Journal of Econometrics 145 (2008) 121–133

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

# Panel data methods for fractional response variables with an application to test pass rates

Leslie E. Papke, Jeffrey M. Wooldridge *

*Department of Economics, Michigan State University, East Lansing, MI 48824-1038, United States*

ARTICLE INFO

*Article history:*
Available online 19 June 2008

*JEL classification:*
C23
C25

*Keywords:*
Fractional response
Panel data
Unobserved effects
Probit
Partial effects

ABSTRACT

We revisit the effects of spending on student performance using data from the state of Michigan. In addition to exploiting a dramatic change in funding in the mid-1990s and subsequent nonsmooth changes, we propose nonlinear panel data models that recognize the bounded nature of the pass rate. Importantly, we show how to estimate average partial effects, which can be compared across many different models (linear and nonlinear) under different assumptions and estimated using many different methods. We find that spending has nontrivial and statistically significant effects, although the diminishing effect is not especially pronounced.



## Executive summary

Determining the effects of school inputs on student performance in the United States is an important policy issue. Discussions of increased funding for K-12 education, as well as the implications for equalized funding across schools, rely on benefits measured in terms of student performance. In many states, including Michigan, success is measured – and reported widely in the press – in terms of pass rates on statewide standardized tests. Because pass rates, when measured as a proportion, are necessarily bounded between zero and one, standard linear models may not provide an accurate picture of the effects of spending on pass rates throughout the entire distribution of spending. In particular, if pass rates depend on spending, the relationship must be bounded – otherwise, pass rates are eventually predicted to be greater than one.

Some of the most convincing studies examining the link between student achievement and spending have used panel data, particularly when the time periods straddle a policy change that induces (arguably) exogenous variation in spending. Yet standard linear panel data models are not well suited to pass rates because it is difficult to impose a positive yet bounded effect of spending on pass rates. In this paper, we extend our earlier work on fractional response models for cross section data to panel data. We allow unobserved time-constant district effects – which capture historical differences among districts – to be systematically related to district spending. In some specifications, we also allow spending to be correlated with time-varying unobserved inputs, such as the average quality of the students in a particular grade, or parental effort.

Using data on a fourth-grade math test for Michigan from 1992 through 2001, which includes significant changes in funding that resulted from Proposal A, we use a probit functional form for the mean response to impose a bounded effect of spending on pass rates. Given a 10% increase in four-year averaged spending, the estimated average effect on the pass rate varies from about three to six percentage points, with the higher estimate occurring when spending is allowed to be correlated with unobserved time-varying inputs. In the latter case, as spending varies from the 5th percentile to the 95th percentile, the estimated effect on the pass rate falls by roughly three percentage points — a nontrivial but not overwhelming change. The estimate for the linear model lies between the marginal effects at the extreme values of spending. Therefore, the linear approximation does a good job in estimating the average effect of spending on pass rates, even though it misses some of the nonlinear effects at more extreme spending levels.

## 1. Introduction

In 1994, voters in Michigan passed Proposal A, which led to major changes in the way K-12 education is financed. The system went from one largely based on local property tax revenues to funding at the state level, supported primarily by an increase in the sales tax rate. One consequence of this change is that the lowest spending districts were provided with a foundation

* Corresponding author. Tel.: +1 517 353 5972; fax: +1 517 432 1068.
*E-mail address:* wooldri1@msu.edu (J.M. Wooldridge).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.05.009

allowance significantly above their previous per-student funding. As described in Papke (2005), the change in funding resulted in a natural experiment that can be used to more precisely estimate the effects of per-student spending on student performance.

Papke (2005) used building-level panel data, for 1993 through 1998, and found nontrivial effects of spending on the pass rate on a statewide fourth-grade math test. One potential drawback of Papke's analysis is that she used linear functional forms in her fixed effects and instrumental variables fixed effects analyses, which ignore the bounded nature of a pass rate (either a percentage or a proportion). Papke did split the sample into districts that initially were performing below the median and those performing above the median, and found very different effects. But such sample splitting is necessarily arbitrary and begs the question as to whether linear functional forms adequately capture the diminishing effects of spending at already high levels of spending.

Empirical studies attempting to explain fractional responses have proliferated in recent years. Just a few examples of fractional responses include pension plan participation rates, industry market shares, television ratings, fraction of land area allocated to agriculture, and test pass rates. Researchers have begun to take seriously the functional form issues that arise with a fractional response: a linear functional form for the conditional mean might miss important nonlinearities. Further, the traditional solution of using the log-odds transformation obviously fails when we observe responses at the corners, zero and one. Just as importantly, even in cases where the variable is strictly inside the unit interval, we cannot recover the expected value of the fractional response from a linear model for the log-odds ratio unless we make strong independence assumptions.

In Papke and Wooldridge (1996), we proposed direct models for the conditional mean of the fractional response that keep the predicted values in the unit interval. We applied the method of quasi-maximum likelihood estimation (QMLE) to obtain robust estimators of the conditional mean parameters with satisfactory efficiency properties. The most common of those methods, where the mean function takes the logistic form, has since been applied in numerous empirical studies, including Hausman and Leonard (1997), Liu et al. (1999), and Wagner (2001). (In a private communication shortly after the publication of Papke and Wooldridge (1996), in which he kindly provided Stata® code, John Mullahy dubbed the method of quasi-MLE with a logistic mean function "fractional logit", or "flogit" for short.)

Hausman and Leonard (1997) applied fractional logit to panel data on television ratings of National Basketball Association games to estimate the effects of superstars on telecast ratings. In using pooled QMLE with panel data, the only extra complication is in ensuring that the standard errors are robust to arbitrary serial correlation (in addition to misspecification of the conditional variance). But a more substantive issue arises with panel data and a nonlinear response function: How can we account for unobserved heterogeneity that is possibly correlated with the explanatory variables?

Wagner (2003) analyzes a large panel data set of firms to explain the export-sales ratio as a function of firm size. Wagner explicitly includes firm-specific intercepts in the fractional logit model, a strategy suggested by Hardin and Hilbe (2007) when one observes the entire population (as in Wagner's case, because he observes all firms in an industry). Generally, while including dummies for each cross section observation allows unobserved heterogeneity to enter in a flexible way, it suffers from an incidental parameters problem under random sampling when $T$ (the number of time periods) is small and $N$ (the number of cross-sectional observations) is large. In particular, with fixed $T$, the estimators of the fixed effects are inconsistent as $N \to \infty$, and this inconsistency transmits itself to the coefficients on the common slope coefficients. The statistical properties of parameter estimates and partial effects of so-called "fixed effects fractional logit" are largely unknown with small $T$. (Hausman and Leonard (1997) include team "fixed effects" in their analysis, but these parameters can be estimated with precision because Hausman and Leonard have many telecasts per team. Therefore, there is no incidental parameters problem in the Hausman and Leonard setup.)

In this paper we extend our earlier work and show how to specify, and estimate, fractional response models for panel data with a large cross-sectional dimension and relatively few time periods. We explicitly allow for time-constant unobserved effects that can be correlated with explanatory variables. We cover two cases. The first is when, conditional on an unobserved effect, the explanatory variables are strictly exogenous. We then relax the strict exogeneity assumption when instrumental variables are available.

Rather than treating the unobserved effects as parameters to estimate, we employ the Mundlak (1978) and Chamberlain (1980) device of modeling the distribution of the unobserved effect conditional on the strictly exogenous variables. To accommodate this approach, we exploit features of the normal distribution. Therefore, unlike in our early work, where we focused mainly on the logistic response function, here we use a probit response function. In binary response contexts, the choice between the logistic and probit conditional mean functions for the structural expectation is largely a matter of taste, although it has long been recognized that, for handling endogenous explanatory variables, the probit mean function has some distinct advantages. We further exploit those advantages for panel data models in this paper. As we will see, the probit response function results in very simple estimation methods. While our focus is on fractional responses, our methods apply to the binary response case with a continuous endogenous explanatory variable and unobserved heterogeneity.

An important feature of our work is that we provide simple estimates of the partial effects averaged across the population – sometimes called the "average partial effects" (APEs) or "population averaged effects". These turn out to be identified under no assumptions on the serial dependence in the response variable, and the suspected endogenous explanatory variable is allowed to arbitrarily correlate with unobserved shocks in other time periods.

The rest of the paper is organized as follows. Section 2 introduces the model and assumptions for the case of strictly exogenous explanatory variables, and shows how to identify the APEs. Section 3 discusses estimation methods, including pooled QMLE and an extension of the generalized estimating equation (GEE) approach. Section 4 relaxes the strict exogeneity assumption, and shows how control function methods can be combined with the Mundlak–Chamberlain device to produce consistent estimators. Section 5 applies the new methods to estimate the effects of spending on math test pass rates for Michigan, and Section 6 summarizes the policy implications of our work.

## 2. Models and quantities of interest for strictly exogenous explanatory variables

We assume that a random sample in the cross section is available, and that we have available $T$ observations, $t = 1, \ldots, T$, for each random draw $i$. For cross-sectional observation $i$ and time period $t$, the response variable is $y_{it}$, $0 \leq y_{it} \leq 1$, where outcomes at the endpoints, zero and one, are allowed. (In fact, $y_{it}$ could be a binary response.) For a set of explanatory variables $\mathbf{x}_{it}$, a $1 \times K$ vector, we assume

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i), \quad t = 1, \ldots, T, \tag{2.1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (cdf). Assumption (2.1) is a convenient functional form assumption. Specifically, the conditional expectation is assumed to be of the index form, where the unobserved effect, $c_i$, appears additively inside the standard normal cdf, $\Phi(\cdot)$.

The use of $\Phi$ in (2.1) deserves comment. In Papke and Wooldridge (1996), we allowed a general function $G(\cdot)$ in place of $\Phi(\cdot)$ but then, for our application to pension plan participation rates, we focused on the logistic function, $\Lambda(z) \equiv \exp(z)/[1+\exp(z)]$. As we proceed it will be clear that using $\Lambda$ in place of $\Phi$ in (2.1) causes no conceptual or theoretical difficulties; rather, the probit function leads to computationally simple estimators in the presence of unobserved heterogeneity or endogenous explanatory variables. Therefore, in this paper we restrict our attention to Eq. (2.1).

Because $\Phi$ is strictly monotonic, the elements of $\boldsymbol{\beta}$ give the directions of the partial effects. For example, dropping the observation index $i$, if $x_{tj}$ is continuous, then

$$\frac{\partial E(y_t|\mathbf{x}_t, c)}{\partial x_{tj}} = \beta_j \phi(\mathbf{x}_t\boldsymbol{\beta} + c). \tag{2.2}$$

For discrete changes in one or more of the explanatory variables, we compute

$$\Phi(\mathbf{x}_t^{(1)}\boldsymbol{\beta} + c) - \Phi(\mathbf{x}_t^{(0)}\boldsymbol{\beta} + c), \tag{2.3}$$

where $\mathbf{x}_t^{(0)}$ and $\mathbf{x}_t^{(1)}$ are two different values of the covariates.

Eqs. (2.2) and (2.3) reveal that the partial effects depend on the level of covariates and the unobserved heterogeneity. Because $\mathbf{x}_t$ is observed, we have a pretty good idea about interesting values to plug in. Or, we can always average the partial effects across the sample $\{\mathbf{x}_{it} : i = 1, \ldots, N\}$ on $\mathbf{x}_t$. But $c$ is not observed. A popular measure of the importance of the observed covariates is to average the partial effects across the distribution of $c$, to obtain the average partial effects. For example, in the continuous case, the APE with respect to $x_{tj}$, evaluated at $\mathbf{x}_t$, is

$$E_c[\beta_j \phi(\mathbf{x}_t\boldsymbol{\beta} + c)] = \beta_j E_c[\phi(\mathbf{x}_t\boldsymbol{\beta} + c)], \tag{2.4}$$

which depends on $\mathbf{x}_t$ (and, of course, $\boldsymbol{\beta}$) but not on $c$. Similarly, we get APEs for discrete changes by averaging (2.3) across the distribution of $c$.

