{R}^{G},\rho_0\in\mathbb{R},\rho_1\in\mathbb{R}} E\left[\log P\left(y_1, y_3|\mathbf{y}_2,\mathbf{z},\mathbf{v}_2\right)\right] \\
= \; & E\Big[y_1 y_3 \log P\left(y_1^{(1)} = 1, y_3 = 1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) \\
&+ (1-y_1)\, y_3 \log P\left(y_1^{(1)} = 0, y_3 = 1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) \\
&+ y_1 (1-y_3) \log P\left(y_1^{(0)} = 1, y_3 = 0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) \\
&+ (1-y_1)(1-y_3) \log P\left(y_1^{(0)} = 0, y_3 = 0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right)\Big],
\end{aligned} \tag{23}$$

is equivalent to solving

$$
\begin{aligned}
& \max_{\mathbf{b}_1\in\mathbb{R}^{K_1},\mathbf{d}\in\mathbb{R}^{L},\mathbf{r}_1\in\mathbb{R}^{G},\mathbf{g}\in\mathbb{R}^{G},\rho_1\in\mathbb{R}} E\left[\log P\left(y_1, y_3|\mathbf{y}_2,\mathbf{z},\mathbf{v}_2\right)\right] \\
= \; & E\left[y_1^{(1)} y_3 \log P\left(y_1^{(1)}=1, y_3=1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right)\right. \\
& + \left(1-y_1^{(1)}\right) y_3 \log P\left(y_1^{(1)}=0, y_3=1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) \\
& \left. + \left(1-y_3\right)\log P\left(y_3=0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right)\right], \qquad (24)
\end{aligned}
$$

and

$$
\begin{aligned}
& \max_{\mathbf{b}_0\in\mathbb{R}^{K_1},\mathbf{d}\in\mathbb{R}^{L},\mathbf{r}_0\in\mathbb{R}^{G},\mathbf{g}\in\mathbb{R}^{G},\rho_0\in\mathbb{R}} E\left[\log P\left(y_1, y_3|\mathbf{y}_2,\mathbf{z},\mathbf{v}_2\right)\right] \\
= \; & E\left[y_1^{(0)} \left(1-y_3\right) \log P\left(y_1^{(0)}=1, y_3=0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right)\right. \\
& + \left(1-y_1^{(0)}\right) \left(1-y_3\right) \log P\left(y_1^{(0)}=0, y_3=0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) \\
& \left. + y_3\log P\left(y_3=1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right)\right], \qquad (25)
\end{aligned}
$$

where

$$P\left(y_1^{(1)}=1, y_3=1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) = \int_{-q_3}^{\infty} \Phi\left(d_1\right)\phi\left(\upsilon_3\right)d\upsilon_3 \qquad (26)$$

$$P\left(y_1^{(1)}=0, y_3=1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) = \int_{-q_3}^{\infty} \left[1-\Phi\left(d_1\right)\right]\phi\left(\upsilon_3\right)d\upsilon_3 \qquad (27)$$

$$P\left(y_1^{(0)}=1, y_3=0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) = \int_{-\infty}^{-q_3} \Phi\left(d_0\right)\phi\left(\upsilon_3\right)d\upsilon_3 \qquad (28)$$

$$P\left(y_1^{(0)}=0, y_3=0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) = \int_{-\infty}^{-q_3} \left[1-\Phi\left(d_0\right)\right]\phi\left(\upsilon_3\right)d\upsilon_3 \qquad (29)$$

$$P\left(y_3=1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) = \Phi\left(q_3\right) \qquad (30)$$

$$P\left(y_3=0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2\right) = 1-\Phi\left(q_3\right) \qquad (31)$$

$$d_1 \equiv \frac{\mathbf{x_1}\mathbf{b}_1+\mathbf{v}_2\mathbf{r}_1+\rho_1\upsilon_3}{\sqrt{1-\rho_1^2}} \qquad (32)$$

$$d_0 \equiv \frac{\mathbf{x_1}\mathbf{b}_0+\mathbf{v}_2\mathbf{r}_0+\rho_0\upsilon_3}{\sqrt{1-\rho_0^2}} \qquad (33)$$

$$q_3 \equiv \mathbf{z}\mathbf{d}+\mathbf{v}_2\mathbf{g}. \qquad (34)$$

Similar to (12), the ASF for the endogenous switching model is a combination of the ASFs for the two regimes:

$$\text{ASF}(\mathbf{x}_1) = \int_{-\infty}^{\infty} \left[y_3 \Phi\left(\mathbf{x_1}\boldsymbol{\beta}_1 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_1\right) + (1 - y_3)\,\Phi\left(\mathbf{x_1}\boldsymbol{\beta}_0 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_0\right)\right] \phi\left(\boldsymbol{v}_{i2}\right) d\boldsymbol{v}_{i2}. \tag{35}$$

APEs for a continuous EEV y_2 and binary EEV y_3 are defined as follows respectively:

$$\begin{aligned}
\text{APE}_{y_2}(\mathbf{x}_1) &= \int_{-\infty}^{\infty} \Big[\beta_{y_2^{(1)}} y_3 \phi\left(\mathbf{x_1}\boldsymbol{\beta}_1 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_1\right) \\
&\quad + \beta_{y_2^{(0)}} (1 - y_3)\, \phi\left(\mathbf{x_1}\boldsymbol{\beta}_0 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_0\right)\Big] \phi\left(\boldsymbol{v}_{i2}\right) d\boldsymbol{v}_{i2}, && \text{(36a)} \\
\text{APE}_{y_3}(\mathbf{x}_1) &= \int_{-\infty}^{\infty} \left[\Phi\left(\mathbf{x_1}\boldsymbol{\beta}_1 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_1\right) - \Phi\left(\mathbf{x_1}\boldsymbol{\beta}_0 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_0\right)\right] \phi\left(\boldsymbol{v}_{i2}\right) d\boldsymbol{v}_{i2}, && \text{(36b)}
\end{aligned}$$

where $\beta_{y_2^{(1)}}$ is the coefficient for y_2 in (21) and $\beta_{y_2^{(0)}}$ is the coefficient for y_2 in (20).

Notice that the APE for a binary exogenous variable z_1 is defined nontrivially as

$$\begin{aligned}
\text{APE}_{z_1}(\mathbf{x}_1) = \int_{-\infty}^{\infty} &\Big\{ y_3 \Big[\Phi\left(\mathbf{x}_1^{(1)}\boldsymbol{\beta}_1 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_1\right) - \Phi\left(\mathbf{x}_1^{(0)}\boldsymbol{\beta}_1 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_1\right)\Big] \\
&+ (1 - y_3) \Big[\Phi\left(\mathbf{x}_1^{(1)}\boldsymbol{\beta}_0 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_0\right) - \Phi\left(\mathbf{x}_1^{(0)}\boldsymbol{\beta}_0 + \boldsymbol{v}_{i2}\boldsymbol{\theta}_0\right)\Big]\Big\} \phi\left(\boldsymbol{v}_{i2}\right) d\boldsymbol{v}_{i2}, \qquad \text{(37)}
\end{aligned}$$

where $\mathbf{x}_1^{(1)}$ denotes explanatory variables at a particular fixed value with $z_1 = 1$ and $\mathbf{x}_1^{(0)}$ denotes the same fixed value of the explanatory variables except that now $z_1 = 0$.

Correspondingly, a consistent estimate of the APEs is a sample analog of (36a) and (36b) with consistent estimates for the parameters plugged in:

$$\begin{aligned}
\widehat{\text{APE}}_{y_2}(\mathbf{x}_1) &= N^{-1}\sum_{i=1}^{N} \Big[\widehat{\beta}_{y_2^{(1)}} y_3 \phi\left(\mathbf{x_1}\widehat{\boldsymbol{\beta}}_1 + \widehat{\mathbf{v}}_{i2}\widehat{\boldsymbol{\theta}}_1\right) \\
&\quad + \widehat{\beta}_{y_2^{(0)}} (1 - y_3)\, \phi\left(\mathbf{x_1}\widehat{\boldsymbol{\beta}}_0 + \widehat{\mathbf{v}}_{i2}\widehat{\boldsymbol{\theta}}_0\right)\Big], && \text{(38a)} \\
\widehat{\text{APE}}_{y_3}(\mathbf{x}_1) &= N^{-1}\sum_{i=1}^{N} \Big[\Phi\left(\mathbf{x}_1\widehat{\boldsymbol{\beta}}_1 + \widehat{\mathbf{v}}_{i2}\widehat{\boldsymbol{\theta}}_1\right) - \Phi\left(\mathbf{x}_1\widehat{\boldsymbol{\beta}}_0 + \widehat{\mathbf{v}}_{i2}\widehat{\boldsymbol{\theta}}_0\right)\Big]. && \text{(38b)}
\end{aligned}$$

As before, instead of deriving complicated analytical formulas for standard errors for estimates of APEs, bootstrap standard errors can be easily applied to account for the sampling variation in the generated regressor $\widehat{\mathbf{v}}$.

Despite the fact that the switching model brings in additional flexibility by allowing the structural error $u \equiv (1 - y_3)\, u_0 + y_3 u_1$ to depend not only on $\mathbf{v}_2$ but also on interactions between $\mathbf{v}_2$ and y_3, assuming that the reduced forms for $\mathbf{y}_2$ remain unchanged across two regimes is restrictive in empirical applications.

3 Test for Endogeneity from a Binary Explanatory Variable

This section focuses on variable addition tests for additional endogeneity from a binary explanatory variable, conditioning on $\widehat{\mathbf{v}}_2$, the residuals from reduced forms for continuous EEVs. As we have seen in equations (1) and (15), the only consistent approach to deal with a binary EEV is to make distributional assumptions and conduct a joint estimation. In real applications, we always want to avoid a joint MLE estimation due to its sensitivity to the distributional assumption and computational difficulty in arriving at a numerical solution. A variable addition test (VAT), as proposed in Wooldridge (2014), helps us determine whether such a joint estimation is necessary by testing on generalized residuals before proceeding to a joint estimation. Especially if we have already controlled for endogeneity from other continuous EEVs by conditioning on $\widehat{\mathbf{v}}_2$, the generalized residual is less likely to be correlated with the remaining unobservable. The following shows that the VAT on the generalized residual is a valid test for endogeneity from a binary explanatory variable because it is asymptotically equivalent to an LM test under the null hypothesis of no endogeneity.

More formally, in the basic model (1), we are interested in testing the following null hypothesis: $\mathbf{H}_0 : \rho = 0$. First, we begin by showing an infeasible Lagrange multiplier (score) test that has the asymptotic distribution of χ^2_1. Then, we show that, conditional on $\mathbf{v}_2$, a VAT test of the generalized residual is asymptotically equivalent to the infeasible LM test and thus has the same asymptotic χ^2_1 distribution. In practice, in order to account for the sampling error in $\widehat{\mathbf{v}}_2$, we bootstrap the two-step procedure to obtain the p-value of the test. Let $\boldsymbol{\gamma} \equiv (\boldsymbol{\beta}, \boldsymbol{\theta})$ and $\mathbf{w}_i \equiv (\mathbf{x_{i1}}, \mathbf{v}_{i2})$. Let $\widetilde{d}_i$ be d_i in (8) evaluated at $\rho = 0$ and $\widetilde{\boldsymbol{\gamma}}$ be the estimates of $\boldsymbol{\gamma}$ obtained from the restricted model. The restricted model is one where $\rho = 0$ so we treat y_3 as an exogenous explanatory variable. Let $\widehat{q}_{3i}$ be q_{3i} in (9) evaluated at the parameters $\left(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\eta}}\right)$ from a reduced-form probit estimation.

