Analysis of School Absenteeism for Single and Full Parent Families: Finite Mixture Roy Approach

Murat K. Munkin
Department of Economics
University of South Florida
Tampa, FL, 33620, U.S.A.
Email: mmunkin@usf.edu
Phone: +1 (813) 974 6517

David Zimmer
Department of Economics
Western Kentucky University
Bowling Green, KY 42101
Email: david.zimmer@wku.edu
Phone: (270) 745 2880

Abstract

This paper analyzes factors affecting school absenteeism due to an injury or illness among the US school student population between 6 and 15 years of age. The number of missed school days is an overdispersed count, modeled in a flexible semiparametric way, using the Finite Mixture Roy (FMR) model for count variables, developed by Munkin (2022). The full/single parent family status (treatment) is potentially endogenous to the dependent variable. The Roy modeling structure captures observed heterogeneity defined by mother's marital status. The FMR model further controls for unobserved heterogeneity using finite mixtures. The objective is to identify components within the data in both states, where the assumption of homogeneity in marginal and treatment effects is more realistic. The considered application motivates two additional features of the model. First, to better understand the structures of the latent components their probabilities are modeled as functions of regressors. Secondly, mother's income is allowed to enter the treatment equation nonparametrically. The FMR model is estimated with two components in each states, interpreted as healthy and unhealthy students. Marital status decreases annual missed school days by about 13 percent for a randomly selected individual, but increases it by 9 percent for those families who actually select to have a single parent, which is evidence of adverse selection.

Key words: School Absenteeism; Marital Status; Finite Mixtures.

JEL Codes: C11, C14

1. Introduction

This paper analyzes factors affecting school absenteeism using 2015 and 2016 waves of the Medical Expenditure Panel Survey (MEPS). School absenteeism is measured as the number of missed school days due to health related reasons during the most recent school year. The dependent variable displays overdispersion and it is analyzed with the Finite Mixture Roy (FMR) model developed by Munkin (2022). A special focus is given to the marital status of the child's mother, the treatment variable, allowed to be potentially endogenous, since its unobservable factors at the family level can also affect school absenteeism. Another feature of the model, influenced by the application, is the fact that mother's income is likely to affect marital status nonlinearly, so that it is modeled nonparametrically. We find evidence that marital status is endogenous, and that the data set fits two mixing components in each state, interpreted as healthy and unhealthy students. Further, marital status decreases annual missed school days by about 13 percent for a randomly selected individual, but increases it by 9 percent for those families who actually select to have a single parent, which is evidence of adverse selection.

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, 2012; 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 family structure 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). Another strand of research points to the role of health, with medical problems strongly linked to higher absenteeism (Holbert, Wu, and Stark, 2002; Basch, 2011). Even district infrastructure appears important, with better quality school buildings linked to lower absenteeism (Duran-Narucki, 2008).

Those explanations notwithstanding, this paper explores an alternative possibility. Specifically, to what extent is school absenteeism influenced by a child's family structure, specifically parental marital status? Existing evidence on this subject is sparse. A handful of studies provide evidence that children from single-parent households miss more school days (Keller, 1983; Sandefur, McLanahan, and Wojtkiewicz, 1992; Vos, 2001; Bock, 2002), but that existing evidence mostly relies on descriptive measures, with little consideration of the more structural, economic nuances explored in the present paper.

It seems like the presence of two parents could affect school attendance, although the direction is not clear *a priori*. On one hand, if the presence of two parents helps facilitate preparation for and transportation to schools, then the presence of two parents might engender increased school attendance. On the other hand, the presence of two parents likely increases a family's daytime childcare options, thus reducing the costs associated with missed school days, and in turn increasing absenteeism.

The *a prior* ambiguity of the effect of parental marital status on school absenteeism points to several econometric challenges. First, children might possess unobserved characteristics that simultaneously relate to both family structure *and* absenteeism, implying that family structure 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, and the main emphasis of this paper, is that whereas standard econometric models capture the effect a treatment (family structure) on an outcome (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 family structure. Further, those conditional mean functions, themselves, are comprised of separate components that correspond to different "types" of children. Table 1 gives the summary statistics of the data set partitioned by mother's marital status (treatment variable), which calls for the Roy modeling structure, an endogenous switching regression framework. To understand data heterogeneity better, the component probabilities are modeled as functions of observed covariates using the smoothly mixing regressions (SMR) approach by Geweke and Keane (2007).