Without further assumptions, neither $\boldsymbol{\beta}$ nor the APEs are known to be identified. In this section, we add two assumptions to (2.1). The first concerns the exogeneity of $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$. We assume that, conditional on $c_i$, $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$ is strictly exogenous:

$$E(y_{it}|\mathbf{x}_i, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i), \quad t = 1, \ldots, T, \tag{2.5}$$

where $\mathbf{x}_i \equiv (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ is the set of covariates in all time periods. Assumption (2.5) is common in unobserved effects panel data models, but it rules out lagged dependent variables in $\mathbf{x}_{it}$, as well as other explanatory variables that may react to past changes in $y_{it}$. Plus, it rules out traditional simultaneity and correlation between time-varying omitted variables and the covariates.

We also need to restrict the distribution of $c_i$ given $\mathbf{x}_i$ in some way. While semiparametric methods are possible, in this paper we propose a conditional normality assumption, as in Chamberlain (1980):

$$c_i|(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}) \sim \text{Normal}(\psi + \bar{\mathbf{x}}_i\boldsymbol{\xi}, \sigma_a^2), \tag{2.6}$$

where $\bar{\mathbf{x}}_i \equiv T^{-1}\sum_{t=1}^{T}\mathbf{x}_{it}$ is the $1 \times K$ vector of time averages. As we will see, (2.6) leads to straightforward estimation of the parameters $\beta_j$ up to a common scale factor, as well as consistent estimators of the APEs. Adding nonlinear functions in $\bar{\mathbf{x}}_i$ in the conditional mean, such as squares and cross products, is straightforward. It is convenient to assume only the time average appears in $D(c_i|\mathbf{x}_i)$ as a way of conserving on degrees-of-freedom. But an unrestricted Chamberlain (1980) device, where we allow each $\mathbf{x}_{it}$ to have a separate vector of coefficients, is also possible.

Another way to relax (2.6) would be to allow for heteroskedasticity, with a convenient specification being $\text{Var}(c_i|\mathbf{x}_i) = \text{Var}(c_i|\bar{\mathbf{x}}_i) = \sigma_a^2 \exp(\bar{\mathbf{x}}_i\boldsymbol{\lambda})$. As shown in Wooldridge (2002, Problem 15.18), the APEs are still identified, and a similar argument works here as well. The normality assumption can be relaxed, too, at the cost of computational simplicity. More generally, one might simply assume $D(c_i|\mathbf{x}_i) = D(c_i|\bar{\mathbf{x}}_i)$, which, combined with assumption (2.5), applies that $E(y_{it}|\mathbf{x}_i) = E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i)$. Then, we could use flexible models for the latter expectation. In this paper, we focus on (2.6) because it leads to especially straightforward estimation and allows simple comparison with linear models.

For some purposes, it is useful to write $c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i$, where $a_i|\mathbf{x}_i \sim \text{Normal}(0, \sigma_a^2)$. (Note that $\sigma_a^2 = \text{Var}(c_i|\mathbf{x}_i)$, the *conditional* variance of $c_i$.) Naturally, if we include time-period dummies in $\mathbf{x}_{it}$, as is usually desirable, we do not include the time averages of these in $\bar{\mathbf{x}}_i$.

Assumptions (2.1), (2.5) and (2.6) impose no additional distributional assumptions on $D(y_{it}|\mathbf{x}_i, c_i)$, and they place no restrictions on the serial dependence in $\{y_{it}\}$ across time. Nevertheless, the elements of $\boldsymbol{\beta}$ are easily shown to be identified up to a positive scale factor, and the APEs are identified. A simple way to see this is to write

$$E(y_{it}|\mathbf{x}_i, a_i) = \Phi(\psi + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i) \tag{2.7}$$

and so

$$\begin{aligned} E(y_{it}|\mathbf{x}_i) &= E[\Phi(\psi + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i)|\mathbf{x}_i] \\ &= \Phi[(\psi + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi})/(1 + \sigma_a^2)^{1/2}] \end{aligned} \tag{2.8}$$

or

$$E(y_{it}|\mathbf{x}_i) \equiv \Phi(\psi_a + \mathbf{x}_{it}\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a), \tag{2.9}$$

where the subscript $a$ denotes division of the original coefficient by $(1 + \sigma_a^2)^{1/2}$. The second equality in (2.8) follows from a well-known mixing property of the normal distribution. (See, for example, Wooldridge (2002, Section 15.8.2) in the case of binary response; the argument is essentially the same.) Because we observe a random sample on $(y_{it}, \mathbf{x}_{it}, \bar{\mathbf{x}}_i)$, (2.9) implies that the scaled coefficients, $\psi_a$, $\beta_a$, and $\xi_a$ are identified, provided there are no perfect linear relationships among the elements of $\mathbf{x}_{it}$ and that there is some time variation in all elements of $\mathbf{x}_{it}$. (The latter requirement ensures that $\mathbf{x}_{it}$ and $\bar{\mathbf{x}}_i$ are not perfectly collinear for all $t$.) In addition, it follows from the same arguments in Wooldridge (2002, Section 15.8.2) that the average partial effects can be obtained by differentiating or differencing

$$E_{\bar{\mathbf{x}}_i}[\Phi(\psi_a + \mathbf{x}_t\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\xi_a)], \tag{2.10}$$

with respect to the elements of $\mathbf{x}_t$. But, by the law of large numbers, (2.10) is consistently estimated by

$$N^{-1}\sum_{i=1}^{N}\Phi(\psi_a + \mathbf{x}_t\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a). \tag{2.11}$$

Therefore, given consistent estimators of the scaled parameters, we can plug them into (2.11) and consistently estimate the APEs.

Before turning to estimation strategies, it is important to understand why we do not replace (2.1) with the logistic function $E(y_{it}|\mathbf{x}_{it}, c_i) = \Lambda(\mathbf{x}_{it}\boldsymbol{\beta} + c_i)$ and try to eliminate $c_i$ by using conditional logit estimation (often called fixed effects logit). As discussed by Wooldridge (2002, Section 15.8.3), the logit conditional MLE is not known to be consistent unless the response variable is binary and, in addition to the strict exogeneity assumption (2.5), the $y_{it}, t = 1, \ldots, T$, are independent conditional on $(\mathbf{x}_i, c_i)$. Therefore, even if the $y_{it}$ *were* binary responses, we would not necessarily want to use conditional logit to estimate $\boldsymbol{\beta}$ because serial dependence is often an issue even after accounting for $c_i$. Plus, even if we could estimate $\boldsymbol{\beta}$, we would not be able to estimate the average partial effects, or partial effects at interesting values of $c$. For all of these reasons, we follow the approach of specifying $D(c_i|\mathbf{x}_i)$ and exploiting normality.

## 3. Estimation methods under strict exogeneity

Given (2.9), there are many consistent estimators of the scaled parameters. For simplicity, define $\mathbf{w}_{it} \equiv (1, \mathbf{x}_{it}, \bar{\mathbf{x}}_i)$, a $1 \times (1 + 2K)$ vector, and let $\boldsymbol{\theta} \equiv (\psi_a, \boldsymbol{\beta}_a', \boldsymbol{\xi}_a')'$. One simple estimator of $\boldsymbol{\theta}$ is the pooled nonlinear least squares (PNLS) estimator with regression function $\Phi(\mathbf{w}_{it}\boldsymbol{\theta})$. The PNLS estimator, while consistent and $\sqrt{N}$-asymptotically normal (with fixed $T$), is almost certainly inefficient, for two reasons. First, $\mathrm{Var}(y_{it}|\mathbf{x}_i)$ is probably not homoskedastic because of the fractional nature of $y_{it}$. One possible alternative is to model $\mathrm{Var}(y_{it}|\mathbf{x}_i)$ and then to use weighted least squares. In some cases – see Papke and Wooldridge (1996) for the cross-sectional case – the conditional variance can be shown to be

$$\mathrm{Var}(y_{it}|\mathbf{x}_i) = \tau^2\Phi(\mathbf{w}_{it}\boldsymbol{\theta})[1 - \Phi(\mathbf{w}_{it}\boldsymbol{\theta})], \tag{3.1}$$

where $0 < \tau^2 \leq 1$. Under (2.9) and (3.1), a natural estimator of $\boldsymbol{\theta}$ is a pooled weighted nonlinear least squares (PWNLS) estimator, where PNLS would be used in the first stage to estimate the weights. But an even simpler estimator avoids the two-step estimation and is asymptotically equivalent to PWNLS: the pooled Bernoulli quasi-MLE (QMLE), which is obtained my maximizing the pooled probit log-likelihood. We call this the "pooled fractional probit" (PFP) estimator. The PFP estimator is trivial to obtain in econometrics packages that support standard probit estimation – provided, that is, the program allows for nonbinary response variables. The explanatory variables are specified as $(1, \mathbf{x}_{it}, \bar{\mathbf{x}}_i)$. Typically, a generalized linear models command is available, as in Stata®. In applying the Bernoulli QMLE, one needs to adjust the standard errors and test statistics to allow for arbitrary serial dependence across $t$. The standard errors that are robust to violations of (3.1) but assume serial independence are likely to be off substantially; most of the time, they would tend to be too small. Typically, standard errors and test statistics computed to be robust to serial dependence are also robust to arbitrary violations of (3.1), as they should be. (The "cluster" option in Stata® is a good example.)

A test of independence between the unobserved effect and $\mathbf{x}_i$ is easily obtained as a test of $H_0 : \boldsymbol{\xi}_a = \mathbf{0}$. Naturally, it is best to make this test fully robust to serial correlation and a misspecified conditional variance.

Because the PFP estimator ignores the serial dependence in the joint distribution $D(y_{i1}, \ldots, y_{iT}|\mathbf{x}_i)$ – which is likely to be substantial even after conditioning on $\mathbf{x}_i$ – it can be inefficient compared with estimation methods that exploit the serial dependence. Yet modeling $D(y_{i1}, \ldots, y_{iT}|\mathbf{x}_i)$ and applying maximum likelihood methods, while possible, is hardly trivial, especially for fractional responses that can have outcomes at the endpoints. Aside from computational difficulties, a full maximum likelihood estimator would produce nonrobust estimators of the parameters of the conditional mean and the APEs. In other words, if our model for $D(y_{i1}, \ldots, y_{iT}|\mathbf{x}_i)$ is misspecified but $E(y_{it}|\mathbf{x}_i)$ is correctly specified, the MLE will be inconsistent for the conditional mean parameters and resulting APEs. (Loudermilk (2007) uses a two-limit Tobit model in the case where a lagged dependent variable is included among the regressors. In such cases, a full joint distributional assumption is very difficult to relax. The two-limit Tobit model is ill-suited for our application because, although our response variable is bounded from below by zero, there are no observations at zero.) Our goal is to obtain consistent estimators under assumptions (2.1), (2.5) and (2.6) only. Nevertheless, we can gain some efficiency by exploiting serial dependence in a robust way.

Multivariate weighted nonlinear least squares (MWNLS) is ideally suited for estimating conditional means for panel data with strictly exogenous regressors in the presence of serial correlation and heteroskedasticity. What we require is a parametric model of $\mathrm{Var}(\mathbf{y}_i|\mathbf{x}_i)$, where $\mathbf{y}_i$ is the $T \times 1$ vector of responses. The model in (3.1) is sensible for the conditional variances, but obtaining the covariances $\mathrm{Cov}(y_{it}, y_{ir}|\mathbf{x}_i)$ is difficult, if not impossible, even if $\mathrm{Var}(\mathbf{y}_i|\mathbf{x}_i, c_i)$ has a fairly simple form (such as being diagonal). Therefore, rather than attempting to find $\mathrm{Var}(\mathbf{y}_i|\mathbf{x}_i)$, we use a "working" version of this variance, which we expect to be misspecified for $\mathrm{Var}(\mathbf{y}_i|\mathbf{x}_i)$. This is the approach underlying the generalized estimating equation (GEE) literature when applied to panel data, as described in Liang and Zeger (1986). In the current context we apply this approach after having modeled $D(c_i|\mathbf{x}_i)$ to arrive at the conditional mean in (2.9).

