2020

State Exit Exams and Graduation Rates: A Hierarchical SLX Modelling Approach

Joshua Hall

Donald Lacombe

Shree B. Pokharel

State Exit Exams and Graduation Rates: A Hierarchical SLX Modelling Approach

Joshua C. Hall
Department of Economics
West Virginia University
joshua.hall@mail.wvu.edu

Donald J. Lacombe
Personal Financial Planning
Texas Tech University
donald.lacombe@ttu.edu

Shree B. Pokharel*
Joint Committee on Government and Finance
West Virginia Legislature
shreebaba@gmail.com

April 10, 2020

Abstract

The literature on high school exit exams has found both positive and negative effects of these high stake exams on high school graduation rates. To this point the literature has not taken into account the embedded nature of school districts within state education systems. We employ a Bayesian Hierarchical SLX model to account for the hierachical nature of education data in the United States. Our approach also allows us to account for spatial spillovers that influence graduation rates across districts and states. Using school district and state-level data for 45 states and 8194 school districts in the United States in 2015, we generally find no statistically significant effect of state exit exams on high school graduation rates. Random effect coefficients, however, point towards high school exit exams being negatively associated with graduation rates in a handful of states.

Keywords: Spatial dependence, Bayesian statistics, hierarchical modelling, state exit exams
JEL Codes: C11, C21, C30

*Pokharel's views are her own and do not represent the views of her employer. This research was primarily conducted while she was employed at West Virginia University.

1 Introduction

The widespread consensus that high school diplomas displayed low academic skills and standards in the latter half of the twentieth century led to policies favoring high–stakes school exit exams in the US. The deficiency of job skills and college preparedness in high school graduates were mostly attributed to social promotion (Reardon and Galindo, 2002) and to 'watered–down' curriculum (Bond and King, 1995). This view was supported when evidence of students' lack of proficiency in primary subjects were compared to those of other countries during the Cold War era and was further emphasized with the publication of *A Nation at Risk: The Imperative for Educational Reform* (Warren et al., 2006).

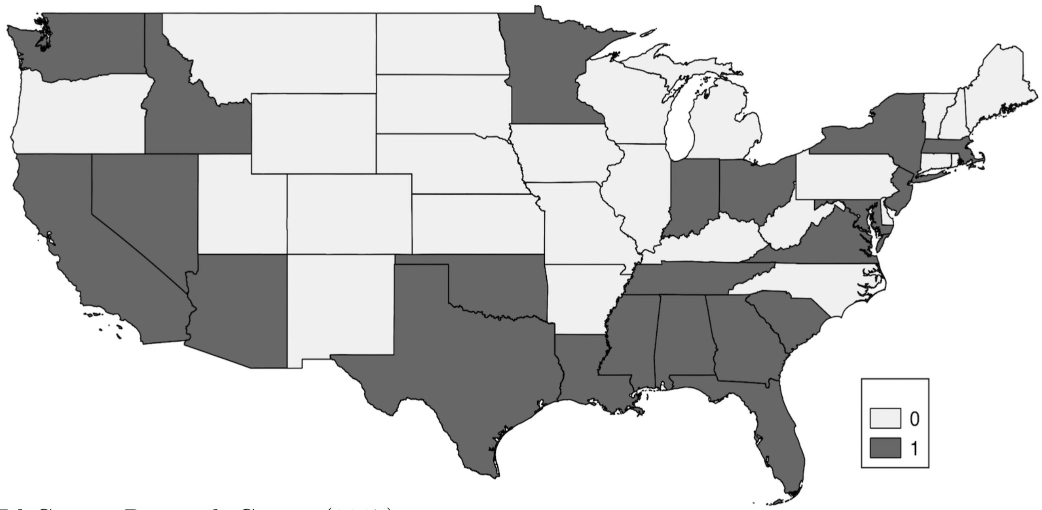
The *Nation at Risk* report stated that "the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation" (National Commission on Excellence in Education, 1983, p. 112) and that "more and more young people emerge from high school ready neither for college nor for work" (National Commission on Excellence in Education, 1983, p. 117). As a remediation measure, the report also recommended, among other things, regular use of standardized tests of achievement that "should be administered at major transition points from one level of schooling to another and particularly from high school to college or work" (National Commission on Excellence in Education, 1983, p. 125). These exams are considered to be 'high stakes' "if they carry serious consequences for students or for educators" (American Educational Research Association, 2000). For students an example would be if failing to pass an exam meant no high school diploma. For educators, it might bring public scrutiny and less financial rewards.

Widespread implementation of these high-stakes exit exams by states started as early as 1980s and has increased over time. Fourteen states enforced these exams in 1990 and the number grew to 18 in 2000 (Warren et al., 2006). As of 2013, 23 out of 50 US states have implemented this policy (Ed Counts Research Center, 2017). Figure 1 shows states with and without state exit exams in 2013. The color grey represents states without the state exit exam requirement, whereas the color black represents states with state exit exam requirement.

The stated goal of these exams was to encourage students and school districts to demonstrate that they had achieved competency in certain areas prior to graduation. Theoretically, the effect of exit exams on graduation rates is ambiguous. Exit exams could combat social promotion by some districts, leading to lower graduation rates. On the other hand, the stigma and competitive effect of reported statewide exam scores could lead to an increase in graduation rates.