As in Semykina and Wooldridge (2017), using the likelihood function $L_i \equiv P\left(y_{i1}, y_{i3} | \mathbf{y}_{i2}, \mathbf{z}_i, \mathbf{v}_{i2}\right)$ for one observation, the LM statistic plugs the estimates from the restricted model into the score from the unrestricted model:

$$LM = \left(\sum_{i=1}^{N} \widetilde{\mathbf{S}}_{i,\rho}\right)' \widetilde{\mathbf{A}}^{22} \left[\widetilde{\mathbf{V}}_{22}\right]^{-1} \widetilde{\mathbf{A}}^{22} \left(\sum_{i=1}^{N} \widetilde{\mathbf{S}}_{i,\rho}\right) / N, \quad (39)$$

where $\widetilde{\mathbf{S}}_{i,\rho} \equiv \frac{\partial \ln L_i}{\partial \rho} |_{\boldsymbol{\gamma}=\widetilde{\boldsymbol{\gamma}}, \rho=0} = \frac{y_{i1} - \Phi\left(\widetilde{d}_i\right)}{\Phi\left(\widetilde{d}_i\right)\left[1-\Phi\left(\widetilde{d}_i\right)\right]} \phi\left(\widetilde{d}_i\right) \widehat{gr}_{i3}$

$\widetilde{\mathbf{A}}$

$$\equiv -\frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N} E\left(\frac{\partial^2 \ln L_i}{\partial\gamma\partial\gamma'}|y_{i3},\mathbf{y}_{i2},\mathbf{z}_i,\mathbf{v}_{i2}\right)|_{\gamma=\widetilde{\gamma},\rho=0} & \sum_{i=1}^{N} E\left(\frac{\partial^2 \ln L_i}{\partial\rho\partial\gamma'}|y_{i3},\mathbf{y}_{i2},\mathbf{z}_i,\mathbf{v}_{i2}\right)|_{\gamma=\widetilde{\gamma},\rho=0} \\ \sum_{i=1}^{N} E\left(\frac{\partial^2 \ln L_i}{\partial\gamma\partial\rho}|y_{i3},\mathbf{y}_{i2},\mathbf{z}_i,\mathbf{v}_{i2}\right)|_{\gamma=\widetilde{\gamma},\rho=0} & \sum_{i=1}^{N} E\left(\frac{\partial^2 \ln L_i}{\partial\rho\partial\rho}|y_{i3},\mathbf{y}_{i2},\mathbf{z}_i,\mathbf{v}_{i2}\right)|_{\gamma=\widetilde{\gamma},\rho=0} \end{pmatrix}$$

$$= \frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N} \frac{\phi\left(\widetilde{d}_i\right)^2}{\Phi\left(\widetilde{d}_i\right)\left[1-\Phi\left(\widetilde{d}_i\right)\right]}\mathbf{w}_i'\mathbf{w}_i & \sum_{i=1}^{N} \frac{\phi\left(\widetilde{d}_i\right)^2}{\Phi\left(\widetilde{d}_i\right)\left[1-\Phi\left(\widetilde{d}_i\right)\right]}\mathbf{w}_i'\widehat{gr}_{i3} \\ \sum_{i=1}^{N} \frac{\phi\left(\widetilde{d}_i\right)^2}{\Phi\left(\widetilde{d}_i\right)\left[1-\Phi\left(\widetilde{d}_i\right)\right]}\widehat{gr}_{i3}\mathbf{w}_i & \sum_{i=1}^{N} \frac{\phi\left(\widetilde{d}_i\right)^2}{\Phi\left(\widetilde{d}_i\right)\left[1-\Phi\left(\widetilde{d}_i\right)\right]}\widehat{gr}_{i3}^2 \end{pmatrix}$$

$$\begin{aligned} \widetilde{\mathbf{A}}^{-1} &= \begin{pmatrix} \widetilde{\mathbf{A}}^{11} & \widetilde{\mathbf{A}}^{12} \\ \widetilde{\mathbf{A}}^{21} & \widetilde{\mathbf{A}}^{22} \end{pmatrix} \\ \widetilde{\mathbf{V}} &= \widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{B}}\widetilde{\mathbf{A}}^{-1} = \begin{pmatrix} \widetilde{\mathbf{V}}_{11} & \widetilde{\mathbf{V}}_{12} \\ \widetilde{\mathbf{V}}_{21} & \widetilde{\mathbf{V}}_{22} \end{pmatrix} \\ \widetilde{\mathbf{B}} &\equiv \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\mathbf{S}}_{i,\rho}\widetilde{\mathbf{S}}_{i,\rho}'\right) \\ \widehat{gr}_{i3} &\equiv y_{i3}\frac{\phi\left(\widehat{q}_{3i}\right)}{\Phi\left(\widehat{q}_{3i}\right)} - \left(1-y_{i3}\right)\frac{\phi\left(-\widehat{q}_{3i}\right)}{\Phi\left(-\widehat{q}_{3i}\right)}. \end{aligned} \quad (40)$$

Matrix $\widetilde{\mathbf{A}}$ above is an estimator of the expected value of the negative Hessian matrix that uses the expected Hessian form. The outer product of scores or usual Hessian form of the matrix could be used. $\widehat{gr}_{i3}$ is a consistent estimator of $gr_{i3} \equiv E\left(v_{i3}|y_{i3},\mathbf{y}_{i2},\mathbf{z}_i,\mathbf{v}_{i2}\right)$.

A VAT can be carried out by the following procedure of testing on generalized residuals:

1. Use OLS to estimate the reduced-form equations for $\mathbf{y}_{i2}$ (1b) to obtain $\widehat{\mathbf{v}}_{i2}$.

2. Use probit to estimate the augmented reduced-form for y_{i3} in (5b), and construct $\widehat{gr}_{i3}$ according to the formula in equation (40).

3. Augment equation (5a) by $\widehat{gr}_{i3}$ and estimate by probit. Use the t statistics for testing single hypotheses.

Under the null hypothesis the coefficient on $\widehat{gr}_{i3}$ is zero, and so estimation of the parameters in $\widehat{gr}_{i3}$ does not affect the $\sqrt{N}$-asymptotic distribution of the test statistic. There is no need to account for the first-step estimation of $\widehat{gr}_{i3}$ when performing the test. However, as in Wooldridge (2010, Section 12.5.2), we need to adjust for the first-step estimation of $\mathbf{v}_{i2}$, by stacking the moment conditions or by bootstrapping the two-step procedure.

The following shows that, conditional on $\mathbf{v}_{i2}$, the variable addition test is asymptotically equivalent to the LM test. Write the second-step log likelihood function as

$$L_i = \Phi\left(\mathbf{x}_{i1}\boldsymbol{\beta} + \mathbf{v}_{i2}\boldsymbol{\theta}+\tau gr_{i3}\right)^{y_{i1}} \left[1 - \Phi\left(\mathbf{x}_{i1}\boldsymbol{\beta} + \mathbf{v}_{i2}\boldsymbol{\theta}+\tau gr_{i3}\right)\right]^{1-y_{i1}}. \tag{41}$$

As mentioned above, we ignore the fact that gr_3 is estimated consistently at the first step. The score vector of (41) is

$$\mathbf{S}_i = \begin{pmatrix} \frac{\partial \ln L_i}{\partial \boldsymbol{\gamma}} \\ \frac{\partial \ln L_i}{\partial \tau} \end{pmatrix} = \frac{y_{i1} - \Phi\left(\mathbf{w}_i\boldsymbol{\gamma}+\tau gr_3\right)}{\Phi\left(\mathbf{w}_i\boldsymbol{\gamma}+\tau gr_3\right)\left[1-\Phi\left(\mathbf{w}_i\boldsymbol{\gamma}+\tau gr_3\right)\right]}\phi\left(\mathbf{w}_i\boldsymbol{\gamma}+\tau gr_3\right)\begin{pmatrix} \mathbf{w}_i \\ gr_{i3} \end{pmatrix}. \tag{42}$$

Summing the score vector over all i and using a mean-value expansion about the true parameter vector gives

$$N^{-1/2}\sum_{i=1}^{N}\widehat{\mathbf{S}}_i = N^{-1/2}\sum_{i=1}^{N}\mathbf{S}_i - \mathbf{A}\sqrt{N}\begin{pmatrix} \widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} \\ \widehat{\tau}-\tau \end{pmatrix} + o_p(1) = 0 \tag{43}$$

where $\widehat{\mathbf{S}}_i$ is the score vector evaluated at the estimated parameters $\left(\widehat{\gamma}', \widehat{\tau}\right)'$, and $\mathbf{A}$ is the expected value of the negative Hessian matrix:

$$\sqrt{N}\begin{pmatrix} \widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} \\ \widehat{\tau}-\tau \end{pmatrix} = \mathbf{A}^{-1}\left[N^{-1/2}\sum_{i=1}^{N}\mathbf{S}_i\right] + o_p(1). \tag{44}$$

When testing $\mathbf{H}_0 : \tau = 0$, the robust Wald test statistic is given by

$$W = (\widehat{\tau}-\tau)'\left(\widehat{\mathbf{V}}_{22}/N\right)^{-1}(\widehat{\tau}-\tau) = \sqrt{N}(\widehat{\tau}-\tau)'\,\widehat{\mathbf{V}}_{22}^{-1}\sqrt{N}(\widehat{\tau}-\tau) \tag{45}$$

where

$$\widehat{\mathbf{V}} = \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1} = \begin{pmatrix} \widehat{\mathbf{V}}_{11} & \widehat{\mathbf{V}}_{12} \\ \widehat{\mathbf{V}}_{21} & \widehat{\mathbf{V}}_{22} \end{pmatrix}, \tag{46}$$

$$\widehat{\mathbf{B}} = \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\mathbf{S}}_{i,\rho}\widetilde{\mathbf{S}}_{i,\rho}'\right), \tag{47}$$

$$\widehat{\mathbf{A}} = \frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N}\frac{\phi(\widehat{p}_i)^2}{\Phi(\widehat{p}_i)[1-\Phi(\widehat{p}_i)]}\mathbf{w}_i'\mathbf{w}_i & \sum_{i=1}^{N}\frac{\phi(\widehat{p}_i)^2}{\Phi(\widehat{p}_i)[1-\Phi(\widehat{p}_i)]}\mathbf{w}_i'\widehat{gr}_{i3} \\ \sum_{i=1}^{N}\frac{\phi(\widehat{p}_i)^2}{\Phi(\widehat{p}_i)[1-\Phi(\widehat{p}_i)]}\widehat{gr}_{i3}\mathbf{w}_i & \sum_{i=1}^{N}\frac{\phi(\widehat{p}_i)^2}{\Phi(\widehat{p}_{ii})[1-\Phi(\widehat{p}_i)]}\widehat{gr}_{i3}^2 \end{pmatrix}, \tag{48}$$

$$\widehat{p}_i = \mathbf{w}_i\widehat{\boldsymbol{\gamma}}+\widehat{\tau}\widehat{gr}_3, \tag{49}$$

$$\widehat{\mathbf{A}}^{-1} \xrightarrow{p} \mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{pmatrix}. \tag{50}$$

So the Wald statistic can also be written as

$$W=\left(\sum_{i=1}^{N}\mathbf{S}_{i},\tau\right)'\mathbf{A}^{22}\widehat{\mathbf{V}}_{22}^{-1}\mathbf{A}^{22}\left(\sum_{i=1}^{N}\mathbf{S}_{i},\tau\right)/N. \tag{51}$$

Under the null of no selection bias ($\tau=0,\rho=0$), the score and Hessian matrices used in (39) and (51) are the same when evaluated at the true parameter values. When the null is true, $\widehat{\tau}\xrightarrow{p}0$, $\sqrt{N}\left(\widehat{\gamma}-\gamma\right)$ and $\sqrt{N}\left(\widetilde{\gamma}-\gamma\right)$ converge in distribution. Therefore, $LM-W\xrightarrow{p}0$, so the tests are asymptotically equivalent. Through bootstrapping the two-step procedure, the p-value for the test can be obtained .

