

ESSAYS ON SCHOOL FINANCE AND TEACHER PERFORMANCE

By

Paul N. Thompson

A DISSERTATION

Submitted
to Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics - Doctor of Philosophy

2014

ABSTRACT

ESSAYS ON SCHOOL FINANCE AND TEACHER PERFORMANCE

By

Paul N. Thompson

Chapter 1 analyzes the Ohio fiscal stress labeling system, a statewide financial intervention system that labels school districts with projected deficits in the general fund. Labeled school districts are required by the state to implement a financial recovery plan that balances budgets, with recovery operated by the district or the state depending on the level of projected deficits. This chapter examines the effect of these labels on school district financial behavior and housing prices from 2000-2012. In response to these labels school districts decrease capital and operating expenditures, with larger percentage reductions in capital, and increase local property tax revenue funding operating expenditures. These labeled school districts are in much better financial positions following successful recovery and maintain financial viability well after recovery is complete. House prices fall following a state takeover of the district's financial decision-making, suggesting that the state takeover sends a stronger signal to residents than district-led recoveries. In addition to responding to labels, school districts and homebuyers are responsive to earlier state interventions that alleviate financial problems before labels become necessary.

Chapter 2 consider issues of equality and efficiency in two different school funding systems - a state-level system in Michigan and a foundation system in Ohio. Unlike Ohio, the Michigan system restricts districts from generating property or income tax revenue to fund operating expenditures. In both states, districts fund capital expenditures with local tax revenue. The results indicate that although average revenue and expenditures per pupil in Michigan and Ohio are almost identical, the distributions of the various revenue sources are quite different. Ohio's funding system has greater equality in terms of total revenue, largely due to Ohio redistributing state funds to the least wealthy districts while Michigan does not. This chapter finds that relatively wealthy Michigan

districts spend more on capital expenditures while relatively wealthy Ohio districts spend more on labor and materials. This suggests that constraints on raising local revenue to fund operating expenditures in Michigan could create efficiency issues.

Chapter 3 analyzes Empirical Bayes' (EB) estimation, a popular procedure used to calculate teacher value-added. This estimation strategy is often motivated as a way to make imprecise estimates more reliable. In this paper we review the theory of EB estimation and use simulated and real student achievement data to study its ability to properly rank teachers. This chapter compares the performance of EB estimators with that of other widely used value-added estimators under different teacher assignment scenarios. This chapter finds that, although EB estimators generally perform well under random assignment of teachers to classrooms, their performance generally suffers under nonrandom teacher assignment. Under nonrandom assignment, estimators that explicitly (if imperfectly) control for the teacher assignment mechanism perform the best out of all the estimators we examine. This chapter also finds that shrinking the estimates, as in EB estimation, does not itself substantially boost performance.

This thesis is dedicated to my wife Katie, who was there to lean on during all the ups and downs of the PhD process. I also dedicate this thesis to my best friend, Shiloh, who spent so many days by my side and always there for a walk or a game of fetch when I needed it. I love you both and couldn't have done it with out you.

I also want to thank my parents who inspired me to pursue my PhD and a career in academia.

ACKNOWLEDGEMENTS

I wish to thank my committee members who were more than generous with their expertise and precious time. A special thanks to Dr. Michael Conlin, my committee chairman, for his countless hours of reflecting, reading, encouraging, and most of all patience throughout the entire process. Thank you Dr. Leslie Papke, Dr. Jeff Wooldridge, and Dr. Mark Skidmore for agreeing to serve on my committee and providing helpful comments.

I want to thank the Institute of Education Sciences, whose Pre-Doctoral Training Grant (Award #R305B090011) to Michigan State University helped fund much of this work. The Economics of Education training program helped me find my passion in education research and allowed me the opportunity to meet many in the field through trips to conferences, on campus speakers, etc.

TABLE OF CONTENTS

LIST OF TABLES	viii	
LIST OF FIGURES	ix	
CHAPTER 1	SCHOOL DISTRICTS AND HOUSING PRICE RESPONSES TO FISCAL STRESS LABELS: EVIDENCE FROM OHIO	1
1.1	Introduction	1
1.2	Fiscal Stress Labels in Ohio	4
1.2.1	Overview of Fiscal Stress Labels	4
1.2.2	Fiscal Oversight	5
1.2.3	Fiscal Emergency	6
1.2.4	Trends in Label Receipt	8
1.3	Data	9
1.4	Descriptive Statistics	10
1.4.1	School District Expenditures and Revenues	10
1.4.2	Residential Home Sales	12
1.5	Empirical Specification and Results	13
1.5.1	Difference-in-Differences Results	13
1.5.2	Event Study Results	18
1.5.3	Regression Discontinuity Results	22
1.6	Conclusion	27
CHAPTER 2	MICHIGAN AND OHIO K-12 EDUCATIONAL FINANCING SYSTEMS: EQUALITY AND EFFICIENCY	30
2.1	Introduction	30
2.2	Institutional Details	32
2.2.1	Michigan K-12 Finance	32
2.2.2	Ohio K-12 Finance	34
2.2.3	Similarities Between Ohio and Michigan Financing Systems	36
2.3	Data and Descriptive Statistics	37
2.3.1	Data	37
2.3.2	Descriptive Statistics	38
2.3.2.1	School District Revenue	40
2.3.2.2	School District Expenditures	47
2.4	Conclusion	49
CHAPTER 3	AN EVALUATION OF EMPIRICAL BAYES' ESTIMATION OF VALUE-ADDED TEACHER PERFORMANCE	51
3.1	Introduction	51
3.2	Empirical Bayes' Estimation	52

3.3	Summary of Estimation Methods	63
3.4	Comparing VAM Methods Using Simulated Data	65
3.4.1	Simulation Design	66
3.4.2	Evaluation Measures	68
3.5	Simulation Results	71
3.5.1	Fixed Teacher Effects versus Random Teacher Effects	71
3.5.1.1	Random Assignment	71
3.5.1.2	Dynamic Grouping and Nonrandom Assignment	73
3.5.1.3	Heterogeneity Grouping and Nonrandom Assignment	74
3.5.2	Shrinkage versus Non-Shrinkage Estimation	75
3.5.3	Sensitivity Analyses	76
3.6	Comparing VAM Methods Using Real Data	77
3.6.1	Data	77
3.6.2	Results	78
3.7	Conclusion	80
APPENDICES		83
APPENDIX A CHAPTER 1 TABLES AND FIGURES		84
APPENDIX B CHAPTER 2 TABLES AND FIGURES		99
APPENDIX C CHAPTER 3 TABLES AND FIGURES		110
BIBLIOGRAPHY		115

LIST OF TABLES

Table A.1	Fiscal Stress Label Transitions from 2000-2012	84
Table A.2	Table of Means - School District Demographic Characteristics	85
Table A.3	Table of Means - School District Financial Characteristics	86
Table A.4	Table of Means - Housing and Parcel Characteristics	88
Table A.5	Effect of Fiscal Stress Receipt on Expenditures per pupil, by severity level	89
Table A.6	Effect of Fiscal Stress Receipt on Revenues per pupil, by severity level	89
Table A.7	Effect of School District Fiscal Stress Label Receipt on Housing Prices	90
Table A.8	Estimated Discontinuity at Year 3 Ratio Cutoff	93
Table A.9	Regression Discontinuity Results	96
Table A.10	Variable Names and Definitions	97
Table A.11	Comparison of Full and Analytic Samples	98
Table B.12	Property Tax Rates and Taxable Values 2002-2010	99
Table B.13	Revenue and Demographic Characteristics	100
Table B.14	Revenue Regression Results Without District Fixed Effects	106
Table B.15	Revenue Regression Results With District Fixed Effects	108
Table C.16	Simulation Results: Comparing Fixed and Random Teacher Effects Estimators	110
Table C.17	Simulation Results: Comparing Shrunken and Unshrunken Estimators	111
Table C.18	Fraction of Teachers Ranked in Same Quintile by Estimator Pairs	112
Table C.19	Description of Value-Added Estimators	113
Table C.20	Definitions of Grouping-Assignment Mechanisms	113
Table C.21	Description of Evaluation Measures of Value-Added Estimator Performance	113

LIST OF FIGURES

Figure A.1	Geographic Distribution of Labeled School Districts Across Ohio	84
Figure A.2	Yearly Number of Labels, by Label Severity	85
Figure A.3	Operating and Capital Per Pupil Expenditures and Revenues (2010 \$)	87
Figure A.4	Event Study Results - District Enrollment	90
Figure A.5	Event Study Results - Total Operating Expenditures PP	91
Figure A.6	Event Study Results - Total Capital Expenditures PP	91
Figure A.7	Event Study Results - Local Property Tax Revenue and Millage Rates	92
Figure A.8	Event Study Results: Housing Prices	93
Figure A.9	Distribution of Projected General Fund Balance to Revenue Ratios	94
Figure A.10	Change in District Finances and Housing Prices, by Year 3 Projected Ratio	95
Figure B.11	Total Enrollment of Quintiles	101
Figure B.12	Total Revenue Per Pupil	102
Figure B.13	State Revenue Per Pupil	103
Figure B.14	Local Revenue Per Pupil	104
Figure B.15	Local Operating Tax Revenue Per Pupil	105
Figure B.16	Local Capital Property Tax Revenue Per Pupil	107
Figure B.17	Total Expenditures Per Pupil	109
Figure C.18	Spearman Rank Correlations Across Different VAM Estimators	112

CHAPTER 1

SCHOOL DISTRICTS AND HOUSING PRICE RESPONSES TO FISCAL STRESS LABELS: EVIDENCE FROM OHIO

1.1 Introduction

In the current financial climate, rising budget deficits have burdened many school districts and local governments, leading some to the brink of bankruptcy. To help address these growing deficits, many states have developed financial intervention systems that identify financially troubled school districts or local governments and provide varying levels of state intervention.¹ On one end of the spectrum, some states heavily intervene and overhaul financial practices. One example is the state of Michigan, where an emergency manager took over the financial decision-making for the city of Detroit after bankruptcy and building nearly \$20 billion of debt. In contrast, other states monitor financial behavior of districts, but provide little, if any, intervention into financial practices. The California financial monitoring system is an example of this approach, where one-fifth of all school districts in the state had FY2013 deficits, but beyond short-term loans there is not much state intervention.

I analyze one of these systems, the Ohio fiscal stress labeling system, which labels school districts with projected deficits in their general fund and requires these districts to implement a financial recovery plan that balances budgets. Districts with less severe deficits receive a fiscal oversight label, under which districts are placed in charge of developing and implementing these recovery plans. Districts in more severe financial trouble receive a fiscal emergency label, under which the state takes over the financial decision making of the district. As part on this financial takeover, the state assumes the responsibility of developing and implementing the recovery plan.

¹A 2013 Pew Charitable Trusts report found that 19 states currently have a financial intervention system in place. Earlier work by Honadle (2003) and Kloha, et al. (2005) found that 15 states had some fiscal health evaluation system in place and nearly a third more states are considering using these indicators.

Despite the growing use of these financial intervention systems, little is known about the effects of these programs. My paper fills this gap in the literature and provides the first evidence on the effect of these fiscal stress labels on school district financial behavior and housing prices. Given the distinction between whether financial recovery is operated by the district or by the state, the Ohio system allows me to identify separate effects depending on the type of label received. I compile a balanced panel of all 613 Ohio school districts from 2000-2012; collecting data on dates of label receipt and removal, school district expenditures and revenues, projected deficits, tax rates, taxable values, local tax election outcomes, and housing transactions. To estimate the effects of these labels, I use three different identification strategies: difference-in-differences, an event study, and a regression discontinuity design. Using these three approaches, I find that these recovery plans do change financial behavior. I find that districts decrease capital and operating expenditures following receipt of fiscal oversight, with larger percentage reductions in capital. Districts increase local property tax revenue funding operating expenditures during both fiscal oversight and fiscal emergency, while shifting their tax mix away from new capital projects. In addition to these effects of label receipt, I also analyze the response of districts and residents to earlier state interventions that try to alleviate financial problems before fiscal stress labels are necessary. I find that districts with projected deficits are decreasing both capital and operating expenditures in response to these early interventions. Residential home sale prices fall following receipt of fiscal emergency, but find no statistically significant effects following fiscal oversight. But the regression discontinuity and event study results suggest that this lack of an effect during fiscal oversight may be due to housing prices falling in response to the financial changes made during these early interventions.

In addition to being the first paper to analyze the effects of this policy, this study also contributes to the literature on the capitalization of school quality into housing prices. While much of this literature has focused on the effect academic quality has on housing prices,² my results suggest

²Previous literature has focused on the effect of more transitory changes in school inputs and outputs on housing prices. Most notably, Black (1999) finds that a \$500 increase in per-pupil expenditures increases house prices by 2.2 percent. Cellini, et al. (2010) examine the effect of bond referenda passage on housing prices. They find that marginal

that residents also care about what these financial recoveries may mean for future academic quality or expected tax rates. This paper also contributes to the literature on school district responses to budgetary issues. Although previous literature³ has examined how districts react to budgetary problems, these studies primarily focus on voluntary changes to finances as a result of budget deficits. These voluntary changes are likely to be smaller and less expansive than those made under the scrutiny of the mandated recovery plans examined here.

For states considering these types of policies, the structure of the state school funding system is likely to play a role in how districts respond. In Ohio, school districts are able to supplement state aid through local tax revenue funding operating expenditures. This setup allows fiscal stress districts to increase taxes along with expenditure cuts to offset deficits. Given that fourteen states have a similar funding structures to Ohio and eleven states place no restrictions on local tax revenue, districts in these other states may respond similarly to these types of labels. Districts in states where local operating tax revenue is restricted, such as Michigan and California, would likely focus more heavily on reductions in expenditures in response to these labels and recovery plans.

homebuyers are willing to pay \$1.50 for an additional dollar of capital expenditure, largely due to increases in safety and aesthetics of new and renovated buildings. Numerous studies of United States and international school districts (Bayer et al., 2007; Black, 1999; Clapp et al., 2008; Dougherty et al., 2009; Gibbons and Machin, 2003, 2006; Gibbons, Machin, and Silva, 2009; Fack and Grenet, 2010; Davidoff and Leigh, 2008) find around a three percent increase in home prices resulting from a one standard deviation increase in test score levels. There is little evidence, however, of capitalization of test score gains (Brasington and Haurin, 2006; Downes and Zabel, 2002; Kane et al., 2006). Another set of papers has examined the effect of school report card grades, which combine multiple sources of test score and district characteristics into a single rating for each school district. Figlio and Lucas (2004) examine the capitalization effects of school district report card grades in Florida and find that report card grades do provide valuable information about school quality to homebuyers. Studies examining these ratings in other settings find more mixed results (Kane, Staiger, and Samms, 2003; Fiva and Kirkebøen, 2009; Zahirovic-Herbert and Turnbull, 2009). The fiscal stress labels may capture aspects of school quality that these school district academic quality rankings fail to signal, aspects that the housing market may value differently than achievement levels. For a more expansive review of this literature, see the Black and Machin chapter in the "Handbook of the Economics of Education" or Nguyen-Hoang and Yinger (2011).

³Previous literature examining the responses of school districts to budgetary issues has analyzed how districts exhibiting characteristics of fiscal distress behave. This literature largely finds, as expected, that districts in fiscal distress respond by increasing revenue and/or decreasing expenditures. For a thorough review of this literature, see Trussel and Patrick (2012).

1.2 Fiscal Stress Labels in Ohio

1.2.1 Overview of Fiscal Stress Labels

Ohio school districts with projected deficits in the general fund receive either a fiscal oversight label or fiscal emergency label depending on the level of the deficit.⁴ Labeled districts are required to develop financial recovery plans that outline changes to financial behavior that achieve balanced budgets. These proposed changes include reductions in expenditures and/or increases in local tax revenues. Under fiscal oversight, district school boards are required to develop and implement recovery plans, incorporating recommendations made by the Auditor and the Ohio Department of Education (ODE). Successful implementation of these recovery plans will result in removal of the label, but failure to adopt or adhere to these recovery plans results in districts being placed into fiscal emergency. Under this most severe label, a state commission assumes the role of financial decision maker for the district, including handling the development and implementation of the financial recovery plan.

After a district is notified of label receipt, a press release is issued on the Auditor's website that details to the public why the district received the label.⁵ This press release is often reported in local newspapers in the days following its issuance. Similarly, a press release is posted to inform residents when a label is changed or removed. Given the lack of local media coverage prior to the publication of the press release, the initial label receipt is likely unanticipated.⁶ This

⁴Given that the main variation of interest is between district-led recovery and state-led recovery, I choose to combine fiscal caution and fiscal watch into one indicator for fiscal oversight. Differentiating between fiscal caution and fiscal watch does not change the overall results. The fiscal emergency and fiscal watch labels were introduced in 1996, while the fiscal caution label was not instituted until 2001. For more detailed information regarding the history of these labels, the various criteria used to select these districts, and the requirements associated with these labels see the Ohio Auditor of the State website (<http://www.auditor.state.oh.us/services/lgs/fiscalwatch/schools.htm>) or the Ohio Department of Education Website (<http://www.ode.state.oh.us/gd/gd.aspx?page=2TopicRelationID=1012>).

⁵[www.auditor.state.oh.us/newscenter/press/releases/category/Fiscal Caution, Watch, and Emergency](http://www.auditor.state.oh.us/newscenter/press/releases/category/Fiscal%20Caution,%20Watch,%20and%20Emergency)

⁶Examinations of newspaper reports gives no indication that residents are aware that the Auditor is examining school district financial records or that a fiscal stress determination was made prior to the publication of the press release. News reports covering districts that have been further downgraded to fiscal watch due to the failure to create a recovery plan speculate about whether the district will fall into fiscal emergency, but no official word is given until the press release is posted. The speculation is also likely due to the information these downgrades provide regarding

unanticipated change in the fiscal stress rating of the district serves as an exogenous informational shock regarding the fiscal health of the district and is the main source of identification in this analysis.

Data on the school districts labeled in these fiscal stress categories are obtained from the Ohio Auditor and the ODE. Information collected includes the date the school district initially received a fiscal stress label and, if no longer labeled, the date the label was removed. The data also include information on whether a fiscal oversight or fiscal emergency label was received and dates of transition from one label to another. Since 2000, 99 of the 613 school districts in Ohio have received at least one of these labels. Including districts that are still currently labeled, districts are labeled for 3.58 years on average. The average duration is 3.23 years for the subset of districts where label removal is observed within the panel.

1.2.2 Fiscal Oversight

The process of receiving a fiscal oversight label starts with the yearly five-year forecast. Each October, school districts must submit a five-year financial forecast to the Ohio Department of Education. This forecast projects the expenditures and revenues of the district for the current fiscal year and the next four fiscal years based on predicted changes in taxable property values, tax rates, state aid, teaching and healthcare contracts, and other operating costs.⁷ Districts with projected current year deficits that exceed two percent of general fund revenue receive a fiscal oversight label, while districts with projected deficits in the second or third projected years receive other, less

the ability of school officials to handle these financial problems.

⁷Districts are flagged if a forecast predicts a deficit (i.e. negative balance in the general fund) for the current fiscal year or projects positive general fund balances that are less than two percent of projected revenue for that year. In addition to current year deficits or low fund balances, district forecasts are flagged if deficits that exceed two percent of projected revenue are present in the second or third projected years. Districts with projected deficits are given an opportunity to make reductions in expenditure and/or increases in revenue in order to alleviate the deficit and avoid receiving the fiscal caution label. Districts with current year projected deficits may avoid fiscal caution if they immediately eliminate the deficit by reducing expenditures or through a tax advance from the county auditor to cover the amount of the deficit, but otherwise do not have sufficient time to levy new taxation to cover the amount of the deficit. Districts with deficits in the second or third projected year may eliminate the deficit through passage of new tax referenda and/or reductions in expenditure.

invasive state interventions. Districts with year 2 projected deficits must submit a letter explaining how they will eliminate the projected deficit, while districts with year 3 projected deficits are only notified of the deficit with no formal requirement to submit any proposed financial changes to the state. The main goal of these early interventions is to help districts with deficits to solve their financial problems before fiscal stress labels become necessary.

After receiving a fiscal oversight label, the districts must develop a financial recovery plan. To help districts in the development of these plans, the Auditor may conduct a performance audit of the district that provides recommendations to help guide the district's recovery plan.⁸ In addition, the Department of Education may visit and inspect the district's monthly financial situation and provide technical assistance and training to school officials. After implementing a recovery plan, a district is removed from fiscal oversight when the district projects positive fund balances for the next two projected years. However, a district cannot be removed from fiscal oversight in the same fiscal year in which the fiscal oversight label was received. On average, districts complete successful financial recovery under fiscal oversight in 2.72 years.⁹

1.2.3 Fiscal Emergency

Districts that are unable to develop recovery plans on their own or fail to adhere to the recovery plan while under fiscal oversight are placed in fiscal emergency. A school district may also receive fiscal emergency if the projected current year general fund deficit exceeds fifteen percent of the general fund revenue and the school district has failed to gain voter approval on a levy that would provide sufficient revenue to eliminate the deficit. School districts in fiscal emergency are placed under the control of a financial planning and supervision commission, which assumes the

⁸This performance audit identifies areas, in which the district can cut costs and raise additional revenue. These identified areas serve as a guide for districts or the state commission in the development of financial recovery plans. Recommendations often cited in these audits include freezing step increases for teacher salaries, reducing staff, increasing employee health insurance contributions, implement fee-based extracurricular activities or reduce extracurricular expenditures, and eliminating unnecessary busing, among others.

⁹The average duration rises to 3.42 years if districts that have not been observed completing financial recovery under fiscal oversight (i.e. those with a fiscal oversight label at the end of the panel) are included in the calculation.

responsibility of developing and implementing a financial recovery plan for the school district.¹⁰ The commission reviews and assumes responsibility for all projected revenues and expenditures and has final approval on any tax levies and debt issuances the school district wishes to propose to voters. The commission also has the authority to reduce the number of employees irrespective of current employment contracts and collective bargaining agreements in order to achieve balanced budgets.¹¹

The Auditor is required to conduct a performance audit for districts in fiscal emergency and these districts are eligible to receive two-year loans from the state. These loans, which are up to the amount of a two year advance on the district's state foundation aid, are intended to keep the district solvent during its recovery.¹² Once the district meets the objectives of the recovery plan and eliminates the deficits that led to fiscal emergency, the district is removed from the label. In rare cases where the district still has outstanding debt at the time of release from fiscal emergency, the district is placed into fiscal watch until the debt is paid. On average, districts that are downgraded to fiscal emergency spend 0.81 years in fiscal oversight. Once in fiscal emergency, successful recovery takes an average of 3.62 years to complete, but in some cases has taken upwards of nine

¹⁰This commission consists of five voting members including the Director of the Office of Budget and Management and the State Superintendent. The other three members of the commission include a business person appointed by the Governor, a business person appointed by the mayor or county auditor, and a parent with a child in the district appointed by the State Superintendent. The business people must have at least five years experience in the public or private sector in business management, public accounting, or another related field. The commission must include at least one female member and one minority member if minorities constitute at least twenty percent of district enrollment. If the commission fails to come to a consensus on a financial recovery plan within 120 days, the commission is dissolved and a fiscal arbitrator is put in charge of recovery. The fiscal arbitrator assumes all powers and duties of the commission, including creation and implementation of a financial recovery plan.

¹¹These reductions first eliminate administrative and non-teaching employees, giving preference to employees with continuing contracts and those with greater seniority. If the budget is still not balanced following this first round of layoffs, the commission may layoff teachers in order to achieve a balanced budget. The commission also may remove the district superintendent and/or treasurer if they fail to comply with the commission's recovery plan.

¹²These loans are paid out of the School District Solvency Assistance Fund, which consists of two accounts, the Shared Resource Account and the Catastrophic Expenditure Account. The Shared Resource Account provides districts with a two year, interest free advance on the district's state foundation payment in order to help the district to remain solvent while it implements its recovery plan, but loans must be repaid within two years. The Catastrophic Expenditures Account is used to issue grants to districts that suffer a catastrophic event that depletes the district's finances, but can also be used for solvency assistance if all of the Shared Resource Account funds have been allocated.

years.¹³

1.2.4 Trends in Label Receipt

Table A.1 provides the number of transitions from one fiscal stress state to another that occurred between 2000 and 2012. At the start of 2000, six districts were in fiscal oversight and nine were in fiscal emergency. Between 2000 and 2010, 80 districts transitioned from no label to fiscal oversight and four transitioned from no label to fiscal emergency. Of the 86 districts labeled in fiscal oversight, 19 transitioned to fiscal emergency because of the failure to adopt a financial recovery plan during fiscal oversight. There were also 71 successful financial recoveries completed during this time period, with 46 completed under fiscal oversight and the other 25 completed under fiscal emergency.

Figure A.1 graphically depicts where these labeled districts are located across the state. The map indicates that many of these labeled districts are clustered around the state's various urban centers. This pattern in the location labeled districts suggests that regional economic changes or spillover effects among adjacent districts may play a role in the deteriorating finances in these districts. Figure A.2 depicts the yearly number of districts labeled as fiscal oversight or fiscal emergency from 2000-2012. The number of school districts in fiscal stress began increasing in 2002, primarily due to the sharp increase in the number of fiscal oversight labels. In 2001, six districts were in fiscal oversight, but by 2005 there were 42 districts holding a fiscal oversight label. While the number of fiscal emergency labels has remained relatively constant since 2005, there has been a noticeable reduction in the number of fiscal oversight labels. Largely due to this decline in the number of fiscal oversight labels, the total number of labels has fallen from 54 in 2005 to 37 in 2012.