The FMR model has several computational challenges described in details in Munkin (2022). Here we give a brief description of those issues. In general, estimation of finite mixtures has some subtle points (Celeux et al., 2019). Given the invariance of the likelihood to label switching of k components, the sampler fails to visit all of $k!$ regions in the support of the posterior distribution. The label switching problem was studied by Celeux, Hurn and Robert (2000), Frühwirth-Schnatter (2001) and Jasra, Holmes and Stephens (2005). It is important to identify the correct numbers of components. This paper uses the random permutation sampler developed by Frühwirth-Schnatter (2004) applied to the method of Chib (1995) to calculate marginal likelihoods to identify the numbers of components. Once the correct numbers of components in the model are identified this paper applies the method of Geweke (2007), in which separation of the components is done by imposing a valid inequality constraint on the draws.

The rest of the paper is organized as follows. Section 2 specifies the FMR model and makes reference to the choice of priors, MCMC algorithm and identification of the numbers of components in the mixtures. Section 3 analyzes factors affecting the annual missed school days including the nonparametric effect of mother's income, and calculates the corresponding treatment effects. Section 4 concludes. The details of the MCMC algorithm and steps in calculating the marginal likelihoods are given in Munkin (2022).

2. The FMR Model

This section defines the Finite Mixture Roy model based on Munkin (2022). Marital status d_i is generated by latent difference in utility D_i in the treated and untreated states for N independent individuals $(i = 1, ..., N)$ such that

$$d_i = I_{[0,+\infty)}\left(D_i\right), \tag{2.1}$$

where $I_{[0,+\infty)}$ is the indicator function for the set $[0, +\infty)$. Continuos instrumental variable, mother's income, s_i enters D_i in a flexible nonparametric way as

$$D_i = f(s_i) + \mathbf{W}_i\boldsymbol{\alpha} + u_i, \tag{2.2}$$

where $\mathbf{W}_i$ is a vector of exogenous regressors, $\boldsymbol{\alpha}$ is a conformable vector of parameters, which does not include an intercept, function $f(.)$ is unknown, and $u_i \overset{iid}{\sim} N\left(0,1\right)$.

Missed school days Y_i assumes two potential outcomes Y_i^1 and Y_i^2 and the observability condition is

$$Y_i = \begin{cases} Y_i^1 & \text{if } d_i = 1 \\ Y_i^2 & \text{if } d_i = 0 \end{cases}.$$

Y_i^1 and Y_i^2 are distributed as finite mixtures of Poisson densities with conditional means $\exp(\mu_{ij}^1)$ and $\exp(\mu_{ij}^2)$, where subscript j indicates that observation i belongs to component j, and

$$\mu_{ij}^1 = \mathbf{X}_i\boldsymbol{\beta}_{1j} + \delta_{1j}u_i + \varepsilon_{ij}^1,$$
$$\mu_{ij}^2 = \mathbf{X}_i\boldsymbol{\beta}_{2j} + \delta_{2j}u_i + \varepsilon_{ij}^2,$$

where $\mathbf{X}_i$ is a vector of exogenous regressors, $\boldsymbol{\beta}_{1j}$ and $\boldsymbol{\beta}_{2j}$ are component j specific conformable vector of parameters, $\varepsilon_{ij}^1 \sim N\left(0, \sigma_{1j}^2\right)$ and $\varepsilon_{ij}^2 \sim N\left(0, \sigma_{2j}^2\right)$ represent unobserved heterogeneity. Random variable u_i is introduced in the conditional means to control for endogeneity.

