Missing School to Visit the Doctor? Analysis Using a Copula-Based Endogenous Switching Regressions Model

David M. Zimmer*
Western Kentucky University

February 21, 2021

Abstract

This paper investigates the question: Does visiting a doctor in order to obtain curative health care services cause children to miss school? The question has important education-related policy implications in light of evidence of the negative short- and long-run implications of absenteeism. To answer that question, this paper constructs and estimates a copula-based endogenous switching regressions model that addresses potential endogeneity of health care usage and accommodates the discrete count nature of missed school days. The main finding is that visiting a doctor causes children to miss approximately 85 percent more school days. That finding suggests that schools might consider programs that allow students to obtain necessary health care without disrupting their daily schedules.

JEL Codes: I21; I11; C51

Keywords: regime switching; count data; health care utilization

*david.zimmer@wku.edu; Grise Hall 426, Western Kentucky University, Bowling Green, KY 42101.

1 Introduction

This paper seeks to answer the question: Does visiting a doctor in order to obtain curative health care services cause children to miss school? To some extent, that relationship is mechanical, in the sense that *any* activity done during school hours might cause a student to miss school. Instead, the paper focuses on the magnitude of that relationship, with particular focus on the role that endogeneity plays in that link. To address potential endogeneity bias, the econometric approach employs an endogenous switching regressions framework assembled from copula functions. The main finding is that children who visit a doctor miss approximately 85 percent more days compared to their counterparts who do not. Based on that finding, the paper advocates in its concluding section that schools might consider programs that allow students to obtain necessary health care without disrupting their daily schedules.

More than 5 million U.S. children miss at least 1 month of school during an academic year, with absenteeism alarmingly high at the elementary school level (Balfanz and Byrnes, 2012; Jordan and Chang, 2015). Of course, absenteeism, itself, is only a concern to the extent that it leads to other problems. Unfortunately, a large swath of research, scattered across several different academic disciplines, has documented the negative consequences of absenteeism. In the short run, missing school appears to lower performance on standardized tests (Goodman, 2014; Balfanz and Byrnes, 2015; Gershenson, Jacknowitz, and Brannegan, 2017). Moreover, those short-run effects remain even after employing more "structural" estimation approaches that attempt to address potential endogeneity bias (Gottfried, 2010; Gottfried, 2011; Gottfried and Kirksey, 2017). In the longer run, missing school serves as a reliable predictor of failure to graduate high school, as well as early struggles at the college level (Balfanz and Byrnes, 2015; Cabus and De Witte, 2015; Coelho et al., 2015). This paper does not explore those negative consequences of absenteeism; rather, it zeros in on the role that seeking medical care might play in causing absenteeism.

In light of the seemingly robust negative consequences of absenteeism, policy makers have enacted numerous interventions that seek to curb missed school days. At the federal level, the Obama administration launched the "Every Student, Every Day" initiative, which, at its core, calls for better data collection on absenteeism. California enacted a similar program at around the same time, entitled "In School + On Track." Specific school districts have launched smaller, more localized efforts (Ginsburg, Jordan, and Chang, 2014).

At the root of those various initiatives is the question of what, specifically, leads students to miss school. Not surprisingly, a fairly sizable literature has sought to identify such causes. Low household income seems to positively correlate with missed days (Epstein and Sheldon, 2002; Coelho et al., 2015).

Relatedly, children who reside in more crime-infested areas tend to miss more days (Bowen and Bowen, 1999; Gottfried, 2014). Boredom and frustration at school seem to increase absenteeism (Kearney, 2008). Even district infrastructure appears important, with better quality school buildings linked to lower absenteeism (Duran-Narucki, 2008). Of particular interest to this paper, of all the reasons for missing school, the most important determinant seems to be health, with medical problems strongly linked to higher absenteeism (Holbert, Wu, and Stark, 2002; Basch, 2011).

But investigating the link between medical care usage and school absenteeism requires confronting several econometric challenges. First, children might possess unobserved characteristics that simultaneously relate to both medical care usage *and* absenteeism, implying that health care consumption is potentially endogenous with respect to missed days. Second, the main outcome variable, missed school days, is recorded as a discrete count integer, which calls for a formal count model. Third, whereas standard econometric models capture the effect of the decision to obtain medical care on missed school days via a simple intercept shift, this paper offers evidence that the *entire* conditional mean function of missed school days differs according to whether a child visits a doctor or not.

The remainder of the paper is structured as follows. After introducing the estimation sample, drawn from the Medical Expenditure Panel Survey, the paper outlines a basic control function estimator. The paper then specifies an endogenous switching regressions setup based on copula functions, before moving to results and concluding remarks.