¹³The average duration is 3.93 years, 0.82 in fiscal oversight and 3.11 in fiscal emergency, if districts that have not been observed completing financial recovery under fiscal emergency (i.e. those with a fiscal emergency label at the end of the panel) are included in the calculation.

1.3 Data

The fiscal stress data described in Section 2 are augmented with data on school district finances, tax rates, taxable values, election outcomes, test scores, and residential home sales. Data on school district general fund revenues and expenditures from 2000-2012¹⁴ come from the Ohio Auditor and include the current balance in the general fund, an observable measure of district financial health. Data on projected deficits, the main criteria used to select these districts, come from the ODE and are discussed in more detail in Section 5.3. Detailed information on expenditures and revenues is obtained from the NCES Common Core of Data. This information contains total revenues and expenditures per year and disaggregated data for the specific components that make up these totals. This allows me to analyze exactly which types of revenues and expenditures districts are targeting in these recovery plans. In addition, student test score proficiency levels, school report card grades, and district accountability status are collected from the Ohio Department of Education. Finally, residential home sales data are collected for 63 of the 88 counties in Ohio.¹⁵

In Ohio, districts fund much of their operating and capital expenditures through local property and income taxation. Tax revenue for operating expenditures is used for general school expenses (e.g. teachers, supplies, etc.), while tax revenue for capital expenditures is largely restricted for use only in funding the specific capital project. Given the flexibility of operating revenue, districts with budget deficits may find it advantageous to adjust their tax mix away from capital projects and towards greater operating taxation in order to offset these deficits. To examine this, I collect data from the Ohio Department of Taxation on all tax rates levied by school districts. These data include the specific purpose of each tax and the yearly tax rate. These tax data are supplemented with yearly taxable values for real property, tangible personal property, and tangible public utility property for each school district. Combining these two data sets allows me to separately calculate

¹⁴The year used in this paper refers to the school year, so this data spans the 1999-2000 school year to the 2011-2012 school year.

¹⁵As shown in Table A.11 in the Appendix, the 63 counties included in the analytic sample are quite representative of the entire state. The districts in the analytic sample have slightly larger enrollments and spend slightly more per student, but are nearly identical on all of the other covariates of interest.

the yearly tax revenue collected for operating and capital expenditures.

Renewal of existing taxation and implementation of new taxation is subject to voter approval. Thus, I obtain data from the Ohio Secretary of State on school tax election outcomes for all property and income tax referenda from 2004-2012. The information collected includes the proposed tax rate, the proposed dollar amount of the debt issuance (if a bond referenda), the number of yes and no votes, the duration of the proposed tax, and the purpose of the tax.

Finally, housing sales data are collected from individual Ohio county auditors. These data include the location of the property, the date of sale, the sale price, numerous characteristics of the home, and the tax district in which the property resides.¹⁶ The data from these various sources are linked to each parcel using the associated tax district. To ease the constraints in collecting the housing transaction data, the sample is restricted to include only single-family homes with sale prices that exceed \$10,000. The analytic sample contains 1,011,726 parcel sales with all the relevant parcel characteristics. For a full description of the data sets used in this analysis and an explanation of the relevant variables see Appendix Table A.10.

1.4 Descriptive Statistics

1.4.1 School District Expenditures and Revenues

Tables A.2 and A.3 provide means and standard deviations for the school district financial and demographic characteristics, broken down by the various stages of fiscal stress. Panel A provides descriptive statistics for the 509 districts that never receive a fiscal stress label during the time period examined in this study. Panel B provides descriptive statistics for the 80 labeled districts that are observed both prior to receiving the label (Pre-Fiscal Stress) and during the time the label is applied to the district (Fiscal Stress). Panel C provides the descriptive statistics for the 72 labeled districts that are observed both during the time period the label is applied to the district and after

¹⁶A tax district corresponds to a unique county-township-city-school combination. Within a school district there may be a number of different tax districts. This means that the taxes faced by residents within a given school district may vary depending on which tax district the resident resides.

the label is removed (Post-Fiscal Stress). While Panel A is mutually exclusive from panels B and C, districts for which the whole fiscal stress duration is observed are found in both panels B and C.

Districts that never receive a label are smaller, less likely to reside in urban areas, have better school report card grades, and have better financial situations than districts that are labeled. In the pre-fiscal stress period, 35 percent of labeled districts have general fund deficits that are greater than two percent of revenue, compared with only six percent of districts that never receive a label. Districts that are never labeled also collect nearly \$1,000 more per pupil in local revenue than districts that are eventually labeled. In addition, as observed in Table A.2, enrollment is falling in these labeled districts following receipt of the label.

Figure A.3 depicts yearly average expenditures and revenues per pupil for operating and capital for each year relative to label receipt.¹⁷ In panel (a), operating expenditures per pupil in labeled districts are rising more quickly than non-labeled districts¹⁸ prior to label receipt. Immediately following label receipt, operating expenditures per pupil for labeled districts fall from about \$9,250 to \$9,075, after which labeled and non-labeled districts have nearly the same levels of operating expenditures. The difference in trend between labeled and non-labeled districts is more noticeable when examining capital expenditures. As observed in panel (b) of Figure A.3, capital expenditures per pupil for districts that are never labeled is relatively constant over time, around an average of \$1,500. Conversely, capital expenditures for labeled districts fall from an average of \$2,000 per pupil two years prior to label receipt to around \$900 per pupil during the year of label receipt and fluctuates around that level in subsequent years. Districts also appear to alleviate these deficits through increases in local operating property tax revenue. As observed in panel (c), labeled districts

¹⁷Since the composition of districts may change (i.e. not observing years prior for districts receiving label prior to 2000 and not observing years post for districts receiving labels later in the time span of interest), I also generate these figures using the same districts observed over the entire duration that I calculate these means. The results obtained from these figures are largely unchanged when the composition of districts is kept constant.

¹⁸Yearly means are calculated for non-labeled districts so that each labeled district observation is compared to the non-labeled districts in the same year. I then calculate means for the variables of interest in each year relative to label receipt for both the labeled districts observed in that time period and the yearly mean for the non-labeled districts associated with those labeled district observations.

receive around \$3,400 per pupil in operating revenue prior to label receipt, which is nearly \$700 to \$800 per pupil less than is received by districts that are never labeled. Following label receipt, operating tax revenue per pupil received by labeled districts increases substantially, before leveling off at around \$4,000 per pupil three years after label receipt. Districts also are shifting away from taxes for capital projects, as evidenced in panel (d) although this seems to be more of a steady decline than a sharp change after the label is received.

Following successful financial recovery, as shown in Panel C of Table A.3, the general fund balances, total expenditures, and total revenues of labeled districts are very similar to those of districts that never receive a label. The only substantial difference between recovered districts and those that are never labeled is in capital expenditures, where successfully recovered districts spend \$845 per pupil on capital compared with \$1,482 for districts that never receive a label.¹⁹

1.4.2 Residential Home Sales

Table A.4 provides means and standard deviations for the characteristics of residential home sales, again broken down by the various stages of fiscal stress. Examining Panels A and B of Table A.4, homes in districts that never receive a label sell for an average of \$147,171 while homes sold in the pre-fiscal stress period in labeled districts sell for an average of \$139,979. The homes sold in districts that never receive a label also tend to be slightly larger, with living areas averaging 1,704 square feet compared with 1,587 square feet in homes sold in pre-fiscal stress labeled districts. The sale price of housing in labeled districts falls from an average of \$139,979 to \$133,267 following the receipt of the label. Given that the average characteristics of the homes sold are nearly identical between the pre-fiscal stress and fiscal stress periods, this reduction in sale prices is likely not attributable to changes in the composition of sold homes. Following the removal

¹⁹Although much of this difference in capital expenditures can be attributed to the sharp drop in capital expenditures following label receipt in panel (b) of Figure A.4, some may be attributed to funding received from the Ohio School Facilities Commission (OSFC). Created in 1997, the OSFC has funded capital projects in over 300 school districts, starting with the lowest wealth districts. Since 2000, 43 fiscal stress districts have been eligible to receive funding through the program, while 217 non-labeled districts have become eligible.

of the label, the average sale price in labeled districts falls from \$140,687 to \$130,080, but again the average characteristics of the homes sold during these two periods are very similar. In many districts, however, a majority of the post-fiscal stress period occurs during the Great Recession, so some aspects of the post-housing bubble may be causing these much lower sale prices during the post-fiscal stress period.

1.5 Empirical Specification and Results

1.5.1 Difference-in-Differences Results

Districts receiving a fiscal stress label are required by the state to implement financial recovery plans that achieve balanced budgets. The descriptive analysis provided in Table A.3 and Figure A.3 suggests that these recovery plans eliminate expenditures on capital projects and increase local tax revenue for operating expenditures. However, the conclusions of this descriptive analysis may be driven by other factors (e.g. size of general fund deficits, enrollment, regional economic trends, etc.). Thus, to analyze the impact of transitioning from one fiscal stress state to another, I estimate the following specification:

$$\ln(y_{dt}) = \alpha + \beta_1 postFO_{dt} + \beta_2 postFE_{dt} + \beta_3 endFO_{dt} + \beta_4 endFE_{dt} + \gamma \mathbf{X}_{dt} + \phi \mathbf{D}_{dt} + \lambda_d + \theta_t + \varepsilon_{st} \quad (1.1)$$

where $\ln(y_{dt})$ is the natural log of a per-pupil expenditure or revenue variable for school district d in year t ; \mathbf{X}_{dt} is a vector including the general fund balance to revenue ratio and indicators for whether a district falls within one of the ratio cutoff values; \mathbf{D}_{dt} is a vector of other time-varying district characteristics; λ_d is a vector of school district fixed effects and θ_t is a vector of year fixed effects; and ε_{st} is an idiosyncratic error.

The main variables of interest are the four fiscal stress indicators for the various transitions into and out of these labels. The $postFO_{dt}$ variable is an indicator for the time period following the receipt of a fiscal oversight label, which identifies the effect of initial label receipt on school district

financial behavior. The $postFE_{dt}$ variable is an indicator for the time period following the receipt of a fiscal emergency label, which identifies the differential effect of transitioning from district-led recovery to state-led recovery.²⁰ Given that changes to finances are likely to be different during district-led recovery than changes resulting from state-developed recovery plans, we might expect $\beta_2 \neq 0$. For districts that complete financial recovery under fiscal oversight, the $endFO_{dt}$ variable is an indicator for the time period following removal of the fiscal oversight label and identifies the differential effect of transitioning from fiscal oversight to no label. Similarly, for districts that complete financial recovery under fiscal emergency, the $endFE_{dt}$ variable is an indicator for the time period following removal of the fiscal emergency label, identifying the differential effect of transitioning from fiscal emergency to no label.

Table A.5 provides results of equation (1) when the specification considers per-pupil expenditure variables. Across each of the expenditure types, I find statistically significant reductions in the period following initial label receipt (i.e. no label to fiscal oversight). As observed in column 1, total expenditures per-pupil are 9.6 percent lower in districts following receipt of fiscal oversight than non-labeled districts with the same general fund balance to revenue ratio. Total operating expenditures per-pupil (column 2) are 3.7 percent lower in fiscal oversight districts than operating expenditures in similar non-labeled districts. Expenditures on salaries (column 3) are 5.4 percent lower following fiscal oversight and continue to decline by 2.9 after receipt of fiscal emergency. Expenditures on employee benefits (column 4) are 2.6 percent lower in fiscal oversight districts than in non-labeled districts with the same general fund balance to revenue ratio. In addition, employee benefits fall by an additional 5.5 percent following the removal of the fiscal emergency label, suggesting that employee contracts renegotiated during the commission's recovery plan may have large effects on employee benefits even after the financial recovery has been completed. These

²⁰Since nearly all of the fiscal emergency districts in the 2000-2010 time period were initially placed in fiscal oversight, this indicator identifies the differential effect of being downgraded into fiscal emergency. As observed in Table A.1, four districts transition immediately from no label to fiscal emergency. The choice of whether to include these districts in the postFE variable makes little difference to the overall results and therefore are included.

long run effects on benefits may suggest that these districts may be able to renegotiate contracts at lower rates during financial recovery, with these contracts remaining in effect well after recovery is complete.

Fiscal oversight districts also make substantial reductions to capital expenditures. As observed in columns 5 and 6 of Table A.5, total capital expenditures per pupil are 55.8 percent lower and per pupil expenditures on new construction are 73.3 percent lower following receipt of fiscal oversight compared to non-labeled districts with the same general fund balance to revenue ratio. I do not find, however, any significant differential effects on capital expenditures following fiscal emergency receipt or after removal of these labels. The larger percentage changes in capital expenditures may suggest that capital projects are relatively easier for districts to eliminate or districts may be eliminating capital projects at higher rates in order to minimize the reductions in operating expenditures.

Table A.6 provides results of equation (1) when the specification considers revenue variables. Although I find no statistically significant increases or reductions in total revenue (column 1) following label receipt, I find a 8.9 percent reduction in total revenue following the removal of the fiscal emergency label, largely due to reductions of 7.5 to 8.1 percent in local, state, and federal revenue. This drop in revenue following removal of fiscal emergency may suggest that districts do not require as much revenue after successful completion of financial recovery. In addition, federal revenue per pupil is 5.3 percent lower in fiscal oversight districts.²¹

Changes in local revenue in these districts are driven primarily by changes local property tax revenue, as districts shift their tax mix away from taxes funding capital projects and towards taxes that generate operating revenue that can be used to offset deficits. Local property tax revenue per pupil increases by 4.9 percent following receipt of fiscal oversight compared to non-labeled districts with the same general fund balance to revenue ratio. While I find no statistically significant

²¹These reductions in federal revenue come from reductions in Title I revenue, IDEA revenue, and non-specified federal aid. Results of various federal sources are available upon request. These results suggest transitory changes in federal aid, especially for districts in urban and low-income communities may have a large impact on district budgets.

differential change in local revenue per pupil following fiscal emergency receipt, the point estimate suggests about a 6.7 percent increase in total local property tax revenue per pupil following fiscal emergency. This increase in local property tax revenue, the sum of revenue collected for operating and capital, is largely driven by additional property taxation to fund operating expenditures.²² As observed in column 6, local property tax revenue for operating expenditures increases by 6.5 percent following fiscal oversight and an additional 9.2 percent following receipt of fiscal emergency. While I find no statistically significant changes in property tax revenue for capital expenditures, the point estimates suggest that capital property tax revenue is falling substantially in these districts. Given that operating revenue can be used to offset deficits while capital revenue is largely earmarked for the specific capital project, it is not surprising that districts in fiscal stress would shift their tax mix towards operating revenue and away from revenue for capital projects.

These changes to revenues and expenditures may positively or negatively impact housing prices depending on the trade off between the reductions in the tax-services bundle these districts offer residents and the increases in the financial viability of these fiscally stressed districts following successful recovery. These labels may also provide a signal that the district is of lower quality, which would likely be reflected in lower sale prices. Although Table A.4 suggested that home prices fall following label receipt, these descriptive statistics do not account for changes in prices due to the Great Recession and the collapse of the housing market, which severely impacted home prices in Ohio. While the average characteristics of homes do not appear to change across different fiscal stress time periods, controlling for the characteristics of individual homes will lead to more reliable estimates of the effect of these labels on housing prices. Therefore, to better understand how these labels may be impacting the housing market, consider a stylized hedonic model of the

²²Districts in Ohio can also use income taxes to fund operating expenditures. I find no statistically significant effects on income tax revenue, a result that is not all that surprising given that many of the districts generating income tax revenue are in rural areas of Ohio. As shown in Figure A.1, a majority of the labeled districts come from the areas surrounding urban areas. In fact, of the 111 districts that have received at least one of these labels only 28 districts implement an income tax.

differential effect of these labels on house prices:

$$\ln(P_{idjt}) = \alpha + \beta_1 FO_{dt} + \beta_2 FE_{dt} + \beta_3 endFO_{dt} + \beta_4 endFE_{dt} + \gamma \mathbf{Z}_{idjt} + \omega \mathbf{X}_{dt} + \lambda_d + \phi_j + \theta_t + \varepsilon_{idjt} \quad (1.2)$$

where $\ln(P_{idjt})$ represents the natural log of the sale price of parcel i in school district d and municipality/township²³ j at month-year t ; \mathbf{Z}_{idjt} is a vector of observable housing and parcel characteristics (e.g. number of rooms, number of bedrooms, number of bathrooms, indicators for the amount of acreage²⁴, and the square footage of the home's living area); \mathbf{X}_{dt} is a vector of observable district-level financial characteristics (e.g. the general fund balance to revenue ratio, which is one of the primary selection variables used by the Auditor when deciding to apply a fiscal stress label, and the various eligibility cutoff points); λ_d is a vector of school district fixed effects; ϕ_j is a set of county fixed effects; θ_t is a set of month-by-year fixed effects; and ε_{idjt} is an idiosyncratic error.

Table A.7 presents the results of the various specifications of equation (2). When I include district and month-by-year fixed effects (column 1), I find that home prices fall by 5.6 percent following the receipt of fiscal emergency,²⁵ but find no statistically significant effects following receipt of fiscal oversight. I also find that housing prices rise by 7.1 percent following the removal of the fiscal emergency label, suggesting that the end of the state takeover may be sending a positive signal about the quality of the district. This is in contrast to the statistically insignificant 6.5 percent drop in home prices following successful recovery during fiscal oversight. The point estimates of these removal variables suggest that districts may end up in a better financial position when

²³Parcels that lie within incorporated areas are linked to a given city or village j . Parcels in unincorporated areas are linked to a given township j .

²⁴Intervals of acreage are used instead of total acreage due to the fact that actual acreage is missing for a large percentage of the sample. In sensitivity checks, total acreage was used in place of these intervals and the general results remained the same, although the power was reduced greatly due to the reduction in the sample size.

²⁵The magnitude of this effect is slightly below that of the 6.1 percent reduction in sale prices resulting from a below average school report card grade, a traditional measure of overall academic quality. These school report card grades are a rating that is based on student performance on standardized tests and other state accountability measures. See Figlio and Lucas (2004) for more information on these report card grades.

recovery is undertaken by the state instead of the district. This may also suggest that the large cuts in expenditures undertaken during fiscal oversight may have long-lasting effects on educational services and academic quality in these districts, which would likely be capitalized into lowered housing valuations after recovery is complete.

As suggested in Figure A.1, regional factors may influence district financial troubles, so controlling for county-level effects may account for region-specific factors that are influencing district revenues and costs. Including county fixed effects (column 2) to the analysis yields nearly identical results to those without the inclusion of these fixed effects. Including a county-by-year fixed effect (column 3) to capture yearly economic changes across different parts of the state, yields similar results although the point estimate is no longer statistically significant for the transition from fiscal oversight to fiscal emergency. Even when including the county-by-year fixed effects, I continue to find a 6.6 percent increase in the sale price following the removal of fiscal emergency.

1.5.2 Event Study Results

To account for the possibility that districts may be changing their behavior prior to receiving the label, which would bias the above results, I also use a flexible event study framework. This model also allows me to examine how soon after label receipt these recovery plans take effect and for how long the effects of these recovery plans persist. If financially recovered districts are maintaining balanced budgets in the future, at least some of these changes are expected to persist over many years. To perform this analysis, I estimate the following model:

$$\ln(y_{dt}) = \alpha + \sum_{k=-6}^9 \beta_k FS_{dtk} + \gamma \mathbf{X}_{dt} + \phi \mathbf{D}_{dt} + \lambda d + \theta_t + \varepsilon_{dt} \quad (1.3)$$

where FS_{dtk} is an indicator for k years before or after initial label receipt, with $k = 0$ signifying the year of label receipt. Note that since I only observe outcomes starting in 2000, the FS_{dtk} indicators with $k < 0$ will be identified primarily off of districts that receive labels later in the

panel.²⁶ Figures 4-8 depict the estimates of β_k from equation (3) for enrollment, total expenditures per pupil for capital and operating, an total local property tax revenue per pupil collected for capital and operating. In each figure the black line represents the point estimates of the β coefficients for each year relative to label receipt dummy variable and the grey shaded region represents the 95% confidence interval from the standard errors clustered at the school district level.²⁷

Since I examine per-pupil revenues and expenditures, changes in enrollment in these labeled districts as a response to these labels will influence the changes observed in per-pupil revenues and expenditures. Therefore, I first estimate equation (3) using district enrollment as the dependent variable, with the results presented graphically in Figure A.4. Although I find no statistically significant changes in district enrollments prior to initial label receipt, I find around a seven percent reduction in enrollment within the first two years after label receipt, with enrollment continuing to decline upwards of seven years after label receipt. These results suggest that parents may be responding negatively to these labels and may be moving out of the district or sending their kids via open enrollment to nearby, non-fiscally stressed districts. In addition, fewer people may be moving into these financially troubled districts and/or fewer open enrollment children may choose districts in fiscal stress. Given this large drop in district enrollment following label receipt, the increases in local property tax revenue per pupil found estimating equation (1) are likely to overstated, while the reductions in per-pupil expenditures are likely to be understated.

Reductions in per-pupil operating and capital expenditures occur immediately after label receipt, with reductions in per-pupil operating expenditures persisting much longer than reductions

²⁶Given that not all districts receive these labels at the same time, the β_k are identified off of a potentially different set of labeled school districts, which may cause the estimates to be biased if there are unobserved heterogenous treatment effects. To test for this bias, I first estimate equation (3) separately for only those districts for which I observe outcomes for the year of label receipt. As the year of label receipt is observed for 87 districts out of the 102 that received a label between 2000-2010, restricting the estimation to this subset of the data yields results that are qualitatively and quantitatively similar to those using all of the data. I also estimate equation (3) using a balanced panel of school districts that I observe in all years. Although the results are similar to those using the unbalanced panel, the standard errors become much larger due to the reductions in the sample size.

²⁷To ease interpretation of the coefficients, I drop the fiscal stress indicator for $k = -1$ (i.e. year prior to label receipt). Thus the β_k coefficients identify treatment effects relative to the effect for the year prior to label receipt. The figures signify that this zero is imposed, by including a zero point estimate without an associated confidence interval.

in per-pupil capital expenditures. Figures 5 and 6 graphically depict results from equation (3) using total operating or total capital expenditures per-pupil variables as dependent variables. Transitory spikes are observed in both capital and operating expenditures per-pupil two or three years prior to label receipt. These spikes could lead to label receipt if these rises in expenditures increase the five-year forecast projections. In addition, there appear to be some marginally statistically significant reductions in per-pupil expenditures occurring the year prior to label receipt. These reductions likely result from districts responding to earlier state interventions (see Sections 2.2 and 5.3 for a greater discussion of these interventions). I find that total operating expenditures per-pupil (Figure A.5) fall by 0.8 percent in the year before label receipt, while total capital expenditures per-pupil (Figure A.6) fall by nearly 10 percent.

Even after the reductions made in the year prior, labeled districts continue to reduce per-pupil expenditures in the years following label receipt. Total operating expenditures per-pupil fall by 5.4 percent in the year following label receipt, with reductions of between 3 and 5 percent persisting upwards of nine years after label receipt. Reductions in per-pupil capital expenditures are much more transitory, with statistically significant reductions occurring only in the first few years after label receipt. Total capital expenditures per pupil fall by 75 percent in the first couple years after label receipt compared to non-labeled districts with the same general fund balance to revenue ratio. Despite the lack of statistical significance in later years, the point estimates suggest that about a 30 percent reduction in capital expenditures per-pupil remains as far out as eight years after initial label receipt. This large reduction in capital expenditures in the first few years of fiscal stress suggests that capital expenditures may be relatively easier to adjust and/or districts may eliminate large capital projects at higher rates in an attempt to minimize cuts to operating expenditures. These results are consistent with the results found in the earlier difference-in-differences analysis, where reductions in capital expenditures were much larger than those made to operating expenditures.

Using local property tax revenue and millage rates as dependent variables in equation (2), the results of which are depicted in Figure A.7, I find that districts are increasing local property tax

revenue funding greater operating expenditures immediately after label receipt. Labeled districts experience a reduction in property tax rates for both capital and operating in the years leading up to label receipt, suggesting that reductions in tax rates may lead to budget deficits in these districts. Unsurprisingly, there are no statistically significant changes to local tax revenue in the year prior to label receipt, as districts that pass new property or income taxes likely avoid receiving the label all together. In addition, many districts do not have adequate time to raise new taxes in order to avoid a label after being flagged by the ODE for deficits in the five-year forecast.

Following label receipt, per-pupil property tax revenue collected for operating expenses increases, while tax revenue funding capital expenditures falls. Operating property tax revenue per-pupil, shown in Panel (a), rises by 9.8 percent in the first two years after label receipt. Although operating property tax revenue per pupil begins to decline in the third year after label receipt, I still observe a 5.7 to 8.6 percent increase relative to the year prior to label receipt that persists upwards of nine years after initial label receipt. Property tax revenue for capital expenditures, depicted in Panel (c), falls by 6 percent in the year of label receipt, but this reduction becomes statistically insignificant between the year after label receipt and five years after label receipt. These changes in property tax revenues are largely driven by changes in the millage rates levied in these districts.²⁸ Examining Panel (b), the millage rate for operating taxes increases by nearly 1.75 mills in the first few years after label receipt. The millage rate begins to fall two or more years after initial label receipt, but significant effects persist as far as six years after label receipt. There are no statistically significant changes in the millage rate for capital projects, as shown in Panel (d), following label receipt.