It is important to understand that the GEE and MWNLS are asymptotically equivalent whenever they use the same estimates of the matrix $\mathrm{Var}(\mathbf{y}_i|\mathbf{x}_i)$. In other words, GEE is quite familiar to economists once we allow the model of $\mathrm{Var}(\mathbf{y}_i|\mathbf{x}_i)$ to be misspecified. To this end, let $\mathbf{V}(\mathbf{x}_i, \boldsymbol{\gamma})$ be a $T \times T$ positive definite matrix, which depends on a vector of parameters, $\boldsymbol{\gamma}$, and on the entire history of the explanatory variables. Let $\mathbf{m}(\mathbf{x}_i, \boldsymbol{\theta})$ denote the conditional mean function for the vector $\mathbf{y}_i$. Because we assume the mean function is correctly specified, let $\boldsymbol{\theta}_o$ denote the value such that $E(\mathbf{y}_i|\mathbf{x}_i) = \mathbf{m}(\mathbf{x}_i, \boldsymbol{\theta}_o)$. In order to apply MWNLS, we need to estimate the variance parameters. However, because this variance matrix is not assumed to be correctly specified, we simply assume that the estimator $\hat{\boldsymbol{\gamma}}$ converges, at the standard $\sqrt{N}$ rate, to some value, say $\boldsymbol{\gamma}^*$. In other words, $\boldsymbol{\gamma}^*$ is defined as the probability limit of $\hat{\boldsymbol{\gamma}}$ (which exists quite generally) and then we assume additionally that $\sqrt{N}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*)$ is bounded in probability. This holds in regular parametric settings.

Because (3.1) is a sensible variance assumption, we follow the GEE literature and specify a "working correlation matrix". The most convenient working correlation structures, and those that are programmed in popular software packages, assume correlations that are not a function of $\mathbf{x}_i$. For our purposes, we focus on a particular correlation matrix that is well-suited for panel data applications with small $T$. In the GEE literature, it is called an "exchangeable" correlation pattern, where we act as if the standardized errors have a constant correlation. To be precise, define, for each $i$, the errors as

$$u_{it} \equiv y_{it} - E(y_{it}|\mathbf{x}_i) = y_{it} - m_t(\mathbf{x}_i, \boldsymbol{\theta}_o), \quad t = 1, \ldots, T, \tag{3.2}$$

where, in our application, $m_t(\mathbf{x}_i, \boldsymbol{\theta}) = \Phi(\mathbf{w}_{it}\boldsymbol{\theta}) = \Phi(\psi_a + \mathbf{x}_{it}\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a)$. Generally, especially if $y_{it}$ is not an unbounded, continuous variable, the conditional correlations, $\mathrm{Corr}(u_{it}, u_{is}|\mathbf{x}_i)$, are a function of $\mathbf{x}_i$. Even if they were not a function of $\mathbf{x}_i$, they would generally depend on $(t, s)$. A simple "working" assumption is that the correlations do not depend on $\mathbf{x}_i$ and, in fact, are the same for all $(t, s)$ pairs. In the context of a linear model, this working assumption is identical to the standard assumption on the correlation matrix in a so-called random effects model; see, for example, Wooldridge (2002, Chapter 10).

If we believe the variance assumption (3.1), it makes sense to define standardized errors as

$$e_{it} \equiv u_{it}/\sqrt{\Phi(\mathbf{w}_{it}\boldsymbol{\theta}_o)[1 - \Phi(\mathbf{w}_{it}\boldsymbol{\theta}_o)]}; \tag{3.3}$$

under (3.1), $\mathrm{Var}(e_{it}|\mathbf{x}_i) = \tau^2$. Then, the exchangeability assumption is that the pairwise correlations between pairs of standardized errors are constant, say $\rho$. Remember, this is a working assumption that leads to an estimated variance matrix to be used in MWNLS. Neither consistency of our estimator of $\boldsymbol{\theta}_o$, nor valid inference, will rest on exchangeability being true.

To estimate a common correlation parameter, let $\tilde{\boldsymbol{\theta}}$ be a preliminary, consistent estimator of $\boldsymbol{\theta}_o$ — probably the pooled Bernoulli QMLE. Define the residuals as $\tilde{u}_{it} \equiv y_{it} - m_t(\mathbf{w}_{it}, \tilde{\boldsymbol{\theta}})$ and the standardized residuals as $\tilde{e}_{it} \equiv \tilde{u}_{it}/\sqrt{\Phi(\mathbf{w}_{it}\tilde{\boldsymbol{\theta}})[1-\Phi(\mathbf{w}_{it}\tilde{\boldsymbol{\theta}})]}$. Then, a natural estimator of a common correlation coefficient is

$$\tilde{\rho} = [NT(T-1)]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}\tilde{e}_{it}\tilde{e}_{is}. \tag{3.4}$$

Under standard regularity conditions, without any substantive restrictions on $\text{Corr}(e_{it}, e_{is}|\mathbf{x}_i)$, the plim of $\hat{\rho}$ is

$$\text{plim}(\tilde{\rho}) = [T(T-1)]^{-1}\sum_{t=1}^{T}\sum_{s\neq t}E(e_{it}e_{is}) \equiv \rho^*. \tag{3.5}$$

If $\text{Corr}(e_{it}, e_{is})$ happens to be the same for all $t \neq s$, then $\tilde{\rho}$ consistently estimates this constant correlation. Generally, it consistently estimates the average of these correlations across all $(t, s)$ pairs, which we simply define as $\rho^*$.

Given the estimated $T \times T$ working correlation matrix, $\mathbf{C}(\tilde{\rho})$, which has unity down its diagonal and $\tilde{\rho}$ everywhere else, we can construct the estimated working variance matrix:

$$\mathbf{V}(\mathbf{x}_i, \tilde{\boldsymbol{\gamma}}) = D(\mathbf{x}_i, \tilde{\boldsymbol{\theta}})^{1/2}\mathbf{C}(\tilde{\rho})\mathbf{D}(\mathbf{x}_i, \tilde{\boldsymbol{\theta}})^{1/2}, \tag{3.6}$$

where $\mathbf{D}(\mathbf{x}_i, \tilde{\boldsymbol{\theta}})$ is the $T \times T$ diagonal matrix with $\Phi(\mathbf{w}_{it}\tilde{\boldsymbol{\theta}})[1 - \Phi(\mathbf{w}_{it}\tilde{\boldsymbol{\theta}})]$ down its diagonal. (Note that dropping the variance scale factor, $\tau^2$, has no effect on estimation or inference.) We can now proceed to the second-step estimation of $\boldsymbol{\theta}_o$ by multivariate WNLS. The MWNLS estimator, say $\hat{\boldsymbol{\theta}}$, solves

$$\min_{\theta}\sum_{i=1}^{N}[\mathbf{y}_i - \mathbf{m}_i(\mathbf{x}_i, \boldsymbol{\theta})]'[\mathbf{V}(\mathbf{x}_i, \tilde{\boldsymbol{\gamma}})]^{-1}[\mathbf{y}_i - \mathbf{m}_i(\mathbf{x}_i, \boldsymbol{\theta})], \tag{3.7}$$

where $\mathbf{m}_i(\mathbf{x}_i, \boldsymbol{\theta})$ is the $T \times 1$ vector with $t$th element $\Phi(\mathbf{w}_{it}\boldsymbol{\theta}) \equiv \Phi(\psi_a + \mathbf{x}_{it}\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a)$. Rather than pose the estimation problem as one of minimizing a weighted sum of squared residuals, the GEE approach works directly from the first-order conditions, but this leads to the same estimator.

Asymptotic inference without any assumptions on $\text{Var}(\mathbf{y}_i|\mathbf{x}_i)$ is straightforward. As shown in Liang and Zeger (1986) – see also Wooldridge (2003, Problem 12.11) for the case where the regressors are explicitly allowed to be random – a consistent estimator of $\text{Avar}\sqrt{N}(\hat{\boldsymbol{\theta}} - \theta_o)$ has the "sandwich" form,

$$\left(N^{-1}\sum_{i=1}^{N}\nabla_\theta\hat{\mathbf{m}}_i'\tilde{\mathbf{V}}_i^{-1}\nabla_\theta\hat{\mathbf{m}}_i\right)^{-1} \times \left(N^{-1}\sum_{i=1}^{N}\nabla_\theta\hat{\mathbf{m}}_i'\tilde{\mathbf{V}}_i^{-1}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\tilde{\mathbf{V}}_i^{-1}\nabla_\theta\hat{\mathbf{m}}_i\right) \times \left(N^{-1}\sum_{i=1}^{N}\nabla_\theta\hat{\mathbf{m}}_i'\tilde{\mathbf{V}}_i^{-1}\nabla_\theta\hat{\mathbf{m}}_i\right)^{-1}, \tag{3.8}$$

where $\nabla_\theta\hat{\mathbf{m}}_i$ is the $T \times P$ Jacobian of $\mathbf{m}(\mathbf{x}_i, \boldsymbol{\theta})$ ($P$ is the dimension of $\boldsymbol{\theta}$) evaluated at $\hat{\boldsymbol{\theta}}$, $\tilde{\mathbf{V}}_i \equiv \mathbf{V}(\mathbf{x}_i, \tilde{\boldsymbol{\gamma}})$, and $\hat{\mathbf{u}}_i \equiv \mathbf{y}_i - \mathbf{m}(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$ is the $T \times 1$ vector of residuals. The matrix used for inference about $\boldsymbol{\theta}_o$ is simply (3.8) but without the terms $N^{-1}$.

Expression (3.8) is fully robust in the sense that only $E(\mathbf{y}_i|\mathbf{x}_i) = \mathbf{m}(\mathbf{x}_i, \boldsymbol{\theta}_o)$ is assumed. For fractional responses with unobserved effects, there are no plausible assumptions under which $\mathbf{V}(\mathbf{x}_i, \boldsymbol{\gamma})$ is correctly specified for $\text{Var}(\mathbf{y}_i|\mathbf{x}_i)$ (up to the scale factor $\tau^2$), so we do not consider a nonrobust variance matrix estimator. Other working correlation matrices can be used – in particular, we can estimate a different correlation for each nonredundant $(t, s)$ pair – but we use only the exchangeability structure in our empirical work.

Given any consistent estimator $\hat{\boldsymbol{\theta}}$, we estimate the average partial effects by taking derivatives or changes with respect to the elements of $\mathbf{x}_t$ of

$$N^{-1}\sum_{i=1}^{N}\Phi(\hat{\psi}_a + \mathbf{x}_t\hat{\boldsymbol{\beta}}_a + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}_a). \tag{3.9}$$

For example, if $x_{t1}$ is continuous, its APE is

$$N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{a1}\phi(\hat{\psi}_a + \mathbf{x}_t\hat{\boldsymbol{\beta}}_a + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}_a), \tag{3.10}$$

and we can further average this across $\mathbf{x}_{it}$, if desired, or plug in the average value of $\mathbf{x}_t$ (or other interesting values). An asymptotic standard error for (3.10) is given in the Appendix. For a quick comparison of the linear model estimates and the fractional probit estimates (whether estimated by pooled QMLE or GEE), it is useful to have a single scale factor (at least for the roughly continuous elements of $\mathbf{x}_t$ ). This scale factor averages out $\mathbf{x}_{it}$ across all time periods and is given by

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(\hat{\psi}_a + \mathbf{x}_{it}\hat{\boldsymbol{\beta}}_a + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}_a); \tag{3.11}$$

we can the fractional probit coefficients, $\hat{\beta}_{aj}$ by the scale factor to obtain APEs (again, for explanatory variables where it makes sense to compute a derivative). We might obtain a different scale factor for each $t$, particularly if we have allowed estimated effects to change across time in estimating a linear model.