2 Data

Data for this study come from the 2015 wave of the Medical Expenditure Panel Survey (MEPS), conducted and published by the Agency for Healthcare Research and Quality, a unit of the U.S. Department of Health and Human Services. Along with the parent database from which it comes, the MEPS enjoys a reputation as the most detailed and complete survey on household-level health and healthcare usage.

This paper focuses on all 2015 respondents ages 6-13, an age range during which schooling is compulsory in most U.S. jurisdictions. The final sample size includes 4,320 unique children. Socioeconomic information for each child comes from the main Household Component file. Information on each child's health care usage during the previous year comes from the Household Component Event files. From those files, this paper focuses on office-based visits to medical providers to obtain "curative" services, defined as visits in which "diagnosis or treatment" was administered. Approximately 33 percent of children in the estimation sample had a curative visit, according to that definition. This paper does not consider visits for "preventive" reasons, such as routine checkups, as those are presumably easier to schedule around school hours.

Table 1 reports sample means partitioned according to whether the child had a curative visit during the previous year. The top of the table reports the mean number of missed school days for health-related issues (not other reasons). Children with curative visits appear to miss almost two more school days, annually, compared to children without curative visits.

However, that difference in mean missed days cannot be interpreted as *causally* linked to medical care usage, as other socioeconomic measures also appear to differ across the two sample partitions. For example, younger children and children from smaller families appear more likely to have curative visits. Blacks and Hispanics appear less likely to have curative visits, compared to their nonblack/nonHispanic counterparts. Similarly, children who speak a language at home other than English have fewer office visits. Household income positively associates with office-based usage, and there also appear to be some regional differences in healthcare usage. Of particular concern, the fact that those *observed* socioeconomic measures appear to differ across the two sample partitions raises the possibility of differences across other *unobserved* dimensions, an issue that subsequent sections of this paper attempt to address.

The methods described in the following sections either require, or are made more robust by, the availability of an "exclusion restriction" that affects a child's propensity to have an office-based curative visit, but does not directly affect a child's number of missed school days. Shown near the bottom of Table 1, being able to travel to one's usual source of care (USC) in less than 30 minutes appears to strongly correlate with a child's likelihood of having an office visit. The likeliest explanation centers on transportation costs, with closer USCs being less "expensive" to transport to, both in terms of direct travel costs and inconveniences of parents rearranging their schedules. At the same time, as argued below, distance from one's USC plausibly remains unrelated to missed school days. Consequently, being able to travel to one's USC in less than 30 minutes serves as an exclusion restriction in the methods outlined below.

3 Control Function Estimator

Let y_i denote the number of days during the past calendar year that child i missed school for health-related reasons. That outcome, recorded as a non-negative discrete count integer, calls for a formal count regression setup. Because the empirical distribution of y_i does not appear highly "overdispersed," in the sense that the variance of missed school days does not substantially exceed its mean, this paper models y_i using a simple Poisson distribution. (Estimates obtained from a Negative Binomial specification did not alter the main conclusions.) Therefore, assuming a Poisson distribution, let the conditional mean of y_i follow

$$\lambda_i = \exp(\rho s_i + \mathbf{X}_i' \boldsymbol{\beta}) \tag{1}$$

where s_i is a binary indicator equaling 1 if the child had any office-based curative visit during the past calendar year, and 0 otherwise. The vector $\mathbf{X}_i$ includes control variables with estimable coefficients $\boldsymbol{\beta}$. The empirical goal is to ascertain the effect of s_i on y_i, captured by the parameter ρ.

Results of a simple Poisson regression, discussed below, show that children who have curative visits miss approximately 70 percent more days, compared to similar students without such visits. However, unobserved factors that affect a child's likelihood of seeking medical care might also relate to his or her propensity to miss school, implying the possible presence of endogeneity bias. For example, health-conscience families might be predisposed toward seeking medical care while also keeping their children home when illnesses spread at school. That concern calls for a method capable of accommodating an endogenous explanatory variable in a Poisson regression.

A popular, and simple, method for accommodating endogenous covariates in nonlinear models is the "control function" approach, which seeks to mimic linear two-stage least squares (Heckman and Robb, 1985; Terza, Basu, and Rathouz, 2008). The method involves first estimating a regression for the endogenous regressor, and then calculating residuals from that regression. Those calculated residuals then appear in the outcome count equation as an additional regressor that "controls" for the endogeneity problem.