²⁸Millage rates increase primarily due to the frequency at which tax referenda are placed on the ballot and the frequency at which these referenda are approved by voters. Districts propose and pass more tax referenda funding operating expenditures after receiving one of these labels, with the largest increase in the number of referenda approved occurring under fiscal emergency. In order to focus more heavily on increasing taxes for operating expenditures, districts appear to shift away from new proposals for capital projects. Following label receipt, districts propose fewer capital projects to voters, an effect that persists as far as five years after label receipt. While omitted here, results of this analysis are available upon request.

To assess how quickly housing prices respond to these labels, I estimate the following model:

$$\ln(P_{idjt}) = \alpha + \sum_{k=-6}^9 \beta_k FS_{dtk} + \gamma Z_{idjt} + \omega \mathbf{X}_{dt} + \lambda_d + \phi_j + \theta_t + \varepsilon_{idjt} \quad (1.4)$$

The results of this analysis are presented in Figure A.8. In the year prior to label receipt housing prices fall by 3.4 percent in labeled districts relative to non-labeled districts. This drop in housing prices may be a response to residents observing the district’s five year forecast and capitalizing the label receipt into valuations before it is received or residents may be capitalizing some of the reductions we see in expenditures occurring in the year before label receipt. The point estimates, although statistically insignificant, suggest housing prices continue to fall following label receipt, declining by a total of 4.7 percent in the six years following label receipt.

There are two main explanations for the drop in housing prices. First, these labels may signal negative information to residents and potential homebuyers about the financial health of the district, lowering the perceived quality of the district. This helps explain the difference-in-differences finding that residents respond more strongly to fiscal emergency labels. This suggested that a stronger signal may be sent by the fiscal emergency label compared to the fiscal oversight label due to the uncertainty surrounding the state takeover. The second explanation for the drop in housing prices is that prices may capitalize the changes in district expenditures and revenues made in response to the recovery plans associated with these labels. This helps explain the drop observed in housing prices in the year before the label, as districts are making reductions to expenditures but residents have not yet been informed about the district receiving a label.

1.5.3 Regression Discontinuity Results

While these changes to district finances and housing prices in the year before label receipt may suggest that districts and homebuyers are anticipating label receipt, it is likely that these changes are a result of the state intervention districts receive prior to label receipt. Recall from Section 2.2 that districts with projected deficits exceeding two percent of general fund revenue in

the second or third projected years are subject to state intervention. Although these interventions are not as stringent as those associated with the label, districts with deficits in the second or third projected years do have the opportunity to work with state financial consultants to help solve their financial problems before fiscal labels become necessary. To assess the effect of these early interventions, I collect five-year financial forecasts of projected deficits from 2001-2012. Given that all districts with projected deficits exceeding two percent of revenue in any of the first three projected years receive some level of state intervention, I use a regression discontinuity design comparing districts just above and below this two percent deficit to revenue ratio cutoff to estimate the effect of receiving the state intervention on district finances and housing prices.

As with any regression discontinuity design, manipulation of the cutoff is a primary concern regarding the validity of the design. Figure A.9 depicts the density of the projected general fund balance to revenue ratio for each of the first three projected years. Of the 7,872 district years in the sample, only 74 and 489 district-years are observed falling below the two percent cutoff in projected year 1 (Panel a) and projected year 2 (Panel b), respectively. While these large discontinuities exist in the density at the two percent cutoff in the first two projected years, there does not appear to be a noticeable discontinuity in the density at the cutoff in the projected year 3 ratio (Panel c). The differences in these densities is likely due to differences in the incentive to manipulate the cutoff, which is based on the relative benefits and costs of manipulation. The benefit of manipulation is the avoidance of the sanction associated with failing to meet the given two percent deficit cutoff, while the main costs are the costs of reducing expenditures and/or increasing taxes or the probability of detection if districts manipulate the assumptions of the forecast to avoid the sanction.²⁹ While the costs are likely to be the same regardless of the projected year used, the benefit grows as the the deficit gets closer to being realized. The benefit of avoidance is higher in

²⁹As discussed previously in section 2.2, districts submit these five year forecasts and may manipulate the projections of revenues and/or expenditures in order to show a positive fund balance. While districts have much more discretion over the projections for expenditures (e.g. projecting smaller increase in salary and/or other operating costs), they may also exaggerate projected property value growth to increase projected property tax revenue.

projected year 1 than projected year 3 because the severity of the projected year 1 intervention (i.e. label receipt) is more severe than the projected year 3 sanction (i.e. notified of deficit and given assistance from fiscal consultant). In addition, some districts may want to fall below the cutoff in projected year 3, in order to qualify for the assistance from the state.

While the endogenous sorting around the cutoff makes use of the year 1 and year 2 projected ratio problematic in an RD design, the year 3 ratio provides a viable option for use in this type of design. Although the density of year 3 projected ratio is continuous,³⁰ the validity of the design can also be called into question if districts on either side of the cutoff differ across other covariates. To test this, I create bins of size 0.02 and run a regression of yearly changes in school district characteristics to observe whether there are statistically significant differences in these covariates on either side of the cutoff. The estimated discontinuities at the cutoff are reported in Table A.8, but I find little evidence of endogenous sorting around the cutoff along these dimensions.

To get a better sense of how school districts are reacting to these early interventions, I examine expenditures and tax revenues for districts just above and below the two percent general fund deficit to revenue ratio from the year 3 projection. It is important to note that because I am using the year 3 projection, I am examining how current expenditures and revenues are influenced by deficits that are two years away from being realized. However, districts that fall below this cutoff in year t need to make any changes in expenditures or revenues in year t or $t+1$ to avoid the year 2 sanction in year $t+1$ and the fiscal stress label in year $t+2$. Thus, it is expected that districts below the cutoff will have yearly changes in expenditures and revenues that are different than the changes of districts above this cutoff.

Figure A.10, which fits a local polynomial to the data on either side of the cutoff, suggests that districts below this cutoff do behave differently between year t and year $t+1$ than districts above the cutoff. In panel (a), we observe a change in operating expenditures per pupil for districts just

³⁰When applying the McCrary test to this data (McCrary 2008), I find a log difference in the density at the two percent cutoff of -0.002, which suggests that there is no noticeable discontinuity in the density at the cutoff.

below the cutoff that is nearly \$130 per pupil less than the change in operating expenditures for districts just above the cutoff. There is a smaller, \$50 discontinuity at the cutoff in terms of capital expenditures per pupil (panel b). In addition to drops in expenditures, we observe a change in operating tax revenue for districts just below the cutoff that is nearly \$75 per pupil more than the change in operating expenditures for districts just above the cutoff (panel c). Districts with larger deficits (i.e. moving farther left away from the cutoff), make much larger reductions in both capital and operating expenditures than those at the cutoff.

I also examine whether homebuyers are capitalizing these financial changes into housing prices. As depicted in panel (d), the change in the average district housing price between year t and year $t+1$ is nearly \$200 lower in districts just below the cutoff than those just above. But as the deficit grows (i.e. as the ratio gets more negative), the reduction in the average district sale price falls substantially. This may indicate that homebuyers are responding more strongly to observing large drops in per-pupil capital and operating expenditures than to the small financial changes made by districts near the cutoff.

To estimate the effect of these early interventions on school district expenditures, local property tax revenues, and housing prices, I estimate the following equation for districts that did not currently have a label in year t :

$$\Delta Y_{id,t+\tau} = \alpha + 1[\text{ratio3}_t < -0.02]_t \beta + f(\text{ratio3}_t, \gamma) + \Delta \mathbf{X}_{d,t+\tau} \delta + \varepsilon_{id,t+\tau} \quad (1.5)$$

where $\Delta Y_{id,t+\tau}$ represents yearly changes in expenditures per pupil, operating property tax revenue per pupil, or the average sale price³¹ in school district d in year $t + \tau$ ($\tau = 0, 1$); $\Delta \mathbf{X}_{d,t+\tau}$ is a vector of changes in district and/or parcel controls; and $\varepsilon_{id,t+\tau}$ is an idiosyncratic error. The main parameter of interest, β identifies the discontinuity in the outcomes around the two percent cutoff

³¹To create this variable, I aggregate the parcel level housing transaction data to the school district level. This gives the average price of all sold homes in the district in year t . I also create yearly averages of the parcel characteristics of these sold homes to use as controls.

in the third projected year general fund balance to revenue ratio (ratio3). Districts that have year 3 projected deficits in year t are likely to make changes to expenditures and revenues during either year t or $t+1$ in order to avoid additional sanctions in year $t+1$ and $t+2$.

Table A.9 reports estimates obtained from the above specification. Districts that are just below the two percent deficit to revenue cutoff in year t do not differentially decrease operating expenditures per pupil (Panel A) compared to districts just above the cutoff between year $t-1$ and year t . However, the change in operating expenditures per pupil between year t and year $t+1$ is \$195 lower for districts just below the cutoff, possibly as districts make a more concerted effort to avoid label receipt in year $t+2$. I find no statistically significant differences in the yearly changes in capital expenditures (Panel B) or operating property tax revenues (Panel C) for districts just below the cutoff. As these deficits grow (i.e. as the projected year 3 ratio becomes more negative), however, districts make much larger reductions to operating expenditures, reduce capital expenditures, and increase local property taxes to fund operating expenditures, as evidenced by the coefficients on the projected year 3 ratio variable. These results are very similar to what was observed in the year before label receipt in the event study results. Recall from Section 3.2, where districts that received a label were found to slightly reduce operating and capital expenditures in the year prior to label receipt. Given that a majority of labeled districts have projected year 3 deficits that are much larger than two percent of revenue,³² taking the event study and regression discontinuity results together suggests that the changes in the year prior to label receipt are likely a response to the early intervention program. The lack of a significant increase in local property taxes for labeled districts in the year before label receipt (see Figure A.7), suggests that these labeled districts are likely districts that tried to offset deficits through increasing local taxes, but failed to secure voter approval.

Although districts just below the cutoff are making small cuts to per-pupil expenditures, hous-

³²Of the 99 labeled districts during the panel, 89 districts had projected deficits greater than eight percent of revenue and 63 had projected deficits greater than fifteen percent of revenue.

ing prices in these districts appear to increase by about \$2,000 between year t and $t+1$. This may suggest that these small cuts in expenditures are more than offset by the positive benefits of maintaining financial viability and avoiding further financial intervention. While homebuyers are responding positively to expenditure changes in districts near the cutoff, they appear to react negatively to the larger financial changes observed for districts farther away from the cutoff. Although housing prices are not very responsive to the projected ratio in general, as evidenced by the lack of statistical significance on the year three projected ratio variable, they are responsive to large deficits. When interacting the ratio with the indicator for whether the ratio is below the cutoff, we see as the ratio gets more negative the housing price falls substantially. Again since a majority of the districts receiving these labels have ratios that are well away from the cutoff, these results are similar to the nearly three percent reduction in housing prices in the year before label receipt found in the event study results. The similarity of the event study and regression discontinuity results suggests that much of the drop in housing prices in labeled districts, at least in the year before the label is received, is attributable to the changes in district finances and not likely due to foresight into label receipt.

1.6 Conclusion

This paper examines the effects of Ohio fiscal stress labeling system on school district financial behavior and housing prices from 2000-2012. I find statistically significant changes in school district finances in response to the mandated recovery plans associated with these labels. Districts increase taxes for operating in conjunction with cuts in expenditures. Labeled districts reduce both operating and capital expenditures, with the largest percentage cuts undertaken in capital expenditures. I also find that home prices fall following the receipt of a fiscal emergency label, but find no statistically significant effects following fiscal oversight. This suggests that homebuyers are responding more strongly to the uncertainty surrounding the state takeover or what the state takeover means for future academic quality. In addition to these labels, districts appear responsive to earlier

state interventions that help districts alleviate financial issues before labels become necessary.

To further inform states considering these types of policies, additional evidence is needed on the effect these policies have on students and teachers in these districts. Given the housing price results, these state financial takeovers may be sending a negative signal to residents. If parents think the district is of lower quality, they may use open enrollment policies to switch their children into more financially stable districts or move out of the district catchment area. The enrollment results suggest that these districts are losing students following the receipt of these labels. If these students are high-achieving, their exodus from the district could have negative effects on their former peers (i.e. loss of a positive peer effect) and losses of many high-achieving students could severely impact whether districts meet certain NCLB targets.

In addition, the number of full-time equivalent teachers is falling in districts following the receipt of these labels. While some of this reduction may be attributable to layoffs made in conjunction with these recovery plans, likely some of this reduction is due to teachers fleeing financially-troubled school districts. If a majority of these are high-quality teachers, overall teacher quality in these districts may be falling after label receipt and given the wide teacher value-added literature we would expect a decline in teacher quality to negatively impact student test scores. In addition to changes to student and teacher compositions, expansive cuts to programs in order to meet the requirements of financial recovery may lower the quality of educational services offered by these labeled districts, in turn lowering student achievement. Although the housing price results hint that educational quality may have been reduced, future work will need to more carefully assess how these financial recoveries are impacting students and educational quality in these districts.

While this policy has changed the financial behavior of financially troubled school districts in Ohio, how districts respond to these policies in other states is likely to depend on the structure of the state funding system. Given the institutional structures in place in Ohio, primarily that districts can raise local operating tax revenue, districts appear to respond as one would expect – increasing local taxes and making expenditure reductions to solve these budgetary issues. Districts

in other institutional settings, especially those in which local operating tax revenue is restricted, may respond differently and focus more heavily on the expenditure cuts. As these intervention systems also apply to local governments in many states, the effects of these systems on local government financial behavior is also of interest. One might expect that residents may respond more strongly to interventions into local governments than school districts, as the average resident is more likely to use public services offered by local governments than school districts. Thus, this study pushes future work to analyze the effects of financial intervention systems in other institutional settings.

CHAPTER 2

MICHIGAN AND OHIO K-12 EDUCATIONAL FINANCING SYSTEMS: EQUALITY AND EFFICIENCY

2.1 Introduction

In the current financial climate, reductions in property values and poor state fiscal conditions have led to budgetary issues for local school districts. How districts choose to respond to these financial pressures depends on the level of state funding and the different types of local taxes available to school districts. Districts that receive greater state aid will likely be sensitive to reductions in state budgets, while districts with greater ability to raise revenue locally may respond to declining state and local economic conditions through the use of voter approved tax referenda.¹ In addition, districts with the ability to tax income may be better able to respond to declines in property values by switching their tax mix away from property taxation and towards greater income taxation. Districts that are restricted from collecting local tax revenue for operating expenditures may use other avenues, such as expenditure cuts and alternative revenue sources, to alleviate these financial pressures.² The feasibility of these other avenues will clearly depend on the wealth of the districts. Preventing districts from generating local tax revenue for operating expenditures, but not imposing these restrictions on capital expenditures, may also affect how districts allocate resources across capital and operating (i.e., labor and materials) expenditures. Therefore, in addition to equality considerations, these restrictions may also have efficiency implications.

Much of the previous literature on equality in school funding has focused on law changes, usu-

¹Dye and Reschovsky (2008) find that local school districts increased property taxes by 37 cents for every dollar lost in state aid.

²Reschovsky (2004) argues that funding constraints limit the ability of districts to respond to reductions in state aid and reductions in taxable values, exacerbating the financial problems facing these districts. Evidence from California suggests that these types of restrictions are often circumvented by increases in non-restricted revenue and non-traditional funding sources, such as private donations (Brunner and Sonstalie, 2003; Brunner and Imazeki, 2005; Hoane, 2004).

ally as a result of court rulings. As expected, these studies largely find that inequality in revenues and expenditures is reduced as a result of these court-mandated reforms.³ Instead of focusing on law changes, we consider how different funding systems impact inequality.⁴ We assess the degree of inequality in per pupil revenue and expenditures across districts under two distinct funding systems – a state-level system (i.e. minimal local control) in Michigan and a foundation system (i.e. greater local control) in Ohio.⁵ Both states use different mechanisms to adjust for inequality resulting from differences in tax base size. Michigan keeps state aid roughly the same for all districts regardless of tax base size, but attempts to minimize inequality in revenue by restricting the collection of local revenue to fund operating expenditures for general education students. The Michigan system does allow residents to vote for property tax millages that finance capital projects and expenditures on special and vocational education. Ohio allows districts to raise unrestricted funds for operating expenditures through local property and income taxes, but accounts for the large inequality this creates by giving a disproportionately large amount of state aid to districts with the smallest tax bases. This paper assesses whether the degree of inequality in revenues and expenditures varies across these two states. We also examine whether the allocation of resources between capital and operating expenditures varies across the different funding systems.

Section II provides an overview of the Michigan and Ohio K-12 education financing systems. Section III first describes the data and contains descriptive statistics on tax rates, taxable values and district demographics. It then contains a comparison across different revenue sources and whether this varies across districts based on property values per pupil. Finally, Section III compares expenditures across districts and discusses whether the composition of these expenditures varies

³Murray, Evans, and Schwab (1998) and Corcoran et al. (2004) find that inequality was reduced by 19 to 34 percent follow these reforms, relative to non-reform states. Also, see Springer, Liu, and Guthrie (2009), Roy (2011), and Berry (2007). For a larger review of the state role on equity and adequacy, see Corcoran and Evans chapter in the Handbook of Research in Education Finance and Policy (2008).

⁴For an overview of different state aid formulas, see Loeb (2001), Hoxby (2001), Fernandez and Rogerson (2003), and Yinger (2004).

⁵Theoretical work by Fernandez and Rogerson (2003) suggests that the foundation system dominates the state-level system in terms of total welfare.

across the states. Section IV concludes.

2.2 Institutional Details

The Michigan and Ohio K-12 education financing systems are quite complicated and differ along several significant dimensions. Perhaps the most important difference is the restriction on local tax revenue collection for operating expenditures in Michigan. This section provides an overview of the Michigan and Ohio funding systems and discusses the differences and similarities of the two systems.

2.2.1 Michigan K-12 Finance

In response to high property taxes and an unequal distribution of school funding across districts, the Michigan K-12 finance system changed dramatically with the implementation of Proposal A in 1996. Proposal A changed the power equalization finance system to a state level system and reduced the funding obtained from local property taxes while increasing the funding from state income and sales taxes. School districts with a millage rate over 18 in fiscal year 1993 had their millage rate reduced to 18 and the new law stipulated that districts only impose this millage on non-homestead properties.^{6,7} In addition to this non-homestead millage, the state allowed 32 of the highest-spending school districts to levy a “hold harmless” millage. The state also imposed a state level property tax of six mills along with increases in the sales, tobacco product, and real estate transfer taxes. Proceeds from these taxes are deposited in the Michigan School Aid Fund and distributed to the school districts through a per-pupil grant. This grant varies depending on which school district the student resides, but does not depend on the amount of local property taxes collected. The amount of the per-pupil grant is the difference between the amount of local revenue that would be collected if the non-homestead millage is at its cap, usually 18 mills, and

⁶Homestead properties are primary resident homes while non-homestead properties are mainly businesses and rental properties.

⁷Only 13 of the 552 school districts imposed a millage of less than 18 in fiscal year 1993. For these districts, the non-homestead millage is capped at the 1993 rate.

the amount required to achieve the school district's per-pupil grant allocation. The funds from the grant are deposited into the school district's general fund to pay for labor, material, utilities and maintenance costs. Along with property tax millages to finance capital projects, local school districts can propose a "sinking fund" property tax millage, the revenues of which fund certain capital expenditures and repairs.⁸ Local school districts can also generate funds using voter-approved referenda for a recreational millage, which provide revenue for the operation of public recreation facilities and playgrounds.

Each of the 552 local school districts in Michigan belongs to one of 57 intermediate school districts (ISDs) that provide special and vocational education. Funding of ISDs did not appreciably change due to Proposal A and a significant portion of ISD funding comes from local property taxes. The variation across ISDs in regards to property tax revenue results in vastly different services being offered across the ISDs. Some of these services are offered directly by the ISD, while others are provided by the local school districts using ISD funds. An ISD can levy three types of property tax millage: operational, special education and vocational education. Voters must pass a referendum to change any of these millage rates and there are caps associated with all three tax millages.⁹ The amount of revenue that is passed on to the local school districts varies across ISDs.¹⁰ Along with these property taxes, voters may approve referenda for enhancement millages, the proceeds of which are distributed by the ISD to their local school districts on a per student basis.¹¹

There have been a few post-2002 changes to the Michigan K-12 education financing system. One noteworthy change occurred in 2008, when industrial personal property was exempted from

⁸The maximum sinking fund millage is five mills for twenty years.

⁹The operating millage rate cannot exceed 1.5 times the number of mills allocated to the ISD in 1993. The special education millage rate cannot exceed 1.75 times the number of mills allocated to the ISD in 1993. The vocational education millage rate is capped at 1 mill for those ISDs that did not levy this tax in 1993 and is capped at 1.5 times the number of mills allocated to the ISD in 1993 for all other ISDs.

¹⁰The majority of ISDs base the distribution of the revenue from the special education millage on the difference between the special education costs of the district and the amount received in state aid. Other ISDs base the distribution on average cost measures and the number of special education students.

¹¹The maximum enhancement millage is three mills for twenty years.

both the non-homestead property tax millage (which is usually 18 mills) and the state level property tax of six mills. During the same time commercial personal property was exempted from 12 of the 18 non-homestead property tax mills.¹² Since 2002, the number of students attending charter schools has increased substantially in Michigan. Charter schools are operated as nonprofit corporations that, like public schools, are provided with the state per pupil foundation allowance but are prohibited from levying taxes which results in capital expenses being paid for by the foundation allowance or independent contributions.

2.2.2 Ohio K-12 Finance

The 613 public school districts and 49 joint vocational school districts¹³ in Ohio are primarily funded through state aid and local property or income taxation. The role of the state in financing education is to ensure that each district receives the necessary funds to provide an “adequate” level of educational services to the students of the district. To do this, the state first determines the amount of per pupil expenditures that are necessary to achieve this adequate education level.¹⁴ School districts are required to raise at least 20 mills of property tax revenue in order to cover some (or all) of this adequacy amount. Currently, the state calculates the local share of the adequacy amount by assuming school districts levy 23 mills of property taxes.¹⁵ After netting out this local “charge-off,” the state provides districts with the remaining revenue needed to achieve the adequacy amount.

Ohio school districts also have the option to supplement state aid through additional property

¹²Personal property is tangible assets of a business such as computers, machinery and equipment. Real property refers to land and buildings.

¹³These vocational schools span one or more counties within the state. They provide vocational training to students from public school districts within the counties the vocational school operates. High school students from these public school districts can opt to pursue this vocational training in place of a traditional public high school education, if accepted into the vocational program.

¹⁴For documentation on the adequacy formula, see the Ohio Legislative Service Commission “School Funding Complete Resource.” (<http://www.lsc.state.oh.us/schoolfunding/edufeb2011.pdf>)

¹⁵There are a few districts that impose less than 23 mills. The state supplements these districts with enough funds to meet the revenue that would have been received had the district levied 23 mills. This additional state supplement is called gap aid.

and income taxation, subject to voter approval. Districts are able to tax both real and tangible personal property.¹⁶ Similar to Michigan, districts can issue debt through bonds for capital projects and improvements to classroom facilities. In addition to debt issuances, permanent improvement property tax levies fund short-term, at most five year, capital improvements. In contrast to Michigan, Ohio districts also have the option to propose additional taxes financing operating expenditures. Revenue for operating expenditures is generated through either current expense millages or emergency operating levies. These current expense taxes can either be property or income taxes that raise revenue over a period of five or more years.¹⁷ Emergency operating taxes collect a district-specified amount of revenue for a period of, at most, five years. In addition to taxes approved by voters, each district is allocated a set amount of property tax millage that is levied without voter approval. Districts primarily allocate the revenue from this “inside millage” towards either current expenses or permanent improvements.

There have been a few significant policy changes since 2002 that have changed the way schools are funded. In 2010, Ohio adopted an evidence-based funding formula that determined the adequacy amount based on the number and type of staff needed to provide a basic level of education. This funding formula also changed how the local share of funding is calculated. Prior to 2010, the local “charge-off” was calculated using recognized valuation.¹⁸ Starting in 2010, the local “charge-off” for districts at the 20-mill floor was calculated using total valuation, while the charge-off of all other districts continued to be calculated using recognized valuation. Beginning

¹⁶There are two classes of real property in Ohio: class I includes all real residential and agricultural property and class II includes all real commercial, industrial, mineral, and railroad property.

¹⁷Prior to 2006, Ohio school districts could levy income taxes on the traditional income tax base (adjusted gross income net personal and dependent exemptions). After 2006, districts were given the option of only taxing the earned income of residents. This earned income is not subject to personal and dependent exemptions.

¹⁸Recognized valuation spreads out the inflationary increase in real property from reappraisal over three years to prevent the state share of funding to fluctuate greatly from one year to the next. Thus, the recognized valuation of a district in the year of reappraisal is the total valuation - (2/3)*Inflationary Increase. A year after reappraisal, recognized valuation becomes the total valuation - (1/3)*Inflationary Increase. Two years after reappraisal, recognized valuation is equal to total valuation. Also, since the school fiscal year runs from July to June and the tax year runs from January to December, valuation from two years prior is used in the calculation of the local charge-off. For example, valuation from 2008 is used in the calculation for the 2009-2010 school year.

in 2006, Ohio gradually phased out taxation on business tangible personal property, with the state providing districts with funds to offset the loss in revenue resulting from this phase out.