A number of papers have been written on the effect of exit exams on educational outcomes

Figure 1: States with and without State Exit Exams



Source: Ed Counts Research Center (2017).

such as high school completion rates, dropout rates and dropout likelihood, earnings, and graduation rate. Greene and Winters (2004), Warren and Jenkins (2005), and Warren et al. (2006) find no effect of state exit exams on high school completion rates. Warren and Edwards (2005) find no effect of high school exit exams on dropout rates. In contrast, Hemelt and Marcotte (2013) report that high school exit exams increase dropout rates among twelfth graders. Beardsley and Berliner (2002) finds high stakes exams increase dropout rates, while Papay et al. (2010) finds a negative influence on dropouts. Ou (2010) finds mixed results of exit exams on dropouts. With respect to earnings, Warren et al. (2008) reports no effect of high school exit exams on earnings.

In terms of graduation rates, studies such as Beardsley and Berliner (2002) and Marchant and Paulson (2005) find state exit exams to negatively affect graduation rates. Baker and Lang (2013) finds no statistically significant effect of high school exit exams on graduation rates. While Beardsley and Berliner (2002)'s study was a qualitative study, Marchant and Paulson (2005)'s and Baker and Lang (2013)'s study was quantitative. However, these studies only take state factors into account without accounting for school districts that are embedded within states.

While these studies look at the impact of high–stakes exit exams on graduation rates, they do not empirically account for the embedded nature of school districts in state education systems. In addition to the hierarchical nature of school data in the United States, school districts have spatial spillovers in terms of policies, teacher labor markets, and student flows. For example, one district in a region raising teacher salaries will likely influence other districts to raise their salaries or risk losing teachers on the margin. To our knowledge, no studies have empirically accounted for spatial spillovers across school districts and the fact that school districts are embedded in states that have different school-related policies. We fill this hole in the literature by employing a Bayesian Hierarchical SLX model. While we find that non–spatial papers miss important spatial spillovers across school districts, we generally find no statistically significant evidence that states with exit exams have higher or lower graduation rates than states without such exams. There does appear to be heterogeneity across states, with the random effects showing a handful of states having a negative and statistically significant relationship between exit exams and graduation rates.

There are three major advantages of using a hierarchical model that we highlight. First, these models help to represent data structures that are close to the real world. These models help to separate individual effects, in the case of this study, school–district level effects on high school graduation rates, from the state level effects. This makes it a closer representation of the real world data structure than a normal linear model. Second, by acknowledging that Ordinary Least Squares (OLS) violates the independence assumption in hierarchical data, it helps correct biased estimates that OLS doesn't. In our case, OLS would take the school districts within a same state as independent from one another, when they clearly are not as they have to take the same state exit exam. This is one of the main identifying features of this paper compared to previous studies in that it takes this key violation into account. Third, it allows for the use of state-level variables to control for state-level variation in policy, in this case whether or not there is a state exit exam. While our results are not causal, they allow us to use a single cross–section of states and still obtain measures of heterogeneity across those states that can be used to identify situations for other more appropriate causal inference approaches.

The remainder of the paper proceeds as follows. Section 2 describes the empirical specification and statistical methodology used in this paper. Section 3 describes the data in detail. Section 4 describes the results and Section 5 concludes.

2 Statistical Methodology

2.1 Hierarchical Models

In their natural state, some data have a hierarchical structure to them. For example, students nested within a classroom, counties nested within states, and school districts nested within states. Hierarchical structure violates the independence assumption since school districts within the same states are exposed to the same set of state laws. In case of this study, school districts within a state are affected by a state exit exam, hence, violating the independence assumption as they are exposed to the same set of information. If not used in their natural state, data can provide biased estimates.

In addition to the three advantages mentioned in the introduction, hierarchical models have at least two additional benefits. The fourth advantage of hierarchical models is that they nest classical regression models and, therefore, account for the fact that each upper level unit intercept is different (intercept of each state in our paper) but also have some similarities. To explain this concept more clearly, we refer to Gelman and Hill (2006) who state that "classical regression model can be viewed as special cases of multilevel models." Here, $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha)$. Ignoring any heterogeneity and assuming a common intercept for all "upper–level" units (i.e., 45 separate states in our paper), the first model is called a fully–pooled model. It assumes that all states are homogeneous and should have a common intercept. On the other hand, a no–pooling model assumes heterogeneity among the states and allows for including a dummy variable for each state. Basically, this model allows all states in our paper to be different from one another.

To explain this further, the matrix form of the hierarchical SLX model representation is given by the following equations, following Chib (2008):

$$\begin{aligned} y = \tilde{X}\beta + \Delta u + \varepsilon \quad \varepsilon\,|\sigma^2 &\sim N(0, \sigma^2 I_N) \\ u\,|\tau^2 &\sim N(0, \tau^2 I_J) \\ \tau^2 &\sim IG(a, b) \end{aligned} \tag{1}$$

where $\tilde{X} = \begin{bmatrix} X & W_1X & \Delta Z\gamma & \Delta W_2 Z\delta \end{bmatrix}$. The dependent variable y is a $N \times 1$ vector of observations and represents graduation rates. X represents the matrix of explanatory variables at the district level and has dimension $N \times k$. The β is a $k \times 1$ vector of coefficients associated with X. The ε is the error term and is normally distributed with a 0 mean and variance of $\sigma^2 I_n$ and has dimension $N \times 1$.

The u represents the $J \times 1$ vector of individual intercepts (state–level intercepts in our study). Z is the vector of explanatory variables (that also includes a constant term) with dimension $J \times m$. γ is a $J \times m$ vector of coefficients associated with the Z term. τ^2 is normally distributed with a 0 mean and variance of $\sigma^2 I_J$ and is dimension $J \times 1$. As is standard in hierarchical models, we assume that ε and u, u and X, and u and Z are not correlated (Raudenbush and Bryk, 2002).