In the current context, the first stage involves a linear probability model for whether the child had a curative visit,

$$s_i = \mathbf{Z}_i'\boldsymbol{\gamma} + u_i. \tag{2}$$

Estimates from that regression are used to calculate residuals, $\widehat{u}_i = s_i - \mathbf{Z}_i'\widehat{\boldsymbol{\gamma}}$, where circumflexes denote estimates values. The second stage involves a Poisson regression where the first-stage residuals appear as an additional control. The mean for that second-stage Poisson regression follows

$$\lambda_i = \exp(\rho s_i + \mathbf{X}_i'\boldsymbol{\beta} + \sigma\widehat{u}_i). \tag{3}$$

For identification, the control function approach requires that at least one explanatory variable appear in $\mathbf{Z}_i$ but not in $\mathbf{X}_i$. For such an exclusion restriction, this paper uses the aforementioned binary indicator – which this paper labels "USC quick" – that indicates whether the child can reach the location of his or her usual source of care in less than 30 minutes. In findings not assembled into a table, linear probability estimates of equation (2) suggest that "USC quick" increases the probability of having a curative visit by 14 percentage points, which represents a 42 percent increase relative to mean curative usage. Consequently, the exclusion restriction appears to significantly and nontrivially correlate with medical care usage.

Instrument validity also requires that "USC quick" *not* affect absenteeism, other than indirectly through its effect on health care consumption. Consequently, valid identification rests on the assumption that families choose the location of their residence for reasons *other than* proximity to a primary care provider, such as nearness to employers or quality of schools. The most serious threat to that assumption, then, stems from the possibility that families of children with chronic health conditions – who are likely to miss school for those conditions – choose to locate close to health care providers in order to provide easier access to treat those conditions. Relatedly, instrument validity would come into question if parents of unhealthy children select into certain types of employment in order to afford scheduling flexibility to help children obtain medical services. To explore whether these concerns affect the paper's main findings, alternative estimates, discussed below, include as an additional explanatory variable an indicator of whether the child is in "fair or poor" health. Those alternative estimates also include several variables related to material employment and education. Although those variable are omitted from the paper's baseline models, mostly due to endogeneity concerns, including those additional measures did not appear to alter the paper's main conclusions.

4 Endogenous Switching Regressions

Despite being a popular method for addressing endogeneity in nonlinear settings, the control function approach outlined in the previous section has several drawbacks. First, the method encounters problems when the endogenous explanatory variable shows discreteness, as in this paper, because discreteness gives rise to several possible definitions of "residuals" in the first stage, with different choices potentially leading to conflicting findings (Wooldridge, 2010, p. 746). A second drawback is that the control function approach only permits the endogenous explanatory variable (office visit) to affect the outcome (missed school days) via an intercept shift. That is, from equation (3), a child with a doctor visit will see his conditional mean shift proportional to the constant ρ.

The endogenous switching regressions model, sometimes referred to as a "switching regimes" model, relaxes the two drawbacks mentioned in the previous paragraph. First, it explicitly handles discreteness in office visits. Second, it allows the *entire* outcome response function of children with and without office visits to differ, not just the intercept. This paper demonstrates that avoiding those drawbacks associated with the control function approach uncovers a far larger (positive) effect of office visits on missed school days.

Envisioned by Roy (1951) and formalized by Borjas (1987), the endogenous switching regressions model has a trio of latent random variables $(y_{1i}^*, y_{2i}^*, y_{3i}^*)$ which connect with observable random variables (s_i, y_{2i}, y_{3i}) via the observation rule

$$\begin{aligned} s_i &= \mathbf{1}\{y_{1i}^* > 0\} \\ y_{2i} &= s_i y_{2i}^* \\ y_{3i} &= (1 - s_i) y_{3i}^* \end{aligned} \tag{4}$$

where $\mathbf{1}\{\ \}$ is the indicator function, equaling 1 if the condition inside the braces holds, and 0 otherwise. That observation rule states that when $s_i = 1$, the outcome y_{2i} is observed, but when $s_i = 0$, the outcome y_{3i} is observed.

Stated in terms of the topic studied in this paper, children who have curative visits ($s_i = 1$) have missed school days (y_{2i}), and children who do *not* have a curative visit ($s_i = 0$) have missed school days (y_{3i}). But the statistical properties of y_{2i} and y_{3i} might differ by more than can be captured by a simple intercept shift.

The standard setup for an endogenous switching regressions model assumes joint normality for the trio $(y_{1i}^*, y_{2i}^*, y_{3i}^*)$, with a probit mechanism for y_{1i}^*. But joint normality does not work in the present context, because missed school days follow discrete count distributions. The following section outlines a copula-based approach, originally developed by Smith (2005), for accommodating settings where each member of that latent trio follows a different distribution.