2.2.3 Similarities Between Ohio and Michigan Financing Systems

There are several similarities between the two school funding systems. Michigan and Ohio both have school choice programs, which allow students to attend a school district even if they do not reside in that district. The state provides the attending district with additional funds to educate these students. For school districts that have capacity and elect to enroll students residing outside the district boundaries, Michigan requires the district to hold a lottery to determine which students are able to attend - with preference given to students with siblings who are already school of choice students in the district. Unlike Michigan, school districts in Ohio do not hold lotteries to determine which students are able to attend via open enrollment. Instead, the school superintendent determines which students are allowed to enroll in the school district when the number of open enrollment applications exceeds the number of slots available. School districts in Ohio also have the option to limit what students are eligible to enroll in the district through open enrollment.¹⁹

Another similarity is that both states have legislation that restricts the growth of local property taxes. The 1978 Headlee Amendment in Michigan requires that property tax rates be decreased (i.e. “rolled back”) if the growth in assessed values, excluding new construction and improvements, exceeds the growth in inflation. County, municipality, intermediate school district and school district property tax rates are “rolled back” so that the same amount of property taxes, in real terms, are collected from the old base.²⁰ For parcels that do not transfer ownership, Proposal A also restricts the growth of individual parcel property assessments to the lesser of the inflation rate and five percent. For parcels that do transfer ownership, the taxable values are assessed at 50 percent of the true cash value. Ownership transfers usually result in a significant increase in the

¹⁹Currently, 431 Ohio school districts allow any student from the state to apply for open enrollment. Of the remaining districts, 62 only accept students from adjacent districts and 120 have no open enrollment policy in place.

²⁰Millage used to retire debt, along with the 6 mill state level property tax, are not subject to these Headlee roll backs.

amount of property tax paid on these properties. Residents can “override” these Headlee rollbacks by voting on referenda that returns the tax rate back to its prior level or by voting on renewal specifying the prior tax rates.²¹

In Ohio, property taxes for current expenses (provided total current expense millage is greater than 20 mills), classroom facilities and permanent improvements are subject to property tax rollbacks. Bonds, emergency operating levies, and all inside millage are exempt from these property tax rollbacks. All real property (both class I and class II) is assessed at 35 percent of true value. Tangible personal property is assessed at a rate between 23 percent and 100 percent of true value. Similar to Michigan, these property tax rollbacks require that class I and class II property tax rates be reduced in proportion to the increase in assessed values. Tangible personal property is exempt from these rollbacks. Since changes in assessed valuation of class I and class II property differ, the rollback factor is different for both classes of real property. Therefore, often class I and class II real property are taxed at different rates, while tangible personal property is taxed at the voted millage rate.

2.3 Data and Descriptive Statistics

2.3.1 Data

We obtained school district specific variables from the National Center for Education Statistics (Local Education Agency Finance Survey and Common Core of Data), the Michigan Department of Education, the Ohio Department of Education, the Michigan Department of Treasury, and the Ohio Department of Taxation. Detailed information on school district revenues and expenditures is obtained from the Local Education Agency Finance Survey. These data consist of annual school district information from 2002 through 2010.²² These data include information on total revenues,

²¹Many districts pass referenda stipulating a “maximum” non-homestead property tax millage of over 18. While the districts are restricted to only levy 18 mills, passing this type of referenda allows a district facing a Headlee rollback to maintain an 18 millage rate without voting on another referendum.

²²Year corresponds to the school fiscal year from July 1st to June 30th. Thus, we analyze data from the 2001-2002 school year to the 2009-2010 school year.

total federal revenues, total state revenues, total local revenues, and total expenditures. These data also include a measure of total enrollment in the district, which we supplement with data on the number of general education and special education students from the Ohio and Michigan Departments of Education. School district demographic data are obtained from the NCES Common Core of Data and the Small Area Income and Poverty Estimates. These data include the number of free and reduced priced lunch students, enrollments broken down by race, number of schools in the district, and the number of school-aged children in poverty.

We also collect annual tax rates and taxable property valuations from the Michigan Department of Treasury and the Ohio Department of Taxation. The tax information provides all the taxes levied in each year from 2002 through 2010 and includes information for whether the tax funds operating or capital expenditures. We then aggregate these tax rates at the district level, which provides us with a total district tax rate to fund operating expenditures and a total tax rate to fund capital expenditures. We then multiply these two aggregate tax rates by the total taxable value in the district for each year, which gives an estimate of the total local property tax revenue collected to fund operating and capital expenditures. We also obtain a measure of the level of income taxes collected by Ohio districts for operating expenditures from the Local Education Agency Finance Survey.

2.3.2 Descriptive Statistics

Table B.1 contains the average school district property tax rates for Michigan and Ohio along with taxable value information. There are interesting differences to note across the states. While both the local school districts and the intermediate school districts in Michigan collect property taxes, the local school districts in Ohio have a much higher overall millage rate from which to generate unrestricted operating revenue. This is due to the fact that Michigan imposes a dollar for dollar “tax” on revenue generated from the non-homestead millage and, therefore, this non-homestead revenue should be and is considered state revenue in our analysis. For Ohio, the local

revenue generated to cover the adequacy amount defined by the state is considered local revenue. While we do not have district level information on these adequacy amounts, for most Ohio districts, especially the relatively wealthy districts, we expect that the current tax rates more than cover the adequacy amount and a nominal change in the tax rates will not change the revenue obtained from the state. Table B.1 also indicates that the taxable values, on which these local property taxes are imposed, are greater for Michigan. Since the average number of students in a school district is 2,774 for Michigan and 2,870 for Ohio (Table B.2), the taxable value difference between Michigan and Ohio is not the result of Ohio having larger school districts. Taxable values differ across the two states because Ohio properties are assessed at only 35 percent of true value compared with a 50 percent assessment rate in Michigan; resulting in the taxable value per pupil being over twice as large in Michigan.

Table B.2 indicates that, along with slightly larger enrollments, Ohio school districts have slightly greater total expenditures and total revenue on average than Michigan.²³ This results in almost identical total expenditures per student and total revenue per student, on average, for the two states. Total revenue is slightly greater in Ohio than Michigan because while Ohio districts average \$8.3 million (23.1-14.8) less in state revenue, they average \$10.6 million (16.8-6.22) more in local revenue than Michigan. Local revenue is comprised primarily of local property and income taxes but there are other sources of local revenue including school lunch receipts, student activity receipts, student fees and transfers from other school districts.²⁴ Greater local tax revenue by Ohio districts is expected because the larger overall millage rates more than offset the lower taxable val-

²³Using the Consumer Price Index, dollar figures in all tables and figures have been converted into 2010 dollars.

²⁴Excluding local property and income tax revenue, Ohio districts average \$2.7 million annually from other local revenue sources while Michigan districts average \$3.0 million annually. This includes revenue from other school districts (primarily payments associated with school choice programs and transfers from the ISDs), which average \$1.4 million and \$0.4 million annually for Michigan and Ohio districts, respectively. While we include these transfers from other school districts as local revenue, it may be more appropriate to classify some as state revenue.

ues.^{25,26} This greater local tax revenue collection in Ohio is mainly attributable to Ohio's current expense millage and is not appreciably affected by local income tax revenue, which averages only \$400,000 annually for Ohio school districts.

While Ohio generates more local revenue for operating expenses, Michigan generates significantly more for capital expenditures. The reason the taxation for capital expenditures differ so dramatically is in part due to the Ohio School Facilities Commission (OSFC). The OSFC is a state agency that provided funds for capital expenditures (e.g. school building construction and renovations) to the school districts. Besides the number of special education students and number of Title I schools, the averages of the other school district demographic characteristics are similar across the states.²⁷ While the difference in the number of Title I schools is surprising, the fewer number of special education students in Michigan local school districts is partly attributable to the fact that many Michigan special education students attend schools operated by the intermediate school districts.²⁸

2.3.2.1 School District Revenue

Figures B.1 through B.6 present yearly student enrollment and per pupil revenue from different sources across different quintiles based on average annual taxable property values per pupil. To compare the inequality in these various revenue sources, separate graphs are provided for Michigan

²⁵The local taxes collected as reported by the districts in the Local Education Agency Finance Survey are less than the amounts calculated using the millage rates and taxable value information. Part of this difference is attributable to the fact that when we use the millage rates and taxable value information from the Michigan Department of Treasury and the Ohio Department of Taxation, we do not take into account the various tax exemptions received by some property owners nor delinquent payments.

²⁶It is important to note that the difference in local revenue per pupil would be less if the non-homestead millage in Michigan was considered local revenue or if the local revenue generated by Ohio districts to achieve the adequacy amount was considered state revenue.

²⁷It should be noted that the number of free and reduced lunch students is not available for Ohio in 2008, but removing this variable from our regressions does very little to change the main conclusions. In addition, the median income by school district is available in the 2000 U.S. Census and 5-year estimates are available from the Census for all Ohio, and almost all Michigan school districts from 2005-2009, 2006-2010 and 2007-2011. We construct annual median income by district by interpolating this information.

²⁸In addition, the structure of special education financing in Ohio provides districts more incentive to classify a student as special education relative to Michigan.

and Ohio.^{29,30} Figure B.1 depicts how total enrollment in the different quintiles changes across years. It is interesting to note that the poorer and poorest quintiles are the smaller, more rural, school districts. While all Michigan quintiles saw an enrollment decrease between five and ten percent, enrollment of the wealthier and wealthiest quintiles in Ohio remained stable across the years while the others, especially the median quintile, experienced a significant decline.³¹ Figure B.2 presents total revenue per pupil obtained from the Local Education Agency Finance Survey. While the more wealthy quintiles in Michigan have greater total revenue per pupil than the other quintiles, this relationship does not hold for Ohio. The Ohio quintile with the highest average total revenue per pupil in 2010 is the median quintile. The dramatic increase in this quintile across years is in part attributable to its 14 percent drop in enrollment, as depicted in Figure B.1. Also note that in 2002, all Michigan quintiles, besides the poorest, had greater total revenue per pupil than their corresponding Ohio quintiles. By 2010, Ohio districts in all quintiles had total revenue per pupil that exceeded their corresponding Michigan quintiles. This is the result of total revenue per pupil declining for all Michigan quintiles from 2002 to 2010, while total revenue per pupil for Ohio quintiles has increased over the period. For several Michigan quintiles, the largest decrease in total revenue per pupil occurred in the last year or last several years of our data when the United States experienced a significant recession. Interestingly, total revenue per pupil for most Ohio quintiles increased during this Great Recession, suggesting that perhaps Ohio districts are better able to respond to economic downturns because of their ability to generate local revenue for operating expenditures. Thus, the structure of the state school financing system may greatly influence how economic downturns impact school districts.

²⁹The wealthiest quintile are the 20 percent of school districts with the largest taxable values per pupil, wealthier quintile are the districts from 20 percent to 40 percent, median quintile are the districts from 40 percent to 60 percent, poorer quintile are the districts from 60 percent to 80 percent, and poorest quintile are the 20 percent of school districts with the smallest taxable values per pupil.

³⁰The quintiles are based on the average annual taxable property values per pupil so that a district remains in a given quintile across all years. To ensure that the composition of each quintile does not change across years, the seven districts with at least one year of missing information are dropped (48 district-year observations).

³¹Because Detroit and other districts around Detroit are dropped due to missing at least one year of data, the actual decline in enrollment for the poorest Michigan quintile is likely larger than depicted in Figure B.1.

As demonstrated in panel (a) of Figure B.3, Michigan's slight decrease in total revenue per pupil is primarily attributable to a decrease in state revenue. Given that Michigan restricts how much revenue can be raised locally to fund operating expenditures, it is not unexpected that Michigan districts usually receive more in state revenue per pupil than comparable Ohio districts. Interestingly, Michigan distributes state revenue relatively evenly across school districts, but districts in the wealthiest quintile receive nearly \$600 more in state revenue per pupil than districts in the other quintiles. This distribution of state aid, which is provided irrespective of district wealth, is consistent with what would be expected given the framework of the Michigan funding laws, which targets inequality by restricting collection of local property taxes. In contrast, Ohio attempts to address inequality by allocating significantly more state revenue to less wealthy districts. Districts in the wealthiest quintile in Ohio receive between \$3,000 and \$5,800 less state revenue per pupil than districts in the poorest quintile, but this difference has decreased across years.

Unlike Ohio, which addresses inequality through disproportionate state aid to the poorest districts, Michigan addresses inequality by placing restrictions on the use of local property taxes for operating expenditures. These restrictions explain the large differences in local revenue we observe in Figure B.4 between Ohio and Michigan districts in all taxable value quintiles. Michigan districts obtain approximately two and a half times less local revenue than similar districts in Ohio. Due to unrestricted property and income taxation at the local level in Ohio, the gap in local revenue between the richest and poorest districts is much larger in Ohio than in Michigan. While the wealthiest Michigan school districts raise approximately \$2,000 per pupil more in local revenue than the poorest districts, the wealthiest districts in Ohio raise \$5,000 more per pupil than the poorest districts. This disparity between the top and bottom is so large in Ohio that even though the state gives disproportionately more state revenue to the relatively poor districts it is not enough to offset this \$5,000 per pupil gap.

In addition to the disparity in total local revenue, the restrictions placed on taxes for operating expenditures in Michigan may result in different mixes of taxes for capital and operating expen-

ditures across the two states. Figure B.5 focuses strictly on local operating tax revenue, which is generated exclusively through property taxes in Michigan, while Ohio districts generate local operating revenue through property and income taxes. The minimal local property tax revenue collected by Michigan districts for operating expenditures is obtained primarily from the hold harmless millage of wealthy districts and the special education millage of the intermediate school districts. In contrast to Michigan, a significant portion of local revenue in Ohio is generated from local taxes and almost all is unrestricted, as state revenue is often used towards funding special education. As expected based on Figure B.4, the amount of local operating tax revenue generated is greater in the relatively wealthy districts and significantly greater in the wealthiest districts. In Ohio, a relatively small fraction of local operating revenue is from income taxes but use of these taxes has increased since 2006 – especially in poorer, agricultural districts that have low taxable value bases, but relatively higher taxable income bases.³²

As state revenue has decreased and local revenue for operating expenditures has largely remained constant from 2002 to 2010, property tax revenue for capital expenditures has increased in Michigan. As depicted in Figure B.6, the Michigan quintiles increased property tax revenue for capital expenditures by between \$200 and \$300 per pupil, with the largest increase occurring in the wealthiest quintile. Ohio districts in all quintiles collect less tax revenue for capital expenditures than districts in the corresponding Michigan quintiles. For the median, poorer, and poorest quintiles, this difference is partly attributable to the Ohio School Facilities Commission (OSFC), which designated \$7 billion for capital expenditures to, primarily, the relatively poor school districts between 2002 and 2010.³³ The OSFC made these capital funds available to the least wealthy districts

³²As found in Spry (2005) and Hall and Ross (2010), a majority of districts using these income taxes are districts in rural areas. Ross and Nguyen-Hoang (2013) show that Ohio districts use the income tax as a supplement to property taxation.

³³The decision to use a significant portion of the tobacco settlement funds to subsidize school capital investments was, in part, a response to a General Accounting Office report entitled “School Facilities: Profiles of School Condition by State.” The 1996 report summarized results from a national survey of school buildings and concluded that the physical condition of school buildings in Ohio was worse than the school building condition in, if not all, almost all other states. For example, the report indicates that 61 percent of the 3,600 Ohio schools surveyed indicate at least one on-site building in inadequate condition compared to only 34 percent for the 3,325 Michigan schools surveyed.

first by ranking the districts based on a weighted average of taxable value per pupil and median income.³⁴ This capital expenditure subsidy from the OSFC is likely to have crowded out some capital expenditures by the school districts. Interestingly, the wealthier and wealthiest quintiles in Ohio, mainly comprised of school districts that did not have access to these OSFC funds, also obtain significantly less tax revenue for capital expenditures than comparably wealthy Michigan school districts. Perhaps these wealthy Ohio districts preferred to have local tax revenue fund operating expenditures rather than capital expenditures. Because of the constraints Proposal A places on raising local operating revenue, the wealthy Michigan districts do not have this choice.

Although the above figures give some sense of the inequality of revenue in both states, these figures do not account for differences in district size and demographics that may contribute to differences in revenue across the two states. To better account for these other factors we estimate the following regression equation:

$$\ln(y_{st}) = \beta_0 \ln(\text{TaxableValue}_{st}) + X_{st}\beta + \lambda_s + \theta_t + \varepsilon_{st} \quad (2.1)$$

where $\ln(y_{st})$ is the natural log of either total revenue, state revenue, or local revenue of district s in year t ; $\ln(\text{TaxableValue}_{st})$ is the total taxable value for the district in year t ; X_{st} is a vector of school district demographics that includes median income, total and special education enrollment counts, the number of white students, the number of students eligible for free/reduced priced lunch, the total number of schools and Title I schools, total district population and total school-aged population in poverty; λ_s is a vector of school district fixed effects; θ_t is a vector of year fixed

³⁴The OSFC began offering funds to the most needy school districts in 1997 and by 2010 funds had worked up to the 450th spot in the rankings. The state has contributed a certain percentage of the capital funds and this percentage depends on the district rankings with poorer districts having lower rankings receiving a greater percentage. For most districts, the state's percent contribution is one minus the ranking list percentile. For example, a district ranked 61 out of the 613 school districts would be at the tenth percentile and the state contribution would be 90 percent of the total project cost. Most districts raise their share by passing a referendum that generates property taxes so that the district can sell bonds. A few districts use existing property taxes, cash on hand and voluntary contributions to raise the local share. In recent years especially, many districts chose not to pursue the available funding or were unable to pass referenda to fund the district's portion of the costs.

effects; and ε_{st} is an idiosyncratic error term. Tables B.3 and B.4 contain estimates when school district fixed effects are not and are included in the specification, respectively.

The estimates from the specification without fixed effects uses primarily across school district variation to identify the relationship between taxable value and revenue. These estimates in Table B.3 suggest that a ten percent increase in total taxable value is associated with approximately a 1.3 percent increase in total revenue for both states, while a ten percent increase in median income is associated with a 2.12 percent increase in total revenue for Michigan and a negligible change for Ohio. When median income is not included as a covariate, the coefficient on taxable value when total revenue is the dependent variable increases to 0.146 for Michigan and to 0.139 for Ohio. As expected based on the graphs in Figure B.3, the coefficient estimates associated with taxable value indicate that state revenue is larger in Michigan districts with greater taxable values. In Ohio, however, state revenue is significantly less in high taxable value school districts compared to low taxable value districts, since the state funding laws give these low taxable value districts disproportionately more state aid. In terms of local revenue, the coefficient estimates indicate that low taxable value districts obtain less in local revenue than high taxable value districts and this difference is greater in Ohio. The results in Table B.3, along with the greater variance in taxable value and median income for Michigan (see Tables B.1 and B.2), suggest that there is greater inequality in Michigan than in Ohio.

The estimates from the fixed effects specifications in Table B.4 use within-school district, across-year variation to identify the relationship between changes in taxable value and changes in revenue. These estimates indicate that an increase in taxable value is associated with a significant increase in state, local and total revenues for Michigan school districts. For Ohio school districts, an increase in taxable value is associated with an increase in local revenue, a decrease in state revenue, and a slight increase in total revenue. Table B.4 also suggests that changes in median income are associated with relatively small increases in revenue for Michigan school districts and

a negligible change in revenue for Ohio districts.³⁵

The graphs in Figures B.2 and B.3 suggest that total and state revenues decreased for Michigan and increased for Ohio school districts during the Great Recession. It is also interesting to consider how this recession affected taxable values, inequality and the relationship between taxable value and revenue. Due in part to differences in assessment procedures, Ohio's assessed values started to decrease in 2007 (affecting school district's 2007-8 taxable values) while Michigan did not experience this decline in taxable values until the 2009-10 school year (denoted as year 2010). The taxable value per pupil decreased less than taxable values in both states because of the decline in the number of students. However, the variance of the taxable value per pupil in both states was much larger after 2007 compared to the prior years. In addition, based on the estimates in Table B.5, the relationship between total taxable value and revenue does appear to change slightly as the result of the Great Recession. Table B.5 estimates similar specifications as in Tables B.3 and B.4 except an interaction term, $\ln(\text{Total Taxable Value}) * \text{Great Recession indicator}$, is included as a covariate to allow for the relationship between taxable value and revenue from 2002 to 2007 to differ from this relationship from 2008 to 2010. The coefficient estimates associated with this interaction term indicate that the positive relationships between taxable value and total/state revenues are stronger during the Great Recession for Michigan school districts but not for Ohio school districts. While these positive coefficient estimates are statistically significant in the Michigan regressions, the magnitudes suggest a relatively small increase in these positive relationships during the Great Recession.

The results in Tables B.3 and B.4 suggest that the Ohio funding system leads to more equitable funding than the Michigan system. Targeting equality through disproportionate state aid to the poorest districts appears to be more effective at promoting equality than imposing constraints on what local revenue can be used to fund. A problem with the Michigan system is that even with

³⁵Because median income is not available annually by district (requiring us to interpolate when constructing these annual measures - see footnote 27), the coefficient estimates pertaining to median income should be viewed with caution when using within-district, across-year variation for identification.

constraints that limit how much local revenue can be used for operating expenditures, local revenue is greater in high taxable value districts and does not offset the fact that state aid is not targeted to relatively poor districts. These constraints may also hinder the ability of Michigan school districts to react to economic downturns, exacerbating this inequality. While wealthy districts in Michigan may be able to counteract these downturns with funding from other local sources, the relatively poor districts in Michigan do not get much funding from other sources, making them the least able to respond to these downturns. The Table B.5 results suggest that the inequality in Michigan may be growing during the Great Recession. Another problem is that these constraints could lead to inefficiencies in terms of how local school districts allocate resources between capital, labor and materials.

2.3.2.2 School District Expenditures

Figure B.7 demonstrates that the pattern in total expenditures per pupil is quite similar to total revenue per pupil both in terms of level and across quintiles. As with total revenue per pupil, total expenditures per pupil on average are similar across the states with expenditures in Michigan decreasing slightly across years while increasing in Ohio. Also similar to total revenue per pupil, Michigan's decrease in total expenditure per pupil was more prevalent during the Great Recession while Ohio districts experienced an increase during this time period. The main difference in per pupil revenue and expenditures is that there are a number of quintile-years where total expenditures per pupil are slightly greater than total revenue per pupil. This is most often the case for the more wealthy quintiles in the earlier years of our data. Perhaps more wealthy Ohio and Michigan districts were able to draw on reserve funds to supplement expenditure beyond what they raise in yearly revenue and that these reserve funds were less available in later years.

Figure B.7 does not provide details on the composition of these expenditures – specifically capital versus operating. How capital expenditures of Ohio school districts compare to those in Michigan school districts depends on whether the comparison is between relatively poor or rela-

tively wealthy districts. Defining relatively poor districts in Ohio as those below the median OSFC ranking and in Michigan as below the median taxable value per pupil, the average annual capital expenditure per pupil is \$1,922 in the relatively poor Ohio districts and \$999 in the relatively poor Michigan districts. Interestingly, over half of these Ohio expenditures are state funds distributed by the OSFC. In terms of the relatively wealthy districts (i.e. districts above the medians), annual capital expenditures per pupil are \$1,027 for Ohio and \$1,321 for Michigan.³⁶ Not only do Michigan's relatively wealthy districts spend 29 percent more in capital expenditures per pupil, over ten percent of the capital expenditures made by Ohio's relatively wealthy districts are state funds from the OSFC.³⁷

In terms of operating expenditures, there are difficulties in making a comparison because the structure of special and vocational education, along with input prices, differs across Ohio and Michigan. However, it is interesting to note that the student-teacher ratio in Ohio averaged 16.88 from 2002 to 2010 while averaging 18.64 in Michigan.³⁸ In summary, these descriptive statistics suggest that while Michigan districts spend more per pupil on capital expenditures than Ohio districts not eligible for state capital funds, Ohio districts spend a larger fraction on teachers. The constraints on Michigan districts that limit how much local revenue can be raised to pay for operating expenditures may result in an inefficient resource allocation between capital, labor and materials.

³⁶A possible explanation for these greater apparent capital expenditures is that wealthy Michigan districts may be able to circumvent the Proposal A restrictions on supplementing operating expenditures by classifying some ambiguous expenditures as capital. While wealthy Michigan districts do attempt to circumvent these restrictions, this is unlikely the reason for the significant difference in capital expenditures as well as capital taxation. As for capital expenditures, over 72 percent is accounted for by new construction in both Ohio and Michigan. It is likely difficult to supplement operating expenditures with expenditures on new construction. Expenditures on new construction is 28 percent more per pupil in relatively wealthy Michigan districts than in relatively wealthy Ohio districts. As for capital taxation, we include classroom facility millages in Ohio as capital taxation while it is clear that much of this revenue would be classified as operating expenditures in Michigan districts. If Michigan districts are circumventing Proposal A, including Ohio tax revenue from classroom facility millages as capital revenue in Figures B.5 and B.6 should make the comparison across states more comparable.

³⁷Stone (2014) documents the dramatic increase in capital spending by Michigan school districts immediately after the implementation of Proposal A and how this increase was larger for wealthier school districts.

³⁸Even after adjusting for possible differences in special education expenditures by local school districts, the student-teacher ratio is significantly higher in Michigan than in Ohio.

2.4 Conclusion

This paper considers equality and efficiency issues of two different school funding systems - a state-level system and a foundation system. The state-level system in Michigan provides all districts with nearly the same amount of state revenue for general operating expenditures and places restrictions on the ability of districts to raise operating revenue through local property taxes. Most notably, Michigan districts have restrictions on raising additional property tax revenue to fund operating expenditure for general education students but do not have these restrictions on capital expenditures. In contrast, Ohio districts are able to raise unrestricted property and income taxes to fund operating and capital expenditures. To account for the large differences in local tax revenue generated by poor versus wealthy districts, Ohio allocates state aid based on the district tax base size.