If $x_{t1}$ is, say, a binary variable, the average partial effect at time $t$ can be estimated as

$$N^{-1}\sum_{i=1}^{N}[\Phi(\hat{\psi}_a + \hat{\beta}_{a1} + \mathbf{x}_{it(1)}\hat{\boldsymbol{\beta}}_{a(1)} + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}_a) - \Phi(\hat{\psi}_a + \mathbf{x}_{it(1)}\hat{\boldsymbol{\beta}}_{a(1)} + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}_a)], \tag{3.12}$$

where $\hat{\beta}_{a1}$ is the coefficient on $x_{t1}$, $\mathbf{x}_{it(1)}$ denotes all covariates except $x_{it1}$, and $\hat{\boldsymbol{\beta}}_{a(1)}$ is the corresponding vector of coefficients. In other words, for each unit we predict the difference in mean responses with ($x_{t1} = 1$) and without ($x_{t1} = 0$) "treatment", and then average the difference in these estimated mean responses across all units. Again, we can also average (3.12) across $t$ if we want an effect averaged across time as well as cross-sectional unit.

## 4. Models with endogenous explanatory variables

In our application to studying the effects of spending on test pass rates, there are reasons to think spending could be correlated with time-varying unobservables, in addition to being correlated with district-level heterogeneity. For example, districts have latitude about how to spend "rainy day" funds, and this could depend on the abilities of the particular cohort of students. (After all, districts are under pressure to obtain high pass rates, and improving pass rates, on standardized tests.) In this section, we propose a simple control function method for allowing a continuous endogenous explanatory variable, such as spending.

How should we represent endogeneity in a fractional response model? For simplicity, assume that we have a single endogenous explanatory variable, say, $y_{it2}$; the extension to a vector is straightforward provided we have sufficient instruments. We now express the conditional mean model as

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, c_{i1}, v_{it1}) = E(y_{it1}|y_{it2}, \mathbf{z}_{it1}, c_{i1}, v_{it1}) = \Phi(\alpha_1 y_{it2} + \mathbf{z}_{it1}\boldsymbol{\delta}_1 + c_{i1} + v_{it1}), \quad (4.1)$$

where $c_{i1}$ is the time-constant unobserved effect and $v_{it1}$ is a time-varying omitted factor that can be correlated with $y_{it2}$, the potentially endogenous variable. Equations similar to (4.1) have been employed in related cross-sectional contexts, particularly by Rivers and Vuong (1988) for the binary response case. Wooldridge (2005) showed how the Rivers and Vuong control function (CF) approach extends readily to the fractional response case for cross-sectional data; in effect, we would set $t = 1$ and drop $c_{i1}$ from (4.1). The exogenous variables are $\mathbf{z}_{it} = (\mathbf{z}_{it1}, \mathbf{z}_{it2})$, where we need some time-varying, strictly exogenous variables $\mathbf{z}_{it2}$ to be excluded from (4.1). This is the same as the requirement for fixed effects two-stage least squares estimation of a linear model.

As before, we model the heterogeneity as a linear function of all exogenous variables, including those omitted from (4.1). This allows the instruments to be systematically correlated with time-constant omitted factors:

$$c_{i1} = \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + a_{i1}, a_{i1}|\mathbf{z}_i \sim \text{Normal}(0, \sigma_{a_1}^2). \quad (4.2)$$

(Actually, for our application in Section 5, we have more specific information about how our instrument is correlated with historical factors, and we will exploit that information. For the development here, we use (4.2).) Plugging into (4.1) we have

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, a_{i1}, v_{it1}) = \Phi(\alpha_1 y_{it2} + \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + a_{i1} + v_{it1}) \equiv \Phi(\alpha_1 y_{it2} + \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + r_{it1}). \quad (4.3)$$

All we have done in Eq. (4.3) is to replace the unobserved heterogeneity $c_{i1}$ with $a_{i1}$, with the key difference being that $a_{i1}$ can be assumed to be independent of $\mathbf{z}_i$. Next, we assume a linear reduced form for $y_{it2}$:

$$y_{it2} = \psi_2 + \mathbf{z}_{it}\delta_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + v_{it2}, \quad t = 1, \ldots, T, \quad (4.4)$$

where, if necessary, we can allow the coefficients in (4.4) to depend on $t$. The addition of the time average of the strictly exogenous variables in (4.4) follows from the Mundlak (1978) device. The nature of endogeneity of $y_{it2}$ is through correlation between $r_{it1} = a_{i1} + v_{it1}$ and the reduced form error, $v_{it2}$. Thus, $y_{it2}$ is allowed to be correlated with unobserved heterogeneity and the time-varying omitted factor. We also assume that $r_{it1}$ given $v_{it2}$ is conditionally normal, which we write as

$$r_{it1} = \eta_1 v_{it2} + e_{it1}, \quad (4.5)$$

$$e_{it1}|(\mathbf{z}_i, v_{it2}) \sim \text{Normal}(0, \sigma_{e_1}^2), \quad t = 1, \ldots, T. \quad (4.6)$$

Because $e_{it1}$ is independent of $(\mathbf{z}_i, v_{it2})$, it is also independent of $y_{it2}$. Again, using a standard mixing property of the normal distribution,

$$E(y_{it1}|\mathbf{z}_i, y_{it2}, v_{it2}) = \Phi(\alpha_{e1} y_{it2} + \mathbf{z}_{it1}\boldsymbol{\delta}_{e1} + \psi_{e1} + \bar{\mathbf{z}}_i\boldsymbol{\xi}_{e1} + \eta_{e1} v_{it2}), \quad (4.7)$$

where the subscript $e$ denotes division by $(1 + \sigma_{e_1}^2)^{1/2}$. This equation is the basis for CF estimation.

The assumptions used to obtain (4.7) would not typically hold for $y_{it2}$ having discreteness or substantively limited range in its distribution. In our application, $y_{it2}$ is the log of average per-student, district-level spending, and so the assumptions are at least plausible, and might at least provide a good approximation. It is straightforward to include powers of $v_{it2}$ in (4.7) to allow greater flexibility. Following Wooldridge (2005) for the cross-sectional case, we could even model $r_{it1}$ given $v_{it2}$ as a heteroskedastic normal. In this paper, we study (4.7).

In obtaining the estimating Eq. (4.7), there is an important difference between this case and the case covered in Section 3. As can be seen from Eq. (2.9), which is the basis for the estimation methods discussed in Section 3, the explanatory variables are, by construction, strictly exogenous. That is, if $\mathbf{w}_{it} = (1, \mathbf{x}_{it}, \bar{\mathbf{x}}_i)$ as in Section 3, then $E(y_{it}|\mathbf{w}_{i1}, \mathbf{w}_{i2}, \ldots, \mathbf{w}_{iT}) = E(y_{it}|\mathbf{w}_{it})$. In (4.7) we make no assumption about how the expectation would change if we condition on $y_{is2}$, $s \neq t$, which is a strength of our approach. In other words, we explicitly allow contemporaneous endogeneity in (4.7) while also allowing for possible feedback between unobserved idiosyncratic changes in $y_{it1}$, as captured by $v_{it1}$, and future spending, say $y_{i,t+h}$ ,2 for $h \geq 1$. Because we do not have strict exogeneity of $\{y_{it2} : t = 1, \ldots, T\}$ in (4.7), the GEE approach to estimation is generally inconsistent. Therefore, we focus on pooled fractional probit in the second-stage estimation.

We can now summarize how Eq. (4.7) leads to a simple two-step estimation procedure for the scaled coefficients:

**Procedure 4.1.** (i) Estimate the reduced form for $y_{it2}$ (pooled across $t$, or maybe for each $t$ separately; at a minimum, different time period intercepts should be allowed). Obtain the residuals, $\hat{v}_{it2}$ for all $(i, t)$ pairs.

(ii) Use the pooled probit QMLE of $y_{it1}$ on $y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, \hat{v}_{it2}$ to estimate $\alpha_{e1}, \boldsymbol{\delta}_{e1}, \psi_{e1}, \boldsymbol{\xi}_{e1}$ and $\eta_{e1}$.

Because of the two-step nature of Procedure 4.1, the standard errors in the second stage should be adjusted for the first-stage estimation, regardless of the estimation method used; see the Appendix. Alternatively, bootstrapping can be used by resampling the cross-sectional units. We rely on bootstrapping in our empirical work.

Conveniently, if $\eta_{e1} = 0$, the first-stage estimation can be ignored, at least using first-order asymptotics. Consequently, a test for endogeneity of $y_{it2}$ is easily obtained as an asymptotic $t$ statistic on $\hat{v}_{it2}$; it should be make robust to arbitrary serial correlation and misspecified variance. Adding first-stage residuals to test for endogeneity of an explanatory variables dates back to Hausman (1978). In a cross-sectional context, Rivers and Vuong (1988) suggested it for the probit model.

How do we interpret the scaled estimates that are obtained from Procedure 4.1? The coefficients certainly give directions of effects and, because they are scaled by the same factor, relative effects on continuous variables are easily obtained. For magnitudes, we again want to estimate the APEs. In model (4.1), the APEs are obtained by computing derivatives, or obtaining differences, in

$$E_{(c_{i1}, v_{it1})}[\Phi(\alpha_1 y_{t2} + \mathbf{z}_{t1}\boldsymbol{\delta}_1 + c_{i1} + v_{it1})] \quad (4.8)$$

with respect to the elements of $(y_{t2}, \mathbf{z}_{t1})$. From Wooldridge (2002, Section 2.2.5), (4.8) can be obtained as

$$E_{(\bar{\mathbf{z}}_i, v_{it2})}[\Phi(\alpha_{e1} y_{t2} + \mathbf{z}_{t1}\boldsymbol{\delta}_{e1} + \psi_{e1} + \bar{\mathbf{z}}_i\boldsymbol{\xi}_{e1} + \eta_{e1} v_{it2})]; \quad (4.9)$$

that is, we "integrate out" $(\bar{\mathbf{z}}_i, v_{it2})$ and then take derivatives or changes with respect to the elements of $(\mathbf{z}_{t1}, y_{t2})$. Because we are not making a distributional assumption about $(\bar{\mathbf{z}}_i, v_{it2})$, we instead estimate the APEs by averaging out $(\bar{\mathbf{z}}_i, \hat{v}_{it2})$ across the sample, for a chosen $t$:

$$N^{-1}\sum_{i=1}^{N}\Phi(\hat{\alpha}_{e1} y_{t2} + \mathbf{z}_{t1}\hat{\boldsymbol{\delta}}_{e1} + \hat{\psi}_{e1} + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_{e1} + \hat{\eta}_{e1}\hat{v}_{it2}). \quad (4.10)$$

For example, since $y_{t2}$ is continuous, we can estimate its average partial effect as

$$\hat{\alpha}_{e1} \cdot \left[N^{-1}\sum_{i=1}^{N}\phi(\hat{\alpha}_{e1} y_{t2} + \mathbf{z}_{t1}\hat{\boldsymbol{\delta}}_{e1} + \hat{\psi}_{e1} + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_{e1} + \hat{\eta}_{e1}\hat{v}_{it2})\right]. \quad (4.11)$$

**Table 1**
District foundation allowances, 1995–2001, nominal dollars (501 Michigan school districts)

| Year | Minimum | % at Minimum | Basic | % Below basic | Maximum | % Above maximum | Median |
|---|---|---|---|---|---|---|---|
| 1995 | 4200 | 5.8 | 5000 | 56.5 | 6500 | 7.4 | 4894 |
| 1996 | 4506 | 5.6 | 5153 | 52.1 | 6653 | 7.4 | 5129 |
| 1997 | 4816 | 5.6 | 5308 | 47.1 | 6808 | 7.4 | 5308 |
| 1998 | 5170 | 8.2 | 5462 | 36.7 | 6962 | 7.4 | 5462 |
| 1999 | 5170 | 8.2 | 5462 | 36.7 | 6962 | 7.4 | 5462 |
| 2000 | 5700 | 56.9 | 5700 | 0 | 7200 | 6.8 | 5700 |
| 2001 | 6000 | 56.7 | 6000 | 0 | 7500 | 6.4 | 6000 |

If desired, we can further average this across the $\mathbf{z}_{it1}$ for selected values of $y_{t2}$. We can average across $t$, too, to obtain the effect averaged across time as well as across the cross section. As in the strictly exogenous case, we can also compute a single scale factor,

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(\hat{\alpha}_{e1}y_{it2}+\mathbf{z}_{it1}\hat{\boldsymbol{\delta}}_{e1}+\hat{\psi}_{e1}+\bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_{e1}+\hat{\eta}_{e1}\hat{v}_{it2}), \quad (4.12)$$

which gives us a number to multiply the fractional probit estimates by to make them comparable to linear model estimates.