This model is also called an "intercepts–as–outcome" model. It is called so because the "*Level 2* equation has the *Level 1* intercept as its *dependent variable* (as its *outcome*)"

(Adewale et al., 2007). It is also called a "random intercepts" model. Understandably, it is also called so as for each state, it sets a baseline of graduation rates. The individual school districts' graduation rates then varies around this baseline for the state it is embedded in due to, for instance, differences in spending per pupil.

Gelman and Hill (2006) argue that data could estimate the level 2 error variance and that they "see no reason (except for convenience) to accept estimates that arbitrarily set this parameter to one of these two extreme values." Here, we can assume that each state is different but also share similarities (Lacombe and Flores, 2017).

The fifth advantage of the hierarchical model is that it corrects for any potential bias that might arise from choosing either only the "fully–pooled" model or the "non–pooling" models. Intercepts in a hierarchical model are a linear combination of the "fully–pooled" model and "no–pooling" model and can be written as (Subramanian, 2010; Luke, 2004):

$$\begin{aligned} \hat{\alpha}_j^{EB} &= \lambda_j \hat{\alpha}_j^{NP} + (1 - \lambda_j)\, \hat{\alpha}^{FP} \\ \lambda_j &= \frac{\tau^2}{(\tau^2 + \sigma^2/n_j)} \end{aligned}$$

where, $\hat{\alpha}_j^{NP}$ represents the "no–pooling" intercept estimate, $\hat{\alpha}_j^{FP}$ represents the "fully–pooled" intercept estimate, and $\hat{\alpha}_j^{EB}$ represents the "empirical Bayes" or "shrinkage" estimate of the linear combination of the "no–pooling" and the "fully–pooled" models. λ_j represent the weights assigned to each aforementioned models and are a function of both level error variance (ie., variance of the school–district level and the state level) and, n_j, the number of level 1 observation in each level 2 unit.

The empirical Bayes works in the following manner. If n_j is small, λ_j is small, which means that $\hat{\alpha}_j^{EB}$, the empirical Bayes, moves close towards the fully–pooled estimate, $\hat{\alpha}_j^{FP}$. Similarly, if n_j is large (such as the number of school districts in the state of Texas, ie., 808 in our sample), λ_j is large, which means that $\hat{\alpha}_j^{EB}$, the empirical Bayes, moves closer towards the no–pooling estimate, $\hat{\alpha}_j^{NP}$. Here, more weight is placed on the "no–pooling" intercept estimate. Hence, the advantage in using the empirical Bayes is that it corrects for any possible bias from choosing either "no–pooling" or "fully–pooled" model at random.

We now extend this basic intercept hierarchical model by adding a spatial factor to it. We use the Spatial Lag of X (SLX) model at both levels of the hierarchy. A SLX model provides a much richer set of results as it allows for local spillovers. As it includes spatially–lagged independent variables (which capture local spillovers), it calculates the direct effects (own effects) as well as the indirect effects. The direct(own) effects calculate the effects of the explanatory variables on the dependent variables; the indirect effects (spillover effects) capture the effect of neighbors on the dependent variable.

2.2 The Hierarchical SLX Model

The matrix form of hierarchical Spatial Lag of X model representation is given by the following equations:

$$\begin{array}{rcl} y = \tilde{X}\beta + \Delta u + \varepsilon \quad \varepsilon\,|\sigma^2 & \sim & N\left(0, \sigma^2 I_N\right) \\ u\,|\tau^2 & \sim & N\left(0, \tau^2 I_J\right) \\ \tau^2 & \sim & IG\left(a, b\right) \end{array} \tag{2}$$

where $\tilde{X} = \begin{bmatrix} X & W_1X & \Delta Z\gamma & \Delta W_2 Z\delta \end{bmatrix}$ as illustrated before. The difference between the previous model and this model is the addition of the W_1X matrix and is what makes this model a spatial econometric one. W_1X is a spatially–weighted explanatory variable matrix at the district level. Again, as before, the addition of spatially–weighted explanatory variable matrix W_2Z with $J \times m$ dimension allows for spillovers at the state level. γ is a $m \times 1$ vector of coefficients associated with the W_2Z term, and we assume that ε and u, u and X, and u and Z are not correlated (Raudenbush and Bryk, 2002).

We use Bayesian econometric methods for our analysis and the priors used in this study are independent, hence, can be multiplied with one another. They are also proper priors and also are conjugates. Priors for β and γ are multivariate normal whereas priors for σ^2 and τ^2 are inverse–Gamma. Since we use uninformative values even though we use proper priors, we use a multivariate normal prior of mean 0_K vector and covariance $1000 \times I_K$ for β and γ where K represents the number of explanatory variables used in the study. The shape and scale parameters for the inverse–Gamma prior are set to 0.001.

We rely on a Gibbs sampling method to obtain our estimates since obtaining closed form solutions of the parameters analytically can only occur under special circumstances. Since the Gibbs sampler only requires that the conditional distributions be available, we rely on this method to obtain estimates.

The model we are estimating in this study is a local spatial econometric model. It is one among the two types of spatial econometric models, the other one being a global model. In a local model, spillovers in the independent variables are allowed. In case of our study, they are represented by the WX and WZ terms. Unlike a local model, global model, however, also contains a Wy term in addition to WX and WZ, allowing the spillover effect to disseminate across the entire sample.

We use a local spatial econometric model because school districts are closely situated to each other and it is unlikely that any spillovers are going to propagate across the entirety of the United States. It is most likely to be contained within a specific geographical range. In addition, we use a local model because the structure of our data is hierarchical in its natural state. Since, each state is different in terms of socioeconomic factors, allocating all states as homogeneous would lead to biased estimates of graduation rates. Moreover, local spillovers are a common occurrences in modelling regional patterns than global spillovers (LeSage, 2014). To put this statement in perspective in relation to our study, one would assume graduation rates in a given school district in Maine to be affected by its close neighbors than to be influenced by school districts in Florida.