The posterior distribution is augmented with latent variables z_{ij}, defined as

$$z_{ij}^t = \begin{cases} 1 & \text{if observation } i \text{ belongs to component } j \\ 0 & \text{otherwise} \end{cases}.$$

where superscript t differentiates between the treated ($t = 1$) and untreated ($t = 2$) states. We allow the corresponding probabilities $\Pr\left(z_{ij}^t = 1\right)$ to depend on covariates specifying latent variables R_{ij}^t ($j = 2, ..., k_t$) defined as

$$R_{ij}^t = \mathbf{V}_i\boldsymbol{\gamma}_{tj} + \xi_{ij}^t, \tag{2.3}$$

where $\mathbf{V}_i$ is a set of covariates (it can be different from $\mathbf{X}_i$), $\boldsymbol{\gamma}_{tj}$ is a conformable vector of parameters, $\boldsymbol{\xi}_i^t \overset{iid}{\sim} N\left(\mathbf{0}, \mathbf{I}_{k_t-1}\right)$ and R_{i1}^t is restricted to zero. Then the components are identified as

$$z_{ij}^t = 1 \text{ if and only if } R_{ij}^t \geqslant R_{il}^t \text{ (for } \forall l,\ l = 1, ..., k_t). \tag{2.4}$$

For further details on the choice of parameter priors and the MCMC estimation algorithms, see Munkin (2022).

3. Application

3.1. Data and Instrument

Data for this study come from the 2015 and 2016 waves 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 respondents ages 6-15, an age range during which schooling is required in most of the United States. Socioeconomic information for each child comes from the main Household Component files, with details of a child's health status coming from the Medical Conditions files. After linking children to their mothers and deleting records with missing information on key variables, the final sample size includes 1,871 unique children. We deleted 3 observations in which the mother's income was negative, perhaps a measurement error since all of them had $-\$268,000$.

The main outcome variable on interest is *misseddays*, a discrete count of the number of missed school days due to health-related reasons during the most recent school year. The main treatment variable is *married*, a binary indicator for whether the child's mother is currently married with her spouse present in the household. Note that, according to the definition, the marital partner need not be the child's biological father. Thus, the treatment should be interpreted simply as capturing the presence in the household of a second adult, not for that adult's biological or emotional relation to the child.

The top row of Table 1 reports the mean number of missed school days, partitioned by mothers' marital status. Children of single mothers report approximately 0.5 more missed school days than children of two-parent households. That difference is statistically significant according to a standard two-sample t-test.

However, that difference in mean missed days cannot be interpreted as *causally* linked to family structure, as other socioeconomic measures also appear to differ across the two sample partitions. Most of the variable names are self-explanatory. For example, mean child age divided by 10, *agekid,* differs across the two partitions, as does child self-reported health, with children of married mothers appearing to report better health and lower BMI. Because poor health was reported by less than one percent of the sample this category was merged with fair health group, *fairkid*.

Furthermore, married mothers are more likely to be non-black/non-Hispanic and more likely to be employed. Married mothers also have larger families and higher income. 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.

Mother's annual personal income divided by 10,000, *incomemom*, is likely to affect the marital status, but it is not clear in what functional form. Therefore, *married* enters the treatment equation nonparametrically. Since the dependent count variable relates to the child, some variables related to the mother can serve as instruments. In fact, *incomemom* enters both equations but in the treatment equation it is allowed to enter nonparametrically.

Summary statistics of the variables used in the application are given in Table 1. The vector of covariates $\mathbf{X}$ in the outcome equations consist of self-perceived health status variables *vegoodkid, goodkid, fairkid* (excellent health status is the excluded category), geographical location variables *northeast, midwest, south*, variables that proxy for socioeconomic status, *incomemom, employedmom, famsize, agekid, femalekid*, child's body mass index, *bmikid,* and year dummy, *year* (year dummy for 2016 is excluded). Vector $\mathbf{W}$ of the insurance equation includes variables that describe socioeconomic status of the mother, *famsize, agemom, employedmom, blackmom, hispmom,* self-perceived health status variables*, vegoodmom, goodmom, fairmom, poormom,* geographical location variables*, northeast, midwest, south,* year dummy, *year,* mother's body mass index, *bmimom,* and *incomemom,* entering the equation nonparametrically.