Obtaining a standard error for (4.11) is a challenge because of the two-step estimation, averaging across $i$, and the nonlinear nature of the function. The Appendix derives a valid standard error using the delta method, and the bootstrap can also be used.

## 5. Application to test pass rates

We apply the methods described above to the problem of estimating the effects of spending on math test outcomes for fourth graders in Michigan. Papke (2005) describes the policy change in Michigan in 1994, where funding for schools was changed from a local, property-tax based system to a statewide system supported primarily through a higher sales tax (and lottery profits). For her econometric analysis, Papke used building-level data for the years 1993 through 1998. Here we use district-level data, for a few reasons. First, we extended the sample through 2001 and, due to changes in the government agencies in charge of collecting and reporting the data, we were not able to obtain spending data at the building level. Second, the district-wide data had many fewer missing observations over the period of interest. The nonlinear models we apply are difficult to extend to unbalanced panel data – a topic for future research. Third, the instrumental variable that we use for spending, the so-called "foundation grant", varies only at the district level. Consequently, probably little is lost by using district-level data on all variables.

The data set we use in the estimation contains 501 school districts for the years 1995 through 2001. We use the data starting in 1992 to obtain average spending measures, so that spending in previous grades is allowed to affect current performance. The response variable, *math*4, is the fraction of fourth graders passing the Michigan Education Assessment Program (MEAP) fourth-grade math test in the district. Papke (2005) provides a discussion about why this is the most reliable measure of achievement: briefly, its definition has remained the same over time and the nature of the math test has not radically changed. Any remaining systematic differences can be accounted for via aggregate year effects. Unfortunately, unlike the pass rates, which have been reported on the Michigan Department of Education web site since the mid-1990s, a long stretch of data on the raw scores is not readily available. In addition, for better or worse, the lawmakers and the public have focused its attention on the district-level pass rates.

Starting in the 1994/1995 school year, each district was awarded a per-student "foundation grant", which was based on funding in 1993/1994. Table 1 describes the evolution of the foundation grant (in nominal dollars) from 1995 through 2001. Initially, roughly 6% of the very low spending districts were brought up to the minimum grant of \$4200 (or given an extra \$250 per student, whichever resulted in the larger amount). Many of these districts were funded at levels considerably less than \$4200 per student. Districts that were funded at levels between \$4200 and \$6500 received amounts between \$160 and \$250, with lower-funded districts receiving higher amounts. This comprised about 85% of the districts. Districts funded above \$6500 received an extra \$160.

**Table 2**
Percentiles, district-level real expenditures per pupil, 2001\$ (501 districts)

| | 10th | 25th | 50th | 75th | 90th |
|---|---|---|---|---|---|
| 1992 | 4328 | 4567 | 4920 | 5582 | 6685 |
| 1993 | 4493 | 4681 | 5075 | 5757 | 6817 |
| 1994 | 4759 | 5013 | 5388 | 6129 | 7064 |
| 1995 | 5359 | 5664 | 6046 | 6746 | 7718 |
| 1996 | 5500 | 5781 | 6199 | 6806 | 7856 |
| 1997 | 5690 | 5954 | 6358 | 6994 | 7853 |
| 1998 | 5782 | 6050 | 6408 | 6935 | 7901 |
| 1999 | 5981 | 6186 | 6545 | 7115 | 8191 |
| 2000 | 6146 | 6346 | 6690 | 7236 | 8300 |
| 2001 | 6341 | 6556 | 6880 | 7439 | 8522 |

As shown in Table 1, initially, more than half of the districts had foundation grants less than the "basic foundation", which represents the minimum amount the state hoped to eventually provide each district. By 1998, just over one-third of the districts were below the basic amount, and by 2000, the minimum and basic amounts were the same. From 1999 to 2000, the minimum grant increased by over 10%, bringing 8.2% of the lowest spending districts with it. By 2001, all districts were receiving at least \$6000 per student. A nontrivial fraction of districts had their nominal per-student funding more than double from 1994 to 2001. At the upper end, the "maximum" foundation allowance increased only from \$6500 to \$7500, although even by 2001, more than 6% of the districts were "hold harmless" districts that were allowed to receive more than the maximum foundation allowance. See Papke (2008) for further discussion on the mechanisms used to fund the extra amounts received by the hold harmless districts.

Table 2 contains percentiles of real per-student spending starting in 1992. In 1992, three years before the reform, districts in the 90th percentile spent 54% more (in real terms) than districts in the 10th percentile. By 2001, the difference had shrunk to 34%. While serious inequality still remained by 2001, the spending gap had certainly closed.

Table 3 contains summary statistics for the first and last years of the sample used in the econometric methods. The average pass rate on the test grew from about 62% in 1995 to almost 76% in 2001. One can attribute the aggregate improvement to a "teaching to the test" phenomenon, making the test easier over time, increased real spending, or some combination of these. Our goal is to use the exogenous change in the way schools are funded to determine if spending has an effect on district-level pass rates.

Table 3 shows that the standard deviation in the real value of the foundation grant fell substantially over the period, reflecting

**Table 3**
Selected sample means (standard deviations)

| | 1995 | 2001 |
|---|---|---|
| Pass rate on fourth-grade math test | .618 (.131) | .755 (.126) |
| Real expenditure per pupil (2001$) | 6329 (986) | 7161 (933) |
| Real foundation grant (2001$) | 5962 (1031) | 6348 (689) |
| Fraction eligible for free and reduced lunch | .280 (.152) | .308 (.170) |
| Enrollment | 3076 (8156) | 3078 (7293) |
| Number of observations (Districts) | 501 | 501 |

the history of grant determination we discussed above. Average enrollment was essentially flat over this period, but the percentage of students eligible for the Federally sponsored free lunch program grew from about 28%–31%, reflecting the slowing down of the Michigan economy.

In deciding on a sensible econometric model, we need to consider how spending over a student's school history should affect test performance. In using school-level data from 1993–1998, Papke (2005) found that spending lagged one year had as much, if not a larger, impact than current spending (although collinearity makes the individual coefficients much less precise than the estimated total effect). In fact, one can imagine that, if spending in third and fourth grade can affect achievement in fourth grade, why not spending in earlier years? Because we have spending back to 1992 and have added the years 1999, 2000, and 2001, we can obtain average per-student spending in first, second, third, and fourth grade and still have seven years of data for actual estimation. Specifically, we convert spending into real dollars, and use a simple average. In other words, our spending measure is $avgrexp = (rexppp + rexppp_{-1} + rexppp_{-2} + rexppp_{-3})/4$, where *rexppp* is real expenditures per pupil. We use log(*avgrexp*) as the spending variable in our models. (We experimented with a weighted average with weights equal to .4, .3, .2, and .1, as well as a two-year average, and the results were very similar.)

In addition to the spending variable and year dummies, we include the fraction of students eligible for the free and reduced-price lunch programs (*lunch*) and district enrollment (in logarithmic form, log(*enroll*)).

The linear unobserved effects model estimated by Papke (2005) can be expressed as

$$math4_{it} = \theta_t + \beta_1 \log(avgrexp_{it}) + \beta_2 lunch_{it} + \beta_3 \log(enroll_{it}) + c_{i1} + u_{it1} \quad (5.1)$$

where $i$ indexes district and $t$ indexes year. Because spending appears in logarithmic form, a diminishing effect of spending is built in to (5.1). But this model does not ensure that the expected value is between zero and one. Estimating Eq. (5.1) by fixed effects is identical to adding the time averages of the three explanatory variables and using pooled OLS. That makes the fixed effects estimates directly comparable to the quasi-MLE estimates where we add the time averages of the explanatory variables to control for correlation between $c_{i1}$ and the explanatory variables. Let $\mathbf{x}_{it} = [\log(avgrexp_{it}), lunch_{it}, \log(enroll_{it})]$. Then the fractional probit model we estimate has the form

$$E(math4_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}) = \Phi(\psi_{at} + \mathbf{x}_{it}\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a), \quad (5.2)$$

where the presence of $\psi_{at}$ emphasizes that we are allowing a different intercept in each year. (In other words, even if no elements of $\mathbf{x}_{it}$ helped to explain pass rates, we allow the average pass rate to change over time.) Recall from Section 3 that we are able to only identify the scaled coefficients (indexed by $a$). Nevertheless, the APEs are identified and depend precisely on the scaled coefficients, and we compute these to make them comparable to the linear model estimates.

Table 4 contains estimates of the linear model in (5.1) and the fractional probit model in (5.2). For brevity, we report only the coefficients on the three explanatory variables that change across district and time. We use two methods to estimate the fractional probit model. The first is the pooled Bernoulli quasi-MLE, where we treat the leftover serial dependence – due to $a_i$ and possibly correlation in idiosyncratic unobservables – as a nuisance to be addressed through corrected standard errors. To possibly enhance efficiency, and as an informal model specification check, we also estimate the fractional probit model using the generalized estimating equation approach described in Section 3. We report the results using an exchangeable working correlation matrix; the results using an unrestricted correlation matrix were similar. In all cases, the standard errors are robust to general heteroskedasticity and serial correlation; in the GEE estimation we allow for misspecification of the conditional variance matrix.

The three sets of estimates tell a consistent story: spending has a positive and statistically significant effect on math pass rates. The advantage of the linear model is that we can easily obtain the magnitude of the effect. If log spending increases by .10 – meaning roughly a 10% increase in spending – the pass rate is estimated to increase by about .038, or 3.8 percentage points, a practically important effect. The fully robust $t$ statistic on this estimate is 4.95.

In the pooled fractional probit estimation, the estimated coefficient is .881 and it is very statistically significant (fully robust $t$ statistic = 4.26). As we discussed in Section 3, the magnitude of the coefficient is not directly comparable to the estimate from the linear model. The adjustment factor in (3.11) is .337. Therefore, the average partial effect of spending on math pass rates is about .297 (fully robust bootstrap $t = 4.27$), which gives an effect on the pass rate almost one percentage point below the linear model estimate. The estimates from the GEE approach are very similar, with $\hat{\beta}_{a,lavgrexp} = .885$ and the rounded scale factor also equal to .337. Interestingly, for the spending coefficient, the fully robust standard errors for the pooled estimation (.207) and the fully robust standard error for the GEE estimation (.206) are very close. In this application, using an exchangeable working correlation matrix in multivariate weighted nonlinear least squares estimation does not appear to enhance efficiency, even though the estimated "working" correlation is almost .5.

The estimates in Table 4 allow spending to be correlated with time-constant district heterogeneity, but it does not allow spending to react to unobserved idiosyncratic changes that can affect $math_{it}$. For example, a district may experience a bad outcome in year $t$ and decide it must spend more in year $t+1$ to improve its pass rates. Most districts have a "rainy day" fund equal to at least 10% of their total budget on which they can draw. This feedback violates the strict exogeneity discussion used by the methods in Table 4. Or, a district, based on past performance on earlier tests, may realize that a better or worse cohort of students is taking the test in a particular year; this would result in contemporaneous endogeneity and invalidate the previous estimates.

Rather than directly apply the instrumental variables method from Section 4, we use a modification to exploit our particular setup. The change in school funding starting in 1995 brings with it a natural instrumental variable for spending. The initial foundation grant was a nonsmooth function of initial funding, and subsequent increases in the foundation grant for the lowest spending districts introduce additional kinks in the relationship. (See, for example, the entries in Table 1 for the years 1999 and 2000.) In using the foundation grant as an instrumental variable for spending, our identification assumption is that test pass rates depend on unobserved heterogeneity in a smooth fashion, and the relationship between heterogeneity and initial funding is smooth. In effect, we are relying on the nondifferentiability in the relationship between the foundation grant and initial funding. There is no question that per-student spending and the foundation grant are highly (partially) correlated.