5 Copula-Based Approach

For each member of the trio $(y_{1i}^*, y_{2i}^*, y_{3i}^*)$, let the marginal cumulative distribution functions (cdf) and marginal probability mass functions (pmf) be denoted $F_j(y_{ji}^*)$ and $f_j(y_{ji}^*)$ for $j = 1, 2, 3$. Those marginal distributions may depend upon explanatory variables and other ancillary distributional parameters. Furthermore, let the first variable in the trio relate to the other two via the bivariate cdfs $F_{12}(y_{1i}^*, y_{2i}^*)$ and $F_{13}(y_{1i}^*, y_{3i}^*)$.

With those distributional forms, following Heckman (1990), Heckman and

Honore (1990), and Heckman and Hotz (1989), the endogenous switching regressions model has an (unlogged) likelihood contribution for observation i

$$L_i = \prod_0 \frac{\partial}{\partial y_3} F_{13}(0, y_3) \prod_1 \left(f_2 - \frac{\partial}{\partial y_2} F_{12}(0, y_2) \right) \tag{5}$$

where $\prod_0$ and $\prod_1$ represent products over observations for which $s_i = 0$ and $s_i = 1$, respectively. Despite discreteness of the main variables of interest, partial derivatives reflect continuity for the underlying latent variables. Note that, since the likelihood function does not depend upon the bivariate distribution $F_{23}(y^*_{2i}, y^*_{3i})$, links between those two marginals cannot be (directly) recovered.

Standard endogenous switching regression routines available in statistical software packages impose normality on the marginals $F_j(y^*_{ji})$ and $f_j(y^*_{ji})$, as well as bivariate normality on the distributions $F_{12}(y^*_{1i}, y^*_{2i})$ and $F_{13}(y^*_{1i}, y^*_{3i})$. This paper maintains normality for $F_1(y^*_{1i})$ and $f_1(y^*_{1i})$, giving a probit flavor to the decision to seek medical care. But because the other two outcomes follow count distributions, this paper, following Smith (2005), uses copulas to assemble the relevant pieces of the likelihood function.

5.1 Copula basics

Introduced by Sklar (1973), copula are bivariate cdfs with uniformly distributed marginals. Consider a continuous bivariate distribution function $G(x_1, x_2)$ with univariate marginal distributions $G_1(x_1)$ and $G_2(x_2)$ and inverse probability transforms (quantile functions) G_1^{-1} and G_2^{-1}. Then $x_1 = G_1^{-1}(u_1) \sim G_1$ and $x_2 = G_2^{-1}(u_2) \sim G_2$ where u_1 and u_2 are uniformly distributed variates. The transforms of uniform variates are distributed as G_1 and G_2. Hence

$$G(x_1, x_2) = G(G_1^{-1}(u_1), G_2^{-1}(u_2)) = C(u_1, u_2), \tag{6}$$

is the unique copula associated with the distribution function. By Sklar's theorem, the copula parameterizes a multivariate distribution in terms of its marginals. For a bivariate distribution G, the copula satisfies

$$G(x_1, x_2) = C(G_1(x_1), G_2(x_2); \theta), \tag{7}$$

where θ is usually a scalar-valued dependence parameter.

The main practical advantage of copulas is that, with knowledge of the marginal distributions G_1 and G_2 and a copula function C to glue them together, one can assemble a representation of the otherwise difficult-to-know bivariate distribution $G(x_1, x_2)$. Assuming one has a reasonable idea of what form the marginals should assume, the main onus on the researcher centers on picking an appropriate copula function, for which a large menu of potential off-the-shelf options exists (Nadarajah, Afuecheta, and Chan, 2017).

This paper uses the popular Frank copula (Frank, 1979), for two reasons. First, as shown in the following subsection, its functional form leads to an easy-to-code expression for the likelihood function. Second, it is one of the few copulas that allows complete negative or positive dependence, including the limiting Fréchet bounds. Such flexibility is important because, in the context of endogenous switching regressions, researchers usually wish to impose relatively few *a priori* restrictions on θ, which captures the extent (and direction) to which the binary switch is endogenous with respect to the outcome. By contrast, by using a copula with a narrower range of dependence, a research might inadvertently restrict the type of endogeneity allowed in a manner inconsistent with actual patterns in the data.