4 Quasi-LIML and Fractional Response

Based on the literature of Quasi-MLE (White, 1982), the findings above carry through if f_1 is a fractional response with a conditional mean that happens to have a probit form. The key insight from quasi-likelihood estimation is that we do not need to know the true distribution of the entire model to obtain consistent parameter estimates. This likelihood function could also be applied to the case where f_1 is a fractional response, as long as we model the conditional mean of f_1 to have a probit form. With the Bernoulli distribution being in the linear exponential family, quasi-LIML can identify parameters in a correctly specified conditional mean regardless of misspecification in other aspects of the distribution.

Namely,

$$E\left(f_1|\mathbf{x}_1,c_1\right)=\Phi\left(\mathbf{x}_1\boldsymbol{\beta}+c_1\right) \tag{52a}$$

$$\mathbf{y}_2=\mathbf{z}\Pi+\mathbf{v}_2 \tag{52b}$$

$$y_3=1\left[\mathbf{z}\boldsymbol{\delta}+u_3\geq 0\right], \tag{52c}$$

where c_1 is an omitted variable thought to be correlated with $\mathbf{y}_2$ and y_3. By assuming c_1 follows a joint normality distribution with $\mathbf{v}_2$ and u_3, linear projections of c_1 and u_3 onto $\mathbf{v}_2$ have the following error form:

$$c_1=\mathbf{v}_2\boldsymbol{\theta}+a_1 \tag{53a}$$

$$u_3=\mathbf{v}_2\boldsymbol{\eta}+v_3 \tag{53b}$$

where $\boldsymbol{\theta}\equiv E\left(\mathbf{v}_2'\mathbf{v}_2\right)^{-1}E\left(\mathbf{v}_2'c_1\right)$ and $\boldsymbol{\eta}\equiv E\left(\mathbf{v}_2'\mathbf{v}_2\right)^{-1}E\left(\mathbf{v}_2'u_3\right)$ are the $G\times 1$ vectors of the population regresion coefficients. Plugging the linear projections (53a) and (53b) back into (52a) and (52c), we have an augmented equation for the conditional mean of f_1 and the reduced form for y_3:

$$E(f_1|\mathbf{x}_1,\mathbf{v}_2,a_1) = \Phi(\mathbf{x}_1\boldsymbol{\beta}+\mathbf{v}_2\boldsymbol{\theta}+a_1) \tag{54a}$$

$$y_3 = 1[\mathbf{z}\boldsymbol{\delta}+\mathbf{v}_2\boldsymbol{\eta}+v_3 \geq 0], \tag{54b}$$

where a_1 is the remaining unobservable factor that, after conditioning on $\mathbf{v}_2$, captures the additional endogeneity from y_3 through v_3. Again, assume a joint normality assumption between a_1 and v_3 as

$$D\left(\begin{array}{c} a_1 \\ v_3 \end{array}\middle| \mathbf{v}_2\right) \sim \text{Normal}\left[\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} \sigma_a^2 & \rho\sigma_a \\ \rho\sigma_a & 1 \end{array}\right)\right], \tag{55}$$

where $\sigma_a^2 \equiv$Var(a_1) and ρ is the covariance. Further averaging out the unobservable a_1, the conditional mean of the joint distribution of f_1 and y_3 has the exact same form as the probit model with many continuous EEVs and one binary EEV in (1).

$$\begin{aligned} E(f_1, y_3=1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2) &= E(y_1, y_3=1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2) = P(y_1=1, y_3=1|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2) \\ &= \int_{-q_3}^{\infty}\Phi(d)\,\phi(v_3)\,dv_3, \end{aligned} \tag{56a}$$

$$\begin{aligned} E(f_1, y_3=0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2) &= E(y_1, y_3=0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2) = P(y_1=1, y_3=0|\mathbf{z},\mathbf{v}_2,\mathbf{y}_2) \\ &= \int_{-\infty}^{-q_3}\Phi(d)\,\phi(v_3)\,dv_3, \end{aligned} \tag{56b}$$

where

$$d \equiv \frac{\mathbf{x_1 b}+\mathbf{v}_2\mathbf{r}+\rho v_3}{\sqrt{1+(1-\rho^2)\sigma_a^2}}, \tag{57a}$$

$$q_3 \equiv \mathbf{zd}+\mathbf{v}_2\mathbf{g}. \tag{57b}$$

Because the Bernoulli log likelihood belongs to the linear exponential family, the solution from the following maximization problem identifies $(\boldsymbol{\beta}, \boldsymbol{\theta})$:

$$\begin{aligned}
& \max_{\mathbf{b}_1 \in \mathbb{R}^{K_1}, \mathbf{d} \in \mathbb{R}^L, \mathbf{r}_1 \in \mathbb{R}^G, \mathbf{g} \in \mathbb{R}^G, \rho_1 \in \mathbb{R}} E\left[\log P\left(f_1, y_3 | \mathbf{y}_2, \mathbf{z}, \mathbf{v}_2\right)\right] \\
= \; & E\left[f_1 y_3 \log \int_{-q_3}^{\infty} \Phi\left(d\right) \phi\left(v_3\right) dv_3 \right. \\
& + \left(1 - f_1\right) y_3 \log \int_{-q_3}^{\infty} \left[1 - \Phi\left(d\right)\right] \phi\left(v_3\right) dv_3 \\
& + f_1 \left(1 - y_3\right) \log \int_{-\infty}^{-q_3} \Phi\left(d\right) \phi\left(v_3\right) dv_3 \\
& \left. + \left(1 - f_1\right)\left(1 - y_3\right) \log \int_{-\infty}^{-q_3} \left[1 - \Phi\left(d\right)\right] \phi\left(v_3\right) dv_3 \right]. \qquad (58)
\end{aligned}$$

5 Monte Carlo Simulations

In this section, six Monte Carlo experiments are conducted to compare the finite sample behavior of different estimators for the binary response model with both continuous and discrete EEVs. The six Monte Carlo experiments fall into two designs. In the first design, error terms $(u_1, \mathbf{v}_2, u_3)$ are jointly normally distributed. In the second design, conditional on $\mathbf{v}_2$, u_1 and u_3 are assumed to have bivariate normal distribution. Each design contains three cases—namely, a just-identified case, an over-identified case, and an endogenous switching case. Nine estimators are compared in each case, four estimators assuming a linear probability model for the binary outcome and the other five estimators acknowledging the nonlinear functional form.

More specifically, in the first design of joint normality, the DGP for the just identified (Just ID) case is

$$y_1 = 1\left[-y_2 + y_3 + 0.3z_1 + 0.3z_2 + u_1 > 0\right] \qquad (59a)$$

$$y_2 = 0.1z_1 + 0.2z_2 + 0.1z_3 + z_4 + v_2 \qquad (59b)$$

$$y_3 = 1\left[0.2z_1 + 0.1z_2 + z_3 + 0.1z_4 + 0.5v_2 + v_3 > 0\right], \qquad (59c)$$

where

$$u_1 = 0.5v_2 + 0.5v_3 + r_1, r_1 \sim \text{Normal}\left(0, 0.5\right), \qquad (60)$$

$$\begin{pmatrix} u_1 \\ v_2 \\ v_3 \end{pmatrix} \sim \text{Normal}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{pmatrix}\right]. \qquad (61)$$

As before, the first equation (59a) is a structural equation, while (59b) and (59c) are reduced form equations. Note that the continuous EEV y_2 and the binary EEV y_3 are generated to have coefficients of opposite signs. The exogenous variables are generated as:

$$\begin{aligned} z_1 &\sim \text{Normal}(0,1), e_2 \sim \text{Normal}(0,1), z_2 = 1[e_2 > 0] \\ z_3 &\sim \text{Normal}(0,1), e_4 \sim \text{Normal}(0,1), z_4 = 1[e_3 > 0]. \end{aligned}$$

where the continuous exogenous variable z_3 is the instrument mainly for the binary EEV y_3, and the binary exogenous variable z_4 is the instrument mainly for the continuous EEV y_2. Both z_3 and z_4 are excluded from the structural equation.

In this DGP, the true ASF is defined as

$$\text{ASF}(\mathbf{x}_1) = \Phi(-y_2 + y_3 + 0.3z_1 + 0.3z_2). \tag{62}$$

The over identified (Over ID) case has the same parameters except that we have two additional instruments, z_5 and z_6, where

$$z_5 \sim \text{Normal}(0,1), e_6 \sim \text{Normal}(0,1), z_6 = 1[e_6 > 0].$$

The continuous exogenous variable z_5 is mainly for the continuous EEV y_2, and the binary exogenous variable z_6 is mainly for the binary EEV y_3. The true ASF remains the same as in (62).

In the endogenous switching case, the correlations between the reduced form errors v_2 and v_3 and the structural errors u_0 and u_1 are designed to have opposite signs across regimes, namely,

$$\begin{aligned} y_1^{(1)} &= 1[-y_2 + y_3 + 0.3z_1 + 0.3z_2 + u_1 > 0] \\ y_1^{(0)} &= 1[0.3y_2 + y_3 - 0.5z_1 + 0.1z_2 + u_0 > 0] \\ y_2 &= 0.1z_1 + 0.2z_2 + 0.1z_3 + z_4 + v_2 \\ y_3 &= 1[0.2z_1 + 0.1z_2 + z_3 + 0.1z_4 + 0.5v_2 + v_3 > 0], \end{aligned} \tag{63}$$

where in addition to (60) and (61) we have

$$u_0 = -0.5v_2 - 0.5v_3 + r_1, r_1 \sim \text{Normal}(0, 0.5) \tag{64}$$

$$\begin{pmatrix} u_0 \\ v_2 \\ v_3 \end{pmatrix} \sim \text{Normal}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & 0 \\ -0.5 & 0 & 1 \end{pmatrix}\right]. \tag{65}$$

The ASF in this case is

$$\text{ASF}(\mathbf{x}_1) = y_3\Phi(-y_2 + y_3 + 0.3z_1 + 0.3z_2) + (1 - y_3)\Phi(0.3y_2 + y_3 - 0.5z_1 + 0.1z_2). \tag{66}$$

In design 2, the parameterizations are the same as in design 1 except that we assume that v_2 follows a demeaned χ_1^2 distribution with one degree of freedom

$$v_2 \sim \chi_1^2 - 1. \tag{67}$$

Conditional on v_2, u_0 and v_3 still follow a bivariate normal distribution.