**Table 4**
Estimates assuming spending is conditionally strictly exogenous, 1995–2001

| Model: | Linear | Fractional probit | | Fractional probit | |
|---|---|---|---|---|---|
| Estimation method: | Fixed effects | Pooled QMLE | | GEE | |
| | Coefficient | Coefficient | APE | Coefficient | APE |
| log(*arexppp*) | .377 | .881 | .299 | .885 | .298 |
| | (.071) | (.207) | (.069) | (.206) | (.070) |
| *lunch* | −.042 | −.219 | −.074 | −.237 | −.080 |
| | (.073) | (.207) | (.067) | (.209) | (.067) |
| log(*enroll*) | .0021 | .089 | .030 | .088 | .029 |
| | (.0488) | (.138) | (.044) | (.139) | (.045) |
| Working correlation | – | – | | .491 | |
| Scale factor | – | .337 | | .337 | |
| Number of districts | 501 | 501 | | 501 | |

Notes: (i) The variable *arexppp* is the average of real expenditures per pupil for the current and previous three years. (ii) All models contain year dummies for 1996 through 2001. (iii) The fractional probit estimation includes the time averages of the three explanatory variables. (iv) The standard errors for the coefficients, in parentheses, are robust to general second-moment misspecification (conditional variance and serial correlation). (v) The standard errors for the APEs were obtained using 500 bootstrap replications.

To ensure that the foundation grant is exogenous in the pass rate equation, we include log spending in 1994 along with interactions between this spending variable and year dummies. The idea is that spending in 1994 might have a direct outcome on test score performance, or at least an indirect effect due to its correlation with unobserved district heterogeneity. Therefore, we augment the model by including the new set of explanatory variables. We also include the time averages of $lunch_{it}$ and $lenroll_{it}$ to allow them to be correlated with the district unobserved effect.

The reduced form equation for the spending variable is

$$\log(avgrexp_{it}) = \eta_t + \pi_{t1}\log(found_{it}) + \pi_{t2}\log(rexppp_{i,1994}) + \pi_5 lunch_{it} + \pi_6 lenroll_{it} + \pi_5 \overline{lunch}_i + \pi_6 \overline{lenroll}_i + v_{it2}, \quad (5.3)$$

so that there are different year intercepts and different slopes on the foundation grant and 1994 spending variables. We need to test whether the coefficients on the foundation variable are statistically different from zero. We do that via a pooled OLS regression using a fully robust variance matrix. The robust Wald test gives a zero $p$-value to four decimal places. The coefficients on log(*found*) range from about zero to .245. The test that all coefficients are the same also rejects with a $p$-value of zero to four decimal places, so we use the foundation grant interacted with all year dummies as instrumental variables (IVs).

Given the strength of the foundation grant as an instrument for spending, we estimate the model

$$math4_{it} = \theta_t + \beta_1\log(avgrexp_{it}) + \beta_2 lunch_{it} + \beta_3\log(enroll_{it}) + \beta_{4t}\log(rexppp_{i,1994}) + \xi_1\overline{lunch}_i + \xi_2\overline{\log(enroll_i)} + v_{it1} \quad (5.4)$$

by instrumental variables. The results are reported in Column (1) of Table 5. We also report the coefficient on $\hat{v}_{i2}$ to obtain the Hausman (1978) test for endogeneity; its fully robust $t$ statistic is −1.83, providing mild evidence that spending is endogenous. Compared with the estimates that treat spending as strictly exogenous conditional on $c_{i1}$, the spending coefficient increases by a nontrivial amount, to .555 (robust $t = 2.51$). The effect of a 10% increase in spending is now an increase of about 5.6 percentage points on the math pass rate. Papke (2005), using school-level data, and Roy (2003), using very similar district-level data, also found that the IV estimates were above the estimates that treat spending as strictly exogenous, although the effects estimated by Papke are smaller. (Roy (2003) does not include 1994 spending in the model but uses fixed effects IVs, and considers 1996 through 2001 and 1998 through 2001 separately. Roy's spending variable is spending lagged one year; here we average current and three lags of spending.)

Observing an IV estimate above the usual fixed effects estimate is consistent with the idea that districts may allocate more resources in years when it suspects the cohort of students might be lower performers, information that can be obtained from district-level tests given in earlier grades. Plus, as discussed by Papke (2005), the larger IV estimate can be interpreted in light of the local average treatment effect literature (Imbens and Angrist, 1994): the effect of spending on test pass rates is higher among districts that were induced to spend more due to the passage of Proposal A. The districts affected the most were the very-low-spending districts, and one could argue that these districts would benefit most from a large increase in funding. Of course, we cannot dismiss the possibility that the foundation grant is not entirely exogenous in Eq. (5.4).

**Table 5**
Estimates allowing spending to be endogenous, 1995–2001

| Model: | Linear | Fractional probit | |
|---|---|---|---|
| Estimation method: | Instrumental variables | Pooled QMLE | |
| | Coefficient | Coefficient | APE |
| log(*arexppp*) | .555 | 1.731 | .583 |
| | (.221) | (.759) | (.255) |
| *lunch* | −.062 | −.298 | −.100 |
| | (.074) | (.202) | (.068) |
| log(*enroll*) | .046 | .286 | .096 |
| | (.070) | (.209) | (.070) |
| $\hat{v}_2$ | −.424 | −1.378 | – |
| | (.232) | (.811) | – |
| Scale factor | – | .337 | |
| Number of districts | 501 | 501 | |

Notes: (i) The variable *arexppp* is the average of real expenditures per pupil for the current and previous three years. (ii) All models contain year dummies for 1996 through 2001, the log of per pupil spending in 1994, interactions of this variable with a full set of time dummies, and the time averages of the lunch and enrollment variables. (iii) The instrumental variables are the log of the foundation grant and that variable interacted with a full set of year dummies. (iv) The standard errors, in parentheses, are fully robust. For the fractional probit model, standard errors are obtained by bootstrapping the 501 school districts using 500 bootstrap replications.

Finally, we estimate the effect of spending allowing it to be endogenous using the control function approach described in Section 4. We simply add the reduced form residuals, $\hat{v}_{it2}$, obtained from Eq. (5.3), to the pooled fractional probit model, along with the other explanatory variables in (5.4). (For the linear model in (5.4), this would be identical to the 2SLS estimates reported in Table 5.) Ignoring the first-stage estimation, the fully robust $t$ statistic on $\hat{v}_{it2}$ is −1.91, and so we find some evidence against the null hypothesis that spending is conditionally strictly exogenous. Therefore, all standard errors for the fractional probit estimates in Table 5 account for the first-stage estimation as well as being fully robust to serial correlation any conditional variance. We used 500 bootstrap replications to obtain the bootstrap standard errors, where we resample districts but not time periods within districts.

The coefficient estimate on the spending variable is 1.731 (robust bootstrap $t = 2.28$), which is double the coefficient estimate when we assumed spending was strictly exogenous conditional on an unobserved effect. (The $t$ statistic obtained by ignoring the first-stage estimation is about 2.65.) But we must be careful because these estimates are implicitly scaled by different factors. The only sensible comparison is to compute the scaling factor and adjust the coefficient accordingly. The average scale factor across all $i$ and $t$ is .337, so the average partial effect of a 10% increase in spending is about .583 (robust bootstrap $t = 2.28$). This APE estimate is very similar to the coefficient from the linear IV estimate, and suggests that, as in many nonlinear contexts, the linear model does a good job of estimating the average partial effect. As with the linear model estimate, instrumenting for spending makes its effect substantially larger than the estimate that treats spending as strictly exogenous.

We performed several specification tests on the models with and without spending assumed to be exogenous. First, recall that Eq. (5.2) imposes the restriction that $E(c_i|\mathbf{x}_i)$ depends only on the time average, $\bar{\mathbf{x}}_i$, a $1 \times 3$ vector in our application. Following Chamberlain (1980), we can allow $E(c_i|\mathbf{x}_i) = \psi + \mathbf{x}_i\boldsymbol{\lambda}$, where, because $T = 7$ and $K = 3$, $\boldsymbol{\lambda}$ contains 21 unrestricted elements. We can easily test the 18 restrictions imposed by the assumption $E(c_i|\mathbf{x}_i) = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi}$ by using a fully robust Wald statistic. The test soundly rejects the restrictions, yielding a chi-square value of 56.64 and a $p$-value of zero to four decimal places. Nevertheless, the estimated effect of average spending is remarkably similar to when we impose the restriction: the coefficient on *lavgrexp* becomes .928 (robust $t = .210$), the scale factor is .336, and the average partial effect is .312, only slightly above that in Table 4. It is well known that in the linear model pooled OLS using the Chamberlain (1980) device is identical to Mundlak's (1978) more restrictive form, which in turn equals the usual fixed effects estimator. Algebraic equivalence does not hold in nonlinear cases, but it is not surprising that Mundlak and Chamberlain's approaches produce similar spending effects.

We also did some simple checks for functional form misspecification. We added $lavgrexp_{it}^2$ to the specifications in Table 4 and, while its coefficient was negative in all cases, it never had a robust $t$ statistic greater than one in magnitude. In the fractional probit model with exogenous spending – the model estimated in the first column of Table 4 – we applied a RESET-type test suggested by Papke and Wooldridge (1996) in the cross-sectional case. In particular, let $\mathbf{w}_{it}\boldsymbol{\theta} = \psi_{at} + \mathbf{x}_{it}\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a$ denote the index inside the probit function in (5.2). After obtaining the estimates in Table 4, we computed $(\mathbf{w}_{it}\hat{\boldsymbol{\theta}})^2$ and $(\mathbf{w}_{it}\hat{\boldsymbol{\theta}})^3$, and then add these as regressors to the pooled fractional probit estimation. The robust $p$-value for their joint significance is .07, so there is marginal, but not especially strong, evidence against specification (5.2) (especially given that we have several thousand observations). Given our estimates from linear and nonlinear models, and the fact that allowing for the Chamberlain (1980) assumption in $E(c_i|\mathbf{x}_i)$ only slightly changes the spending effect, there is little to suggest that allowing more general functional forms will drastically change estimates of the average partial effects.

As a different functional form check, we also estimated the parameters in Eq. (5.2) by pooled nonlinear least squares. The results are very similar to the pooled QMLE and GEE estimates. For example, the nonlinear least squares coefficient on *lavgrexp* is .905 (standard error = .209), compared with .881 (standard error = .207).

As a check on the overidentifying restrictions obtained by interacting the foundation grant with the year dummies in Table 5, we also obtained the IV and control function estimates when we use the log of the foundation grant as the lone instrumental variable, so that the model is just identified. The results are quite similar (though less precise) than in the overidentified case. Specifically, the coefficient on log(*avgrexp*) in the linear model becomes .534 (robust $t = 2.26$), which is very similar to the .555 estimate using the full set of instruments. For the control function approach using pooled fractional probit in the second stage, the APE for log(*avgrexp*) is about .533 (robust bootstrap $t = 2.01$) — remarkably close to the linear model estimate. The pooled fractional probit APE estimate is slightly below that obtained using the full set of instruments, .583, but it is reasonably close. Therefore, we conclude that the choice of instruments has little effect on the estimated partial effects of interest.

## 6. Further policy discussion

In the previous section we reported statistically significant effects of average real spending on test pass rates using a variety of models and estimation methods. We emphasize again that, for comparisons across different models and estimation methods, it only makes sense to compare average partial effects. The estimated effect of a 10% increase in real spending, averaged over the current and three previous years, ranges from 3.0 to 5.8 percentage points on the pass rate of the fourth-grade MEAP test.

Because linear models using fixed effects are commonly applied for policy questions of the kind we have addressed, there are two important questions that arise from our current work. First, how important is it to account for the nonlinearity in the underlying relationship? Second, how important is it to allow spending to be endogenous after allowing for correlation with unobserved district heterogeneity?

It seems evident that, for estimating the marginal effect of a given percentage change in spending, the difference between linear and nonlinear models is not important, although the average partial effect in when we treat spending as exogenous is about one percentage point higher for the linear model. The more important differences are obtained depending on whether we allow spending to be endogenous or not. This finding is common in applications of nonlinear models with endogenous explanatory variables: when the focus is on average effects, nonlinearity seems less important, often much less so, than allowing endogeneity of the key policy variable. We emphasize, however, that in all cases we are allowing spending to be correlated with unobserved district heterogeneity.