Figure 1 presents histograms missed school days for three samples: all observations, married mothers and divorced mothers. The distributions are likely to be truncated by MEPS at 16 days. Figure 2 presents differences in the frequencies across married versus divorced groups placing them against each other. The divorced mother group has a larger mean (4.323 versus 3.812) and larger variance (standard deviation of 3.814 versus 3.318), however, overall the distributions appear to be very similar.

3.2. Results

We estimate the model in which there are two components both the treated and untreated states. The FMR model is estimated, imposing inequality constraints, based on the calculated means of the components, assigning each draw to either the lower or larger mean component. The components are well identified and the corresponding posterior distributions have clear signs of convergence for all parameters, component specific means and weights. In the treated state the conditional means of the estimated components are 3.360 and 6.672, and the estimated probabilities are 0.367 and 0.633 respectively. In the untreated state the conditional means are 4.421 and 8.324, and the component probabilities are 0.340 and 0.660. The chains show good mixing properties, although the covariance parameters are slower to coverage as expected. The Markov chains for them display considerable serial correlations so it is run for 100,000 replications after discarding first 10,000 replications. Posterior means and standard deviations of the parameters are presented in Tables 2, 3 and 4.

Parameters $\boldsymbol{\gamma}_{12}$ and $\boldsymbol{\gamma}_{22}$ in Table 2 must be interpreted as increasing the probabilities of belonging to the higher mean components. These probabilities are likely to be influenced by health status both in the treated and untreated states. We create a measure of health status as the total number of child's medical conditions. However, instead of using a count we create variables *chronic1, chronic2, chronic3, chronic4, chronic5* and *chronic6plus* as indicators of 1, 2, 3, 4, 5 and 6 or more chronic conditions respectively. The excluded category is no chronic conditions. From the results in Table 2 it can be seen that all chronic condition indicators have strong impacts: the larger the numbers of chronic conditions the greater positive impacts they have on the probability of belonging to the higher utilization groups. Overall, it appears reasonable to interpret the components as relatively healthy and unhealthy groups.

It is interesting to notice that almost no variable has a strong impact on the missed school days in the healthy group in both treatment groups. This can be interpreted as having to miss few days for random health shocks uniformly distributed across all children in the sample. However in the higher mean groups interpreted as unhealthy we see that family size has a strong negative effect. Overall, health status variables produce mixed results. Fair and poor health status of the child has a positive impact only for divorced mothers (component 2). Mother's employment status has a strong negative impact on unhealthy children to miss school for the unmarried mothers. Surprisingly, mother's income does not have any significant effect. Midwest and south have a negative impact only for married mothers.

The probability of treatment is strongly and positively affected by family size, age of the mother, mother's employment status and being from the south. Being black and Hispanic decreases the probability of treatment. Poor health status and body mass index also negatively affect the marital status.

Figure 3 plots the estimated function $f(s_i)$ together with the 95% posterior probability intervals indicated by the dashed lines. Variable *incomemom* has a long right tail with 95% of observations not exceeding the annual income of \$85,000. Besides 5% of the observations range between \$85,000 and \$260,000 and are very sparse resulting in very imprecise nonparametric estimates. Therefore, we truncate *incomemom* at \$85,000 and round it up to \$200 which gives $k_\nu = 329$ distinct values. The actual variable *incomemom* is defined as mother's income divided by 10,000. The linear specification predicts that *incomemom* has no impact on the probability of treatment. However, according to the nonparametric model, the probability of treatment is maximized when income is zero, then it monotonically decreases from 0 to $\$15,000$ where it reaches the minimum, after which it stays flat until starting to increase from \$20,000 to \$50,000 after which it stays flat.