5.2 Copula-based endogenous switching regressions

The Frank copula belongs to a larger class of Archimedean copulas that assume the form

$$C = (u_1, u_2) = \vartheta^{-1}(\vartheta(u_1) + \vartheta(u_2))$$

where $\vartheta(\ )$ is a "generator" function. For the Frank copula, the generator takes the form

$$\vartheta(t) = -\log \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}. \tag{8}$$

Following Smith (2005), if both $F_{12}(y^*_{1i}, y^*_{2i})$ and $F_{13}(y^*_{1i}, y^*_{3i})$ are specified as Frank copulas, then the likelihood function in equation (5) simplifies to

$$L_i = \prod\nolimits_0 \frac{\vartheta'(F_3)}{\vartheta'(C_{13})} f_3 \prod\nolimits_1 \left(1 - \frac{\vartheta'(F_2)}{\vartheta'(C_{12})}\right) f_2 \tag{9}$$

where $\vartheta'(t) = \frac{\partial}{\partial t}\vartheta(t)$, and where F_1, which appears in both C_{12} and C_{13}, equals $F_1(0)$. (Recall that, similar to conventional endogenous switching regressions models, the marginal F_1 is normal, giving the decision to seek care a probit mechanism.) Because the generator, shown in equation (8), assumes a simple form, that likelihood expression is fairly easy to code.

The final modeling detail involves the marginal cdfs F_2 and F_3 and their corresponding pmfs f_2 and f_3. As discussed above, with missed school days recorded as a nonnegative discrete integer, and with that outcome showing minimal "overdispersion," this paper assumes Poisson cdfs and pmfs for those expressions, with conditions means $\lambda_{2i} = \exp(\mathbf{X}_i'\boldsymbol{\beta}_2)$ and $\lambda_{3i} = \exp(\mathbf{X}_i'\boldsymbol{\beta}_3)$.

The endogenous switching regressions model achieves formal identification via the nonlinear function forms present in (9). Nonetheless, to achieve more robust identification, and also to facilitate comparisons to the control function approach from Section 3, the marginal for F_1 includes as an additional explanatory variable the aforementioned indicator for whether the child can reach the location of his or her usual source of care in less than 30 minutes.

6 Results

This section first presents results from a simple Poisson regression that makes no attempt to correct for potential endogeneity of health care usage. The section then moves to results obtained from the control function estimator, before finally turning to the endogenous switching regressions setup.

6.1 Simple Poisson regression

The left panel of Table 2 presents results from a simple Poisson regression. Focusing first on the control variables, most of the signs of the coefficient estimates corroborate *a priori* expectations. Age, family size, and family income all negatively correlate with missed school days. Residents of the midwest and south miss fewer days than their counterparts in the west, while residents of the northeast appear to miss more than those in the west. Black children and children who speak non-English at home both appear to miss fewer days.

Turning attention to the top of the table, children who had at least one curative office-based visit to a medical provider during the past year appear to miss more days than children who did not have such a visit. The table also reports the average marginal effect, calculated as the sample average of the difference $\exp(\widehat{\rho}1 + \mathbf{X}_i'\widehat{\boldsymbol{\beta}}) - \exp(\widehat{\rho}0 + \mathbf{X}_i'\widehat{\boldsymbol{\beta}})$. That marginal effect suggests that children with curative visits miss approximately 1.5 more school days than children without curative visits. Compared to the overall sample mean of missed school days (2.1), that marginal effect suggests that curative office visits associate with an approximate 70 percent increase in missed school days.

6.2 Control function approach

The simple Poisson regression makes no attempt to correct for potential endogeneity of medical care usage. The right panel of Table 2 attempts to address that concern using a control function approach. The first step (not shown) estimates a linear probability model for whether the child had a curative visit. That model includes the explanatory variables listed in Table 2, as well as the aforementioned indicator for whether the child can reach the location of his or her usual source of care in less than 30 minutes. Estimated residuals from that first stage then appear as an additional explanatory variable, labeled "control function," in Table 2.

The control function approach does not substantially alter any of the control variables, but it does appear to shrink the estimated coefficient (and marginal effect) of "curative visit." The marginal effect (1.1 days) suggests that curative visits lead to approximately 50 percent more missed days (relative to the overall sample mean). The finding that the marginal effect shrinks would suggest that families inclined to seek medical care are also predisposed to missing school for reasons not necessarily related to visiting a doctor.

However, the estimated coefficient attached to the first-stage residuals, shown near the bottom of Table 2, does not come close to achieving statistical significance (t statistic < 1.0), which, in turn, raises doubts about whether medical care usage actually is endogenous with respect to missed school days.

Based on the lack of statistical significance of the first-stage residuals, one might reasonably conclude that the estimated marginal effect obtained from the simple Poisson regression, while possibly slightly inflated, is closer to the truth.

6.3 Endogenous switching regressions

The control function approach – and other models similar to it – allow medical care usage to affect missed school days via an intercept shift. By contrast, endogenous switching regressions allow the entire conditional mean function to mutate when a child seeks medical care. In doing so, endogenous switching regressions allow for a more complicated form of endogeneity.