If we move beyond average effects, the picture is somewhat different. We can determine the importance of using a nonlinear model to allow for diminishing spending effects – beyond the diminishing effect of an extra dollar already built in by using log of spending – by obtaining partial effects at different percentiles of the spending distribution. Table 6 contains estimated partial effects for the log spending variable at the 5th, 25th, 50th, 75th, and 95th percentiles of the spending distribution for the earliest year, 1995, and the most recent year, 2001. (We average across all of the other explanatory variables appearing in the model, including the time averages, so these are average partial effects but at a fixed spending level.) We use the pooled fractional probit estimates for the cases where spending is treated as exogenous (second column in Table 4) and where spending is allowed to be endogenous (second column in Table 5).

Whether spending is exogenous or not, the largest APE is at the lowest level of spending, as expected. Given the estimates in Tables 4 and 5, it is not surprising that the spending effect is larger at all percentiles when spending is allowed to be endogenous. In 2001, the APE starting at the 5th percentile is .599 (standard error = .273) when spending is endogenous, but only .292 (standard error = .072) when spending is exogenous. Of course, the control function estimate is much less precise. While the estimated difference in the effects, .307, is nontrivial, the bootstrap standard

**Table 6**
Average partial effects of log(average real expenditures) at different percentiles of the average real expenditures distribution

| | 1995 | | 2001 | |
|---|---|---|---|---|
| Percentile | Spending exogenous APE (standard error) | Spending endogenous APE (standard error) | Spending exogenous APE (standard error) | Spending endogenous APE (standard error) |
| 5th | .341 (.081) | .664 (.268) | .292 (.072) | .599 (.273) |
| 25th | .337 (.080) | .660 (.274) | .283 (.069) | .569 (.254) |
| 50th | .332 (.078) | .646 (.267) | .276 (.065) | .543 (.234) |
| 75th | .321 (.072) | .601 (.229) | .264 (.060) | .498 (.198) |
| 95th | .292 (.057) | .471 (.127) | .229 (.043) | .365 (.101) |

Note: These average partial effects are computed for the fractional probit estimates in Tables 4 and 5. The standard errors are obtained from 500 bootstrap replications.

error of the difference is .283, and so the $t$ statistic for testing equality is about 1.1.

The APE of log average real spending decreases as we increase the spending percentile, although it does so modestly when spending is exogenous. When spending is endogenous, the estimated difference between the 5th and 95th percentiles is much larger. For example, in 2001, the estimated effects on pass rates of a 10% increase in average real spending is 6% points at the 5th percentile and 3.7% points at the 95th percentile (with the $t$ statistic of the difference being only 1.26). Because it is difficult to increase pass rates by a substantial amount if they are already high, one can question the practical importance of the estimated difference. However, we should remember that a 10% increase in spending at a high level of spending is much more in dollar terms than at a low level of spending. For example, even in the last year in our sample, a 10% increase in spending at the 95th percentile translates into about a 15% increase in spending at the 5th percentile. Using the partial effects in Table 6, when spending is allowed to be endogenous, we estimate that the pass rate at low-spending districts increases by 9 percentage points on average compared with the gain of 3.7 points for high-spending districts. This difference strikes us as nontrivial.

Our conclusion, which is tempered by the imprecision in some of the estimates – especially when spending is allowed to be endogenous – is that allowing for nonlinearity is moderately important for estimating a diminishing marginal effect of spending on test pass rates. Because the nonlinear models we suggest are simple to estimate, and partial effects with heterogeneity averaged out are easy to obtain, there is little cost to using "fractional probit" for fractional responses with panel data. At a minimum, they can be used to supplement traditional linear models.

Finally, as commonly happens in situations where instrumental variables methods are compared with methods that assume exogeneity, the IV estimates are much less precise. For the problem of estimating the effects of spending on student performance, our cautious conclusion is that allowing for spending to be endogenous seems nominally important, but sampling error makes the differences in the estimated marginal effects, at best, barely statistically significant.

## Acknowledgements

Three anonymous referees, the guest editors, and numerous seminar participants provided comments that substantially improved this work.

## Appendix

In this Appendix we obtain asymptotic standard errors for the two-step estimators in Section 4 and the average partial effects. We use the pooled Bernoulli QMLE, as it is consistent even if, after we account for the contemporaneous endogeneity of spending, spending is not strictly exogenous. In the first stage, we assume that a linear reduced form for $y_{it2}$ has been estimated by pooled OLS. A setup that covers (4.4) as well as the method we apply in Section 5 is

$$y_{it2} = \mathbf{h}_{it}\boldsymbol{\pi}_2 + v_{it2}, \quad t = 1, \ldots, T, \tag{A.1}$$

where $\mathbf{h}_{it}$ can be any $1 \times M$ vector of exogenous variables. Then, under standard regularity conditions,

$$\sqrt{N}(\hat{\boldsymbol{\pi}}_2 - \boldsymbol{\pi}_2) = N^{-1/2}\sum_{i=1}^{N}\mathbf{r}_{i2} + o_p(1), \tag{A.2}$$

where

$$\mathbf{r}_{i2} = \mathbf{A}_2^{-1}\mathbf{H}_i'\mathbf{v}_{i2}, \tag{A.3}$$

$\mathbf{H}_i$ is the $T \times M$ matrix with $t$th row $\mathbf{h}_{it}$, $\mathbf{A}_2 = E(\mathbf{H}_i'\mathbf{H}_i)$, and $\mathbf{v}_{i2}$ is the $T \times 1$ vector of reduced form errors. Next, we write

$$E(y_{it1}|y_{it2}, \mathbf{h}_{it}) = \Phi[\alpha_1 y_{it2} + \mathbf{w}_{it}\boldsymbol{\lambda}_1 + \eta_1(y_{it2} - \mathbf{h}_{it}\boldsymbol{\pi}_2)], \tag{A.4}$$

where we drop the $e$ subscripting on the scaled parameters for notational simplicity. (After all, it is only the scaled parameters we are estimating and that appear in the average partial effects.) The vector $\mathbf{w}_{it}$ is the subset of exogenous variables appearing in the "structural" equation (so $\mathbf{w}_{it} \subset \mathbf{h}_{it}$). In (4.7), $\mathbf{w}_{it} \equiv (1, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i)$. As usual, for identification we need $\mathbf{w}_{it}$ to be a strict subset of $\mathbf{h}_{it}$. Our goal is to obtain the asymptotic variance for the second-step estimators of $\alpha_1, \boldsymbol{\lambda}_1, \eta_1$. We collect these into the parameter vector $\boldsymbol{\theta}_1$. Then, using the pooled Bernoulli QMLE, the first-order condition for $\boldsymbol{\theta}_1$ is

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{s}_{it1}(\hat{\boldsymbol{\theta}}_1; \hat{\boldsymbol{\pi}}_2) \equiv \sum_{i=1}^{N}\mathbf{s}_{i1}(\hat{\boldsymbol{\theta}}_1; \hat{\boldsymbol{\pi}}_2) = \mathbf{0}, \tag{A.5}$$

where $\hat{\boldsymbol{\pi}}_2$ is the first-step pooled OLS estimator and $\mathbf{s}_{it1}(\boldsymbol{\theta}_1; \boldsymbol{\pi}_2)$ is the score of the Bernoulli quasi-log-likelihood for observation $(i, t)$ with respect to $\boldsymbol{\theta}_1$. We can write this score as

$$\mathbf{s}_{it1}(\boldsymbol{\theta}_1; \boldsymbol{\pi}_2) = \frac{\mathbf{g}_{it}'\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)[y_{it1} - \Phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)]}{\Phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)[1 - \Phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)]}, \tag{A.6}$$

where $\mathbf{g}_{it} = (y_{it2}, \mathbf{w}_{it}, v_{it2})$ and we suppress (for now) the dependence of $v_{it2}$ on $\boldsymbol{\pi}_2$. This is simply the score function for a probit log-likelihood, but $y_{it1}$ is not (necessarily) a binary variable. Following Wooldridge (2002, Section 12.5.2), we can obtain the first-order representation for $\sqrt{N}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)$:

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) = \mathbf{A}_1^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{r}_{i1}(\boldsymbol{\theta}_1; \boldsymbol{\pi}_2)\right) + o_p(1), \tag{A.7}$$

where

$$\mathbf{A}_1 = E[-\nabla_{\theta_1}\mathbf{s}_{i1}(\boldsymbol{\theta}_1; \boldsymbol{\pi}_2)] = E\left(\sum_{t=1}^{T}\frac{[\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)]^2\mathbf{g}_{it}'\mathbf{g}_{it}}{\Phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)[1 - \Phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)]}\right), \tag{A.8}$$

$$\mathbf{r}_{i1}(\boldsymbol{\theta}_1; \boldsymbol{\pi}_2) = \mathbf{s}_{i1}(\boldsymbol{\theta}_1; \boldsymbol{\pi}_2) - \mathbf{F}_1\mathbf{r}_{i2}(\boldsymbol{\pi}_2), \tag{A.9}$$

and

$$\mathbf{F}_1 = E[\nabla_{\boldsymbol{\pi}_2}\mathbf{s}_{i1}(\boldsymbol{\theta}_1;\boldsymbol{\pi}_2)] = \eta_1 E\left(\sum_{t=1}^{T}\frac{[\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)]^2\mathbf{g}_{it}'\mathbf{h}_{it}}{\Phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)[1-\Phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)]}\right). \tag{A.10}$$

Therefore,

$$\text{Avar}[\sqrt{N}(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1)] = \mathbf{A}_1^{-1}\text{Var}[\mathbf{r}_{i1}(\boldsymbol{\theta}_1;\boldsymbol{\pi}_2)]\mathbf{A}_1^{-1}. \tag{A.11}$$

Note that when $y_{it2}$ is exogenous, that is, $\eta_1 = 0$, no adjustment is necessary for the first-stage estimation. This is typical when using control function methods to test for endogeneity: under the null hypothesis of exogeneity, it is very easy to compute a valid test statistic because we can ignore the first-stage estimation.

Because $y_{it2}$ is not necessarily strictly exogenous even after we include $v_{it2}$ in the equation, the two scores $\mathbf{s}_{i1}(\boldsymbol{\theta}_1;\boldsymbol{\pi}_2)$ and $\mathbf{r}_{i2}(\boldsymbol{\pi}_2)$ are generally correlated. (The scores for time period $t$ are uncorrelated, but there can be correlation across different time periods.) Generally, a valid estimator of $\text{Avar}[\sqrt{N}(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}_1)]$ is

$$\hat{\mathbf{A}}_1^{-1}\left(N^{-1}\sum_{i=1}^{N}\hat{\mathbf{r}}_{i1}\hat{\mathbf{r}}_{i1}'\right)\hat{\mathbf{A}}_1^{-1}, \tag{A.12}$$

where

$$\hat{\mathbf{A}}_1 = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{[\phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)]^2\hat{\mathbf{g}}_{it}'\hat{\mathbf{g}}_{it}}{\Phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)[1-\Phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)]}, \tag{A.13}$$

$$\hat{\mathbf{g}}_{it} = (y_{it2}, \mathbf{w}_{it}, \hat{v}_{it2}), \qquad \hat{\mathbf{r}}_{i1} = \hat{\mathbf{s}}_{i1} - \hat{\mathbf{F}}_1\hat{\mathbf{r}}_{i2},$$

$$\hat{\mathbf{r}}_{i2} = \hat{\mathbf{A}}_2^{-1}\mathbf{H}_i'\hat{\mathbf{v}}_{i2}, \tag{A.14}$$

and

$$\hat{\mathbf{F}}_1 = \hat{\eta}_1\left[N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{[\phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)]^2\hat{\mathbf{g}}_{it}'\mathbf{h}_{it}}{\Phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)[1-\Phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)]}\right]. \tag{A.15}$$

Note that $\hat{\mathbf{A}}_1$ is just the usual Hessian from the pooled Bernoulli quasi-log-likelihood – that is, the Hessian with respect to $\boldsymbol{\theta}_1$ – divided by the cross-sectional sample size. The asymptotic variance of $\hat{\boldsymbol{\theta}}_1$ is estimated as

$$\widehat{\text{Avar}(\hat{\boldsymbol{\theta}}_1)} = \hat{\mathbf{A}}_1^{-1}\left(N^{-1}\sum_{i=1}^{N}\hat{\mathbf{r}}_{i1}\hat{\mathbf{r}}_{i1}'\right)\hat{\mathbf{A}}_1^{-1}/N. \tag{A.16}$$

Of course the asymptotic standard errors are obtained by the square roots of the diagonal elements of this matrix. As usual, the divisions by $N$ cancel everywhere in (A.16) so that the estimator of $\text{Avar}(\hat{\boldsymbol{\theta}}_1)$ shrinks to zero at the rate $N^{-1}$.