The estimated posterior means and standard deviations of the covariance parameters, δ_{11}, δ_{12}, δ_{21} and δ_{22}, are given in Table 3 as -0.037 (0.088), -0.211 (0.133), -0.091 (0.098) and -0.249 (0.113) respectively. Only δ_{22} is well separated from zero by more than two standard deviations. Therefore, We can formally test the null hypothesis that jointly restricts all the covariance parameters to zero, $H_0 : \delta_{11} = 0$, $\delta_{12} = 0$, $\delta_{21} = 0$, $\delta_{22} = 0$, against the alternative that leaves them unconstrained. The Bayes factor can be calculated using the Savage-Dickey density ratio approach (Verdinelli and Wasserman, 1995) as

$$B_0 = \frac{\pi(\delta^*_{11}, \delta^*_{12}, \delta^*_{21}, \delta^*_{22}|\mathbf{y})}{\pi(\delta^*_{11}, \delta^*_{12}, \delta^*_{21}, \delta^*_{22})}, \tag{3.1}$$

where $\pi(\delta^*_{11}, \delta^*_{12}, \delta^*_{21}, \delta^*_{22}|\mathbf{y})$ is the posterior density and $\pi(\delta^*_{11}, \delta^*_{12}, \delta^*_{21}, \delta^*_{22})$ is the prior density of parameters δ_{11}, δ_{12}, δ_{21} and δ_{22} evaluated at the point $\delta^*_{11} = 0$, $\delta^*_{12} = 0$, $\delta^*_{21} = 0$, $\delta^*_{22} = 0$. The null hypothesis of no endogeneity is readily rejected not only $H_0 : \delta_{22} = 0$, but would be rejected for the joint test.

3.3. Average Treatment Effects

Next the average treatment effect (ATE) and the average treatment effect for the treated (ATET) parameters are calculated for the nonparametric specification of the FMR model with two components both in the treated and untreated states. Definition of dependent variable Y_i establishes the link between the observed and counterfactual outcomes as

$$Y_i = d_i \sum_{j=1}^{2} I_{\left\{z_{ij}^1=1\right\}} Y_{ij}^1 + (1-d_i) \sum_{j=1}^{2} I_{\left\{z_{ij}^2=1\right\}} Y_{ij}^2.$$

ATE is the expected outcome gain from receipt of treatment for a randomly chosen individual and ATET is the expected outcome gain for those who actually receive the treatment. For the computational details on how the ATE and ATET parameters, $E\left[Y^1 - Y^2 | \mathbf{X}\right]$ and $E\left[Y^1 - Y^2 | \mathbf{X}, \mathbf{W}, d=1\right]$, are calculated see Munkin (2022). The estimated ATE is -0.511 (0.118) and ATET is 0.346 (0.413). The size of ATET relative to ATE determines whether adverse or favorable selection is present.

Thus, being from a family with two parents decreases annual missed school days by about 13 percent for a randomly selected individual, but increases it by 9 percent for those families who actually select to have a single parent. The estimated ATET is consistent with adverse selection since the average treatment effect for those who select the treatment is about 0.86 days larger than the ATE for a randomly chosen individual.

4. Conclusion

This paper analyzes factors affecting school absenteeism using the Finite Mixture Roy model developed by Munkin (2022). The observed patterns of missed school days display heterogeneous patterns consistent with finite mixtures. The Roy structure captures observed heterogeneity generated by the mother's marital status. Finite mixtures further control for unobserved heterogeneity generated by the presence of health and unhealthy children in the sample. The assumption is that finite mixtures identify components, in which the marginal and treatment effects are homogeneous. To interpret the components, their probabilities are modeled as functions of covariates using the smoothly mixing regression approach. Mother's income is allowed to enter the treatment equation nonparametrically. We estimate a model specification with two components both in the treated and untreated states. Marital status decreases annual missed school days by about 13 percent for a randomly selected individual, but increases it by 9 percent for those families who actually select to have a single parent, which is evidence of adverse selection.

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.

Bock, J. (2002). Evolutionary demography and intrahousehold time allocation: School attendance and child labor among the Okavango Delta peoples of Botswana. American Journal of Human Biology, 14(2), 206-221.

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.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association,* 90, 1313-1321.

Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of American Statistical Association,* 95, 957-970.

Celeux, G., Kamary, K., Malsiner-Walli, G., Marin, J.-M., & Robert, C. P. (2019). Computational solutions for bayesian inference in mixture models. In Frühwirth-Schnatter, S., Celeux, G., & Robert, C. P. (Eds.), *Handbook of Mixture Analysis,* Chapter 5, 73-96. Boca Raton, FL: CRC Press.