Results for the copula-based endogenous switching regressions specification appear in Table 3. The left panel shows probit estimates for the endogenous "switch"; that is, whether the child has a curative visit. The exclusion restriction, shown near the top of the table, suggests that being able to travel to one's USC in less than 30 minutes significantly and nontrivially increases the probability of having a curative visit. Elsewhere, age and family size negatively correlate with curative visits, while family income positively correlates with such visits. Being black or Hispanic negatively relates to office visits, as does speaking a language other than English at home.

The other two panels report Poisson regression estimates for each part of the endogenous switch. Signs and statistical significance mostly align with what appears in Table 2. However, the finding that some coefficients differ substantially between the two splits – for example, the negative effect of being female is far larger among children without office visits – highlights that the effect of medical care usage on missed school days shows more complexity than can be captured by a simple intercept shift.

The bottom of the table reports estimates for the copula dependence terms, θ_{12} and θ_{13}, which capture the presence and direction of endogeneity. Recalling that the control function approach failed to detect evidence of endogeneity, the switching regressions setup tells a more nuanced story. The estimate of θ_{12} equals -0.651, and is highly statistically significant. The interpretation is that, on average, unobserved traits that induce children to have office visits tend to *reduce* missed school days *among children who have office visits*. Such a pattern would be evident if, for example, parents fortunate enough to have access to medical care or scheduling flexibility also happen to be diligent about school attendance. On the other hand, the other dependence term, θ_{13}, does not differ from zero, indicating a lack of such endogeneity among children without curative visits.

The marginal effect of curative visits on missed school days is calculated as,

$$\frac{1}{n}\sum_{i=1}^{n}(\exp(\mathbf{X}_i'\widehat{\boldsymbol{\beta}}_2)-\exp(\mathbf{X}_i'\widehat{\boldsymbol{\beta}}_3)) \tag{10}$$

where $\widehat{\boldsymbol{\beta}}_2$ and $\widehat{\boldsymbol{\beta}}_3$ refer, respectively, to the coefficient estimates reported in the middle and right panels of Table 3. The standard error for that estimate comes from a parametric bootstrap procedure in which the parameter vectors $\widehat{\boldsymbol{\beta}}_2$ and $\widehat{\boldsymbol{\beta}}_3$ are randomly perturbed by drawing them from normal distributions centered on their converged estimates and variances equal to their converged covariance matrices. Using those perturbed vectors, expression (10) is recalculated. After repeating that process several hundred times, the standard deviation of all the replicated calculations of equation (10) serves as the standard error.

Table 4 reports that marginal effect. For comparison, the table also reports the other marginal effects that appear in Table 2. The simple Poisson regression produced a marginal effect of 1.5 days, an increase of approximately 70 percent. The control function approach shrank that effect to 1.1 days, an increase of approximately 50 percent, albeit with no evidence of endogeneity. The switching regressions method, by contrast, produces the largest marginal effect of the three, 1.8 days. That effect represents an approximate 85 percent increase in missed school days, relative to the sample mean.

As noted in Section 3, the main threat to identification is parents of children with chronic conditions – who likely will miss some school anyway – choosing to locate close to medical providers, or selecting into employment arrangements that afford scheduling flexibility. As an informal check of such behavior, all of the models were re-estimated after including (in all equations) an indicator for whether the child reports “fair or poor” health. In addition to that child health indicator, those additional estimations also included controls for maternal employment status, maternal fulltime employment status, and maternal educational attainment. Those additional controls do not appear in the baseline specification reported in this paper, in part due to endogeneity concerns, and in part due to that fact that approximately 9 percent of children in the estimation sample lack mothers. Nonetheless, the inclusion of those additional controls produced conclusions that were qualitatively, and also mostly quantitatively, similar to those reported here.

The main conclusion, therefore, is that visiting a doctor results in a substantial increase – approximately 85 percent – in the number of missed school days. Moreover, unobserved characteristics that cause children to seek care also tend to reduce missed school days, but only among children who seek care. Simpler, but more restrictive methods, such as the popular control function method, misleadingly suggest a far smaller effect, mostly because simpler methods cannot detect the relatively nuanced form of endogeneity bias that applies to children who seek care, but not to those who don't.

7 Conclusion and Policy Implications

This paper investigates the question: Does visiting a doctor in order to obtain curative health care services cause children to miss school? To answer that question, the paper constructs and estimates a copula-based endogenous switching regressions model that (1) addresses potential endogeneity of health care usage; (2) accommodates the discrete count nature of missed school days; and (3) allows the entire conditional mean function to differ according to whether children visited a doctor.

A simple control function estimator finds a positive link between health care usage and missed school days, but that same estimator does not detect any endogeneity bias. The less restrictive switching regressions setup, by contrast, finds a far larger link between health care usage and missed school days, along with significant evidence of endogeneity bias. The main punchline is that visiting a doctor causes children to miss approximately 85 percent more school days.