Our results indicate that, while the average revenue per pupil and expenditure per pupil of Michigan and Ohio school districts are almost identical between 2002 and 2010, they have slightly decreased over this period for Michigan and increased for Ohio. We also find that Michigan districts receive a significantly larger proportion of total revenue from the state while Ohio districts receive a larger proportion from local property and income taxes. In terms of degree of equality, measured by how revenue and expenditures vary across districts based on taxable value per pupil, there is less variation across Ohio districts. This suggests that the Ohio funding system leads to greater equality than the Michigan funding system. In terms of the distribution of expenditures, we find that wealthy Michigan districts spend more per pupil on capital expenditures while wealthy Ohio districts spend more per pupil on labor and materials. This suggests that the constraints on raising local revenue to fund operating expenditures in Michigan could create efficiency issues.

Despite Ohio and Michigan spending similar amounts per pupil on average from 2002 to 2010, there are noticeable differences in student achievement on the National Assessment of Educational Progress (NAEP) 2003, 2005, 2007 and 2009 math and reading exams (Source: <http://nces.ed.gov/nationsreportcard/states/>). Ohio performs better on these exams than Michigan irrespective of

grade and year.³⁹ Across all states, Ohio most often ranks between 12 and 18 in exam performance while Michigan most often ranks between 30 and 37. Perhaps the efficiency issues associated with the Michigan school funding system are contributing to the difference in student performance across the states. Ohio not only outperforms Michigan on these NAEP exams, but the difference in performance has increased across years. While Ohio's ranking has slightly improved from 2003 and 2009 for almost all grades and subjects, Michigan's ranking has steadily declined across years. Perhaps the reductions in revenue and expenditures have contributed to the decline in student performance in Michigan.

³⁹The NAEP math and reading exams in Michigan and Ohio are taken by fourth and eighth graders.

CHAPTER 3

AN EVALUATION OF EMPIRICAL BAYES' ESTIMATION OF VALUE-ADDED TEACHER PERFORMANCE

3.1 Introduction

Empirical Bayes' (EB) estimation of teacher effects has gained recent popularity in the value-added research community (see, for example, McCaffrey et al. 2004; Kane & Staiger 2008; Chetty, Friedman, & Rockoff 2011; Corcoran, Jennings, & Beveridge 2011; and Jacob & Lefgren 2005, 2008). Researchers motivate the use of EB estimation as a way to decrease classification error of teachers, especially when limited data is available to compute value-added estimates. When there are only a small number of students per teacher, teacher value-added estimates can be very noisy. EB estimates of teacher value-added reduce the variability of the estimates by shrinking them toward the average estimated teacher effect in the sample and, therefore, are often referred to as “shrinkage estimators.” As the degree of shrinkage depends on class size, estimates for teachers with smaller class sizes are more affected, potentially helping with the misclassification of these teachers. EB estimation may also be less computationally demanding than methods that view the teacher effects as fixed parameters to estimate. Finally, EB estimation has been motivated as a way to estimate teacher value added when including controls for peer effects and other classroom-level covariates.

This paper analyzes the performance of EB estimation using both simulated and real student achievement data. We first provide a detailed theoretical derivation of the EB estimator, which has not previously been explicitly derived in the teacher value-added literature. This theoretical discussion also provides the basis for our hypotheses for how EB and other value-added estimators are expected to perform under the different simulation scenarios we examine. We test our theoretical predictions by comparing the performance of EB estimators to estimators that treat the

teacher effect as fixed. We first use a simulation, where the true teacher effect is known, comparing performance under random teacher assignment and various nonrandom assignment simulation scenarios. In addition to the random vs. fixed teacher effects comparison, we also examine whether shrinking the estimates improves performance. Finally, we apply these estimators to real student achievement data to see how the rankings of teachers vary across these estimators in practice.

Despite the potential benefits of EB estimation, we find that the estimated teacher effects can suffer from severe bias under nonrandom teacher assignment. By treating the teacher effects as random, EB estimation assumes that teacher assignment is uncorrelated with factors that predict student achievement – including observed factors such as past test scores. While the bias (technically, the inconsistency) disappears as the number of students per teacher increases – because the EB estimates converges to the so-called fixed effects estimates – the bias still can be important for the kinds of data used to estimate teacher VAMs. This is because the EB estimators of the coefficients on the covariates in the model are inconsistent for fixed class sizes as the number of classrooms grows. By contrast, estimators that include the teacher assignment indicators along with the covariates in a multiple regression analysis are consistent (as the number of classrooms grows) for the coefficients on the covariates. This generally leads to less bias in the estimated teacher VAMs under nonrandom assignment without many students per teacher.

The paper begins in Section 2 with a detailed theoretical derivation of the EB estimator and is followed in Section 3 by a description of the five estimators we examine. Section 4 describes our simulation design and the different student grouping and teacher assignment scenarios we examine, with Section 5 providing the results of this analysis. Section 6 provides an analysis of these estimators using real student achievement data and Section 7 concludes.

3.2 Empirical Bayes' Estimation

There are several ways to derive Empirical Bayes' estimators of teacher value added. We adopt a so-called “mixed estimation” approach, as in Ballou, Sanders, and Wright (2004), because it is

fairly straightforward and does not require delving into Bayesian estimation methods. Our focus is on estimating teacher effects grade by grade. Therefore, we assume either that we have a single cross section or multiple cohorts of students for each teacher. We do not include cohort effects; multiple cohorts are allowed by pooling students across cohorts for each teacher.

Let y_i denote a measure of achievement for student i , randomly drawn from the population. This measure could be a test score or a gain score (i.e. current score minus lagged score). Suppose there are G teachers and the teacher effects are b_g , $g = 1, \dots, G$. In the mixed effects setting, these are treated as random variables as opposed to fixed population parameters. Viewing the b_g as random variables independent of other observable factors affecting test scores has consequences for the properties of EB estimators.

Typically VAMs are estimated controlling for other factors, which we denote by a row vector \mathbf{x}_i . These factors include prior test scores (with the exception of when a gain score used as dependent variable) and, in some cases, student-level and/or classroom-level covariates. We assume the coefficients on these covariates are fixed parameters. We can write a mixed effects linear model as

$$y_i = \mathbf{x}_i\boldsymbol{\gamma} + \mathbf{z}_i\mathbf{b} + u_i, \tag{1}$$

where \mathbf{z}_i is a $1 \times G$ row vector of teacher assignment dummies, \mathbf{b} is the $G \times 1$ vector of teacher effects, and u_i contains the unobserved student-specific effects. Because a student is assigned to one and only one teacher, $z_{i1} + z_{i2} + \dots + z_{iG} = 1$. Equation (1) is an example of a “mixed model” because it includes the usual fixed population parameters $\boldsymbol{\gamma}$ and the random coefficients \mathbf{b} . Even if there are no covariates, \mathbf{x}_i typically includes an intercept. If $\mathbf{x}_i\boldsymbol{\gamma}$ is only a constant, so $\mathbf{x}_i\boldsymbol{\gamma} = \gamma$, then γ is the average teacher effect and we can then take $E(\mathbf{b}) = \mathbf{0}$. This means that b_g is the effect of teacher g net of the overall mean teacher effect.

Equation (1) is written for a particular student i so that teacher assignment is determined by the

vector \mathbf{z}_i . A standard assumption is that, conditional on \mathbf{b} , (1) represents a linear conditional mean:

$$E(y_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{b}) = \mathbf{x}_i\boldsymbol{\gamma} + \mathbf{z}_i\mathbf{b} \quad (2)$$

which follows from

$$E(u_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{b}) = 0. \quad (3)$$

An important implication of (3) is that u_i is uncorrelated with \mathbf{z}_i , so that teacher assignment is not systematically related to unobserved student characteristics once we have controlled for the observed factors in \mathbf{x}_i .

If we assume a sample of N students assigned to one of G teachers we can write (1) in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{b} + \mathbf{u}, \quad (4)$$

where \mathbf{y} and \mathbf{u} are $N \times 1$, \mathbf{X} is $N \times K$, and \mathbf{Z} is $N \times G$. In order to obtain the best linear unbiased estimator (BLUE) of $\boldsymbol{\gamma}$ and the best linear unbiased predictor (BLUP) of \mathbf{b} , we assume that the covariates and teacher assignments satisfy a strict exogeneity assumption:

$$E(u_i|\mathbf{X}, \mathbf{Z}, \mathbf{b}) = 0, i = 1, \dots, N. \quad (5)$$

An implication of assumption (5) is that inputs and teacher assignment of *other* students does not affect the outcome of student i .

Given assumption (5) we can write the conditional expectation of \mathbf{y} as

$$E(\mathbf{y}|\mathbf{X}, \mathbf{Z}, \mathbf{b}) = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{b} \quad (6)$$

In the EB literature a standard assumption is

$$\mathbf{b} \text{ is independent of } (\mathbf{X}, \mathbf{Z}), \quad (7)$$

in which case

$$E(\mathbf{y}|\mathbf{X}, \mathbf{Z}) = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}E(\mathbf{b}|\mathbf{X}, \mathbf{Z}) = \mathbf{X}\boldsymbol{\gamma} = E(\mathbf{y}|\mathbf{X}) \quad (8)$$

because $E(\mathbf{b}|\mathbf{X}, \mathbf{Z}) = E(\mathbf{b}) = \mathbf{0}$. Assumption (7) has the implication that teacher assignment for student i does not depend on the quality of the teacher (as measured by the b_g).

From an econometric perspective, equation (8) means that $\boldsymbol{\gamma}$ can be estimated in an unbiased way by OLS regression of

$$y_i \text{ on } \mathbf{x}_i, \quad i = 1, \dots, N. \quad (9)$$

Consequently, we can estimate the effects of the covariates \mathbf{x}_i by omitting the teacher assignment dummies. Practically, this means we are assuming teacher assignment is uncorrelated with the covariates \mathbf{x}_i .

Under (5) and (7), the OLS estimator of $\boldsymbol{\gamma}$ is unbiased and consistent, but it is inefficient if we impose the standard classical linear model assumptions on \mathbf{u} . In particular, if

$$\text{Var}(\mathbf{u}|\mathbf{X}, \mathbf{Z}, \mathbf{b}) = \text{Var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}_N \quad (10)$$

then

$$\begin{aligned} \text{Var}(\mathbf{y}|\mathbf{X}, \mathbf{Z}) &= E[(\mathbf{Z}\mathbf{b} + \mathbf{u})(\mathbf{Z}\mathbf{b} + \mathbf{u})'|\mathbf{X}, \mathbf{Z}] \\ &= \mathbf{Z}\text{Var}(\mathbf{b})\mathbf{Z}' + \text{Var}(\mathbf{u}) = \sigma_b^2 \mathbf{Z}\mathbf{Z}' + \sigma_u^2 \mathbf{I}_N, \end{aligned}$$

where we also add the standard assumption that the elements of \mathbf{b} are uncorrelated

$$\text{Var}(\mathbf{b}) = \sigma_b^2 \mathbf{I}_G, \quad (11)$$

and σ_b^2 is the variance of the teacher effects, b_g .

Under the assumption that σ_b^2 and σ_u^2 are known – actually, it suffices to know their ratio – the BLUE of γ under the preceding assumptions is the generalized least squares (GLS) estimator,

$$\gamma^* = [\mathbf{X}'(\sigma_b^2 \mathbf{Z}\mathbf{Z}' + \sigma_u^2 \mathbf{I}_N)^{-1} \mathbf{X}]^{-1} \mathbf{X}'(\sigma_b^2 \mathbf{Z}\mathbf{Z}' + \sigma_u^2 \mathbf{I}_N)^{-1} \mathbf{y}. \quad (12)$$

The $N \times N$ matrix $\mathbf{Z}\mathbf{Z}'$ is a block diagonal matrix with G blocks, where block g is an $N_g \times N_g$ matrix of ones and N_g is the number of students taught by teacher g . The GLS estimator γ^* is the well-known “random effects” (RE) estimator popular from panel data and cluster sample analysis. Note that the “random effects” in this case are teacher effects, not student-specific effects.

Before we discuss γ^* further, as well as estimation of \mathbf{b} , it is helpful to write down the mixed effects model in perhaps a more common form. After students have been designated to classrooms, we can write y_{gi} as the outcome for student i in class g , and similarly for \mathbf{x}_{gi} and u_{gi} . Then, for classroom g , we have

$$y_{gi} = \mathbf{x}_{gi}\gamma + b_g + u_{gi} \equiv \mathbf{x}_{gi}\gamma + r_{gi}, \quad i = 1, \dots, N_g, \quad (13)$$

where $r_{gi} \equiv b_g + u_{gi}$ is the composite error term. Equation (13) makes it easy to see that the BLUE of γ is the random effects estimator. It also highlights the assumption that b_g is independent of the covariates \mathbf{x}_{gi} . Further, the assumption $E(u_{gi} | \mathbf{X}_g, b_g) = 0$ implies that covariates from student h do not affect the outcome of student i . We can also see that OLS pooled across i and g is unbiased for γ because we are assuming $E(b_g | \mathbf{X}_g) = 0$.

As shown in, say, Ballou, Sanders, and Wright (2004), the BLUP of \mathbf{b} under assumptions (5),

(7), and (10) is

$$\mathbf{b}^* = (\mathbf{Z}'\mathbf{Z} + \rho\mathbf{I}_G)^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\gamma}^*) \equiv (\mathbf{Z}'\mathbf{Z} + \rho\mathbf{I}_G)^{-1}\mathbf{Z}'\mathbf{r}^*, \quad (14)$$

where $\rho = \sigma_u^2/\sigma_b^2$ and $\mathbf{r}^* = \mathbf{y} - \mathbf{X}\boldsymbol{\gamma}^*$ is the vector of residuals. Straightforward matrix algebra shows each b_g^* can be expressed as

$$\begin{aligned} b_g^* &= (N_g + \rho)^{-1} \sum_{i=1}^{N_g} r_{gi}^* = \left(\frac{N_g}{N_g + \rho} \right) \bar{r}_g^* \\ &= \left(\frac{\sigma_b^2}{\sigma_b^2 + (\sigma_u^2/N_g)} \right) \bar{r}_g^* = \left(\frac{\sigma_b^2}{\sigma_b^2 + (\sigma_u^2/N_g)} \right) (\bar{y}_g - \bar{\mathbf{x}}_g\boldsymbol{\gamma}^*), \end{aligned} \quad (15)$$

where

$$\bar{r}_g^* = N_g^{-1} \sum_{i=1}^{N_g} r_{gi}^* = \bar{y}_g - \bar{\mathbf{x}}_g\boldsymbol{\gamma}^* \quad (16)$$

is the average of the residuals $r_{gi}^* = y_{gi} - \mathbf{x}_{gi}\boldsymbol{\gamma}^*$ within classroom g .

To operationalize $\boldsymbol{\gamma}^*$ and b_g^* we must replace σ_b^2 and σ_u^2 with estimates. There are different ways to obtain estimates depending on whether one uses OLS residuals after an initial estimation or a joint estimation method. With the composite error defined as $r_{gi} = b_g + u_{gi}$ we can write $\sigma_r^2 = \sigma_b^2 + \sigma_u^2$. An estimator of σ_r^2 can be obtained from the usual sum of squared residuals from the OLS regression

$$y_{gi} \text{ on } \mathbf{x}_{gi}, \quad i = 1, \dots, N_g, \quad g = 1, \dots, G. \quad (17)$$

Call the residuals \tilde{r}_{gi} . Then a consistent estimator is

$$\tilde{\sigma}_r^2 = \frac{1}{(N - K)} \sum_{g=1}^G \sum_{i=1}^{N_g} \tilde{r}_{gi}^2, \quad (18)$$

which is just the usual degrees-of-freedom (df) adjusted error variance estimator from OLS.

To estimate σ_u^2 , write

$$r_{gi} - \bar{r}_g = u_{gi} - \bar{u}_g,$$

where \bar{r}_g is the within-teacher average, and similarly for \bar{u}_g . A standard result on demeaning a set of uncorrelated random variables with the same variance gives $\text{Var}(u_{gi} - \bar{u}_g) = \sigma_u^2(1 - N_g^{-1})$ and so, for each g , $E \left[\sum_{i=1}^{N_g} (r_{gi} - \bar{r}_g)^2 \right] = \sigma_u^2(N_g - 1)$. When we sum across teachers it follows that

$$\frac{1}{(N - G)} \sum_{g=1}^G \sum_{i=1}^{N_g} (r_{gi} - \bar{r}_g)^2 \quad (19)$$

has expected value σ_u^2 . To turn (19) into an estimator we can replace the r_{gi} with the OLS residuals, as before, \tilde{r}_{gi} , from the regression in (17). The estimator based on the OLS residuals is

$$\tilde{\sigma}_u^2 = \frac{1}{(N - G)} \sum_{g=1}^G \sum_{i=1}^{N_g} (\tilde{r}_{gi} - \bar{\tilde{r}}_g)^2. \quad (20)$$

With fixed class sizes and G getting large, the estimator that uses N in place of $N - G$ is not consistent. Therefore, we prefer the estimator in equation (20), as it should have less bias in applications where G/N is not small. With many students per teacher the difference should be minor. We could also use $N - G - K$ as a further df adjustment, but subtracting off K does not affect the consistency.

Given $\tilde{\sigma}_r^2$ and $\tilde{\sigma}_u^2$ we can estimate σ_b^2 as

$$\tilde{\sigma}_b^2 = \tilde{\sigma}_r^2 - \tilde{\sigma}_u^2. \quad (21)$$

In any particular data set – especially if the data have been generated to violate the standard assumptions listed above – there is no guarantee that expression (21) is nonnegative. A simple solution to this problem (and one used in software packages, such as Stata) is to set $\tilde{\sigma}_b^2 = 0$ whenever $\tilde{\sigma}_r^2 < \tilde{\sigma}_u^2$. In order to ensure this happens infrequently with multiple cohorts, we compute $\tilde{\sigma}_u^2$ by replacing $\bar{\tilde{r}}_g$ with the average obtained for the particular cohort. This ensures that, for a given cohort, the terms $\sum_{i=1}^{N_g} (\tilde{r}_{gi} - \bar{\tilde{r}}_g)^2$ are as small as possible. In theory, if there are no cohort effects

we could use an overall cohort mean for \bar{r}_g . But using cohort-specific means reduces the problem of negative $\tilde{\sigma}_b^2$ when the model is misspecified.

An appealing alternative is to estimate σ_b^2 and σ_u^2 jointly along with γ , using software that ensures nonnegativity of the variance estimates. The most common approach to doing so is to assume joint normality of the teacher effects, b_g , and the student effects, u_{gi} , across all g and i – along with the previous assumptions. One important point is that the resulting estimators are consistent even without the normality assumption; so, technically, we can think of them as “quasi-” maximum likelihood estimators. The maximum likelihood estimator of σ_u^2 has the same form as in equation (20), except the residuals are based on the MLE of γ rather than the OLS estimator. A similar comment holds for the MLE of σ_b^2 (if we do not constrain it to be nonnegative). See, for example, Hsiao (2003, Section 3.3.3).

Unlike the GLS estimator of γ , the feasible GLS (FGLS) estimator is no longer unbiased (even under assumptions (5) and (7)), and so we must rely on asymptotic theory. In the current context, the estimator is consistent and asymptotically normal provided $G \rightarrow \infty$ with N_g fixed. In simulations, Hansen (2007) shows that the asymptotic properties work well when G and N_g are roughly around 40. In practice, this means that the number of teachers, G , should be substantially larger than the number of students per teacher, N_g . Typically this is the case in VAM studies, which are applied to large school districts or entire states and therefore include many teachers. Often the number of students per teacher is fewer than 100 with several hundred or even several thousand teachers.

When γ^* is replaced with the FGLS estimator and the variances σ_b^2 and σ_u^2 are replaced with estimators, the EB estimator of \mathbf{b} is no longer a BLUP. Nevertheless, we use the same formula as in (15) for operationalizing the BLUPs. Conveniently, certain statistical packages – such as Stata 12 with its “xtmixed” command – allow one to recover the operationalized BLUPs after maximum likelihood estimation. When we use the (quasi-) MLEs to obtain the b_g^* we obtain what are typically called the Empirical Bayes’ estimates.

One way to understand the shrinkage nature of b_g^* is to compare it with the estimator obtained by treating the teacher effects as fixed parameters. Let $\hat{\gamma}$ and $\hat{\beta}$ be the OLS estimators from the regression

$$y_i \text{ on } \mathbf{x}_i, \mathbf{z}_i, i = 1, \dots, N. \quad (22)$$

Then $\hat{\gamma}$ is the so-called “fixed effects” (FE) estimator obtained by a regression of y_i on the controls in \mathbf{x}_i and the teacher assignment dummies in \mathbf{z}_i . In the context of the model

$$y = \mathbf{X}\gamma + \mathbf{Z}\beta + \mathbf{u} \quad (23)$$

$$E(\mathbf{u}|\mathbf{X}, \mathbf{Z}) = \mathbf{0}, \text{Var}(\mathbf{u}|\mathbf{X}, \mathbf{Z}) = \sigma_u^2 \mathbf{I}_N,$$

$\hat{\gamma}$ is the BLUE of γ and $\hat{\beta}$ is the BLUE of β . As is well-known, $\hat{\gamma}$ can be obtained by an OLS regression where y_{gi} and \mathbf{x}_{gi} have been deviated from within-teacher averages (see, for example, Wooldridge 2010, Chapter 10). Further, the estimated teacher fixed effects can be obtained as

$$\hat{\beta}_g = \bar{y}_g - \bar{\mathbf{x}}_g \hat{\gamma}. \quad (24)$$

Equation (24) makes computation of the teacher VAMs fairly efficient if one does not want to run the long regression in (22).

By comparing equations (15) and (24) we see that the EB estimator b_g^* differs from the fixed effects estimator $\hat{\beta}_g$ in two ways. First, and most importantly, the RE estimator γ^* is used in computing b_g^* while $\hat{\beta}_g$ uses the FE estimator $\hat{\gamma}$. Second, b_g^* shrinks the average of the residuals toward zero by the factor

$$\frac{\sigma_b^2}{\sigma_b^2 + (\sigma_u^2/N_g)} = \frac{1}{1 + (\rho/N_g)} \quad (25)$$

where

$$\rho = \sigma_u^2 / \sigma_b^2. \quad (26)$$

Equation (25) illustrates the well-known result that the smaller is the number of students for teacher g , N_g , the more the average residual is shrunk toward zero.

A well-known algebraic result – for example, Wooldridge (2010, Chapter 10) – that holds for any given number of teachers G is that

$$\gamma^* \rightarrow \hat{\gamma} \text{ as } \rho \rightarrow 0 \text{ or } N_g \rightarrow \infty.^1 \quad (27)$$

Equation (27) can be verified by noting that the RE estimator of γ can be obtained from the pooled OLS regression

$$y_{gi} - \theta_g \bar{y}_g \text{ on } \mathbf{x}_{gi} - \theta_g \bar{\mathbf{x}}_g \quad (28)$$

where

$$\theta_g = 1 - \left(\frac{\sigma_u^2}{\sigma_u^2 + N_g \sigma_b^2} \right)^{1/2} = 1 - \left(\frac{1}{1 + (N_g/\rho)} \right)^{1/2}. \quad (29)$$

It is easily seen that $\theta_g \rightarrow 1$ as $\rho \rightarrow 0$ or $N_g \rightarrow \infty$. In other words, with many students per teacher or large teacher effects relative to student effects, the RE and FE estimates can be very close. But they are never identical. Not coincidentally, the shrinkage factor in equation (25) tends to unity if $\rho \rightarrow 0$ or $N_g \rightarrow \infty$. The bottom line is that with a “large” number of students per teacher the shrinkage estimates of the teacher effects can be close to the fixed effects estimates. The RE and FE estimates also tend to be similar when σ_u^2 (the student effect) is “small” relative to σ_b^2 (the teacher effect), but this scenario seems unlikely.

An important point that appears to go unnoticed in applying the shrinkage approach is that in situations where γ^* and $\hat{\gamma}$ substantively differ, γ^* suffers from systematic bias because it assumes teacher assignment is uncorrelated with \mathbf{x}_i . Because γ^* is used in constructing the b_g^* in equation (15), the bias in γ^* generally results in biased teacher effects, and the teacher effects would be biased even if (15) did not employ a shrinkage factor. The shrinkage likely exacerbates the problem:

the estimates are being shrunk toward values that are systematically biased for the teacher effects.²

The expression in equation (15) motivates a common two-step alternative to the EB approach. In the first step of the procedure, one obtains $\tilde{\gamma}$ using the OLS regression in equation (17), and obtains the residuals, \tilde{r}_{gi} . In the second step, one averages the residuals \tilde{r}_{gi} within each teacher to obtain the teacher effect for teacher g . We call this approach the “average residual” (AR) method. After obtaining the averages of the residuals one can, in a third step, shrink the averages using the empirical Bayes’ shrinkage factors in equation (15). Typically the estimates in equations (18) and (20), based on the OLS residuals, are used in obtaining the shrinkage factors. We call the resulting estimator the “shrunk average residual” (SAR) method.

With or without shrinking, the AR approach suffers from systematic bias if teacher assignment, \mathbf{z}_i , is correlated with the covariates, \mathbf{x}_i . In effect, the AR approach partials \mathbf{x}_i out of y_i but does not partial \mathbf{x}_i out of \mathbf{z}_i , the latter of which is crucial if \mathbf{x}_i and \mathbf{z}_i are correlated. The so-called “fixed effects” regression in (22) partials \mathbf{x}_i out of \mathbf{z}_i , which makes it a more reliable estimator under nonrandom teacher assignment – perhaps much more reliable with strong forms of nonrandom assignment.