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.

Deb, P., & Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics,* 12, 313-336.

Deb, P., Munkin, M. K., & Trivedi, P. K. (2006). Private insurance, selection, and the health care use: a Bayesian analysis of a Roy-type model. *Journal of Business and Economic Statistics,* 24, 403-415.

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.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of American Statistical Association,* 96, 194-209.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal,* 7, 143-167.

Frühwirth-Schnatter, S., Celeux, G., & Robert, C. P. (2019). *Handbook of Mixture Analysis.* Boca Raton, FL: CRC Press.

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.

Geweke, J. (2007). Interpretation and inference in mixture models: simple MCMC works. *Computational Statistics & Data Analysis,* 51, 3529-3550.

Geweke, J., & Keane, M. (2007). Smoothly mixing regressions. Journal of Econometrics, 138, 252-291.

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.

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.

Hu, X., Munkin, M. K., & Trivedi, P. K. (2015). Estimating incentive and selection effects in the Medigap insurance market: an application with Dirichlet process mixture model. *Journal of Applied Econometrics,* 30, 1115–1143.

Hurn, M., Justel, A., & Robert, C. P. (2003). Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics,* 12, 55-79.

Jasra, A., Holmes, C. C., & Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science,* 20, 50–67.

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.

Keller, R. T. (1983). Predicting absenteeism from prior absenteeism, attitudinal factors, and nonattitudinal factors. Journal of Applied psychology, 68(3), 536.

Munkin, M. K. (2022). Count Roy Model with Finite Mixtures. *Journal of Applied Econometrics*, 37, 1160-1181

Poirier, D. J., & and Tobias, J. L. (2003). On the predictive distributions of outcome gains in the presence of an unidentified parameter. *Journal of Business and Economic Statistics,* 21, 258-268.

Sandefur, G. D., McLanahan, S., & Wojtkiewicz, R. A. (1992). The effects of parental marital status during adolescence on high school graduation. Social Forces, 71(1), 103-121.

Verdinelli, I., & Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of American Statistical Association,* 90, 614-618.

Vos, S. D. (2001). Family structure and school attendance among children 13-16 in Argentina and Panama. Journal of Comparative Family Studies, 32(1), 99-115.

Table 1. Summary of the Data

	Full Sample (N=1871)		Married (N=1094)		Divorced (N=777)	
	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
misseddays	4.024	3.540	3.812	3.318	4.323	3.814
incomemom	2.541	2.433	2.745	2.615	2.254	2.120
agemom	3.840	0.710	3.925	0.672	3.722	0.744
bmimom	29.153	7.367	28.128	6.728	30.597	7.967
married	0.585	0.493	1.000	0.000	0.000	0.000
famsize	4.441	1.454	4.736	1.196	4.027	1.669
agekid	1.035	0.290	1.020	0.286	1.056	0.295
bmikid	20.521	5.856	19.782	5.170	21.560	6.568
totchr	2.090	2.151	2.075	2.021	2.112	2.323
year	0.519	0.500	0.517	0.500	0.521	0.500
blackmom	0.171	0.377	0.078	0.268	0.302	0.460
hispmom	0.321	0.467	0.304	0.460	0.345	0.476
femalekid	0.500	0.500	0.488	0.500	0.517	0.500
vegoodmom	0.313	0.464	0.350	0.477	0.260	0.439
goodmom	0.330	0.470	0.303	0.460	0.367	0.482
fairmom	0.118	0.323	0.106	0.308	0.135	0.342
poormom	0.025	0.155	0.013	0.112	0.041	0.199
vegoodkid	0.281	0.449	0.299	0.458	0.255	0.436
goodkid	0.174	0.379	0.151	0.358	0.206	0.405
fairkid	0.042	0.200	0.025	0.155	0.066	0.248
employedmom	0.656	0.475	0.681	0.466	0.620	0.486
northeast	0.143	0.350	0.136	0.343	0.153	0.360
midwest	0.217	0.412	0.234	0.424	0.193	0.395
south	0.368	0.482	0.352	0.478	0.391	0.488
chronic	0.777	0.417	0.795	0.404	0.750	0.433
chronic1	0.278	0.448	0.284	0.451	0.269	0.444
chronic2	0.187	0.390	0.197	0.398	0.172	0.378
chronic3	0.113	0.317	0.122	0.327	0.102	0.302
chronic4	0.079	0.269	0.084	0.278	0.071	0.257
chronic5	0.049	0.215	0.047	0.211	0.051	0.221
chronic6plus	0.072	0.260	0.062	0.242	0.085	0.279