This topic has obvious education-related policy implications in light of the aforementioned evidence of the negative short- and long-run implications of absenteeism. But this topic might have even larger implications if child absenteeism causes parents to miss work. Labor economists have long recognized that injury or illness, whether to oneself or family members, represents one of the most common reasons for worker absenteeism (see, "The Causes And Costs Of Absenteeism In The Workplace," Forbes, 7/10/2013). Indeed, the U.S. Centers for Disease Control and Prevention (CDC) estimates that health-related worker absenteeism costs employers $225.8 billion annually, or about $1,685 per worker. (See, "Worker Illness and Injury Costs U.S. Employers $225.8 Billion Annually," CDC Foundation Report, 1/28/2015).

This paper does not explicitly explore those negative consequences of absenteeism. Rather, this paper's findings speak to the possible need for schools and employers to consider programs that allow subjects to obtain necessary medical care without disrupting their daily schedules. Indeed, some under-reported provisions of the Affordable Care Act, called "personal responsibility reforms," offer incentives for employers (and schools) to conduct onsite wellness activities (Wiley, 2014). To the extent that such activities substitute for office-based medical services, personal responsibility reforms might offer a path toward divorcing the link between health care usage and absenteeism. The results of this paper represent an argument for a closer examination of the effectiveness of those programs.

References

Balfanz, R. and V. Byrnes (2012). "Chronic absenteeism: summarizing what we know from nationally available data." Johns Hopkins University Center for Social Organization of Schools, 1, 1-46.

Basch, C. (2011). "Healthier students are better learners: a missing link in school reforms to close the achievement gap. Equity matters." Journal of School Health, 81, 593-598.

Borjas, G. (1987). "Self-selection and the earnings of immigrants." American Economic Review, 77, 531-553.

Bowen, N. and G. Bowen (1999). "Effects of crime and violence in neighborhoods and schools on the school behavior and performance of adolescents." Journal of Adolescent Research, 14, 319-342.

Cabus, S. and K. De Witte (2015). "Does unauthorized school absenteeism accelerates the dropout decision? – evidence from a Bayesian duration model." Applied Economic Letters, 22, 266–271.

Coelho, R., Fischer, S., McKnight, F., Matteson, S., and T. Schwarts (2015). "The effects of early chronic absenteeism on third-grade academic achievement measures." Workshop in Public Affairs, University of Wisconsin-Madison.

Durán-Narucki, V. (2008). "School building condition, school attendance, and academic achievement in New York City public schools: a mediation model." Journal of Environmental Psychology, 28, 278-286.

Epstein, J. and S. Sheldon (2002). "Present and accounted for: improving student attendance through family and community involvement." The Journal of Educational Research, 95, 308-318.

Frank, M. (1979). "On the simultaneous associativity of F(x,y) and x+y - F(x,y)." Aequationes Math, 19, 194-226.

Gershenson, S., Jacknowitz, A., and A. Brannegan (2017). "Are student absences worth the worry in US primary schools?" Education Finance and Policy, 12, 137-165.

Ginsburg, A., Jordan, P., and H. Chang (2014). "Absences add up: how school attendance influences student success." Attendance Works, August.

Goodman, J. (2014). "Flaking out: student absences and snow days as disruptions of instructional time." NBER Working Paper, no. w20221.

Gottfried, M. (2010). "Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. American." Educational Research Journal, 47, 434–465.

Gottfried, M. (2011). "The detrimental effects of missing school: evidence from urban siblings." American Journal of Education, 117, 147–182.

Gottfried, M. (2014). "Can neighbor attributes predict school absences?" Urban Education, 49, 216-250.

Gottfried, M. and J. Kirksey (2017). "'When' students miss school: the role of timing of absenteeism on students' test performance." Educational Researcher, 46, 119–130.

Heckman, J. and R. Robb (1985). "Alternative methods for evaluating the impact of interventions." In J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*, 156-245, Cambridge University Press: New York.

Heckman, J. (1990). "Varieties of selection bias." *American Economic Review Papers and Proceedings*, 90, 313-318.

Heckman, J. and B. Honore (1990). "The empirical content of the Roy model." *Econometrica*, 58, 1121-1149.

Heckman, J. and V. Hotz (1989). "Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training." *Journal of the American Statistical Association*, 84, 862-874.

Holbert, T., Wu, L., and M. Stark (2002). "School attendance initiative: the first 3 years: 1998/99-2000/01." Report prepared for US Dept of Justice, National Institute of Law Enforcement and Criminal Justice, Office of Development and Testing.

Jordan, P. and H. Chang (2015). "Mapping the early attendance gap: charting a course for student success." Attendance Works, September.