It is also important to know that the SAR approach is inferior to the EB approach under nonrandom assignment. The logic is simple. First, the algebraic relationship between RE and FE means that γ^* tends to be closer to the FE estimator, $\hat{\gamma}$, than the OLS estimator, $\tilde{\gamma}$. Consequently, under nonrandom teacher assignment, the estimated teacher effects using the RE estimator of γ will have less bias than the estimates that begin with OLS estimation of γ . Second, if teacher assignment is

²Without covariates, the difference between the EB and fixed effects estimates of the b_g is much less important: they differ only due to the shrinkage factor. In practice, the fixed effects estimates, $\hat{\beta}_g$, are obtained without removing an overall teacher average, which means $\hat{\beta}_g = \bar{y}_g$. To obtain a comparable expression for b_g^* we must account for the GLS estimator of the mean teacher effect, which would be obtained as the intercept in the RE estimation. Call this estimator μ_b^* , which in the case of no covariates is γ^* . Then the teacher effects are

$$b_g^* = \mu_b^* + \eta_g(\bar{y}_g - \mu_b^*) = \eta_g \bar{y}_g + (1 - \eta_g)\mu_b^* = \bar{y}_g - (1 - \eta_g)(\bar{y}_g - \mu_b^*),$$

where η_g is the shrinkage factor in equation (25). Compared with the FE estimate of b_g , b_g^* is shrunk toward the overall mean μ_b^* . When the teacher effects are treated as parameters to estimate, the b_g^* are biased because of the shrinkage factor, even when they are BLUP.

uncorrelated with the covariates, the OLS estimator of γ is inefficient relative to the RE estimator under the standard random effects assumptions (because the RE estimator is FGLS). Thus, the only possible justification for SAR is computational simplicity. But the saving is likely to be minor unless the number of controls in \mathbf{x}_i is very large. For the kinds of data sets widely available, the computational saving from using SAR rather than EB is likely to be minor.

Finally, we must emphasize that fixed effects estimation of the teacher VAMs allows any correlation between \mathbf{z}_i and \mathbf{x}_i , and thus expect it to outperform EB estimation and strongly outperform SAR under nonrandom assignment. The bias due to nonrandom allocation of students to teachers is also discussed in Rothstein (2009, 2010).

3.3 Summary of Estimation Methods

In this paper we examine five different value-added estimators used to recover the teacher effects and apply them to both real and simulated data. Some of the estimators use EB or shrinkage techniques, while others do not. They can all be cast as a special case of the estimators described in the previous section. For clarity, we briefly describe each one, with additional reference to each of these specifications provided in Table C.4.

The estimators can be obtained from a dynamic equation of the form

$$A_{it} = \lambda A_{i,t-1} + \mathbf{X}_{it} \boldsymbol{\delta} + \mathbf{Z}_{it} \boldsymbol{\beta} + v_{it}, \quad (30)$$

in which A_{it} is achievement (measured by a test score) for student i in grade t , \mathbf{X}_{it} is a vector of student characteristics, and \mathbf{Z}_{it} is the vector of teacher assignment dummies. This is similar to equation (1) but with the lagged test score written separately from X_{it} for clarity. Also, \mathbf{X}_{it} is omitted from the estimation of the teacher effects in the simulation analysis below, as student characteristics are not included in the data generating process. The EB estimator we analyzed in Section 2 was for the case of a single cross-section of students. Thus, we use only one grade – fifth

grade – for the analysis.

We first analyze EB LAG, a dynamic MLE version of the EB estimator that treats the teacher effects as random. This technique obtains the estimates of the teacher effects using the normal maximum likelihood in the first stage, regressing A_{it} on its lag, $A_{i,t-1}$, and X_{it} . In the second stage, the shrinkage factor is applied to these teacher effects. As described in Rabe-Hesketh and Skrondal (2012), this two step procedure can be performed in one-step using the “xtmixed” command in Stata 12 with teacher random effects. The predicted random effects of this regression are identical to shrinking the MLE estimates by the shrinkage factor. This procedure is generally justified even if the unobservables do not have normal distributions, in which case we are applying quasi-MLE. A second estimator we consider is the average residual (AR) method described in Section 2. This technique mainly differs from EB LAG in that it uses OLS in the first stage. The residuals of this OLS regression are obtained, and then we average these residuals by classroom to calculate the estimated teacher effects. We expect the EB LAG estimator to outperform the AR estimator in most scenarios, given that MLE is being used in the first-stage instead of OLS.

We compare the estimators that treat the teacher effect as random with an estimator that explicitly controls for the teacher effect through the inclusion of teacher assignment dummy variables. This third estimator applies OLS to (30) by pooling across students and classrooms. This “dynamic OLS,” or DOLS estimator treats the teacher effects as fixed parameters to estimate. The inclusion of the lagged test score accounts for the possibility that teacher assignment is related to students’ most recent test score. Guarino, Reckase, and Wooldridge (forthcoming) discuss the assumptions under which DOLS consistently estimates β when (30) is obtained from a structural cumulative effects model, and the assumptions are quite restrictive. Nevertheless, their simulations show the DOLS estimator often estimates β well even when the assumptions underlying the consistency of DOLS fails.

Given that EB estimation is often motivated as a way to increase precision and decrease misclassification, we also analyze whether shrinking AR and DOLS estimates enhances performance.

Thus, the fourth estimator we analyze is our shrunken average residual (SAR) estimator. This estimator takes the AR estimates and shrinks them by the shrinkage factor described in Section 2 (see Equation (25)). Shrinking the AR estimates does not result in a true EB estimator since AR uses OLS in the first stage, but it is commonly used as a simpler way of operationalizing the EB approach (see, for example, Kane & Staiger, 2008). As discussed in Section 2, with a sufficiently large number of students per teacher, the EB LAG estimator converges to the DOLS estimator but SAR does not. Thus, as the number of students per teacher grows, we would expect EB LAG to perform more similarly to DOLS than SAR. Finally, we consider a shrunken DOLS (SDOLS) estimator, which takes the DOLS estimated teacher fixed effects and shrinks them by the shrinkage factor derived in Section 2. Although SDOLS is rarely done in practice and is not a true EB estimator, we include it as an exploratory exercise in order to better determine the effects of shrinking itself when the number of students per teacher differs. When the class sizes are all the same, the shrunken estimates (SDOLS and SAR) will only differ from the unshrunk estimates by a constant positive multiple. Thus, shrinking the DOLS or AR estimates will have no effect in terms of ranking teachers. It is important to keep in mind that, unlike DOLS and SDOLS, the AR and SAR estimators do not allow for general correlation between the teacher assignment and past test scores (or other covariates).

3.4 Comparing VAM Methods Using Simulated Data

The question of which VAM estimators perform the best can only be addressed in simulations in which the true teacher effects are known. Therefore, to evaluate the performance of EB estimators relative to other common value-added estimators, we apply these methods to simulated data. This approach allows us to examine how well various estimators recover the true teacher effect under a variety of student grouping and teacher assignment scenarios. Using data generated as described in Section 4.1, we apply the set of value-added estimators discussed in Section 3 and compare the resulting estimates with the true underlying teacher effects.

3.4.1 Simulation Design

Much of our main analysis focuses on a base case that restricts the data generating process to a relatively narrow set of idealized conditions. These ideal conditions do not allow for measurement error or peer effects and assume that teacher effects are constant over time, but include these as sensitivity tests of the main results. The data are constructed to represent grades three through five (the tested grades) in a hypothetical school. For simplicity and comparison with the theoretical predictions, we assume that the learning process has been going on for a few years but only calculate estimates of teacher effects for fifth grade teachers – a single cross section.³ We create data sets that contain students nested within teachers, with students followed longitudinally over time in order to reflect the institutional structure of an elementary school. Our simple baseline data generating process is as follows:

$$\begin{aligned}A_{i3} &= \lambda A_{i2} + \beta_{i3} + c_i + e_{i3} \\A_{i4} &= \lambda A_{i3} + \beta_{i4} + c_i + e_{i4} \\A_{i5} &= \lambda A_{i4} + \beta_{i5} + c_i + e_{i5}\end{aligned}\tag{31}$$

in which A_{i2} is a baseline score reflecting the subject-specific knowledge of child i entering third grade; A_{it} is the grade- t test score ($t = 3, 4, 5$); λ is a time constant decay parameter, where lambda is set equal to zero in the simulations for lag scores greater than one year prior; β_{it} is the teacher-specific contribution to growth (the true teacher value-added effect); c_i is a time-invariant student-specific effect (may be thought of as “ability” or “motivation”); and e_{it} is a random deviation for each student. We assume independence of e_{it} over time, a restriction implying that past shocks to student learning decay at the same rate as all inputs (see Guarino, Reckase, & Wooldridge,

³Despite only estimating value-added for grade 5 teachers, we keep the three grade structure when generating the student test scores since the fifth grade achievement is based on more than just the current teacher and prior test score of the student; it is a function of all prior teacher, unobservable student, and random influences. Thus, to ignore that process and generate fifth grade test scores based on a “baseline” fourth grade test score seems inappropriate given this context.

2012 for a more detailed discussion of this “common factor restriction” assumption). In all of the simulations reported in this paper, the random variables A_{i2} , β_{it} , c_i , and e_{it} are drawn from normal distributions with means of zero. The standard deviation of the teacher effect is .25, the standard deviation of the student fixed effect is .5, and the standard deviation of the random noise component is 1. These give relative shares of 5, 19, and 76 percent of the total variance in gain scores (when $\lambda = 1$), respectively. Given that the student and noise components are larger than the teacher effects, we call these “small” teacher effects. We also conduct a sensitivity analysis using “large” teacher effects, where the true teacher effects are drawn from a distribution with a standard deviation of 1. The baseline score is drawn from a distribution with a standard deviation of 1. We also allow for correlation between the time-invariant child-specific heterogeneity, c_i , and the baseline test score, A_{i2} , which we set to 0.5. This correlation reflects that students with better unobserved “ability” likely have higher test scores as well.

Our data are simulated using 36 teachers that teach 720 students per cohort. In order to create a situation in which there is a substantial variation in class size – to showcase the potential disparities between EB/shrinkage and other estimators – we vary the number of students per classroom. Teachers receive classes of varying sizes, but receive the same number of students in each cohort. Of the 36 teachers we simulate, twelve teachers have classes of 10 students, twelve teachers have a class size of 20, and twelve teachers have class sizes of 30. The mean and standard deviation of the true teacher effects of the 36 teachers we estimate are 0.501 and 0.244, respectively. We simulate the data using both one and four cohorts of students to provide further variation in the amount of data from which the teacher effects are calculated. In the case of four cohorts, data are pooled across the cohorts so that value-added estimates are based off of sample sizes of 40, 80, and 120, instead of 10, 20, and 30 as in the one cohort case. Therefore, we would expect that estimates in the four cohort case to be less noisy than those from the one cohort case, possibly mitigating the potential gains from EB estimation.

To create different scenarios, we vary certain key features: the grouping of students into classes

and the assignment of classes of students to teachers within schools. We generate data using each of the seven different mechanisms for the assignment of students outlined in Table C.5. Students are grouped into classrooms either randomly, based on their prior year achievement level (dynamic grouping or DG), or based on their unobserved heterogeneity (heterogeneity grouping or HG). In the random case, students are assigned a random number and then grouped into classrooms of various sizes based on that random number. In the grouping cases, students are ranked by either the prior test score or the student fixed effect and grouped into classrooms of various sizes based on that ranking. Teachers are assigned to these classrooms either randomly (denoted RA) or nonrandomly. Teachers assigned nonrandomly can be assigned positively (denoted PA), meaning the best teachers are assigned to classrooms with the best students, or negatively (denoted NA), meaning the best teachers are assigned to classrooms with the worst students.

These grouping and assignment procedures are not purely deterministic, as we allow for a random component with standard deviation of 1 in the assignment mechanism. As a sensitivity analysis, we also run simulations setting this standard deviation to 0.1, meaning the grouping of students into classrooms is more nonrandom. We use the estimators discussed in Section 3, but with only a constant, teacher dummies (if applicable), and the lagged test score included as covariates. We use 100 Monte Carlo replications per scenario in evaluating each estimator.

3.4.2 Evaluation Measures

For each estimator across each iteration, we save the individual estimated teacher effects and also retain the true teacher effects, which are fixed across the iterations for each teacher. To study how well the methods recover the true teacher effects, we adopt five simple summary measures using the teacher level data. The first is a measure of how well the estimates preserve the rankings of the true teacher effects. We compute the Spearman rank correlation, $\hat{\rho}$, between the estimated teacher effects and the true effects and report the average $\hat{\rho}$ across the 100 iterations. Second, we compute a measure of misclassification. These misclassification rates are obtained as the per-

centage of above average teachers in the true quality distribution (i.e. teachers with true $\beta_g > 0$) who are misclassified as below average in the distribution of estimated teacher effects for the given estimator. Given that this is just an arbitrary cutoff point, we also obtain the fraction of teachers that are misclassified in the tails of the distribution (e.g fraction of above average teachers that are misclassified in the bottom quintile).

In addition to examining rank correlations and misclassification rates, it is also helpful to have a measure that quantifies some notion of the magnitude of the bias in the estimates. Given that some teacher effects are biased upwards and others downwards, it is difficult to capture the overall bias in the estimates in a simple way. We create a statistic, $\hat{\theta}$, that captures how closely the magnitude of the deviation of the estimates from their mean tracks the size of the deviation of the true effects from the true mean. To calculate this measure, we regress the deviation of the estimated teacher effects from their overall estimated means on the analogous deviation of the true effects generated from the simulation for each estimator. We can represent this simple regression as

$$\hat{\beta}_g - \bar{\hat{\beta}} = \hat{\theta}(\beta_g - \bar{\beta}) + residual_g, \quad (32)$$

in which $\hat{\beta}_g$ is the estimated teacher effect and β_g is the true effect of teacher g . From this simple regression, we report the average coefficient, $\bar{\hat{\theta}}$, across the 100 replications of the simulation for each estimator. This regression tells us whether the estimated teacher effects are correctly distributed around the average teacher. If $\hat{\theta} = 1$, then a movement of β_g away from its mean is tracked by the same movement of $\hat{\beta}_g$ from its mean.

When $\hat{\theta} \approx 1$, the magnitudes of the estimated teacher effects can be compared across teachers. If $\hat{\theta} > 1$, the estimated teacher effects amplify the true teacher effects. In other words, teachers above average will be estimated to be even more above average and vice versa for below average teachers. An estimation method that produces $\hat{\theta}$ substantially above one generally does a good job of ranking teachers, but the magnitudes of the differences in estimated teacher effects cannot be

trusted. The magnitudes also cannot be trusted if $\hat{\theta} < 1$; in this case, ranking the teachers becomes more difficult because the estimated effects are compressed relative to the true teacher effects.

In addition to ranking teachers correctly, the magnitude of the estimated teacher effects is also important in policy applications. It is helpful to examine the extent to which shrinking the estimates, as in the EB methods, increases bias in these noisy estimates. Thus, we report the average value of $\hat{\theta}$ across the simulations because it provides evidence of which methods, under which scenarios, produce estimated teacher effects whose magnitudes have meaning. This measure also provides insight into why some methods rank teachers relatively well even when the estimated effects are systematically biased.

The precision of these methods is also a key consideration when evaluating the overall performance. As described in Section 2, EB methods are not unbiased when thinking about the teacher effects as fixed parameters we are trying to estimate. However, if the identifying assumptions hold, these methods should provide more precise estimates. This is one motivation for using such methods, as estimates should be more stable over time, leading to a smaller variance in the teacher effects. As the teacher effect is fixed for each teacher across the 100 iterations, we have 100 estimates of each teacher effect. As a summary measure for the precision of the estimators, we calculate the standard deviation of the 100 teacher effect estimates for each teacher and then take a simple average across all teachers.

To further analyze the variance-bias tradeoff for each of these estimators, we also include average mean squared error (MSE). This measure averages the following across all g teachers and across simulation runs:

$$\widehat{MSE}_g = (\beta_g - \hat{\beta}_g)^2 \tag{33}$$

This provides a simple statistic to determine whether the bias induced by shrinking is justifiable due to gains in precision.

3.5 Simulation Results

Tables C.1 and C.2 report the five evaluation measures described in Section 4.2 for each particular estimator-assignment scenario combination. For ease in interpreting the tables, a quick guide to descriptions of each of these estimators, grouping-assignment mechanisms, and evaluation measures can be found in Appendix tables C.4 through C.6. As these shrinkage and EB estimators are often motivated as a way to reduce noise, one might expect these approaches to be most beneficial with very limited student data per teacher. Thus, we estimate teacher effects using both four cohorts and one cohort of data. The tables show results for the case $\lambda = .5$. Though not reported, we also conducted a full set of simulations for $\lambda = 0.75$ and $\lambda = 1$, and the main conclusions are unchanged. The full set of simulation results is available upon request from the authors.

3.5.1 Fixed Teacher Effects versus Random Teacher Effects

We first compare the performance of the DOLS estimator, which treats teacher effects as fixed parameters to estimate, to the AR and EB LAG estimators that treat teacher effects as random in Table C.1. Under nonrandom assignment of teachers, we expect DOLS, which explicitly controls for teacher assignment through the inclusion of teacher assignment indicators, to perform better than those estimators treating the teacher effects as random. When teacher assignment is based on the lagged test score, DOLS directly controls for the assignment mechanism by including both the lagged score and teacher assignment indicators and should perform particularly well in this case. The simulation results presented here largely support this hypothesis.

3.5.1.1 Random Assignment

We begin with the pure random assignment (RA) case, where EB-type estimation methods are theoretically justified. The results of the random assignment case are given in the top panel of Table C.1 and they suggest very little substantial difference between the performance of the fixed and random effects estimators under this scenario. As the theory suggests, EB LAG performs well in

the four cohort case, with rank correlations between the estimated and the true teacher effects near 0.86, which is nearly the same as the 0.85 rank correlation for DOLS and AR. In addition to very similar rank correlations, the misclassification rates are very similar across the three estimators, with about 15 percent of above average teachers misclassified as below average. The similarities between the three estimators in terms of rank correlation and misclassification rates remains when using only one cohort. Reducing the amount of data used to estimate the teacher effects lowers the performance of all estimators, decreasing the rank correlations and increasing the misclassification rates. With one cohort, rank correlations between the estimated and true teacher effects are about 0.65 to 0.67, and between 25 and 26 percent of above average teachers are misclassified as below average.

In addition to rank correlations and misclassification rates, we also examine the bias and precision of the estimators. While DOLS and AR appear to be unbiased with average $\hat{\theta}$ values close to 1, EB LAG substantially underestimates the magnitudes of the true teacher effects with an average $\hat{\theta}$ value of 0.78 using four cohorts and 0.49 using one cohort. This bias is likely the result of the shrinkage technique that is applied, but this shrinkage does cause EB LAG to be slightly more precise than AR or DOLS. While DOLS and AR both have similar average standard deviations of the estimated teacher effects near 0.13 and 0.27 in the four and one cohort cases, respectively, EB LAG has lower average standard deviations of 0.12 and 0.18, respectively. Given the precision gain in EB LAG, the MSE measure suggests that EB LAG may be preferred to DOLS or AR under random assignment.

We now move to the cases where the students are *nonrandomly grouped* together, but teachers are still *randomly assigned* to classrooms. We allow for nonrandom grouping based on either the prior year test score (dynamic grouping, DG) or student-level heterogeneity (heterogeneity grouping, HG). Under these DG-RA and HG-RA scenarios in Table C.1, we see a fairly similar pattern as in the RA scenario, although the overall performance of all estimators is somewhat diminished, especially in the HG-RA scenario.

3.5.1.2 Dynamic Grouping and Nonrandom Assignment

The performance of these various estimators diverges noticeably under *nonrandom* teacher assignment. We continue to nonrandomly group teachers as described above, but now allow for nonrandom assignment of students to teachers. Classes with high test scores or high unobserved ability can be assigned to either the best (positive assignment) or worst (negative assignment) teachers. A key finding of this analysis is the disparity in performance of estimators that treat teacher effects as random (e.g. AR and EB LAG) compared with the DOLS estimator. These results suggest that when there is nonrandom teacher assignment based on the prior test score estimators explicitly controlling for the teacher assignment should be preferred to those that treat the teacher effects as random.

DOLS substantially outperforms AR and EB LAG under the DG-PA scenario. When using four cohorts, DOLS has a rank correlation of 0.86 under DG-PA, while AR and EB LAG have rank correlations of 0.60 and 0.76, respectively. AR and EB LAG also have large misclassification rates, with 28 to 32 percent of above average teachers being misclassified as below average compared with only 23 percent for DOLS.

In addition to misclassifying and poorly ranking teachers, the AR and EB LAG methods also underestimate the magnitudes of the true teacher effects. While DOLS has an average $\hat{\theta}$ value of 0.99, the AR and EB LAG estimators have average $\hat{\theta}$ values of 0.53 and 0.49, respectively. While some of the bias of the EB LAG estimates can be attributed to shrinkage, the larger issue is the bias caused by the failure of the AR and EB LAG approaches to net out the correlation between the lagged test score (i.e. the assignment mechanism in these DG scenarios) and the teacher assignment, a correlation that DOLS explicitly allows for with the inclusion of teacher dummies in the regression. Just as in the random assignment case, DOLS and EB LAG have similar MSE measures, while the MSE for AR is slightly larger. In the four cohort case, DOLS, EB LAG and AR have MSE values of 0.02, 0.02, and 0.04, respectively. Similar results are found if we instead examine the DG-NA case.

These simulation results also verify an important result of the theoretical discussion: the performance of EB LAG approaches the performance of DOLS as the number of students per teacher grows. We see less of a disparity in the performance of DOLS and EB LAG when computing VAMs using four cohorts compared to one, but the relative performance of AR does not improve with more students per teacher. For example, under DG-PA with one cohort of students, AR and EB LAG have rank correlations of 0.38 and 0.45, respectively, compared to 0.63 for DOLS. With four cohorts of students, the rank correlation for EB LAG is much closer to that for DOLS (0.76 and 0.86, respectively) than is the rank correlation for AR (0.60). This theoretical result is also applicable to the SAR estimator we examine below, which is used as a simpler way to operationalize the EB approach. In summary, EB LAG, which uses random effects estimation in the first stage, is preferred to those using OLS (AR and SAR) under nonrandom teacher assignment, as these estimates approach the preferred DOLS estimates that treat teacher effects as fixed.

3.5.1.3 Heterogeneity Grouping and Nonrandom Assignment

As a final scenario we examine the case of nonrandom teacher assignment to students grouped on the basis of student-level heterogeneity. The results for these HG scenarios are especially unstable: all estimators do an excellent job ranking teachers under positive teacher assignment, and all estimators do a very poor job under negative teacher assignment. In the HG-PA case with four cohorts of students, the magnitudes of the estimated VAMs are amplified as seen by the large average values for $\hat{\theta}$ between 1.43 and 1.61. This improves the ability of the various estimators to rank teachers as evidenced by the high rank correlations of about 0.94 for all estimators. The EB LAG estimator performs the best in this scenario, as it performs as well as the other estimators in terms of ranking and misclassification of teachers but has the smallest MSE measure. Under HG-NA with four cohorts, the performance of all estimators falls substantially, largely caused by severely underestimated teacher effects ($\hat{\theta}$ values between 0.15 and 0.33). These compressed teacher effect estimates make it difficult to rank teachers in this scenario, resulting in low rank correlations for

all estimators between 0.38 and 0.41. Just as in the HG-PA scenario, the performance of the three estimators under HG-NA is very similar across the evaluation measures we examine.

Why is the performance of DOLS, AR, and EB LAG so similar under HG-PA and HG-NA, while differing so greatly under DG-PA and DG-NA? Despite correlation between the baseline test score and the student fixed effect, the lagged test score appears to be a weak proxy for the assignment mechanism in the HG scenarios. Since none of the three estimators do well at allowing for the correlation between the assignment mechanism and the teacher assignment in these cases, the distinction between estimators that include teacher fixed effects and those that treat teacher effects as random is less stark. As found in Guarino, Reckase, and Wooldridge (forthcoming), a gain score estimator with student fixed effects included is the most robust in these HG scenarios, as it does allow for the correlation between the assignment mechanism (i.e. student fixed effect) and the teacher assignment (i.e teacher dummy variables). Their results lend further support for our conclusions here that allowing for this correlation is extremely important in the performance of these value added estimators when there is nonrandom assignment.

3.5.2 Shrinkage versus Non-Shrinkage Estimation

Use of EB and other shrinkage estimators is often motivated as a way to reduce the noise in the estimation of teacher effects, particularly for teachers with a small number of students. Greater stability in the estimated effects is thought to reduce misclassification of teachers. We observed in section 3.1 that EB LAG was generally outperformed by the fixed effects estimator, DOLS. However, under nonrandom teacher assignment, we are unable to tell how much of the bias in the EB LAG estimator is due to treating the teacher effects as random and how much is due to the shrinkage procedure. To examine the effects of shrinkage itself, we compare the performance of unshrunk estimators, DOLS and AR, with their shrunken versions, SDOLS and SAR, in Table C.2. Although SDOLS is not a commonly used or theoretically justified estimator, we explore it here to identify whether shrinking teacher fixed effect estimates could be useful in practice.

Our simulation results show that there is no substantial improvement in the performance of the DOLS or AR estimators after applying the shrinkage factor to the estimates. Using four cohorts of students, the performance measures for DOLS and AR compared to their shrunken counterparts are nearly identical to two decimal places across all grouping and assignment scenarios. Even with very limited data per teacher in the one cohort case, when we would expect shrinkage to have a greater effect on the estimates, we find very little change in the performance of the estimators after the shrinkage factor is applied.