In addition, the coefficients of local spillovers models as the SLX used in this study are easy to interpret as compared to global models such as the Spatial Durbin Model (SDM). In our model, all the coefficients have a straightforward interpretation. At the *district level*, β represents its own partial derivative (direct effects) and θ represents the cross–partial derivatives (indirect/spillover effects). The total effect is represented by $\beta + \theta$. At the *state level*, γ represents the direct effect and δ represents the indirect effect. The total effect is represented by $\gamma + \delta$.

3 Data

3.1 District Level Data

At the school district level we use high school graduation rates as the dependent variable. The dependent variable is the high school graduation rate of public school districts of 45 states and excludes charter schools and private schools. There are 8,194 individual school districts in our study. Since it is a spatial econometric approach, we exclude Alaska and Hawaii in our calculation as they have no contiguous neighbors. We also do not capture public school districts in Ohio, Utah, and Vermont due to missing data. The data for graduation rates is obtained from individual states' Department of Education website.

High school graduation rates have been a subject of debate in the education literature. Heckman and LaFontaine (2010) report that graduation rates differ from anywhere between 66% to 88% depending on the definition, sources, or methods used. The definition of high school graduation rates differ from "dividing the number of public high school diplomas by an estimate of the number of students who would have received diplomas that year if graduation rates were 100 percent" (Greene, 2001) to the government mandated Four-year Adjusted Cohort Graduation Rate (ACGR) implemented by the U.S. Department of Education. The ACGR is calculated "as the number of students who graduate in four years with a regular high school diploma divided by the number of students who entered high school four years earlier (adjusting for transfers in and out, émigrés and deceased students). (US Department of Education, 2017)" Despite the differences in measures of calculating high school graduation rates, we use the government mandated definition and rates as they provide with a uniform measure across all school districts.

We employ an education production function approach to modeling graduation rates. Hanushek (1986) categorizes inputs into education production as either family inputs, school inputs, or peer inputs. There is a large literature on the inputs into education production and we use variables that are standard in this literature (Hanushek et al., 2009). In this study, *Log of Mean Household Income* and *Log of Children from Single Family Household* fall under family inputs category. *Log of Mean Household Income* is the mean household income in a school district in 2012 inflation–adjusted dollars. *Log of Children from Single Family Household* represent the percentage of school-age children in a school district coming from a single parent household where the mother is present but not the father. Palardy (2013) found traditional family structure (consisting of both parents in the household) to positively affect high school graduation rates.

In addition to family inputs over which school districts have no control, there are school district inputs that may affect high school graduation rates. The variables that fall under this category are *Log of Instructional Salary per Pupil*, *Teacher Student Ratio*, *Local Revenue as a % of Total Revenue*, and *Log of Expenditure per Pupil*. *Log of Instructional Salary per Pupil* represents expenditures on salaries of staff categorized as instruction, such as classroom teachers. *Teacher Student Ratio* is also hypothesized to have a positive effect on graduation rates (Krueger, 2003). It is calculated as the number of enrolled students in a public school district divided by the total number of teachers. Reardon and Galindo (2002) find a negative relationship between student–teacher ratio on dropout rates. *Local Revenue as a % of Total Revenue* is calculated by dividing total local revenues by the total revenue. Hoxby (1999) and Hall (2007) provide evidence that school districts where more revenue comes from local taxpayers have better outcomes, other things being equal. We also include *Log of Expenditure per Pupil* as literature such as Jackson et al. (2016) find a positive relationship between increases in per pupil spending and completed years of education.

Finally, racial fractionalization index within a school district is represented by *Racial Fracitionalization Index* and serves as a Peer Input. It measures "the probability that two school district residents drawn randomly will be of different races" (Hall and Leeson, 2010). This measure intends to capture the differences in provision of education that might arise due to disagreement over education production that are correlated with race. Hall and Leeson (2010) find a negative relationship between a racial fractionalization index and school district performance in Ohio.

3.2 State Level Data

While school–district level explanatory variables can be hypothesized to affect graduation rates of school districts the most, state level policies also can have an effect. Therefore, we use state exit exam as our *state* variable.[1] Our *state* variable is a binary variable (*State Exit Exams*) that equals one if the state requires a state exit exam for its high school students, and 0 otherwise.

Table 1: Summary Statistics

	Year	Source	Mean	SD
District Level Variables				
High School Graduation Rates	2015	State Departments of Education	87.944	10.674
Log of Mean Household Income	2015	NCES	11.206	0.344
Log of Children from Single Parent Household	2015	NCES	5.427	1.502
Log of Instructional Salary per Pupil	2015	NCES	8.379	0.312
Teacher Student Ratio	2015	NCES	0.071	0.018
Local Revenue as a % of Total Revenue	2015	NCES	42.611	19.776
Log of Expenditure per Pupil	2015	NCES	9.474	0.336
Racial Fractionalization Index	2015	Own Calculation from NCES data	0.332	0.207
State Level Variable				
State Exit Exams	2013	Education Counts Research Center	0.46	0.50

NCES = National Center for Education Statistics.
N=8,194 for all *Level 1* variables.
N=45 for the *Level 2* variable.

Table 1 gives the summary statistics, year, and source for all variables used in this study.

4 Empirical Results

Tables 2 and 3 report the results from *district level* and *state level* SLX hierarchical model, respectively.

We ran the Gibbs sampling algorithm through the full conditional distribution of each of the parameters in this study. For each model, we ran 100,000 iterations using the Gibbs sampling algorithm to get our parameters estimates. However, we discard the first 50,000 iterations as they are in the "'burn–in" phase. The remaining 50,000 iterations are used to obtain parameter estimates.