Kearney, C. (2008). "School absenteeism and school refusal behavior in youth: a contemporary review." Clinical Psychology Review 28, 451-471.

Nadarajah, S., Afuecheta, E., and S. Chan (2017). "A compendium of copulas." Statistica, 77, 279-328.

Roy, A. (1951). "Some thoughts on the distribution of earnings." Oxford Economic Papers, 3, 135-146.

Sklar, A. (1973). "Random variables, joint distributions, and copulas." Kybernetica, 9, 449–460.

Smith, M. (2005). "Using copulas to model switching regimes with an application to child labour." Economic Record, 81, S47-S57.

Wiley, L. (2014). "Access to health care as an incentive for healthy behavior? an assessment of the Affordable Care Act's personal responsibility for wellness reforms." Indiana Health Law Review, 11, 642-715.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data: Second Edition*. MIT Press: Cambridge, MA.

Table 1: Sample means
(2015 MEPS respondents between ages 6-13)

	Curative visit n = 1,412	No curative visit n = 2,908
Missed school days	3.33**	1.50
Age	9.29**	9.55
Female	0.48	0.47
Black	0.15**	0.25
Hispanic	0.38**	0.44
Northeast	0.13	0.13
Midwest	0.21**	0.17
South	0.38**	0.42
West (omitted)	—	—
Family size	4.41**	4.80
Family income relative to poverty line (1-5)	3.03**	2.79
Language other than English spoken at home	0.42**	0.49
Can travel to USC in < 30 minutes	0.90**	0.80

* column means differ p < .10
** column means differ p < .05

Table 2: Poisson regression estimates of missed school days

	Simple Poisson		With control function	
	Coef.	St. Err.	Coef.	St. Err.
Curative visit	0.712^{**}	0.022	0.516^{**}	0.209
Marginal effect	$[1.496^{**}]$	0.047	$[1.085^{**}]$	0.439
Age	-0.025^{**}	0.005	-0.027^{**}	0.005
Female	-0.028	0.021	-0.026	0.021
Black	-0.363^{**}	0.032	-0.395^{**}	0.047
Hispanic	-0.038	0.031	-0.046	0.033
Northeast	0.147^{**}	0.033	0.145^{**}	0.033
Midwest	-0.060^{*}	0.032	-0.059^{*}	0.032
South	-0.215^{**}	0.028	-0.216^{**}	0.028
West (omitted)	–	–	–	–
Family size	-0.070^{**}	0.008	-0.076^{**}	0.010
Family income (1-5)	-0.052^{**}	0.007	-0.049^{**}	0.008
Not English at home	-0.358^{**}	0.029	-0.369^{**}	0.032
Control function	–	–	0.198	0.210
Constant	1.464^{**}	0.069	1.585^{**}	0.145

* $p < .10$

** $p < .05$

Table 3: Endogenous switching regressions estimates of missed school days

	Curative visit?		Curative visit = yes		Curative visit = no	
	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
Can travel to USC in < 30 minutes	0.444^{**}	0.058	—	—	—	—
Age	-0.034^{**}	0.009	-0.022^{**}	0.006	-0.026^{**}	0.007
Female	0.035	0.041	-0.050^{*}	0.029	-0.118^{**}	0.031
Black	-0.489^{**}	0.059	-0.323^{**}	0.047	-0.372^{**}	0.044
Hispanic	-0.136^{**}	0.059	-0.027	0.042	-0.027	0.045
Northeast	-0.051	0.069	0.131^{**}	0.046	0.163^{**}	0.047
Midwest	-0.003	0.062	-0.084^{*}	0.043	-0.014	0.046
South	-0.020	0.052	-0.130^{**}	0.038	-0.299^{**}	0.041
West (omitted)	—	—	—	—	—	—
Family size	-0.098^{**}	0.014	-0.061^{**}	0.011	-0.071^{**}	0.010
Family income (1-5)	0.033^{**}	0.015	-0.051^{**}	0.010	-0.046^{**}	0.011
Not English at home	-0.154^{**}	0.056	-0.291^{**}	0.041	-0.425^{**}	0.043
Constant	0.073	0.139	2.059^{**}	0.092	1.553^{**}	0.097
			$\theta_{12} = -0.651^{**}$	0.148	$\theta_{13} = 0.008$	0.258

* p < .10

** p < .05

Table 4: Marginal effects of curative visit on missed school days (mean missed days = 2.1)

	Effect	St. Err.	% change relative to mean
Simple poisson	1.496**	0.047	70%
Control function poisson	1.085**	0.439	50%
Endogenous switching regressions	1.764**	0.071	85%

* p < .10
** p < .05