In all experiments, the number of replications is 1000, and the results of the experiments are presented for sample sizes of 1000, 3000 and 5000. Tables 1 and 2 report biases and the root mean squared errors (RMSEs) for estimators of marginal effects for y_2 and y_3, respectively. Figures 1 and 2 depict the empirical distributions of estimators of marginal effects for y_2 with sample size of 1000 under design 1 and design 2, respectively. Similarly, Figures 3 and 4 depict the empirical distributions of estimators of marginal effects for y_3 with sample size of 1000 under designs 1 and 2, respectively.

For each of the above cases, coefficients of linear probability models and APEs of probit models are compared. Further, for probit models, joint estimation methods with the binary EEV are compared with two-step estimation methods with control function terms (residuals or generalized residuals) plugged in. In addition, a switching version of each model is considered to account for the endongenous switching case. More specifically, CF Biprobit is the control function approach inserting the first-stage residual from the reduced-form estimation of y_2 into the second-step joint estimation between y_1 and y_3. CF Biprobit Switching performs the Heckman probit with sample selection for $y_1^{(1)}$ and $y_1^{(0)}$ separately using y_3 as a sample selection indicator. CF Probit avoids the joint estimation with y_3 by inserting a generalized residual from y_3 as a proxy for endogeneity given the residual from y_2. CF Probit Switching performs the CF Probit separately for sub-samples defined by y_3. CF Linear inserts a residual from y_2 and a generalized residual from y_3 into the linear probability model for y_1. CF Linear Switching allows for a full set of interactions between y_3 and other observables and unobservables in the linear probability model. Usual 2SLS uses linear probability models for both y_1 and y_3 and applies the usual two-step IV estimation. Optimal IV uses predicted values from reduced forms for y_1 and y_3 as instruments for a linear probability model of y_1. A probit model is used to predict y_3. Joint MLE is a full joint estimation of y_1, y_2 and y_3.

Under joint normality as in Figure 1, CF Biprobit and Joint MLE are the consistent estimators in the Just ID case and Over ID case while CF Biprobit Switching is the consistent estimator in the Switching case. Their empirical distributions are centered around the true APE depicted by the red vertical line. Besides these consistent estimators, approximations provided by the CF Probit (or CF Probit Switching in the Switching case) outperform, to a great extent, the approximations provided by the linear probability estimators such as CF Linear (or CF Linear Switching in the Switching case), Usual 2SLS and Optimal IV. In fact, in the Switching case, CF Probit Switching and CF Biprobit Switching (the consistent estimator in this case) seem to completely overlap with each other, suggesting a negligible amount of bias. In the Just ID case and Over ID case, CF Probit has a mild amount of upward bias and a slightly lower peak than CF Biprobit and Joint MLE. In contrast, approximations provided by linear probability model estimators (CF Linear, Usual 2SLS, Optimal IV and CF Linear Switching) have a significant amount of downward bias in all cases. The differences in bias within the linear probability model estimators are not noticeable: they all seem to cluster together. In the Switching case, linear probability model estimators are similar to the misspecified CF Biprobit and Joint MLE estimators with a similar amount of downward bias. When CF Biprobit and Joint MLE are consistent, they still completely overlap with each other. This happens not only in the Just ID case but also in the Over ID case, suggesting a negligible amount of efficiency loss by carrying out a two-step procedure. CF Biprobit Switching and CF Probit Switching, however, suffer a slightly flatter peak compared to their counterpart non-switching estimators (CF Biprobit and CF Probit) in the Just ID case and Over ID case, indicating an efficiency loss from a more complex parameterization.

When the error terms follow a conditional normality, the estimators for the marginal effects of y_2 have fairly different finite sample behaviors from those under joint normality. As reflected in Figure 2, Joint MLE lacks robustness and is no longer the consistent estimator in any case. As before, CF Biprobit is the consistent estimator in the Just ID case and Over ID case while CF Biprobit Switching is the consistent estimator in the Switching case. Approximations provided by the CF Probit (or CF Probit Switching in the Switching case) are still the best: they almost overlap with those provided by CF Biprobit (or CF Biprobit Switching in the Switching case), the consistent estimator. Joint MLE is biased upwards to a noticeable degree in the Just ID case and Over ID case. In the Switching case, misspecified Joint MLE is biased downward and is similar to other inconsistent estimators like CF Biprobit and estimators using linear probability models (CF Linear, Usual 2SLS and Optimal IV). CF Linear Switching performs mildly better than other linear estimators in the Switching case but is still more biased compared to the consistent estimator. Overall, linear probability estimators continue to perform poorly in all cases: they are biased far downwards. The CF Biprobit Switching and CF Probit Switching still lead to efficiency loss indicated by flatter peaks in the Just ID case and Over ID case.

Under joint normality, marginal effects of y_3 follow a similar pattern as that of y_2, with some minor differences. As in Figure 3, CF Biprobit and Joint MLE still overlap with each other in all cases, whether as consistent estimators in the Just ID case and Over ID case, or as misspecified estimators in the Switching case. The approximation provided by CF Probit (or CF Probit Switching in the Switching case) is still the best but with a flatter peak than those in Figure 1. The linear probability estimators are biased upwards. The differences in empirical distributions for the linear estimators are more pronounced in the binary EEV y_3 than for continuous EEV y_2. Particularly, CF Linear using the generalized residual is no longer close to Usual 2SLS using linear probability model for y_3. In the Switching case, linear probability estimators all lie between the CF Biprobit Switching and CF Biprobit with varying degrees of bias and precision.

Under conditional normality, marginal effects of y_3 are depicted in Figure 4. Like estimators in Figure 3, they have identical patterns with their counterparts for y_2. More specifically, Joint MLE is biased downwards in the Just ID case and Over ID case but biased upwards in the Switching case. Similarly, linear probability estimators are biased upwards rather than downwards. CF Probit (or CF Probit Switching in the Switching case) still provides the best approximations in all cases, significantly better than linear probability estimators.

Table 1 and Table 2 report the bias and RMSE of y_2 and y_3 for all the estimators in the six cases, respectively. Despite the difference in sign and magnitude, the patterns of estimators for y_2 and y_3 are similar. Methods using CF approaches are listed in Column (1) through (6), followed by conventional methods like IV 2SLS, Opt. IV 2SLS and Joint MLE from Column (7) to (9). With the increase of sample size from 1000 to 5000, the bias of CF Biprobit in the Just ID case and Over ID case (or CF Biprobit Switching in the Switching case) shrinks drastically to zero. Their RMSEs also decrease by about half at the same time. The bias of CF Probit in the Just ID case and Over ID case (or CF Probit Switching in the

Switching case) is small at sample size of 1000 but shrink by a less magnititude as sample size increases to 5000. The RMSEs of CF Probit or CF Probit Switching also decrease by about half as the sample size increases. The bias of the linear probability estimators (CF 2SLS, CF 2SLS Switching, IV 2SLS and Opt. IV 2SLS) and misspecified Joint MLE, however, is huge to start. Their RMSEs does not decrease but even increases in some cases as the sample size increase. The RMSEs of the linear probability estimators also do not decrease.

In summary, the Monte Carlo results show that CF Biprobit does not lose efficiency compared to Joint MLE in the correctly specified case. CF Probit (or CF Probit in the Switching case) provides good approximations, outperforming linear estimators to a great extent.

6 Empirical Illustration

As an empirical illustration, we revisit the empirical example of Murtazashvili and Wooldridge (2015) under different functional form assumptions and estimation methods. Murtazashvili and Wooldridge (2015) study the sensitivity of the budget share of housing expenditure to total expenditure and to home ownership using a linear model in a panel data setting with many sources of heterogeneity. Here, we employ a fractional response model with switching, as in (68). This nonlinear model acknowledges the fractional nature of the budget share and therefore has nonconstant marginal effects. As in Murtazashvili and Wooldridge (2015), total expenditure is considered to be a continuous EEV. The home ownership dummy is the binary EEV in the endogenous switching share equation. The model is

$$
\begin{aligned}
E(\text{HousingShare}|\mathbf{x}_1, c_1, c_0) &= \Phi\left[\beta_0 + \beta_1 \text{Log(Expend.)} + \beta_2 \text{Homeowner}\right. \\
&\quad + \mathbf{z}_1 \boldsymbol{\beta}_3 + \beta_4 \text{Log(Expend.)} \cdot \text{Homeowner} \\
&\quad \left. + \beta_5 \text{Homeowner} \cdot \mathbf{z}_1 + c_0 + \text{Homeowner} \cdot c_1\right], \qquad (68a) \\
\text{Log(Expend.)} &= \zeta_0 + \mathbf{z}\boldsymbol{\zeta}_1 + v_2, \qquad (68b) \\
\text{Homeowner} &= 1\left[\gamma_0 + \mathbf{z}\boldsymbol{\gamma}_1 + u_3 > 0\right]. \qquad (68c)
\end{aligned}
$$

We also use just one cross-sectional period from the sample, which turns out to give fairly close estimates of marginal effects for the variables of interest to the panel linear setting with many sources of heterogeneity. The summary APEs from the nonlinear models are compared to the coefficients from the linear models for the housing share, similar to a linear probability model for a binary response.

The sample used in the estimation is the 2001 wave of the Panel Study of Income Dynamics (PSID), which consists of 2355 owners and 629 renters. As we suspect that home ownership dummy indicates switching into differing regions, we report separate summary statistics for different home ownership statuses as in Table 3. Due to the way the dependent variable housing budget share is constructed, and the increase in the price for homes, 84 out of 2355 home owners face a negative housing budget share. As the dependent variable has to be in the unity interval for a fractional response, the housing budget for these 84 homeowners is set to zero. On average, owners spend smaller budget shares on housing than renters. The total expenditure and income of owners are greater than those of renters. Log price, age of the household head, marital status, whether recently moved and race are the exogenous control variables. The log of income is the instrument primarily for log of expenditure, whereas years of education of the household head and number of children in the household are instruments mainly for home ownership.