[1] We are limited to one variable at the *state* level due to the fact that we are estimating fifty different intercepts for every variable included at this level.

In order to find the statistical significance of the parameters, we also calculate the 95% credible intervals for each parameter as this is a standard practice in Bayesian analyses. We do so to check whether the 95% credible interval contains 0 or not. If it does not, then the parameter estimate has a marginal density away from zero, hence, suggesting that the independent variable is statistically significant.

Determining the proper weight of the spatial weight matrix, W, is very important to our analysis. For global models such as the SAR or SDM, its effect on the estimates are little as long as LeSage and Pace's (2009) recommendation is followed. However, that is not true in the case of local models. Therefore, it is important to determine the most appropriate W matrix in order to get correct estimates. We define the most appropriate W matrix as the one that has the best goodness of fit as defined by the lowest Deviance Information Criterion statistic (DIC).[2]

We find the correct W matrix for both levels in the following manner: we compare nineteen different nearest–neighbor W matrices with neighbors ranging from 2 to 20. The district level consists of a total of 19 different models and the state level consists of a total of 9 different models. Therefore, there are 200 different models to choose from. If they are thought of in a matrix form, in terms of our study, the rows represent school–district level nearest neighbor $W's$ and columns represent the state level nearest–neighbor W matrices.

We then use the Deviance Information Criterion statistic (DIC) to choose from the different models to determine the W matrix that is the most appropriate (Spiegelhalter et al., 2002). The DIC statistic with the lowest number is the most appropriate model to use. In our case, W matrix with 11 nearest–neighbor at the school–district level and with 2 nearest–neighbors at the state level are the most appropriate spatial weight matrices to use.[3] Using the centroid of each school district, each school district's neighbors are defined as the eleven districts in that state whose centroid is closest. The same is done for states, except states are only assigned two neighbors according to the DIC criteria.

4.1 District Level Results

Table 2 shows the average direct, indirect, and total effects of *district level* explanatory variables on high school graduation rates. The results show that most of the variables of

[2] The DIC criteria is similar to a log-likelihood value but for a Bayesian context. It assesses model fit and the number of parameters, penalizing overfitting (Darmofal, 2009).

[3] For those unfamiliar with spatial econometrics, it may not seem intuitive for a state to only have 2 neighbors, when many states have multiple neighbors under simple rook contiguity. It is important to remember that the DIC criteria penalizes overfitting. While Tennessee may have eight neighbors by rook continguity, the DIC criteria says its two nearest neighbors (measured by distance between state centroids) provides the best model fit. Given regional convergence in politics (Heckelman and Dinan, 2013) and incomes (Heckelman, 2013) it is not surprising that the best model fit occurs with a parsimonious number of neighbors.

interest are statistically significant for all (direct, indirect, and total) effects.

Table 2: District Level Results with State Exit Exams as Level 2 Variable with 11 Nearest–Neighbor W Matrix

	Posterior Mean	Lower 95%	Upper 95%
Direct Effects			
Log of Mean Household Income	9.3033*	8.493	10.1245
Log of Children from Single Parent Household	-1.2726*	-1.4602	-1.0838
Log of Instructional Salary per Pupil	4.3249*	2.6365	6.0189
Teacher Student Ratio	-27.436*	-46.8731	-7.8065
Local Revenue as a % of Total Revenue	0.0526*	0.0374	0.0682
Log of Expenditure per Pupil	-4.6486*	-5.7808	-3.5192
Racial Fractionalization Index	-3.5688*	-4.8667	-2.2621
Indirect Effects			
Log of Mean Household Income	-0.0057	-1.5882	1.5663
Log of Children from Single Parent Household	-0.5792*	-0.9559	-0.2028
Log of Instructional Salary per Pupil	0.8295	-2.6425	4.3186
Teacher Student Ratio	-33.9301*	-66.0164	-1.6151
Local Revenue as a % of Total Revenue	-0.0536*	-0.0808	-0.0263
Log of Expenditure per Pupil	0.4016	-2.3622	3.1555
Racial Fractionalization Index	-0.8901	-3.4078	1.6414
Total Effects			
Log of Mean Household Income	9.2976*	7.8166	10.7666
Log of Children from Single Parent Household	-1.8518*	-2.2234	-1.4799
Log of Instructional Salary per Pupil	5.1543*	1.6651	8.6249
Teacher Student Ratio	-61.3661*	-97.8014	-24.8594
Local Revenue as a % of Total Revenue	-0.0010	-0.0264	0.0242
Log of Expenditure per Pupil	-4.247*	-7.0151	-1.4745
Racial Fractionalization Index	-4.4589*	-6.8034	-2.1044
Variance Posterior Mean			
Level 1 Error Variance: σ^2	8.4277		
Level 2 Error Variance: τ^2	8.5092		
DIC	2257805.39		

Note: N=8194. * denotes variables with a 95% credible interval without a 0. The results above are the level 1 (nested) results from a Bayesian Hierarchial SLX model where level 2 is the state. The aggregate state level results are presented in Table 3 while Table 4 presents all of the state-level intercepts.

The direct effects are comparable to previous studies as they represent the raw beta estimates in normal linear models. The own effects (direct effects) of all explanatory variables are statistically significant. *Log of Mean Household Income* is positively related to graduation rates with a coefficient of 9.3033. Since this variable is log transformed, a 1% increase in *Log of Mean Household Income* increases graduation rates by 0.09% (0.01×9.3033). *Log of Children from Single Parent Household* is negatively associated with graduation rates. A 1% increase in *Log of Children from Single Parent Household* decreases graduation rates by 0.01%. As shown in the previous section, high school graduation rates is found to increase in a traditional family structure where both parents are present (Palardy, 2013). *Log of Instructional Salary per Pupil* is associated with an increase in graduation rates as well. A 1% increase in *Log of Instructional Salary per Pupil* increases the dependent variable by 0.04%. Higher levels of *Teacher Student Ratio* are negatively related to graduation rates.