Table 2. Posterior Means and Standard Deviations of Component Parameters $\boldsymbol{\gamma}_{12}$ (treated), $\boldsymbol{\gamma}_{22}$ (untreated)

	Married Mothers		Divorced Mothers	
	Vector γ_{12}		Vector γ_{22}	
	mean	std.dev.	mean	std.dev.
CONST	-1.052	0.289	-1.991	0.583
chronic1	0.401	0.293	1.170	0.562
chronic2	0.754	0.307	1.971	0.552
chronic3	0.754	0.341	1.660	0.644
chronic4	1.672	0.961	1.889	0.655
chronic5	1.142	0.719	2.682	1.015
chronic6plus	3.019	1.560	3.717	1.342
PROB $j = 2$	0.367	0.067	0.340	0.073

Table 3. Posterior Means and Standard Deviations of Parameters
$\boldsymbol{\beta}_{1j}$, $\boldsymbol{\beta}_{2j}$, δ_{1j}, δ_{2j}, σ_{1j}^2, σ_{2j}^2 by State ($d = 0, 1$) and Components ($j = 1, 2$)

	Married Mothers				Divorced Mothers			
	Component 1		Component 2		Component 1		Component 2	
	mean	std.dev.	mean	std.dev.	mean	std.dev.	mean	std.dev.
CONST	0.841	0.347	1.996	0.383	0.112	0.309	1.671	0.412
famsize	−0.021	0.036	−0.104	0.046	0.013	0.029	−0.117	0.040
agekid	−0.115	0.147	0.174	0.175	0.238	0.153	0.304	0.207
femalekid	0.022	0.078	−0.023	0.093	−0.001	0.086	0.172	0.115
bmikid	0.012	0.009	0.010	0.009	0.007	0.007	0.003	0.008
vegoodkid	0.018	0.088	0.161	0.106	−0.034	0.103	0.060	0.140
goodkid	−0.067	0.134	0.242	0.136	0.117	0.120	0.103	0.161
fairkid	0.318	0.398	0.054	0.410	0.242	0.194	0.268	0.133
northeast	0.032	0.141	−0.143	0.154	0.134	0.138	0.301	0.176
midwest	0.032	0.107	−0.269	0.134	0.246	0.126	0.135	0.172
south	−0.012	0.102	−0.270	0.129	0.186	0.111	0.028	0.166
year	0.037	0.082	−0.018	0.095	−0.012	0.088	0.153	0.113
incomemom	−0.008	0.019	0.020	0.022	0.009	0.030	0.034	0.036
employedmom	−0.168	0.101	−0.087	0.125	0.096	0.120	−0.410	0.130
δ_{tj} ($t = 1, 2$)	−0.037	0.088	−0.211	0.133	−0.091	0.098	−0.249	0.113
σ_{tj}^2	0.158	0.034	0.275	0.055	0.201	0.041	0.251	0.056
$E \exp(\mu_j^t)$	3.360	0.163	6.672	0.523	4.421	0.202	8.324	0.708

Table 4. Posterior Means and Standard Deviations of the Treatment Equation Parameter $\boldsymbol{\alpha}$

	Treatment Equation	
	Parameter α	
	mean	std.dev.
famsize	0.271	0.022
agemom	0.284	0.045
bmimom	−0.021	0.004
blackmom	−1.231	0.092
hispmom	−0.405	0.074
vegoodmom	0.066	0.086
goodmom	−0.074	0.086
fairmom	0.027	0.114
poormom	−0.478	0.218
northeast	0.071	0.099
midwest	0.174	0.093
south	0.210	0.081
year	−0.044	0.061
employedmom	0.368	0.094

Figure 1. Histograms of Missed School Days for
All Observations, Married Mothers and Divorced Mothers.