Table 4 reports the first-stage reduced-form estimation for the two EEVs. Linear reduced form regressions are reported for log total expenditure, the continuous EEV, in Columns (1) and (2). Probit regression is reported for home ownership, the binary EEV, in Columns (3) to (5). Slight variations in the specifications are reported in each case. For example, age squared is included in Column (2) in addition to age in its level form as in Column (1), which turns out to be significant but practically unimportant. For any specification, the instruments mentioned above are strong enough. The probit reduced forms for the home ownership dummy are reported with and without the continuous EEV, as in Column (3) and (5). The predicted value of home ownership from Column (3) contains only an exogenous explanatory variable and is used as an instrument in Regression (3) in Table 5. Columns (4) and (5) show that including the residual from the reduced form of the log expenditure is sufficient to control for all the endogeneity from the total expenditure in the home ownership equation.

Table 5 compares the APEs from the fractional response models to the coefficients of linear probability models. Columns (1) to (4) report the coefficients from the linear models for the housing budget share, and Columns (5) to (10) report the APEs from the fractional response models. The same set of estimators as in the Monte Carlo study are compared here, the only difference being that the dependent variable is a fractional response, instead of a binary response. A "Frac" is added to the names to indicate that a quasi-probit is assumed for the housing budget share. When the home ownership equation is jointly estimated with the fractional probit, as in Columns (9) and (10), the biprobit and heckprobit command in Stata are modified to allow for a fractional dependent variable. The standard errors for the estimates of APEs are bootstrapped.

As we can tell from Table 5, first of all, failing to account for endogeneity, whether as in the linear model represented in Column (1) or as in the Frac Probit model as in Column (5), leads to fairly different estimates from those methods that account for endogeneity. Among the linear probability models that have accounted for endogeneity, the estimates from IV 2SLS differ significantly from that Opt. IV 2SLS and CF 2SLS. Both Opt. IV 2SLS and CF 2SLS are close to the APEs in the Frac Probit models that have accounted for endogeneity. This similarity suggests that the relationship between the housing budget share and the covariates of interest may be close to linear in the unit interval so that the two-stage least squares estimator provides a good approximation. Among the Frac Probit models, the estimates from different methods of accounting for endogeneity are fairly close across the board. The difference between conducting joint estimations with the home ownership equation, as in Columns (9) and (10), and plugging in the generalized residual from home ownership, as in Columns (7) and (8), is small. Particularly, if we use home ownership as the switching indicator, the estimates and standard error from Columns (10) and (8) are the same, at least to the third decimal place. The difference between Columns (7) and (9) is also negligibly small.

Table 6 reports the t statistics for testing of endogeneity and their p-values. The p-values are obtained from bootstrapping the test statistics. Only estimators that employ CF approaches are considered in this table. All the test statistics reported are Wald tests. The names refer to the estimators, as in Table 5. Columns (1) to (4) report the variable addition tests (VATs) on control function terms only. Columns (5) and (6) also report Wald tests on the correlation parameter ρ (or ρ_0 and ρ_1 for the two switching regimes), representing the endogeneity from the home ownership equation, given the control function term from the log expenditure equation. In any case, the evidence of endogeneity from log expenditure is strong: the p-values are identically zero in any tests on the significance of $\widehat{v}_2$, the control function term from the log expenditure equation. No test based on the fractional response model confirms the endogeneity from home ownership, whether it is a VAT test on the generalized residual or on the correlation parameter, although the test on the generalized residual in the linear model CF 2SLS turns out to be significant. The test statistics and p-value from the VAT test on generalized residual in Column (3) are quite similar to the Wald test on the correlation parameter ρ in Column (5), suggesting the validity of using VAT on generalized residuals to detect the additional endogeneity from home ownership. The LM or LR test on ρ can also be performed, but the proper method of bootstrapping for p-values has to be determined. Another advantage of performing a VAT test on generalized residuals is its robustness to model specifications. In the case of switching, a joint test concerning two regimes needs to be conducted to detect endogeneity. This can be easily done by performing a joint test on the interaction terms between the control function terms and the switching indicator. However, finding a way to combine correlation parameters ρ_0 and ρ_1 obtained from different regimes is difficult.

7 Conclusion

This paper has shown applications of control function approaches to account for one binary EEV and many continuous EEVs in binary and fractional response models. The control function approach is computationally simple and allows for flexible incorporation of different sources of heterogeneity, as in an endogenous switching model and a fractional response model. Partial effects based on the ASF can be given interpretations as causal effects. Computation is very efficient and so proper standard errors for the average partial effects are easily obtained. A VAT test based on the generalized residual is shown to be a valid test for detecting additional endogeneity from the binary EEV, conditioning on the reduced form errors for the continuous EEVs. The simulation study shows that using generalized residuals to account for endogeneity provides a fairly good approximation to the true APE, significantly better than approximations provided by linear probability models

Applying the CF approach to an empirical example using a fractional response model for the housing budget share, we find little evidence that homeownership is endogenous after allowing for endogeneity of total expenditure, the continuous EEV. This is revealed by performing VATs on the generalized residuals and validated by a Wald test on the correlation parameter. The computationally simple VAT also allows a robustness check on computing APEs and requires only standard probit estimation after estimating linear reduced forms by OLS and obtain a probit reduced form for the binary EEV. We believe this is a very useful addition to the empirical researchers toolkit: it allows researchers to focuse on the substantive issues in their models rather than trying to solve potentially difficult computational problems.

References

Amemiya, T., 1978. The estimation of a simultaneous equation generalized probit model. Econometrica 46, 1193–1205.

Blundell, **R.W.**, **Powell**, **J.L.**, 2003. Endogeneity in nonparametric and semiparametric regression models. Econometric Society Monographs 36, 312–357.

Blundell, R.W., Powell, J.L., 2004. Endogeneity in semiparametric binary response models. Review of Economic Studies 71, 655–679.

Conlin, M., Dickert-Conlin, S., Chapman, G., 2013. Voluntary disclosure and the strategic behavior of colleges. Journal of Economic Behavior & Organization 96, 48–64. doi:10.1016/j.jebo.2013.09.007

Dong, Y., Lewbel, A., 2015. A simple estimator for binary choice models with endogenous regressors. Econometric Reviews 34, 82–105. doi:10.1080/07474938.2014.944470

Edmans, A., **Goldstein, I.**, **Jiang, W.**, 2012. The Real effects of financial markets: The impact of prices on takeovers. The Journal of Finance 67, 933–971. doi:10.1111/j.1540-6261.2012.01738.x

Hausman, J.A., 1975. An instrumental variable approach to full information estimators for linear and certain nonlinear econometric models. Econometrica 43, 727–38.

Heckman, J., 1978. Dummy endogenous variables in a simultaneous equation system. Econometrica 46, 931–959.

Imbens, G., Angrist, J., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467-475.

Lin, W., **Wooldridge, J.M.**, 2015b. On different approaches to obtaining partial effects in binary response models with endogenous regressors. Economics Letters 134, 58–61. doi:10.1016/j.econlet.2015.05.019

Murtazashvili, I., **Wooldridge, J.M.**, 2015. A control function approach to estimating switching regression models with endogenous explanatory variables and endogenous switching. Journal of Econometrics. doi:10.1016/j.jeconom.2015.06.014

Papke, L.E., Wooldridge, J.M., 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates: summary. Journal of Applied Econometrics (1986-1998) 11, 619.

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

Roodman, D., 2015. CMP: Stata module to implement conditional (recursive) mixed process estimator, Statistical Software Components. Boston College Department of Economics.

Semykina A., Wooldridge JM., 2017. Binary response panel data models with sample selection and selfselection. Journal of Applied Econometrics,1–19.

Nam, S., **Wooldridge, J.M.**, 2014. On computing average partial effects in models with endogeneity or heterogeneity. Mimeo. Department of Economics, Michigan State University.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.

Wooldridge, J.M., 2005. Unobserved heterogeneity and estimation of average partial effects. Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg 27–55.

Wooldridge, J.M., 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, Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions 182, 226–234. doi:10.1016/j.jeconom.2014.04.020

Appendix: Figures and Tables

Figure 1: Empirical Distribution of APEs for y2 for the Sample Size of 1000 under Joint Normality

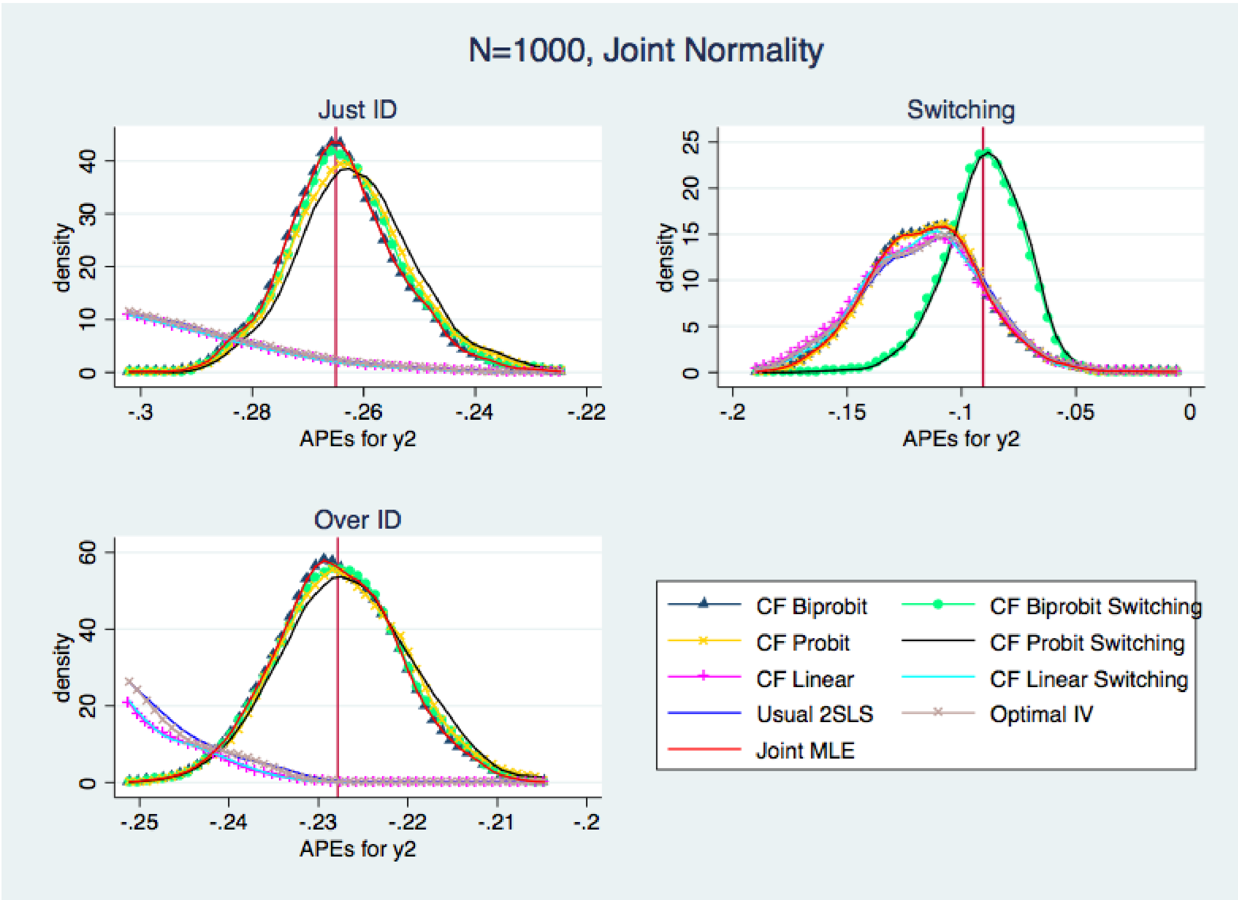

Figure 2: Empirical Distribution of APEs for y2 for the Sample Size of 1000 under Conditional Normality