Local Revenue as a % of Total Revenue is associated with an increase in graduation rates, consistent with Hall (2007). A 1% increase in this variable leads to a 0.05% increase in graduation rates. Another variable used to explain variation in graduation rates is *Log of Expenditure per Pupil*. Surprisingly, this variable bears a negative sign. The effect of per pupil expenditures can be hypothesized to be positive because one can expect students to get access to better resources which might subsequently lead to an increase in graduation rates with increases in per pupil spending. A 1% increase in this variable decreases graduation rates by 0.05%. While the direct effects of this variable on graduation rates is not found, as mentioned in the previous section, Jackson et al. (2016) find a positive relationship between increases in per pupil spending and completed years of education. Finally, *Racial Fractionalization Index*, is also negatively associated with graduation rates. If a racially homogeneous school district, for instance, Macon County School District in Alabama with a *Racial Fractionalization Index* of 0.05 were to become heterogeneous, for instance, like Brewton City School District in Alabama with the index of 0.50, we would expect the graduation rate to drop by 1.78%. Explaining this result in terms of economic significance, a one standard deviation in the *Racial Fractionalization Index* (0.207) in a said school district decreases the graduation rate by 0.73% (0.207×-3.5688) or only about 6.9% ($0.207 \times -3.5688/10.67$) of the standard deviation of the dependent variable. Hall and Leeson (2010) also find diversity to negatively affect educational outcomes.

Since we are using a spatial hierarchical model in our study, we are also able to test the spillover effects or neighborhood effects, called indirect effects. One of the distinguishing factors between this study and previous studies is that we account for indirect effects in our analysis. The average indirect effect captures spillover effects of a change in an explanatory variable in a school district and how that affects observations of the dependent variable in neighboring school districts.

The indirect effects of a 1% increase in *Mean Household Income* in a school district decreases graduation rates in surrounding school districts by 0.000057% and is not statistically robust. *Log of Children from Single Parent Household* in a school district exerts a negative effect on the graduation rates of its neighboring school districts. This variable is statistically significant. An increase in *Log of Instructional Salary per Pupil* in a school district does not spillover to the graduation rates of surrounding school districts in a statistically significant manner. However, it exerts a positive effect. Like the direct effects of *Teacher Student Ratio*, the indirect effect is statistically significant and also negative. This means that an increase in *Teacher Student Ratio* in a school district negatively affects the graduation rates of surrounding school districts. *Local Revenue as a % of Total Revenue* increase in a school district decreases the graduation rates of surrounding school districts. This may be due to the fact that surrounding school district resources may be allocated to the referenced school district. This variable is statistically robust as well. *Log of Expenditure per Pupil* in a school district increases graduation rates in surrounding school districts and is statistically insignificant. It may be the case that an increase in a school district expenditure per pupil increases funding in surrounding school districts due to network effects, all else equal. As a school district becomes more heterogeneous, graduation rates of surrounding school districts decreases.

The final effect is called the total effects and is defined as the sum of direct effects and indirect effects. At this level, all variables of interest are statistically significant except for the *Local Revenue as a % of Total Revenue* variable. While *Log of Mean Household Income* and *Log of Instructional Salary per Pupil* are positively related to high school graduation rates, the remaining other variables exert a negative effect on the dependent variable.

4.2 State Level Results

As mentioned in Section 2, an advantage of using a hierarchical SLX model is its ability to account for heterogeneity at the state level, unlike the standard fixed effects models and also its ability to include spatially lagged independent variables at the *state level*. The inclusion of spatially lagged independent variables provides us with direct, indirect, and total effects at the state level as well. This variable accounts for the fact that states often engage in yardstick competition in the adoption of policies. We take the DIC statistic into account at the state level analysis also and use 2 nearest–neighbor spatially–weighed W matrix to get our estimates. Table 3 reports the results at the state level.

As can be seen, neither (direct, indirect, or total) effects have a statistically significant influence on graduation rates at the state level. State exit exams, while having a positive effect, do not have a statistically significant effect on graduation rates after accounting for state-level heterogeneity and spillovers. The stated goal of increasing competency in core

Table 3: State Exit Exam Results with 2 nearest–neighbor W matrix

	Posterior Mean	Lower 95%	Upper 95%
Constant	-9.0049	-25.0682	4.9573
Direct Effect	2.4626	-2.6976	7.6349
Indirect Effect	5.6681	-1.6215	13.0331
Total Effect	8.1307	-1.6644	17.7764

Note: N=45. * denotes variables with a 95% confidence interval without a 0. The results above are the level 2 (top) results from a Bayesian Hierarchial SLX model where level 1 are school districts. The school district results are are presented in Table 2 while Table 4 presents all of the state-level intercepts.

areas among high school students by conducting these high–stakes exams did not have any statistically significant relationship with increasing graduation rates after controlling for the embedded nature of school districts within states.

The random effects for each state are contained in Table 4 where entries in boldface type represent estimates of the 95% credible interval that do not contain zero. Of note is that the states of Colorado, Montana, Nevada, New Mexico, Oregon, Washington, and Wyoming all have negative estimates of the effect of a state exit exam on graduation rates. These negative results for these states highlight the heterogeneity across states in terms of the subjects tested, standards, and cut scores (Center on Education Policy, 2009).[4] Some states use general core competency exams in subjects such as science and mathematics as their high-stakes test, while other use end-of-course exams (Holme and Heilig, 2012). As Bishop (2004) notes, a third of states in 2004 still used minimum competency exams as a graduation requirement. If the purpose of high-stakes exit exams is to motivate students to achieve more, Bishop (2004) argues that minimum competency exams provide little inducement to try harder as additional work is not necessary for most students to pass.