Next, we obtain a standard error for the average partial effects reported in Section 5. First, we obtain a standard error for the vector of scaled coefficients times the scale factor in (4.12), which we write generically as

$$\hat{\boldsymbol{\tau}}_1 = \left((NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)\right)\hat{\boldsymbol{\theta}}_1, \tag{A.17}$$

where $\hat{\mathbf{g}}_{it} = (y_{it2}, \mathbf{w}_{it}, \hat{v}_{it2})$, as before. In the model with $y_{it2}$ assumed exogenous, we drop the term $\hat{v}_{it2}$. Let $\boldsymbol{\tau}_1$ denote the vector of scaled population coefficients, averaged across time, so

$$\boldsymbol{\tau}_1 = \left(T^{-1}\sum_{t=1}^{T}E[\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)]\right)\boldsymbol{\theta}_1. \tag{A.18}$$

Then we need the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\tau}}_1-\boldsymbol{\tau}_1)$. We use Problem 12.12 in Wooldridge (2003), recognizing that the full set of estimated parameters is $\hat{\boldsymbol{\omega}} = (\hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\pi}}_2')'$ (except when we assume $y_{it2}$ is exogenous). Then, letting $\mathbf{j}(\mathbf{g}_i, \mathbf{h}_i, \boldsymbol{\omega}) \equiv T^{-1}\sum_{t=1}^{T}\phi[\alpha_1 y_{it2} + \mathbf{w}_{it}\boldsymbol{\lambda}_1 + \eta_1(y_{it2}-\mathbf{h}_{it}\boldsymbol{\pi}_2)]\boldsymbol{\theta}_1$, we have

$$\sqrt{N}(\hat{\boldsymbol{\tau}}_1-\boldsymbol{\tau}_1) = N^{-1/2}\sum_{i=1}^{N}\left[\left(T^{-1}\sum_{t=1}^{T}\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)\boldsymbol{\theta}_1\right)-\boldsymbol{\tau}_1\right] + E[\nabla_{\boldsymbol{\omega}}\mathbf{j}(\mathbf{g}_i,\mathbf{h}_i,\boldsymbol{\omega})]\sqrt{N}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}) + o_p(1). \tag{A.19}$$

From above, we have the first-order representation for $\sqrt{N}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega})$,

$$\sqrt{N}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}) = N^{-1/2}\sum_{i=1}^{N}\begin{pmatrix}\mathbf{A}_1^{-1}\mathbf{r}_{i1}\\ \mathbf{r}_{i2}\end{pmatrix} + o_p(1) \equiv N^{-1/2}\sum_{i=1}^{N}\mathbf{k}_i + o_p(1), \tag{A.20}$$

where $\mathbf{A}_1$, $\mathbf{r}_{i1}$ and $\mathbf{r}_{i2}$ are defined earlier. So the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\tau}}_1-\boldsymbol{\tau}_1)$ is

$$\text{Var}\left\{\left[\left(T^{-1}\sum_{t=1}^{T}\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)\boldsymbol{\theta}_1\right)-\boldsymbol{\tau}_1\right]+\mathbf{J}(\boldsymbol{\omega})\mathbf{k}_i\right\}, \tag{A.21}$$

where $\mathbf{J}(\boldsymbol{\omega}) \equiv E[\nabla_{\boldsymbol{\omega}}\mathbf{j}(\mathbf{g}_i,\mathbf{h}_i,\boldsymbol{\omega})]$. We only have left to find the Jacobian $\nabla_{\boldsymbol{\omega}}\mathbf{j}(\mathbf{g}_i,\mathbf{h}_i,\boldsymbol{\omega})$, and then to propose obvious estimators for each term. We find the Jacobian first with respect to $\boldsymbol{\theta}_1$ and then with respect to $\boldsymbol{\pi}_2$. For both terms, we need the derivative of the standard normal pdf, $\phi(z)$, which is simply $-z\phi(z)$. Then

$$\nabla_{\boldsymbol{\theta}_1}\mathbf{j}(\mathbf{g}_i,\mathbf{h}_i,\boldsymbol{\omega}) = T^{-1}\sum_{t=1}^{T}\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)[\mathbf{I}_{P_1}-(\mathbf{g}_{it}\boldsymbol{\theta}_1)\boldsymbol{\theta}_1\mathbf{g}_{it}], \tag{A.22}$$

where $\mathbf{I}_{P_1}$ is the $P_1\times P_1$ identity matrix and $P_1$ is the dimension of $\boldsymbol{\theta}_1$. Similarly,

$$\nabla_{\boldsymbol{\pi}_2}\mathbf{j}(\mathbf{g}_i,\mathbf{h}_i,\boldsymbol{\omega}) = \eta_1 T^{-1}\sum_{t=1}^{T}(\mathbf{g}_{it}\boldsymbol{\theta}_1)\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)\boldsymbol{\theta}_1\mathbf{h}_{it} \tag{A.23}$$

is $P_1\times P_2$, where $P_2$ is the dimension of $\boldsymbol{\pi}_2$. It follows that

$$\nabla_{\boldsymbol{\omega}}\mathbf{j}(\mathbf{g}_i,\mathbf{h}_i,\boldsymbol{\omega}) = \left[T^{-1}\sum_{t=1}^{T}\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1) \times [\mathbf{I}_{P_1}-(\mathbf{g}_{it}\boldsymbol{\theta}_1)\boldsymbol{\theta}_1\mathbf{g}_{it}] \mid \eta_1 T^{-1}\sum_{t=1}^{T}(\mathbf{g}_{it}\boldsymbol{\theta}_1)\phi(\mathbf{g}_{it}\boldsymbol{\theta}_1)\boldsymbol{\theta}_1\mathbf{h}_{it}\right], \tag{A.24}$$

and its expected value is easily estimated as

$$\hat{\mathbf{J}} = \mathbf{J}(\hat{\boldsymbol{\omega}}) = (NT)^{-1}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1) \times [\mathbf{I}_{P_1}-(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)\hat{\boldsymbol{\theta}}_1\hat{\mathbf{g}}_{it}] \mid \hat{\eta}_1\sum_{t=1}^{T}(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)\phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)\hat{\boldsymbol{\theta}}_1\mathbf{h}_{it}\right], \tag{A.25}$$

where $\hat{\mathbf{g}}_{it} = (y_{it2}, \mathbf{w}_{it}, \hat{v}_{it2})$. If $y_{it2}$ is assumed to be exogenous, $\hat{v}_{it2}$ is dropped and $\hat{\mathbf{J}}$ consists of only the first term in the partition. Finally, $\text{Avar}[\sqrt{N}(\hat{\boldsymbol{\tau}}_1-\boldsymbol{\tau}_1)]$ is consistently estimated as

$$N^{-1}\sum_{i=1}^{N}\left[\left(T^{-1}\sum_{t=1}^{T}\phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)\hat{\boldsymbol{\theta}}_1\right)-\hat{\boldsymbol{\tau}}_1+\hat{\mathbf{J}}\hat{\mathbf{k}}_i\right] \times \left[\left(T^{-1}\sum_{t=1}^{T}\phi(\hat{\mathbf{g}}_{it}\hat{\boldsymbol{\theta}}_1)\hat{\boldsymbol{\theta}}_1\right)-\hat{\boldsymbol{\tau}}_1+\hat{\mathbf{J}}\hat{\mathbf{k}}_i\right]', \tag{A.26}$$

where all quantities are evaluated at the estimators given previously. This is the full vector of APEs (assuming continuous explanatory variables). The asymptotic standard error for any particular APE is obtained as the square root of the corresponding diagonal element in (A.26), divided by $\sqrt{N}$.

A similar argument gives the APE when we fix one of the regressors at a specific value and consider a specific time period. In our case, this is the spending variable, $y_{t2}$. We fix this variable at $y_{t2}^{o}$ and, for concreteness, consider the APE for the last time period, $T$. Then

$$\hat{\tau}_1 = \hat{\alpha}_1 \left( N^{-1} \sum_{i=1}^{N} \phi(\hat{\alpha}_1 y_{T2}^{o} + \mathbf{w}_{iT}\hat{\boldsymbol{\lambda}}_1 + \hat{\eta}_1 \hat{v}_{iT2}) \right), \tag{A.27}$$

and it can be shown that a consistent estimator of Avar$[\sqrt{N}(\hat{\tau}_1 - \tau_1)]$ is

$$N^{-1} \sum_{i=1}^{N} \left[ \hat{\alpha}_1 \phi(\hat{\mathbf{g}}_{iT}\hat{\boldsymbol{\theta}}_1) - \hat{\tau}_1 + \hat{\mathbf{J}}\hat{\mathbf{k}}_i \right]^2, \tag{A.28}$$

where $\mathbf{k}_i$ is defined in (A.20) and $\hat{\mathbf{J}}$ is now

$$N^{-1} \sum_{i=1}^{N} \Big[ \phi(\hat{\mathbf{g}}_{iT}\hat{\boldsymbol{\theta}}_1)[\mathbf{e}_1' - \hat{\alpha}_1(\hat{\mathbf{g}}_{iT}\hat{\boldsymbol{\theta}}_1)\hat{\mathbf{g}}_{iT}] \mid \hat{\eta}_1(\hat{\mathbf{g}}_{iT}\hat{\boldsymbol{\theta}}_1)\phi(\hat{\mathbf{g}}_{iT}\hat{\boldsymbol{\theta}}_1)\hat{\boldsymbol{\theta}}_1\mathbf{h}_{iT} \Big], \tag{A.29}$$

where $\mathbf{e}_1' = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \end{pmatrix}$. Note that $\hat{\mathbf{g}}_{iT} = (y_{T2}^{o}, \mathbf{w}_{iT}, \hat{v}_{iT2})$.

## References


Chamberlain, G., 1980. Analysis of variance with qualitative data. Review of Economic Studies 47, 225–238.

Hardin, J.W., Hilbe, J.M., 2007. Generalized Linear Models and Extensions, 2e. Stata Press, College Station, TX.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.

Hausman, J.A., Leonard, G.K., 1997. Superstars in the national basketball association: Economic value and policy. Journal of Labor Economics 15, 586–624.

Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467–476.

Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.

Liu, J.L., Liu, J.T., Hammitt, J.K., Chou, S.Y., 1999. The price elasticity of opium in Taiwan, 1914–1942. Journal of Health Economics 18, 795–810.

Loudermilk, M.S., 2007. Estimation of fractional dependent variables in dynamic panel data models with an application to firm dividend policy. Journal of Business and Economic Statistics 25, 462–472.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46, 69–85.

Papke, L.E., 2005. The effects of spending on test pass rates: Evidence from Michigan. Journal of Public Economics 821–839.

Papke, L.E., 2008. The effects of changes in Michigan's school finance system. Public Finance Review 36, 456–474.

Papke, L.E., Wooldridge, J.M., 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. Journal of Applied Econometrics 11, 619–632.

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

Roy, J., 2003. Impact of school finance reform on resource equalization and academic performance: Evidence from Michigan. Princeton University, Education Research Section Working Paper No. 8.

Wagner, J., 2001. A note on the firm size-export relationship. Small Business Economics 17, 229–337.

Wagner, J., 2003. Unobserved firm heterogeneity and the size-exports nexus: Evidence from German panel data. Review of World Economics 139, 161–172.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.

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

Wooldridge, J.M., 2005. Unobserved heterogeneity and estimation of average partial effects. In: Andrews, D.W.K., Stock, J.H. (Eds.), Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, Cambridge, pp. 27–55.