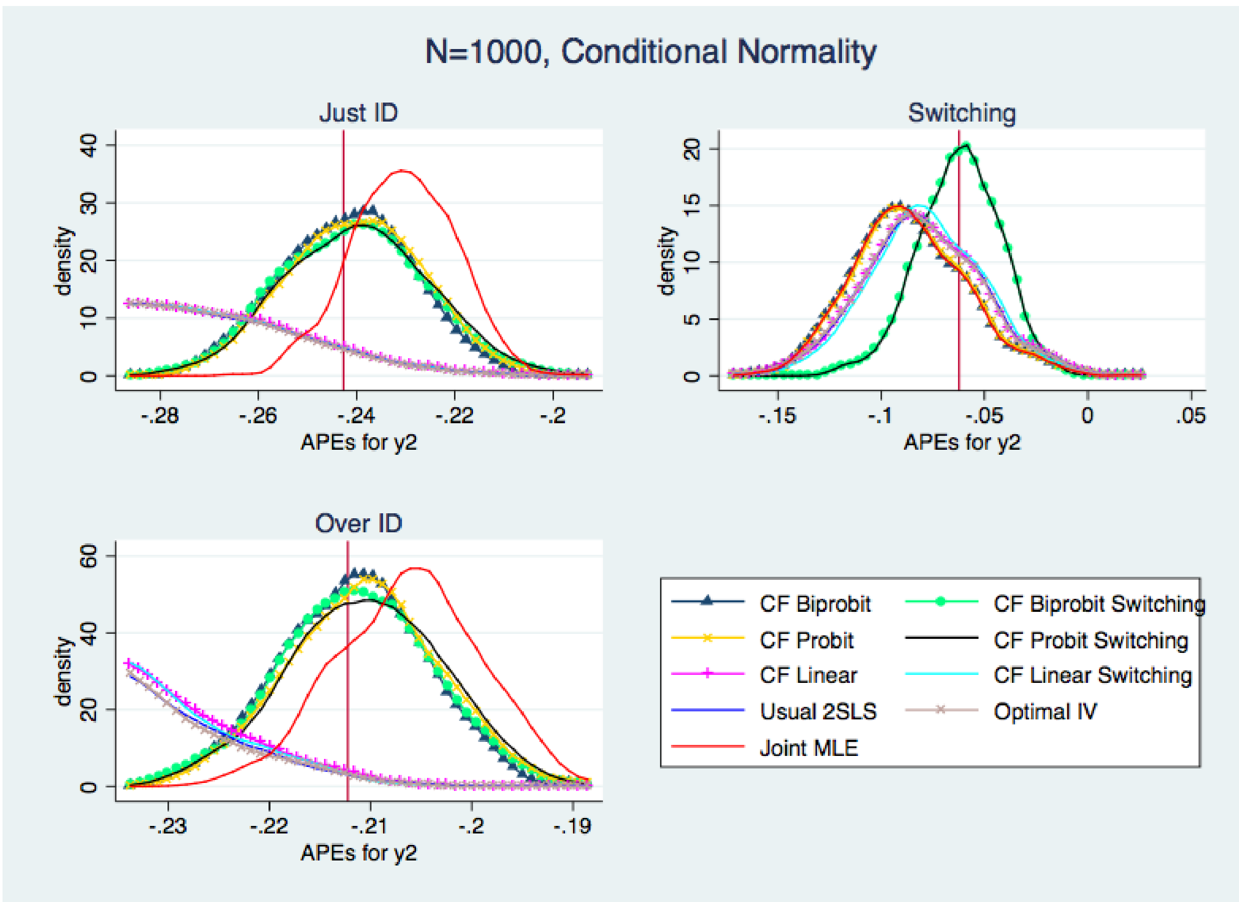

Figure 3: Empirical Distribution of APEs for y3 for the Sample Size of 1000 under Joint Normality

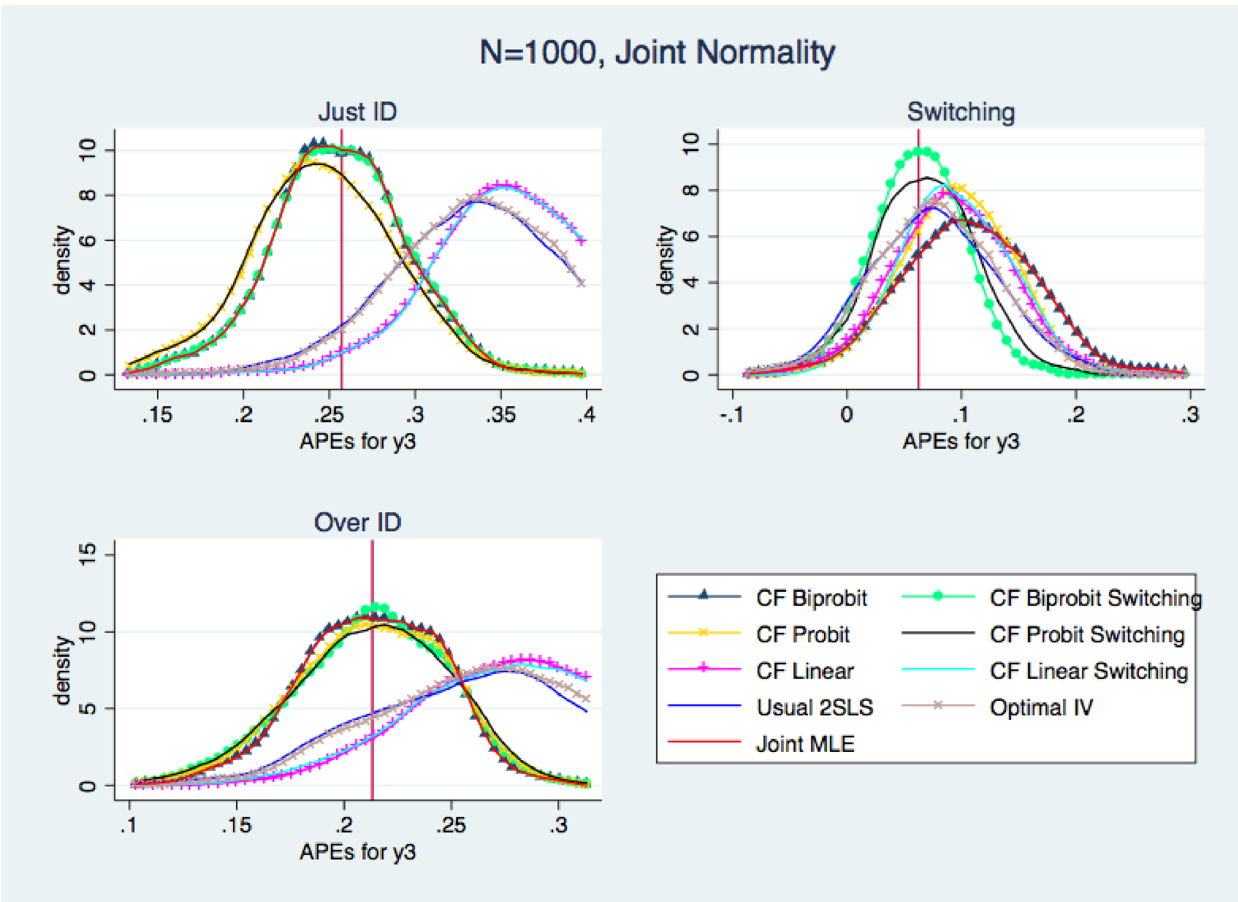

Figure 4: Empirical Distribution of APEs for y3 for the Sample Size of 1000 under Conditional Normality