The difference across states in terms of what is meant by a high-stakes exit exam helps to explain both the differential intercepts in Table 4 but also the different findings in the literature. For example, Baker and Lang (2013) find that minimum competency exams do not affect graduation rates, but standards based exit exams are associated with declines. Of these states, Colorado, Montana, Oregon, and Wyoming do not have high-stake exit exams and thus the negative and statistically-significant coefficient suggests that high stakes high school exit exams in these states would lead to a decline in high school graduation rates, *ceteris paribus*. While our results cannot pinpoint why high-stakes exams in these

[4] They also highlight the heterogeneity in test-taking population, which can influence the effect of high-stake exit exams depending on the type of exam (Bishop et al., 2000).

Table 4: Individual Estimates for Each State

State	Intercept	Lower 95%	Upper 95%
Alabama	-8.06	-19.98	10.92
Arizona	-15.54	-27.32	3.21
Arkansas	-11.88	-23.71	7.05
California	-11.96	-23.96	7.35
Colorado	**-21.04**	**-33.05**	**-1.93**
Connecticut	-14.77	-27.31	5.23
Delaware	-15.17	-28.22	4.69
Florida	-18.82	-30.73	0.06
Georgia	-13.28	-25.10	5.70
Idaho	-17.23	-29.17	1.82
Illinois	-17.15	-29.11	2.07
Indiana	-9.55	-21.52	9.66
Iowa	-11.90	-23.91	7.41
Kansas	-12.61	-24.68	6.76
Kentucky	-7.32	-19.33	11.81
Louisiana	-14.41	-26.39	4.60
Maine	-14.40	-26.70	5.00
Maryland	-14.14	-26.85	5.43
Massachusetts	-16.46	-28.86	3.47
Michigan	-19.18	-31.13	0.01
Minnesota	-15.81	-28.03	3.92
Mississippi	-18.69	-30.46	0.19
Missouri	-8.31	-20.12	10.65
Montana	**-63.50**	**-76.51**	**-43.53**
Nebraska	-13.87	-26.82	5.71
Nevada	**-20.16**	**-33.17**	**-0.63**
New Hampshire	-15.45	-27.90	4.31
New Jersey	-13.03	-25.37	6.75
New Mexico	**-24.02**	**-36.25**	**-4.74**
New York	-17.23	-29.80	2.84
North Carolina	-11.83	-23.74	7.29
North Dakota	-18.38	-30.80	1.41
Oklahoma	-18.86	-30.60	0.02
Oregon	**-22.06**	**-34.17**	**-2.80**
Pennsylvania	-12.83	-25.00	6.72
Rhode Island	-16.52	-29.27	3.26
South Carolina	-12.26	-24.21	6.67
South Dakota	-14.82	-26.77	4.31
Tennessee	-8.64	-20.58	10.40
Texas	-6.35	-18.15	12.68
Virginia	-9.52	-21.56	9.70
Washington	**-21.05**	**-33.19**	**-1.58**
West Virginia	-11.86	-24.16	7.49
Wisconsin	-10.72	-24.68	9.49
Wyoming	**-23.75**	**-36.39**	**-3.93**

Note: Results are the state-level random effects from the Bayesian Hierarchial SLX model estimated in Table 2 and Table 3. Boldface represents estimates of the 95% credible interval that do not contain zero.

states would lower graduation rates, it suggests that looking at these six states would be a fruitful avenue for future research. For example, Washington State's exam was reported to be extremely difficult (Harris, 2000; Shaw, 2002) and thus might encourage students to drop out if they feel they are too far from passing.

5 Conclusion

Given the longstanding history of high stakes state exit exams on high school graduation rates in the United States, it comes as no surprise that various empirical studies have found mixed results. However, these studies do not take into account the hierarchical nature of school districts and their embedded nature within state education systems. In addition, they do not account for spatial spillovers.

We employ a Bayesian Hierarchical Spatial Lag of X Model to draw inferences. Given the direct effects and spillover effects, we find that on average *Mean Household Income* and *Instructional Salary per Pupil* positively affect high school graduation rates, whereas, *Single Parent Household*, *Teacher Student Ratio*, *Local Revenue as a % of Total Revenue*, *Expenditure per Pupil*, and *Racial Fractionalization Index* negatively affects graduation rates at the school district. However, after controlling for hierarchical school–district level characteristics and spatial spillovers, at the state level, we generally find no statistically significant evidence of the effect of these high–stakes exit exams on graduation rates. Looking at the random effects, several states have a negative random effects coefficient and a 95% credible interval that does not contain zero. This heterogeneity in response is an avenue for future research, perhaps using the synthetic control method.

Our contributions to the literature are the following. First, we employ a hierarchical model to correct for results from previous studies which mostly used conventional linear methods. Second, papers with non–spatial estimates do not account for spillover effects of the explanatory variables on graduation rates. We are able to capture spatial spillovers in the explanatory variables to explain differences in high school graduation rates. Third, we are able to highlight heterogeneity in state-level effects of high-stakes exit exams that may be useful for future research.

References

Adewale, A. J., Hayduk, L., Estabrooks, C. A., Cummings, G. G., Midodzi, W. K., and Derksen, L. (2007). Understanding hierarchical linear models: Applications in nursing research. *Nursing Research*, 56(4):S40–S46.

American Educational Research Association (2000). *Position Statement on High-Stakes Testing*. American Educational Research Association, Washington, DC.