In the one cohort case, shrinking either the DOLS or AR estimates slightly decreases (in the second decimal place) both the average $\hat{\theta}$ values and average standard deviation of the estimated teacher effects. This increased bias in the estimates is expected when applying the shrinkage factor and, depending on the scenario and estimator we examine, the effect of this precision-bias tradeoff may increase or decrease the MSE measure when comparing the shrunken and unshrunken estimates. Shrinking the DOLS and AR estimates generally reduces the MSE but makes no substantial difference on the misclassification rate of teachers.

The effect of shrinkage itself does not appear to be practically important for properly ranking teachers or ameliorate the performance of the biased AR estimator found in the DG-PA and DG-NA scenarios. Given that shrinking the AR estimates does little to mitigate the performance drop of AR under DG-PA and DG-NA, our evidence suggests that shrinking teacher fixed effects estimates is preferred over shrinking teacher random effects if such techniques are desired.

3.5.3 Sensitivity Analyses

As mentioned in Section 4.1, we also test the sensitivity of these results by changing some of the parameters of the model. First, we increase the standard deviation of the distribution from which the true teacher effects are drawn. Using these “large” teacher effects, all estimators increase in performance and EB LAG performs similarly to DOLS in the DG-PA and DG-NA cases. The AR method, however, continues to suffer in performance under DG-PA and DG-NA compared

to DOLS. Second, we allow for more non-randomness (i.e. decrease the amount of noise) in the assignment of teachers into classrooms. As the assignment of teachers becomes more nonrandom, the performance of AR and EB LAG suffers even more greatly in terms of lower rank correlations and higher misclassification rates than what is observed in the results in Table C.1 and C.2. Given that some models use multiple prior test scores (e.g. EVAAS, VARC) and we also estimate DOLS, AR, and EB LAG with multiple lagged test scores as a sensitivity analysis. Although adding multiple lags improves the performance of AR and EB LAG in the random assignment case, the performance of these estimators still suffers greatly compared to DOLS in the DG-PA and DG-NA scenarios. As a final sensitivity test we include a peer effect (e.g. avg. c_i of student's classmates) in the underlying DGP. Even when peer effects are included, EB LAG and AR continue to suffer in performance under the DG-PA and DG-NA cases.

3.6 Comparing VAM Methods Using Real Data

We also apply these estimation methods to actual student-level test score data and examine the rank correlations between the estimated teacher effects of the various estimators for each school district. In addition to rank correlations, we also examine whether teachers are being classified in the extremes uniformly across all of the estimators we examine. Although the real data does not allow comparison between the estimated effects and the true teacher effects, we are able to make comparisons between the estimated effects of different estimators. This comparison provides a measure of the sensitivity of the estimated teacher effects to specifications that shrink the estimates and/or treat the teacher effects as random or fixed. The results of this analysis provide some perspective on the impact of shrinking and Empirical Bayes' methods in a real-world setting.

3.6.1 Data

We apply the five methods described in Section 3 to data from an anonymous southern U.S. state. The data span 2001 through 2007 and grades four through six, but test scores are collected

for each student from grades three through six. The data set includes 1,488,253 total students from which we have at least one current year score and one lagged score. Only 482,031 students have test scores for all grades. The data set also contains 43,868 unique teachers that we observe for a varying number of cohorts of students. We observe 39 percent of teachers for only one year, but we do see 20 percent of teachers for four or more years. These teachers, on average, teach about 26 students per year, with only a small percentage (less than two percent) teaching more than 30 students per year. The high percentage of teachers that we observe for only one year could motivate researchers to employ shrinkage and EB estimators as a way to reduce precision problems due to minimal data.

We estimate teacher effects district-by-district using equation (30) with controls for various student characteristics and include dummies for the year. Student characteristics include race, gender, disability status, free/reduced price lunch eligibility, limited English proficiency status, and the number of student absences from school. As discussed above, the teacher effects are estimated using data on multiple cohorts (between one and seven) of students. For simplicity and comparison with the simulation results, we estimate the value-added measures for those teachers with fifth grade students in the 67 districts, but again teachers receive multiple cohorts of students. Overall, we estimate 20,749 teacher effects using test score data from the annual assessment exam administered by the state.

3.6.2 Results

Figure C.1 presents box plots that depict the distributions of the within-district rank correlations between the various lagged score estimators, DOLS, SDOLS, AR, SAR, and EB LAG. The results presented here are from using math scores, but the results are similar when reading scores are used. As in the discussion of the simulation results, we first compare the DOLS estimator, which treats the teacher effects as fixed, with the estimators that treat the teacher effects as random. Comparing DOLS and AR, we find that the median rank correlation is around 0.99, but there are nine districts

with rank correlations below 0.90 and 2 districts with correlations below 0.50. We also observe a slightly lower median rank correlation between DOLS and EB LAG, at around 0.97, with five districts with rank correlations below 0.90 and three below 0.50. These results are not inconsistent with our simulation results: the performance of DOLS, AR, and EB LAG is very similar under cases of random assignment of teachers to classrooms, but the performance of AR and EB LAG is substantially different from DOLS under non-random assignment based on prior test scores. Thus, it could be the case that these outlier districts observed in the left tails of the top two box plots may be composed of schools that engage more heavily in nonrandom assignment of teachers to classrooms.

Comparing the two estimators that treat teacher effects as random, AR and EB LAG, we find that while the median rank correlation is 0.96, nine districts have rank correlations of between 0.82 and 0.92. These results suggest that the estimates are somewhat sensitive to how the teacher effects are calculated in the first stage. This was also the case in the simulated results, where the performance of the AR estimator suffered more than the performance of the EB LAG estimator in cases of non-random assignment based on the prior test score.

For a thorough comparison with the simulation results, we also compare the shrunken and unshrunken estimates of DOLS and AR using the real data. We find median rank correlations of around 0.97 for both the DOLS and SDOLS comparison and the AR and SAR comparison, suggesting that shrinkage has a small impact on the estimates. It appears that in certain cases, shrinkage may have a larger impact on the DOLS estimates, as two districts have rank correlations of 0.50 and 0.72. Our simulation results suggested that shrinking the estimates had very little impact on estimator performance.

In addition to rank correlation comparisons, we also examine the extent to which teachers are classified in the tails of the distribution by the different estimators. If shrinkage is having some effect, we would expect to see some teachers classified in the extremes to be pushed toward the middle of the distribution after applying the shrinkage factor. Table C.3 lists the fraction of

teachers ranked in the same quintile, either the top or bottom, by different pairs of estimators. Comparing across estimators that assume fixed teacher effects to those that assume random teacher effects, we do not see much movement across quintiles. For example, comparing DOLS to EB LAG, we find that about 91 percent of the teachers that are classified in the top quintile using DOLS are also in this quintile using EB LAG. This suggests that teacher assignment may not be largely based on prior student achievement or that the prior test score is a poor proxy for the true assignment mechanism. If the prior test score or other covariates insufficiently proxy for the underlying assignment mechanism, then the choice to include teacher assignment variables will matter little in how teachers are ranked.

Comparing the rankings of unshrunk and corresponding shrunken estimators, we see that about 90 percent of teachers are ranked in the same quintile by both the unshrunk estimators (DOLS and AR) and their shrunken counterparts (SDOLS and SAR). This suggests that shrinking the estimates results in some reclassification of teachers in the tails to quintiles in the middle of the distribution. Using real data, however, we are unable to tell whether this reclassification is appropriate. Our simulated analysis suggested that shrinking the estimates had little impact if any on misclassification rates.

3.7 Conclusion

Using simulation experiments where the true teacher effects are known, we have explored the properties of two commonly used Empirical Bayes' estimators as well as the effects of shrinking estimates of teacher effects in general. Overall, EB methods do not appear to have much advantage, if any, over simple methods such as DOLS that treat the teacher effects as fixed, even in the case of random teacher assignment where EB estimation is theoretically justified. Under random assignment, all estimators perform well in terms of their ability of ranking teachers, properly classifying teachers, and providing unbiased estimates. EB methods have a very slight gain in precision compared to the other methods in this case.

We generally find that EB estimation is not appropriate under nonrandom teacher assignment. The hallmark of EB estimation of teacher effects is to treat the teacher effects as random variables that are independent (or at least uncorrelated) with any other covariates. This assumption is tantamount to assuming that teacher assignment does not depend on other covariates such as past test scores (this is also true for the AR methods). When teacher assignment is not random, estimators that either explicitly control for the assignment mechanism or proxy for it in some way typically provide more reliable estimates of the teacher effects. Among the estimators and assignment scenarios we study, DOLS and SDOLS are the only estimators that control for the assignment mechanism (again, either explicitly or by proxy) through the inclusion of both the lagged test score and teacher assignment dummies. As expected, DOLS and SDOLS outperform the other estimators in the nonrandom teacher assignment scenarios. In the analysis of the real data, we found that the rank correlations between, say, DOLS and EB LAG or DOLS and SAR are quite low for some districts, suggesting that the decision among these estimators is important. Thus, if there is a possibility of nonrandom assignment, DOLS should be the preferred estimator.

As predicted by theory and seen in the simulation results, the random effects estimator, EB LAG, converges to the fixed effects estimator, DOLS, as the number of students per teacher gets large. Therefore, it could be that EB LAG is performing well in large samples simply because the estimates are approaching the DOLS estimates. However, the average residual methods, AR and SAR, do not have this property. Thus, despite the recent popularity, we strongly caution using SAR as a simpler way to operationalize the EB LAG estimator. If EB-type methods are being used, it is important to estimate the coefficients in the first stage using random effects estimation (as in our EB LAG estimator) rather than OLS.

Lastly, we find that shrinking the estimates of the teacher effects does not seem to improve the performance of the estimators, even in the case where estimates are based on one cohort of students. The performance measures are extremely close in our simulations for those estimators that differ only due to the shrinkage factor – DOLS and SDOLS or AR and SAR. The rank correlations

for these two pairs of estimators are also very close to one in almost all districts. Also, we find in the simulations that shrinking the AR estimates, which is a popular way to operationalize the EB approach, does not reduce misclassification of teachers. Thus, our evidence suggests that the rationale for using shrinkage estimators to reduce the misclassification of teachers due to noisy estimates of teacher effects should not be given much weight. Accounting for nonrandom teacher assignment when choosing among estimators is more imperative.

Given the robust nature of the DOLS estimator to a wide variety of grouping and assignment scenarios, it should be preferred to AR and EB methods when there is uncertainty about the true underlying assignment mechanism. If the assignment mechanism is known to be random, applying these AR and EB estimators can be appropriate, especially when the amount of data per teacher is minimal. However, given that the assignment mechanism is not likely known, blindly applying these AR and EB methods can be extremely problematic, especially if teachers are truly assigned nonrandomly to classrooms. Therefore, we stress caution in applying these AR and EB methods and urge researchers to be mindful of the underlying assignment mechanism when choosing between the various value-added methods.

APPENDICES

APPENDIX A CHAPTER 1 TABLES AND FIGURES

Table A.1 Fiscal Stress Label Transitions from 2000-2012

Oversight in 2000	Emergency in 2000	No Label to Oversight	No Label to Emergency	Oversight to Emergency	Oversight to No Label	Emergency to No Label
6	9	92	4	19	57	30

Figure A.1 Geographic Distribution of Labeled School Districts Across Ohio

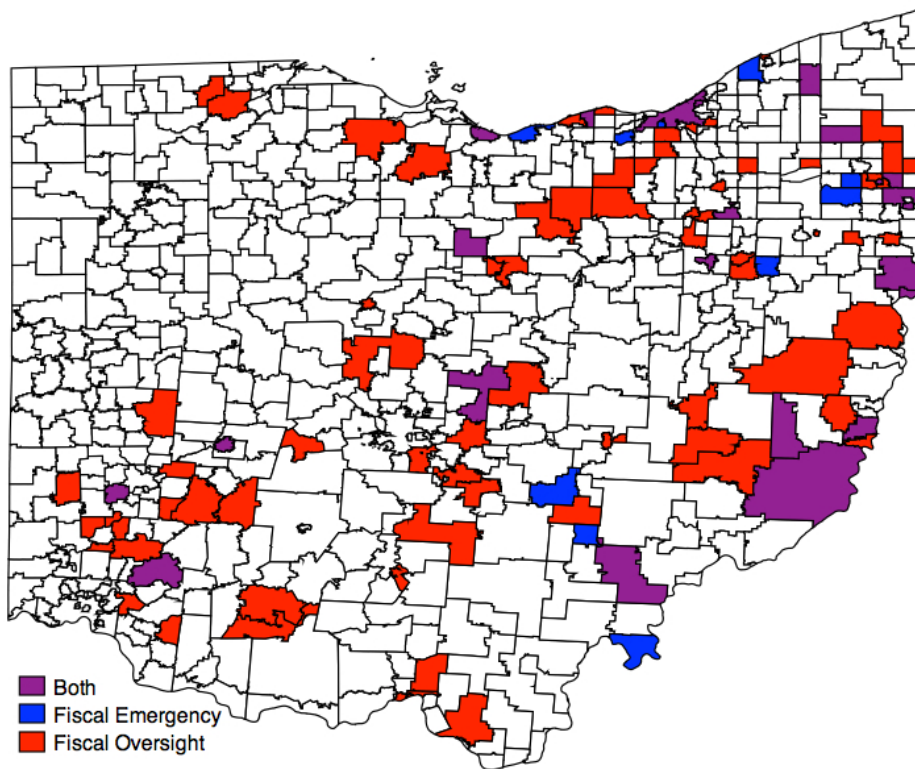


Figure A.2 Yearly Number of Labels, by Label Severity

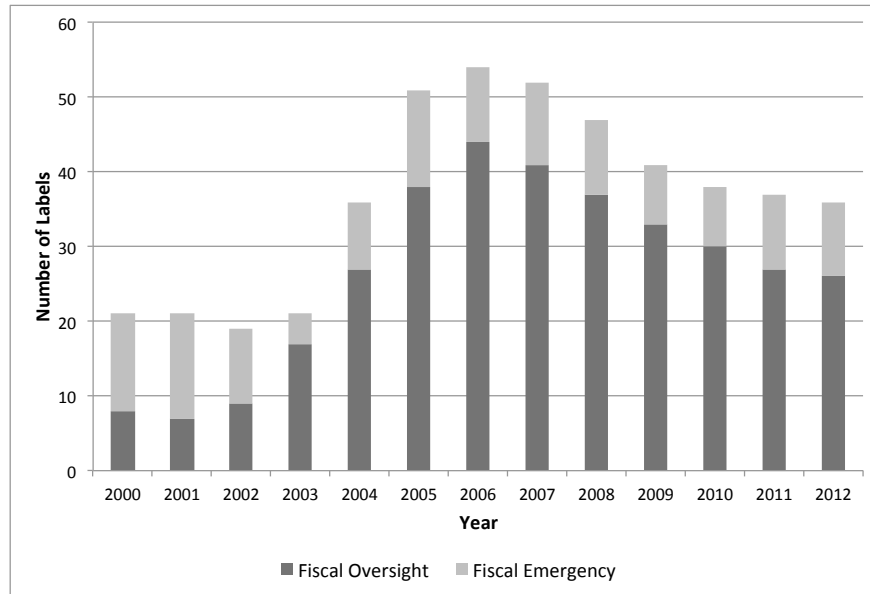


Table A.2 Table of Means - School District Demographic Characteristics

	Panel A Never in Fiscal Stress	Panel B † Pre-Fiscal Stress Fiscal Stress		Panel C ‡ Fiscal Stress Post-Fiscal Stress	
Total Enrollment	2,761 (4,192)	3,007 (2,620)	2,840 (2,383)	3,738 (7,689)	3,411 (6,043)
Proportion Free/Reduce Lunch Students	0.22 (0.13)	0.29 (0.17)	0.32 (0.16)	0.28 (0.16)	0.35 (0.18)
Number of Schools in District	5.57 (8.31)	6.43 (4.73)	5.69 (4.24)	7.28 (13.60)	7.07 (12.49)
FTE Teachers	163 (254)	16.99 (1.81)	17.64 (2.00)	17.36 (2.05)	17.12 (1.73)
Poor School Report Card Grade	0.16 (0.20)	0.22 (0.22)	0.36 (0.39)	0.30 (0.36)	0.23 (0.32)
Urban District	0.16 (0.36)	0.36 (0.48)	0.37 (0.49)	0.35 (0.48)	0.35 (0.48)
School Districts	507	80	80	72	72

Note: Pre-Fiscal Stress corresponds to the time period prior to the declaration of fiscal stress for labeled districts. Fiscal Stress corresponds to the time period a fiscal stress label is attached to the district. Post-Fiscal Stress corresponds to the time period after the label has been removed. Non-Fiscal Stress corresponds to all non-labeled jurisdictions. Standard Deviations are given in parentheses.

†Panel B includes districts that appear in both the pre-FS period and the FS period.

‡Panel C includes districts that appear in both the FS period and the post-FS period.

Table A.3 Table of Means - School District Financial Characteristics

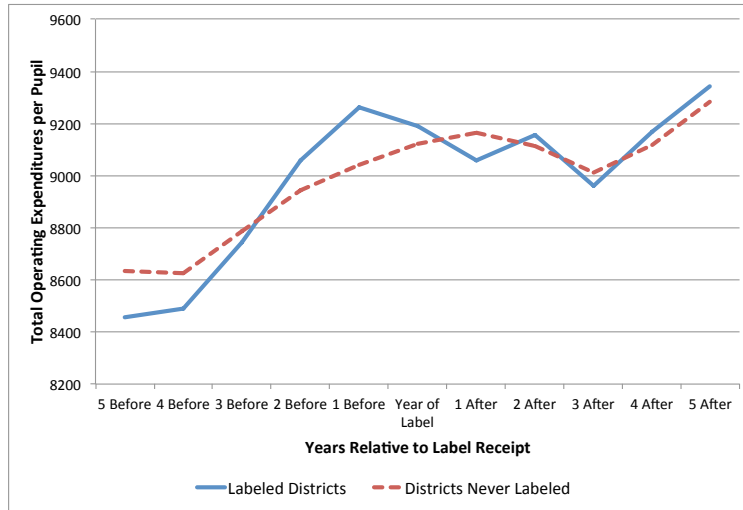
	Panel A Never in Fiscal Stress	Panel B† Pre-Fiscal Stress Fiscal Stress		Panel C‡ Fiscal Stress Post-Fiscal Stress	
Overall District Finances					
Gen Fund Balance/Revenue Ratio	0.21 (0.16)	0.04 (0.10)	0.01 (0.09)	0.02 (0.09)	0.16 (0.12)
Ratio < -0.02	0.06 (0.14)	0.35 (0.37)	0.36 (0.33)	0.34 (0.33)	0.05 (0.15)
District Expenditures					
Total Expenditure PP	11,164 (4,413)	10,906 (2,323)	10,572 (2,064)	10,250 (1,891)	10,986 (1,718)
Operating Expenditures PP	9,140 (3,524)	8,691 (1,189)	9,227 (1,525)	8,917 (1,163)	9,486 (1,130)
Capital Expenditures PP	1,482 (1,034)	1,775 (1,819)	748 (985)	822 (1,107)	845 (902)
Total Salaries PP	5,473 (1,693)	5,397 (789)	5,398 (826)	5,321 (706)	5,564 (754)
Total Employee Benefits PP	1,829 (552)	1,752 (337)	2,060 (436)	1,941 (417)	2,055 (367)
District Revenues					
Total Revenue PP	11,330 (4,665)	10,436 (2,178)	11,295 (2,306)	11,028 (2,182)	11,639 (1,809)
Total Federal Rev PP	636 (331)	654 (394)	1019 (673)	776 (557)	1,104 (708)
Total State Rev PP	5,216 (3,851)	5,408 (2,537)	5,521 (2,197)	5,121 (2,549)	5,209 (1,777)
Total Local Rev PP	5,478 (3,122)	4,373 (1,467)	4,755 (1,684)	5,131 (1,762)	5,327 (1,751)
Local Prop Tax Rev PP	4,758 (3,540)	3,807 (1,626)	4,292 (2,005)	4,644 (1,982)	4,581 (1,891)
Local Oper Prop Tax Rev PP	4,177 (3,205)	3,369 (1,509)	3,807 (1,880)	4,177 (1,932)	4,086 (1,810)
Local Cap Prop Tax Rev PP	567 (432)	426 (320)	479 (373)	460 (368)	495 (368)
Local Income Tax Rev PP	251 (450)	76 (215)	160 (333)	210 (403)	350 (572)
School Districts	507	80	80	72	72

Note: Pre-Fiscal Stress corresponds to the time period prior to the declaration of fiscal stress for labeled districts. Fiscal Stress corresponds to the time period a fiscal stress label is attached to the district. Post-Fiscal Stress corresponds to the time period after the label has been removed. Never in Fiscal Stress corresponds to all non-labeled school districts. Standard Deviations are given in parentheses. Per-pupil variables are given in 2010 \$.

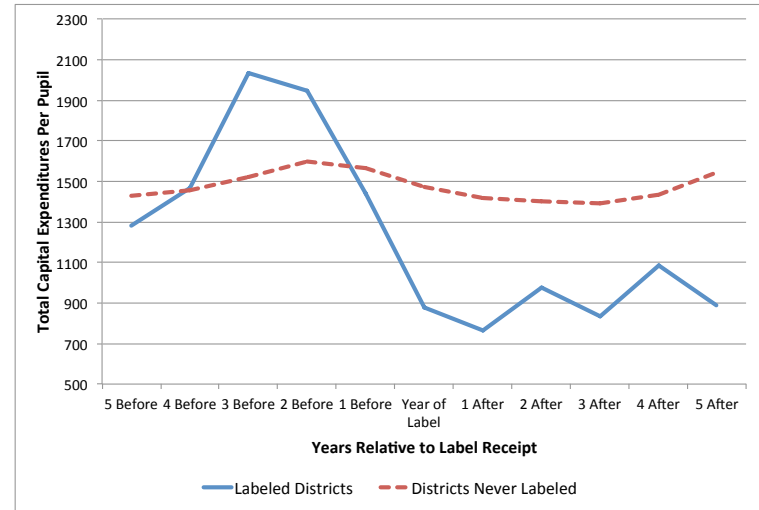
†Panel B includes districts that appear in both the pre-FS period and the FS period.

‡Panel C includes districts that appear in both the FS period and the post-FS period.

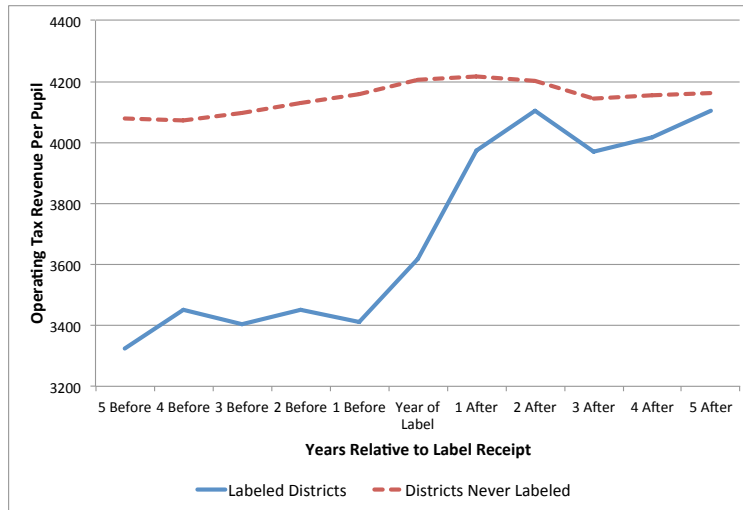
Figure A.3 Operating and Capital Per Pupil Expenditures and Revenues (2010 \$)



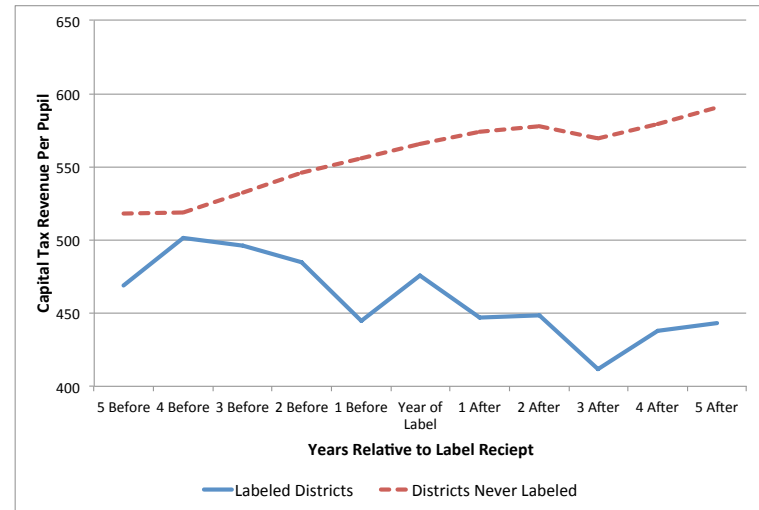
(a) Operating Expenditures Per Pupil



(b) Capital Expenditures Per Pupil



(c) Operating Tax Revenues Per Pupil



(d) Capital Tax Revenues Per Pupil

Table A.4 Table of Means - Housing and Parcel Characteristics

	Panel A	Panel B[†]		Panel C[‡]	
	Never in Fiscal Stress	Pre-Fiscal Stress	Fiscal Stress	Fiscal Stress	Post-Fiscal Stress
Sale Price (2010 \$)	147,171 (64,913)	139,979 (57,058)	133,267 (58,147)	140,687 (57,626)	130,080 (56,211)
Rooms	6.32 (0.62)	6.18 (0.59)	6.15 (0.52)	6.18 (0.52)	6.18 (0.50)
Bedrooms	3.07 (0.25)	3.01 (0.28)	3.03 (0.21)	3.02 (0.21)	3.03 (0.21)
Total Baths	1.66 (0.39)	1.63 (0.41)	1.61 (0.33)	1.62 (0.35)	1.63 (0.36)
Living Area (Sq. Ft.)	1,704 (330)	1,587 (298)	1,585 (335)	1,584 (289)	1,622 (295)
Year Home Built	1961 (17)	1960 (18)	1959 (17)	1960 (18)	1960 (18)
School Districts	409	73	73	60	60

Note: Pre-Fiscal Stress corresponds to period before label receipt. Fiscal Stress corresponds to the time a fiscal stress label is attached to the district. Post-Fiscal Stress corresponds to the period after the label is removed. Never in Fiscal Stress corresponds to all non-labeled jurisdictions. Standard Deviations are given in parentheses.