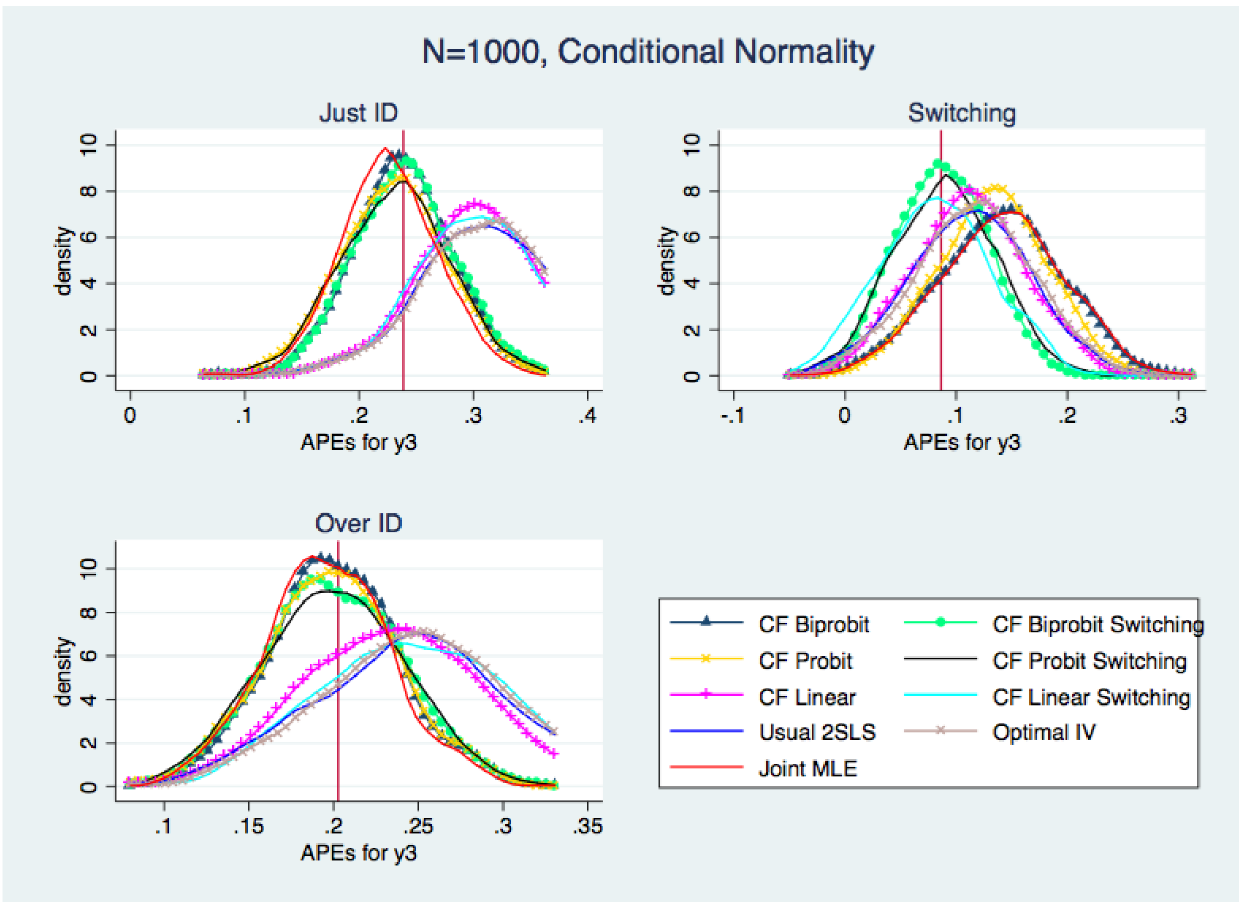

Table 1: Simulation Results for APE of y2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Design 1: Joint Normality								
	CF Biprobit	CF Biprobit Switching	CF Probit	CF Probit Switching	CF 2SLS	CF 2SLS Switching	IV 2SLS	Opt. IV 2SLS	MLE
Case 1:	Just ID, One Region, APE_{y_2}=-.2650								
N=1000									
Bias	.0011	.0020	.0029	.0041	-.0566	-.0568	-.0541	-.0542	.0011
RMSE	.0102	.0105	.0112	.0117	.0634	.0635	.0613	.0614	.0102
N=3000									
Bias	.0003	.0005	.0018	.0022	-.0571	-.0572	-.0548	-.0549	.0003
RMSE	.0054	.0055	.0060	.0061	.0593	.0593	.0571	.0572	.0054
N=5000									
Bias	.0002	.0004	.0018	.0020	-.0573	-.0574	-.0550	-.0551	.0002
RMSE	.0044	.0045	.0050	.0051	.0587	.0587	.0564	.0565	.0044
Case 2:	Over ID, One Region, APE_{y_2}=-.2279								
N=1000									
Bias	.0000	.0003	.0008	.0011	-.0342	-.0342	-.0322	-.0325	.0000
RMSE	.0068	.0069	.0072	.0074	.0359	.0358	.0341	.0343	.0067
N=3000									
Bias	.0001	.0002	.0008	.0009	-.0335	-.0334	-.0315	-.0317	.0001
RMSE	.0039	.0039	.0041	.0041	.0341	.0339	.0321	.0323	.0039
N=5000									
Bias	.0000	.0000	.0007	.0007	-.0338	-.0339	-.0317	-.0320	0000
RMSE	.0029	.0029	.0031	.0032	.0341	.0635	.0321	.0323	.0029
Case 3:	Just ID, Two Regions, APE_{y_2}=-.0902								
N=1000									
Bias	-.0248	-.0001	-.0237	.0010	-.0274	-.0261	-.0252	-.0255	-.0248
RMSE	.0345	.0171	.0337	.0171	.0382	.0365	.0368	.0369	.0346
N=3000									
Bias	-.0252	-.0003	-.0251	-.0003	-.0269	-.0165	-.0249	-.0192	-.0252
RMSE	.0289	.0098	.0295	.0108	.0311	.0225	.0293	.0251	.0289
N=5000									
Bias	-.0256	.0001	-.0245	.0009	-.0271	-.0260	-.0250	-.0253	-.0256
RMSE	.0280	.0076	.0269	.0077	.0298	.0286	.0279	.0282	.0280

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Design 2: Conditional Normality								
	CF Biprobit	CF Biprobit Switching	CF Probit	CF Probit Switching	CF 2SLS	CF 2SLS Switching	IV 2SLS	Opt. IV 2SLS	MLE
Case 1:	Just ID, One Region, APE_{y_2}=-.2426								
N=1000									
Bias	.0009	.0018	.0024	.0028	-.0463	-.0465	-.0474	-.0476	.0121
RMSE	.0133	.0147	.0138	.0151	.0560	.0562	.0570	.0571	.0160
N=3000									
Bias	.0002	.0004	.0016	.0013	-.0458	-.0459	-.0468	-.0469	.0114
RMSE	.0075	.0081	.0078	.0084	.0490	.0490	.0498	.0500	.0128
N=5000									
Bias	-.0001	.0001	.0013	.0009	-.0461	-.0461	-.0472	-.0472	.0112
RMSE	.0058	.0062	.0061	.0064	.0480	.0480	.0490	.0490	.0121
Case 2:	Over ID, One Region, APE_{y_2}=-.2123								
N=1000									
Bias	.0005	.0006	.0015	.0015	-.0269	-.0273	-.0291	-.0294	.0059
RMSE	.0072	.0074	.0078	.0080	.0298	.0301	.0319	.0321	.0092
N=3000									
Bias	.0000	.0000	.0011	.0009	-.0272	-.0277	-.0292	-.0294	.0057
RMSE	.0041	.0042	.0045	.0046	.0281	.0286	.0300	.0302	.0070
N=5000									
Bias	.0000	.0001	.0010	.0009	-.0271	-.0276	-.0290	-.0293	.0057
RMSE	.0032	.0033	.0035	.0036	.0277	.0281	.0295	.0298	.0065
Case 3:	Just ID, Two Regions, APE_{y_2}= -.0624								
N=1000									
Bias	-.0239	-.0001	-.0230	.0001	-.0176	-.0158	-.0176	-.0182	-.0241
RMSE	.0363	.0197	.0357	.0196	.0336	.0317	.0336	.0339	.0363
N=3000									
Bias	.-.0259	-.0004	-.0251	-.0003	-.0187	-.0166	-.0186	-.0192	-.0261
RMSE	.0302	.0108	.0295	.0108	.0247	.0225	.0246	.0251	.0303
N=5000									
Bias	-.0251	.0000	-.0243	.0001	-.0175	-.0159	-.0175	-.0181	-.0253
RMSE	.0278	.0083	.0270	.0083	.0213	.0197	.0214	.0218	.0279

[a] Sequential averaging of the control function term v_2 and $\mathbf{x}$ is applied to compute estimates of APEs.

[b] The bias is defined as the difference between the true APEs and the estimates. RMSE is the root mean squared error.

[c] Estimator (1) is the CF approach inserting the first-stage residual $\widehat{v}_2$ to a second-stage joint biprobit between y_1 , y_3. Estimator (2) is the CF approach inserting the first-stage residual $\widehat{v}_2$ to a second-stage joint biprobit between $y_1^{(1)}$, y_3 and $y_1^{(0)}$, y_3. Estimator (3) is the CF approach inserting first-stage residual $\widehat{v}_2$ and $\widehat{gr}_3$ into the probit model for y_1. Estimator (4) is the CF approach inserting first-stage residual $\widehat{v}_2$ and $\widehat{gr}_3$ into the probit model for y_1 separately for sub-samples defined by y_3. Estimator (5) is the CF approach applied to linear probability model for y_1 by inserting first-stage residual $\widehat{v}_2$ and $\widehat{gr}_3$. Estimator (6) is the CF approach applied to linear probability model for y_1 by inserting first-stage residual $\widehat{v}_2$ and $\widehat{gr}_3$ for sub-samples defined by y_3. Estimator (7) is the 2SLS IV approach for a linear probability model of y_1. Estimator (8) is the 2SLS IV approach using predicted fitted values from the first-stage reduced forms for y_2 and y_3 as instruments. y_2 is predicted using a linear model and y_3 is predicted using probit model. Estimator (9) is the joint estimation of y_1, y_2 and y_3 by maximum likelihood.

Table 2: Simulation Results for APE of y3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Design 1: Joint Normality									Design 2: Conditional Normality								
	CF Biprobit	CF Biprobit Switching	CF Probit	CF Probit Switching	CF 2SLS	CF 2SLS Switching	IV 2SLS	Opt. IV 2SLS	MLE	CF Biprobit	CF Biprobit Switching	CF Probit	CF Probit Switching	CF 2SLS	CF 2SLS Switching	IV 2SLS	Opt. IV 2SLS	MLE
Case 1	Just ID, One Region, APE$_{y_3}$=.2573									Just ID, One Region, APE$_{y_3}$=.2385								
N=1000																		
Bias	-.0013	-.0012	-.0116	-.0119	.1010	.1010	.0809	.0823	-.0014	-.0003	.0013	-.0087	-.0075	.0667	.0656	.0738	.0752	-.0119
RMSE	.0380	.0382	.0440	.0442	.1117	.1119	.0966	.0972	.0380	.0422	.0438	.0472	.0484	.0866	.0869	.0956	.0961	.0432
N=3000																		
Bias	.0001	.0002	-.0090	-.0091	.1032	.1033	.0852	.0857	.0000	-.0001	.0001	-.0084	-.0083	.0675	.0663	.0744	.0757	-.0118
RMSE	.0228	.0229	.0267	.0268	.1069	.1070	.0904	.0906	.0228	.0236	.0241	.0276	.0281	.0742	.0735	.0819	.0830	.0259
N=5000																		
Bias	-.0004	-.0003	-.0096	-.0096	.1024	.1026	.0840	.0849	-.0004	.0008	.0009	-.0073	-.0071	.0683	.0671	.0768	.0815	-.0112
RMSE	.0172	.0172	.0213	.0214	.1047	.1048	.0873	.0880	.0171	.0193	.0198	.0224	.0228	.0727	.0717	.0816	.0815	.0221
Case 2	Over ID, One Region, APE$_{y_3}$=.2132									Over ID, One Region, APE$_{y_3}$=.2026								
N=1000																		
Bias	.0002	-.0008	-.0003	-.0006	.0672	.0653	.0512	.0534	.0001	.0009	.0009	-.0012	.0003	.0353	.0493	.0492	.0513	-.0044
RMSE	.0323	.0349	.0350	.0367	.0813	.0811	.0744	.0729	.0322	.0394	.0394	.0400	.0411	.0622	.0736	.0772	.0762	.0372
N=3000																		
Bias	.0002	.0004	-.0003	.0003	.0663	.0647	.0501	.0521	.0001	.0004	.0008	-.0017	-.0006	.0330	.0339	.0494	.0510	-.0041
RMSE	.0188	.0197	.0198	.0207	.0715	.0705	.0587	.0595	.0188	.0217	.0229	.0237	.0234	.0444	.0455	.0608	.0608	.0226
N=5000																		
Bias	-.0001	-.0002	-.0012	-.0007	.0655	.0635	.0491	.0513	-.0001	.0003	.0006	-.0016	-.0005	.0338	.0478	.0485	.0506	-.0041
RMSE	.0147	.0152	.0158	.0162	.0687	.0670	.0545	.0559	.0146	.0173	.0180	.0185	.0189	.0414	.0540	.0557	.0569	.0174
Case 3	Just ID, Two Regions, APE$_{y_3}$=.0622									Just ID, Two Regions, APE$_{y_3}$=.0864								
N=1000																		
Bias	.0436	.0005	.0339	.0065	.0292	.0327	.0128	.0153	.0437	.0535	.0007	.0462	.0028	.0257	-.0045	.0257	.0301	.0549
RMSE	.0721	.0392	0585	.0444	.0576	.0575	.0559	.0546	.0722	.0766	.0409	.0669	.0455	.0555	.0500	.0597	.0601	.0779
N=3000																		
Bias	.0437	.0001	.0467	.0023	.0292	-.0051	.0129	.0294	.0438	.0536	-.0003	.0467	.0023	.0254	-.0051	.0246	.0294	.0546
RMSE	.0555	.0236	.0547	.0268	.0413	.0294	.0351	.0428	.0555	.0624	.0238	.0547	.0268	.0386	.0294	.0408	.0428	.0634
N=5000																		
Bias	.0447	.0010	.0353	.0076	.0300	.0335	.0130	.0159	.0447	.0536	-.0003	.0466	.0023	.0245	-.0057	.0247	.0291	.0545
RMSE	.0516	.0180	.0413	.0215	.0372	.0395	.0273	.0280	.0516	.0592	.0185	.0516	.0208	.0333	.0237	.0350	.0378	.0601

[a] Sequential averaging of the control function term v_2 and $\mathbf{x}$ is applied to compute estimates of APEs.