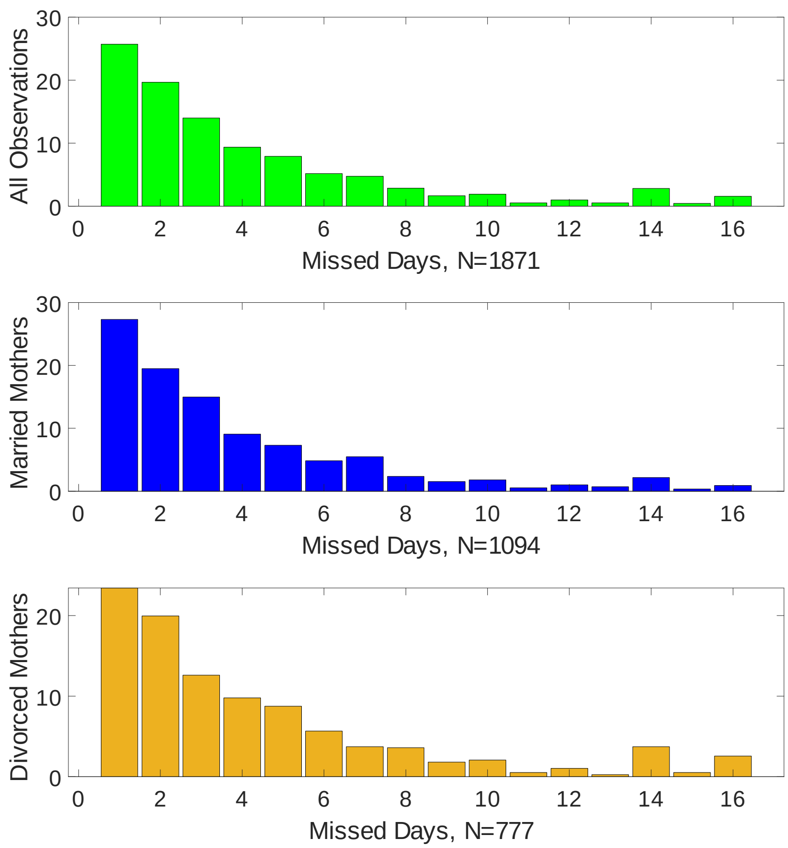

Figure 2. Histograms of missed school days for married mothers versus divorced.

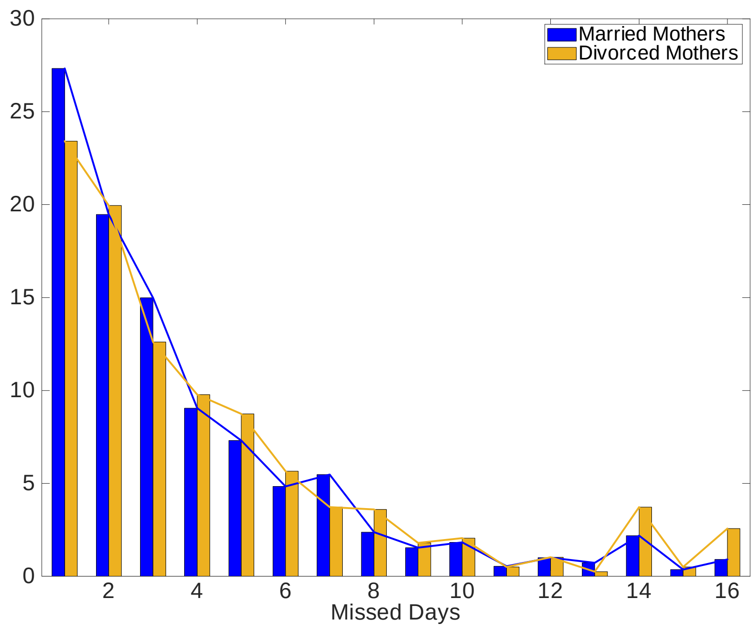

Figure 3. The Effect of mother's income on missed school days.