Baker, O. and Lang, K. (2013). The effect of high school exit exams on graduation, employment, wages and incarceration. Working Paper 19182, National Bureau of Economic Research.

Beardsley, A. and Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18).

Bishop, J. H. (2004). Money and motivation: Other nations elicit better performance from their students through the use of high-stakes graduation exams. *Education Next*, 4(1):62–68.

Bishop, J. H., Moriarty, J. Y., and Mane, F. (2000). Diplomas for learning, not seat time: The impacts of New York Regents examinations. *Economics of Education Review*, 19(4):333–349.

Bond, L. A. and King, D. (1995). *State High School Graduation Testing: Status and Recommendations*. North Central Regional Educational Laboratory, Oak Brook, IL.

Center on Education Policy (2009). *State High School Exit Exams: Trends in Test Programs, Alternate Pathways, and Pass Rates*. Center on Education Policy, Washington DC.

Chib, S. (2008). Panel data modeling and inference: a bayesian primer. In *The econometrics of panel data*, pages 479–515. Springer.

Darmofal, D. (2009). Bayesian spatial survival models for political event processes. *American Journal of Political Science*, 53(1):241–257.

Ed Counts Research Center (2017). State exit exam based on 10th grade standards or higher. `http://http://www.edcounts.org/createtable/viewtable.php`.

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.

Greene, J. P. (2001). *High School Graduation Rates in the United States*. Manhattan Institute for Policy Research, New York, NY.

Greene, J. P. and Winters, M. A. (2004). *Pushed out or pulled up? Exit exams and dropout rates in public high schools*. Manhattan Institute for Policy Research, New York, NY.

Hall, J. (2007). Local school finance and productive efficiency: Evidence from Ohio. *Atlantic Economic Journal*, 35(3):289–301.

Hall, J. C. and Leeson, P. T. (2010). Racial fractionalization and school performance. *American Journal of Economics and Sociology*, 69(2):736–758.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3):1141–1177.

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2009). New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(3):349–383.

Harris, W. (2000). WASL a 'monster' of a test; even state officials don't have all the answers yet. *Spokesman Review*, 11 May.

Heckelman, J. and Dinan, J. (2013). Empirical evidence regarding regional political convergence in the United States. *Journal of Regional Analysis and Policy*, 43(1100-2016-90114):170–177.

Heckelman, J. C. (2013). Income convergence among US states: Cross-sectional and time series evidence. *Canadian Journal of Economics*, 46(3):1085–1109.

Heckman, J. J. and LaFontaine, P. A. (2010). The American high school graduation rate: Trends and levels. *Review of Economics and Statistics*, 92(2):244–262.

Hemelt, S. W. and Marcotte, D. E. (2013). High school exit exams and dropout in an era of increased accountability. *Journal of Policy Analysis and Management*, 32(2):323–349.

Holme, J. J. and Heilig, J. V. (2012). High-stakes decisions: The legal landscape of high school exit exams and the implications for schools and leaders. *Journal of School Leadership*, 22(6):1177–1197.

Hoxby, C. M. (1999). The productivity of schools and other local public goods producers. *Journal of Public Economics*, 74(1):1–30.

Jackson, C. K., Johnson, R. C., and Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, 131(1):157–218.

Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal*, 113(485):F34–F63.

Lacombe, D. J. and Flores, M. (2017). A hierarchical slx model application to violent crime in mexico. *The Annals of Regional Science*, 58(1):119–134.

LeSage, J. P. (2014). What regional scientists need to know about spatial econometrics. *Review of Regional Studies*, 44(1):13.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, FL.

Luke, D. A. (2004). *Multilevel Modeling*. Sage Publications, Thousand Oaks, CA.

Marchant, G. J. and Paulson, S. E. (2005). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives*, 13:6.

National Commission on Excellence in Education (1983). A nation at risk: The imperative for educational reform. *The Elementary School Journal*, 84(2):113–130.

Ou, D. (2010). To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam. *Economics of Education Review*, 29(2):171–186.

Palardy, G. J. (2013). High school socioeconomic segregation and student attainment. *American Educational Research Journal*, 50(4):714–754.

Papay, J. P., Murnane, R. J., and Willett, J. B. (2010). The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis*, 32(1):5–23.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, Thousand Oaks, CA.

Reardon, S. F. and Galindo, C. (2002). Do high-stakes tests affect students' decisions to drop out of school? Evidence from NELS. In *Annual Meeting of the American Educational Research Association, New Orleans, LA*.

Shaw, L. (2002). 10th grade exam harder than the college test? *Seattle Times*, 28 November.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, 64(4):583–639.

Subramanian, S. (2010). Multilevel modeling. In *Handbook of Applied Spatial Analysis*, pages 507–525. Springer.

US Department of Education (2017). *Every Student Succeeds Act: High School Graduation Rate Non-Regulatory Guidance*. US Department of Education, Washington, DC.

Warren, J. R. and Edwards, M. R. (2005). High school exit examinations and high school completion: Evidence from the early 1990s. *Educational Evaluation and Policy Analysis*, 27(1):53–74.

Warren, J. R., Jenkins, K. N., and Kulick, R. B. (2006). High school exit examinations and state-level completion and GED rates, 1975 through 2002. *Educational Evaluation and Policy Analysis*, 28(2):131–152.

Warren, R. W., Grodsky, E., and Lee, J. C. (2008). State high school exit examinations and postsecondary labor market outcomes. *Sociology of Education*, 81(1):77–107.

Warren, R. W. and Jenkins, K. N. (2005). High school exit examinations and high school dropout in Texas and Florida, 1971–2000. *Sociology of Education*, 78(2):122–143.