[b] The bias is defined as the difference between the true APEs and the estimates. RMSE is the root mean squared error.

[c] Estimator (1) is the CF approach inserting the first-stage residual $\widehat{v}_2$ to a second-stage joint biprobit between y_1 , y_3. Estimator (2) is the CF approach inserting the first-stage residual $\widehat{v}_2$ to a second-stage joint biprobit between $y_1^{(1)}$, y_3 and $y_1^{(0)}$, y_3. Estimator (3) is the CF approach inserting first-stage residual $\widehat{v}_2$ and $\widehat{gr}_3$ into the probit model for y_1. Estimator (4) is the CF approach inserting first-stage residual $\widehat{v}_2$ and $\widehat{gr}_3$ into the probit model for y_1 separately for sub-samples defined by y_3. Estimator (5) is the CF approach applied to linear probability model for y_1 by inserting first-stage residual $\widehat{v}_2$ and $\widehat{gr}_3$. Estimator (6) is the CF approach applied to linear probability model for y_1 by inserting first-stage residual $\widehat{v}_2$ and $\widehat{gr}_3$ for sub-samples defined by y_3. Estimator (7) is the 2SLS IV approach for a linear probability model of y_1. Estimator (8) is the 2SLS IV approach using predicted fitted values from the first-stage reduced forms for y_2 and y_3 as instruments. y_2 is predicted using a linear model and y_3 is predicted using probit model. Estimator (9) is the joint estimation of y_1, y_2 and y_3 by maximum likelihood.

Table 3: Summary Statistics of the Estimation Sample (N=2964)

Variable	Owner	Renter
Budget Share on Housing	.20	.41
	(.16)	(.16)
Ln(Expenditure)	10.35	9.82
	(.59)	(.59)
Ln(Income)	10.94	10.29
	(.75)	(.73)
Ln(Price)	8.55	8.93
	(.21)	(.12)
Age	49.88	44.45
	(12.97)	(13.69)
Married	.79	.35
	(.40)	(.48)
Moved	.10	.33
	(.30)	(.47)
Black	.21	.46
	(.41)	(.50)
Years of education	13.27	12.12
	(2.73)	(2.90)
Number of Children	.94	1.05
	(1.14)	(1.24)
Obs.	2355	629

[a] The sample is based on the 2001waves of the Panel Study of Income Dynamics (PSID). All monetary variables were converted to 1998 dollars before they were logged.

[b] Sample standard deviations are in parentheses below the sample means.

Table 4: The Frist Stage Reduced Form Regression for the EEVs

	(1)	(2)	(3)	(4)	(5)
Estimation Method	OLS	OLS	Probit	Probit	Probit
Dependent Variable	Ln(Expenditure)	Ln(Expenditure)	Owner	Owner	Owner
$\widehat{v}_2$				.058***	
				(.009)	
Ln(Expenditure)					.058***
					(.009)
Ln(Income)	.410***	.385***	.055***	.057***	.033***
	(.013)	(.013)	(.006)	(.006)	(.007)
Education	.025***	.024***	.005**	.004**	.002
	(.0031)	(.003)	(.001)	(.0015)	(.0015)
Children	.056***	.062***	.012***	.011***	.008**
	(.0077)	(.008)	(.004)	(.003)	(.004)
Age	-.0025***	.029***	.003***	.003***	.003***
	(.0007)	(.004)	(.0003)	(.0003)	(.0003)
Age^2		-.0002***			
		(.00004)			
Ln(Price)	-.068**	-.067**	-.646***	-.624***	-.620***
	(.0318)	(.0314)	(.017)	(.016)	(.016)
Married	.27***	.26***	.054***	.053***	.037***
	(.021)	(.021)	(.010)	(.009)	(.010)
Moved	.012*	.037*	-.065***	-.063***	-.064***
	(.022)	(.022)	(.009)	(.009)	(.009)
Black	-.058**	-.079***	-.064***	-.060***	-.057***
	(.019)	(.0189)	(.0087)	(.009)	(.009)

[a] $\widehat{v}_2$ denotes the residual from Regression (1), the reduced form for log total expenditure.
[b] Regression (1) and (2) are first-stage regressions for log total expenditure, the continuous EEV. Regression (3)-(5) are first-stage regressions for home ownership, the binary EEV.
[c] * p-value¡10%
** p-value¡5%
*** p-value¡1%

Table 5: Comparing marginal effects in the structural equation of the housing share

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Estimation Method		IV	Opt. IV	CF		CF	CF	CF	CF	CF
Functional form for y_1	Linear No EEVs	2SLS	2SLS	2SLS	Frac Probit No EEVs	Frac Probit One EEV	Frac Probit Two EEVs	Frac Probit Switching	Frac Biprobit	Frac Biprobit Switching
Marginal Effects	Coeff	Coeff	Coeff	Coeff	APE	APE	APE	APE	APE	APE
Ln(Expenditure)	-.109***	-.176**	-.056***	-.061***	-.106***	-.064***	-.061***	-.059***	-.062***	-.059***
	(.005)	(.084)	(.01)	(.01)	(.005)	(.009)	(.010)	(.009)	(.01)	(.009)
Owner	-.094***	.336	-.14***	-.124***	-.074***	-.088***	-.101***	-.124***	-.098***	-.124***
	(.009)	(.330)	(.016)	(.015)	(.0082)	(.012)	(.020)	(.038)	(.020)	(.038)
Age	.002***	.0003	.003***	.003***	.002***	.0025***	.0026***	.0026***	.0025***	.0026***
	(.0002)	(.001)	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)
Ln(Price)	.147***	.54*	.109***	.124***	.151***	.147***	.136***	.138***	.139***	.138***
	(.013)	(.29)	(.017)	(.017)	(.013)	(.016)	(.021)	(.022)	(.021)	(.022)
Married	-.005	-.06**	-.027	-.028***	-.0026	-.027***	-.027***	-.027***	-.027***	-.027***
	(.007)	(.025)	(.008)	(.009)	(.006)	(.008)	(.008)	(.008)	(.008)	(.008)
Moved	.011	.071	.005	.007	.009	.009	.007	.005	.007	.005
	(.007)	(.047)	(.008)	(.007)	(.007)	(.007)	(.008)	(.009)	(.007)	(.009)
Black	-.012**	.041	-.010*	-.009	-.012**	-.006	-.008	-.009	-.007	-.009
	(.006)	(.036)	(.006)	(.006)	(.006)	(.006)	(.006)	(.007)	(.006)	(.007)

[a] The dependent variable is the expenditure share on housing.

[b] Standard errors for the estimated APEs were bootstrap standard errors with 200 replications.

[c] Regression (1) is the OLS for linear probability model that assums no EEVs. Regression (2) is the 2SLS IV estimator for linear probability model that uses a linear probability model for the reduced form of home ownership. Regression (3) is the 2SLS IV estimator for linear probability model that uses the predicted fitted values from the first stage regressions (1) and (3) in Table 4 as IV. Regression (4) is the CF approach for linear probability model using $\widehat{v}_2$ and $\widehat{gr}_3$. Regression (5) is the fractional response model that assumes no EEVs. Regression (6) is the CF approach for the fractional response model by inserting only $\widehat{v}_2$. Regression (7) is the CF approach for the fractional response model by inserting both $\widehat{v}_2$ and $\widehat{gr}_3$. Regression (8) is the CF approach for the fractional response model by inserting both $\widehat{v}_2$ and $\widehat{gr}_3$ applied separately to two sub-samples of households defined by their homeownership status. Regression (9) is the CF approach to a joint estimation between fractional response and homeownership in the second step. Regression (10) is the CF approach applied separately to two-samples of households defined by their homeownership status and a joint estimation between fractional response and homeownership in the second step.

[d] * p-value¡10%
** p-value¡5%
*** p-value¡1%

Table 6: Variable Addition Tests for Endogeneity

	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Method	CF	CF	CF	CF	CF	CF
Functional form for y_1	2SLS	Frac Probit One EEV	Frac Probit Two EEVs	Frac Probit Switching	Frac Biprobit	Frac Biprobit Swiching
Significance of $\widehat{v}_2$	41.79	38.43	38.42		76.91	
	(.000)	(.000)	(.000)		(.000)	
Significance of $\widehat{gr}_3$	4.07		0.71			
	(.036)		(.405)			
Joint significance of $\widehat{v}_2, \widehat{gr}_3$	21.25		38.67			
	(.000)		(.000)			
Joint significance of $\widehat{gr}_3^{(1)}, \widehat{gr}_3^{(0)}$				2.87		
				(0.238)		
Significance of $\widehat{gr}_3^{(1)}$				0.61		
				(.431)		
Significance of $\widehat{gr}_3^{(0)}$				2.26		
				(.154)		
Joint significance of $\widehat{v}_2^{(1)}, \widehat{v}_2^{(0)}$				44.73		
				(0.000)		
Significance of $\widehat{v}_2^{(1)}$				20.54		41.44
				(.000)		(.000)
Significance of $\widehat{v}_2^{(0)}$				24.18		41.36
				(.000)		(.000)
Significance of ρ					0.73	
					(.407)	
Significance of ρ_1						.064
						(.410)
Significance of ρ_0						.207
						(.131)
Joint significance of $\widehat{gr}_3^{(1)}, \widehat{gr}_3^{(0)}, \widehat{v}_2^{(1)}, \widehat{v}_2^{(0)}$				46.56		
				(.000)		

[a] The Wald test statistics for null hypotheses of no endogeneity are reported for control function terms and the correlation parameters.

[b] Bootstrap p-value is reported in the parentheses under the Wald test statistic.

[c] $\widehat{v}_2$ is the residual from first-stage reduced-form regression of log expenditure in Column (1) of 4. $\widehat{gr}_3$ is the generalized residual from first-stage reduced-form regression of home ownership in Column (4) of Table 4. Any superscript (1) indicates that this term appears in the regime with home owernship =1 in a switching regression. Any superscript (0) indicates that this term appears in the regime with home owernship =0 in a switching regression. ρ is the correlation from the joint estimation with homeownship given the control function term $\widehat{v}_2$ has been plugged in. Any subscript 1 indicates it is the correlation with housing budget share in regime 1. Any subscript 0 indicates it is the correlation with housing budget share in regime 0.