[†]Panel B includes districts that appear in both the pre-FS and FS periods.

[‡]Panel C includes districts that appear in both the FS and post-FS periods.

Table A.5 Effect of Fiscal Stress Receipt on Expenditures per pupil, by severity level

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Expend	Operating Expend	Total Salaries	Employee Benefits	Capital Expend	New Construct
No Label to Fiscal Oversight	-0.096*** (0.023)	-0.037*** (0.009)	-0.054*** (0.009)	-0.026** (0.012)	-0.558*** (0.152)	-0.733*** (0.258)
Fiscal Oversight to No Label	0.021 (0.017)	0.017* (0.009)	0.022*** (0.008)	0.015 (0.014)	0.191 (0.137)	0.010 (0.247)
Fiscal Oversight to Fiscal Emergency	-0.020 (0.046)	-0.023 (0.018)	-0.029* (0.017)	0.018 (0.025)	-0.181 (0.328)	-0.042 (0.610)
Fiscal Emergency to No Label	-0.011 (0.025)	0.010 (0.013)	0.017 (0.012)	-0.055** (0.025)	0.078 (0.242)	0.204 (0.477)
GF Balance/Rev Ratio	-0.004 (0.037)	-0.019 (0.015)	-0.035** (0.015)	-0.109*** (0.021)	0.575** (0.280)	0.758** (0.384)
Ratio \leq -0.02	0.026* (0.013)	0.020*** (0.005)	0.017*** (0.006)	0.027*** (0.008)	0.035 (0.089)	0.255* (0.139)
Constant	8.877*** (0.010)	8.720*** (0.004)	8.263*** (0.004)	7.015*** (0.007)	6.098*** (0.063)	5.030*** (0.106)
Observations	7,337	7,337	7,337	7,337	7,337	7,337
R-squared	0.58	0.94	0.94	0.91	0.25	0.30

Dependent variables are natural logs of the listed per-pupil variables. Each specification contains controls for school district time varying demographics, school district fixed effects, and year fixed effects. Robust standard errors, clustered at the school district level given in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table A.6 Effect of Fiscal Stress Receipt on Revenues per pupil, by severity level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total Revenue	Federal Revenue	State Revenue	Local Revenue	Local Prop Tax Rev	Oper Prop Tax Rev	Cap Prop Tax Rev
No Label to Fiscal Oversight	0.001 (0.021)	-0.053** (0.022)	-0.032 (0.031)	0.016 (0.015)	0.049** (0.019)	0.065*** (0.020)	-0.049 (0.061)
Fiscal Oversight to No Label	0.010 (0.020)	-0.050* (0.027)	-0.006 (0.032)	0.023 (0.020)	-0.013 (0.021)	-0.003 (0.025)	-0.102 (0.066)
Fiscal Oversight to Fiscal Emergency	-0.001 (0.041)	0.013 (0.046)	-0.039 (0.063)	0.026 (0.028)	0.067 (0.045)	0.092** (0.047)	-0.043 (0.081)
Fiscal Emergency to No Label	-0.089*** (0.029)	-0.081*** (0.031)	-0.077 (0.053)	-0.075*** (0.027)	-0.070*** (0.025)	-0.054* (0.028)	-0.098 (0.085)
GF Balance/Rev Ratio	0.158*** (0.029)	0.104*** (0.039)	0.138*** (0.044)	0.187*** (0.032)	0.049* (0.028)	0.046* (0.028)	0.144* (0.083)
Ratio \leq -0.02	0.006 (0.010)	0.010 (0.013)	0.005 (0.015)	0.005 (0.010)	0.010 (0.009)	0.011 (0.009)	0.016 (0.033)
Constant	8.875*** (0.009)	5.560*** (0.014)	8.010*** (0.013)	8.118*** (0.008)	7.958*** (0.007)	7.834*** (0.007)	5.690*** (0.027)
Observations	7,336	7,329	7,336	7,336	7,336	7,336	6,749
R-squared	0.70	0.91	0.74	0.93	0.96	0.97	0.77

Dependent variables are natural logs of the listed per-pupil variables. Each specification contains controls for school district time varying demographics, school district fixed effects, and year fixed effects. Robust standard errors, clustered at the school district level given in parentheses.

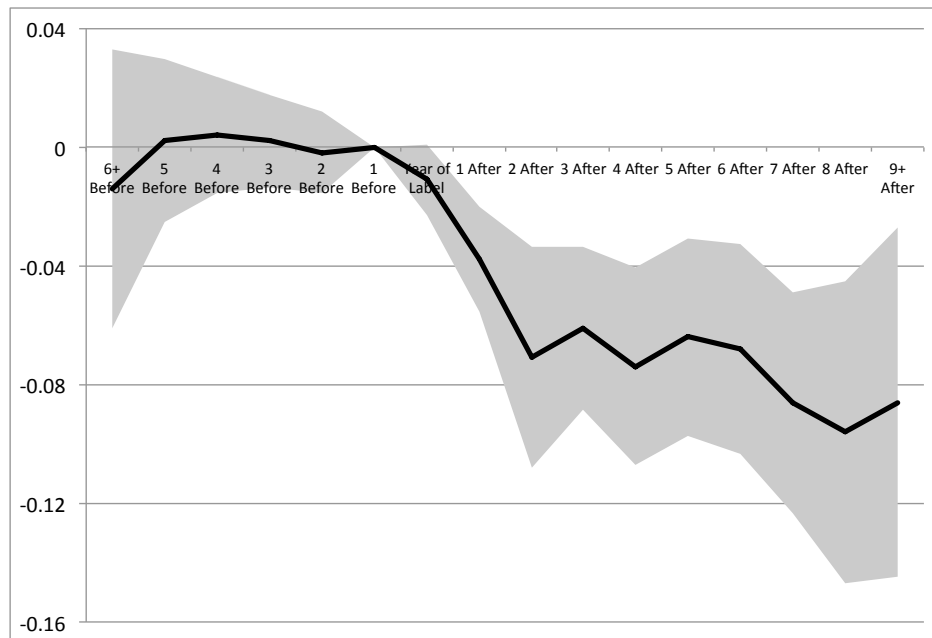
*** p<0.01, ** p<0.05, * p<0.1

Table A.7 Effect of School District Fiscal Stress Label Receipt on Housing Prices

	(1)	(2)	(3)
	ln(Parcel Sale Price)		
No Label to Fiscal Oversight	0.012 (0.035)	0.013 (0.035)	-0.032 (0.031)
Fiscal Oversight to No Label	-0.065 (0.057)	-0.064 (0.057)	-0.029 (0.025)
Fiscal Oversight to Fiscal Emergency	-0.056* (0.029)	-0.056* (0.029)	-0.012 (0.026)
Fiscal Emergency to No Label	0.071*** (0.015)	0.071*** (0.015)	0.066*** (0.013)
GF Balance/Rev Ratio	-0.019 (0.050)	-0.018 (0.050)	-0.003 (0.030)
District FE	X	X	X
Time (Month-Year) FE	X	X	X
County FE		X	
County-Year FE			X
Observations	1,011,726	1,011,726	1,011,726
R-squared	0.54	0.54	0.55

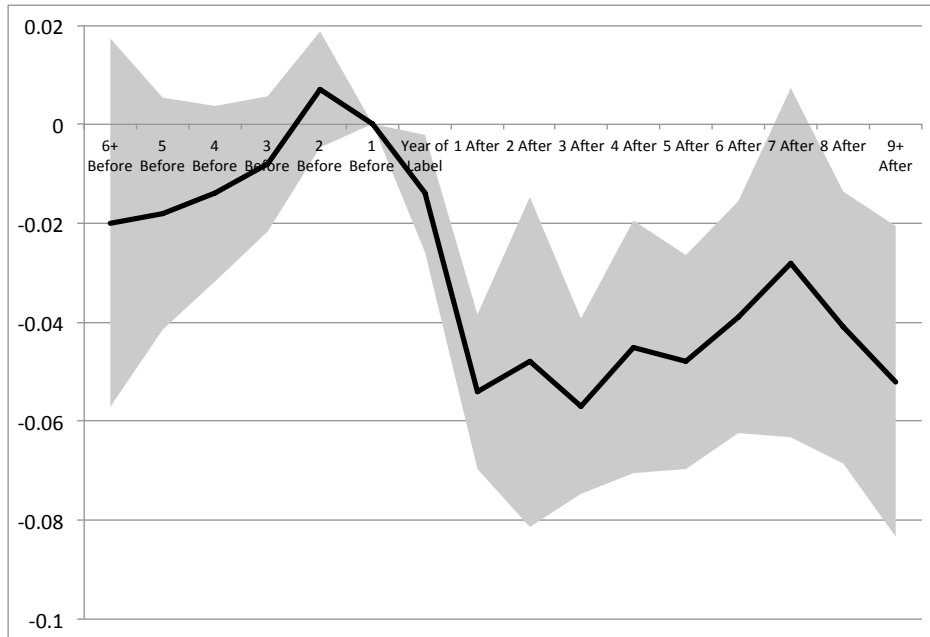
Each specification contains controls for school district financial statistics and demographics, parcel characteristics, and set of fixed effects. Robust standard errors, clustered at the school district level given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure A.4 Event Study Results - District Enrollment



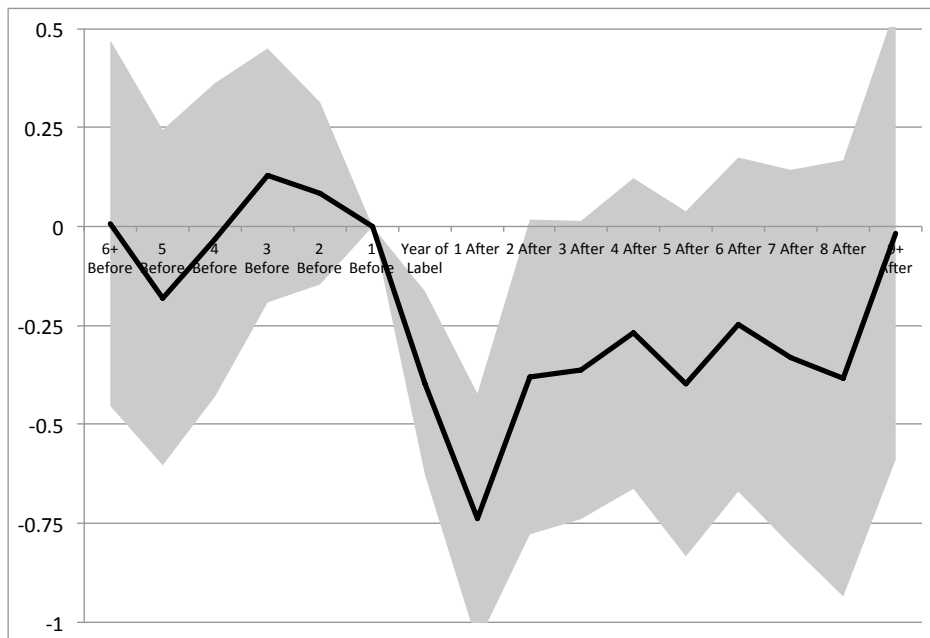
Grey shaded region depicts the 95% confidence interval for the estimated effect in each year relative to label receipt. Y-axis gives point estimate from equation (3) using ln(enrollment) as dependent variable.

Figure A.5 Event Study Results - Total Operating Expenditures PP



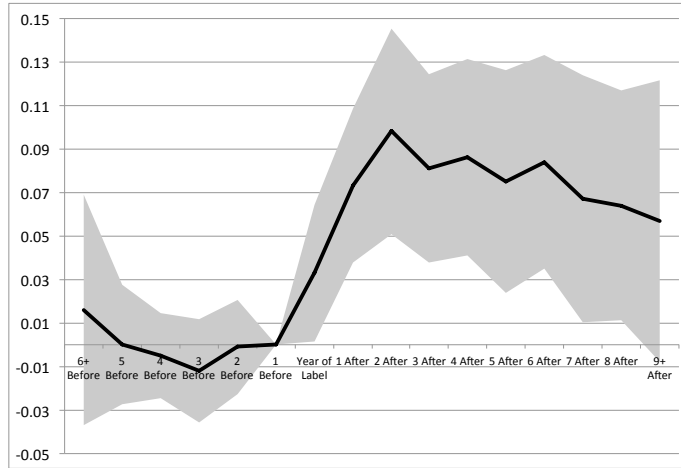
Grey shaded region depicts the 95% confidence interval for the estimated effect in each year relative to label receipt. Y-axis gives point estimate from equation (3) using $\ln(\text{operating expenditures PP})$ as dependent variable.

Figure A.6 Event Study Results - Total Capital Expenditures PP

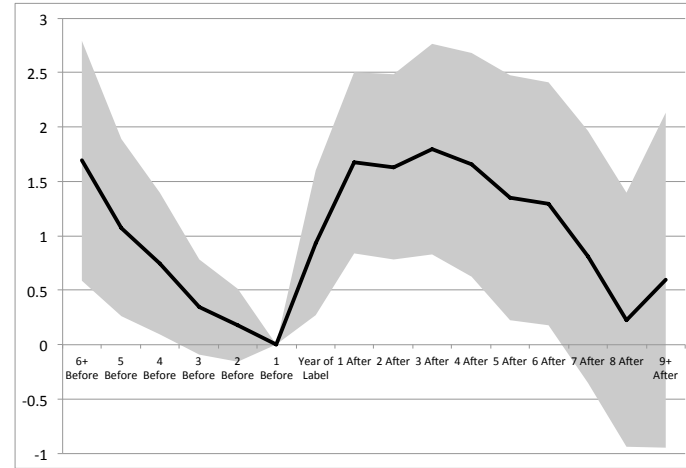


Grey shaded region depicts the 95% confidence interval for the estimated effect in each year relative to label receipt. Y-axis gives point estimate from equation (3) using $\ln(\text{capital expenditures PP})$ as dependent variable.

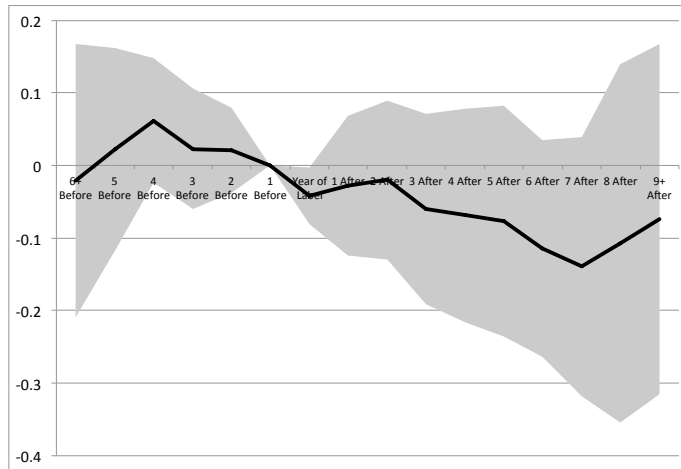
Figure A.7 Event Study Results - Local Property Tax Revenue and Millage Rates



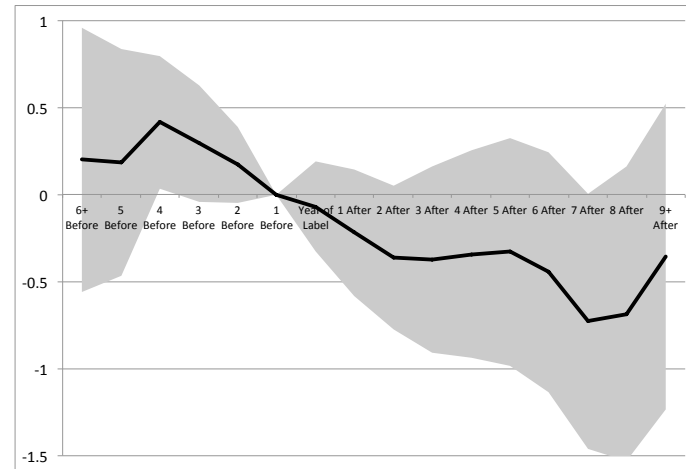
(a) Total Local Operating Property Tax Revenue PP



(b) Total Local Operating Property Tax Millage Rate



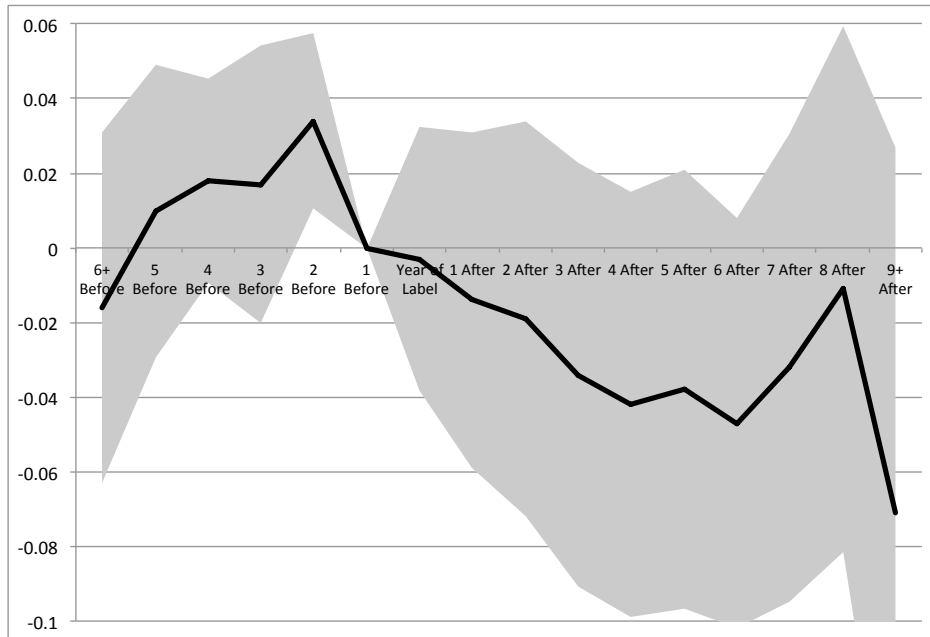
(c) Total Local Capital Property Tax Revenue PP



(d) Total Local Capital Property Tax Millage Rate

Grey shaded region depicts the 95% confidence interval for the estimated effect in each year relative to label receipt. Y-axis gives point estimate from equation (3) using $\ln(\text{local tax revenue})$ or millage rate as dependent variable.

Figure A.8 Event Study Results: Housing Prices



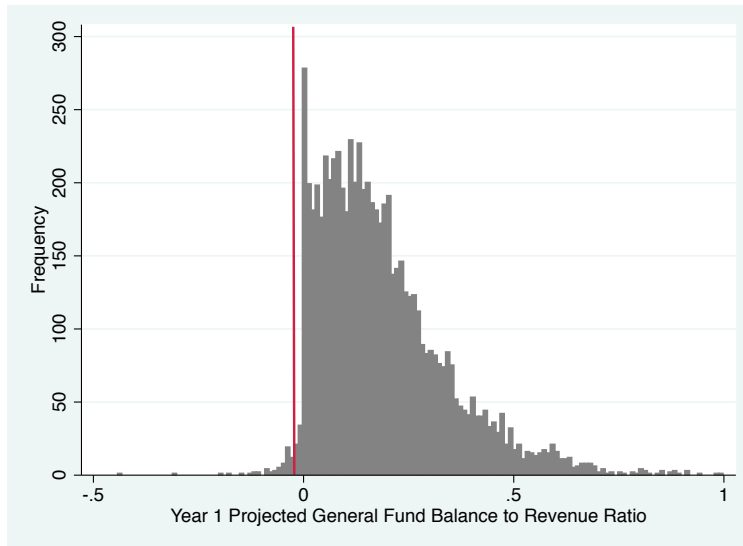
Grey shaded region depicts the 95% confidence interval for the estimated effect in each year relative to label receipt. Y-axis gives point estimate from equation (4) using $\ln(\text{parcel sale price})$ as dependent variable.

Table A.8 Estimated Discontinuity at Year 3 Ratio Cutoff

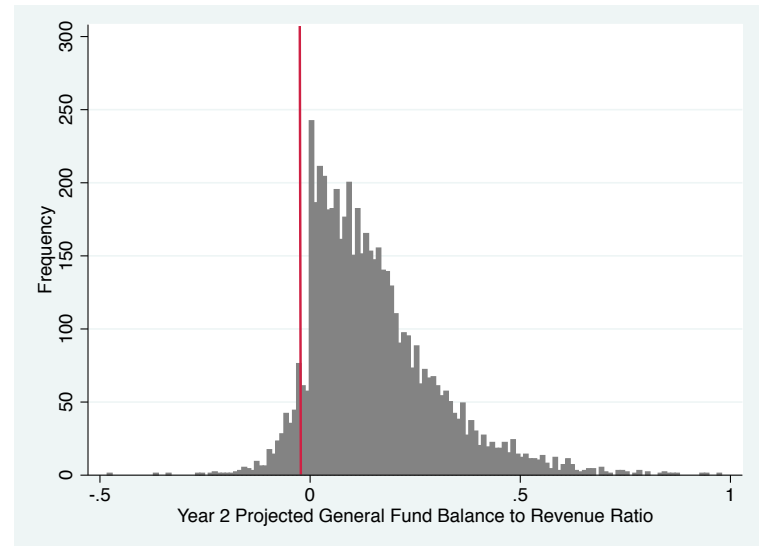
Enrollment	Population	Poverty 5-17
-13.74 (31.85)	24.81 (90.16)	35.48 (30.06)
FRL Students	Black Students	White Students
-284.39 (238.70)	-13.26 (19.30)	-0.61 (21.58)

Dependent variable is change from year t-1 to year t for the listed variable. Estimated discontinuity from regression of the given variable on 0.02 width bins of the year 3 projected ratio around the -0.02 cutoff. Each cell gives the estimate of the -0.02 to 0 bin relative to the -0.02 to -0.04 bin. Robust standard errors given in parentheses.
 *** p<0.01, ** p<0.05, * p<0.1

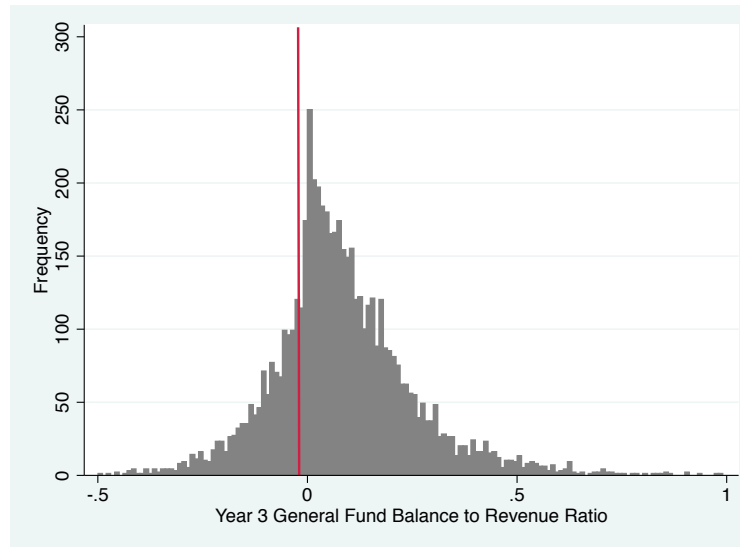
Figure A.9 Distribution of Projected General Fund Balance to Revenue Ratios



(a) Year 1 Projected Ratio

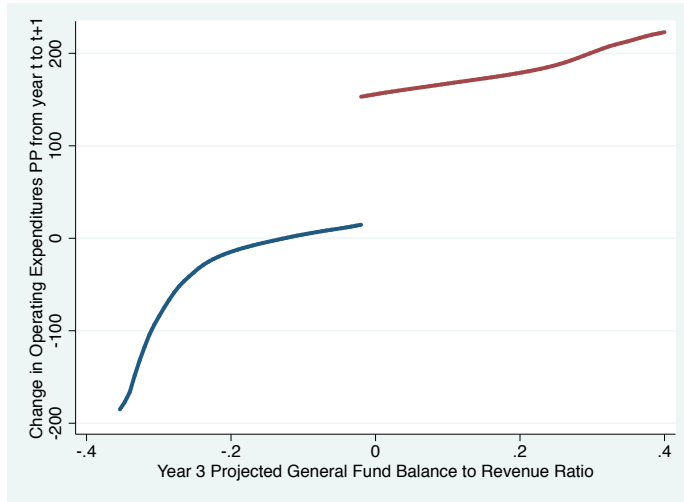


(b) Year 2 Projected Ratio

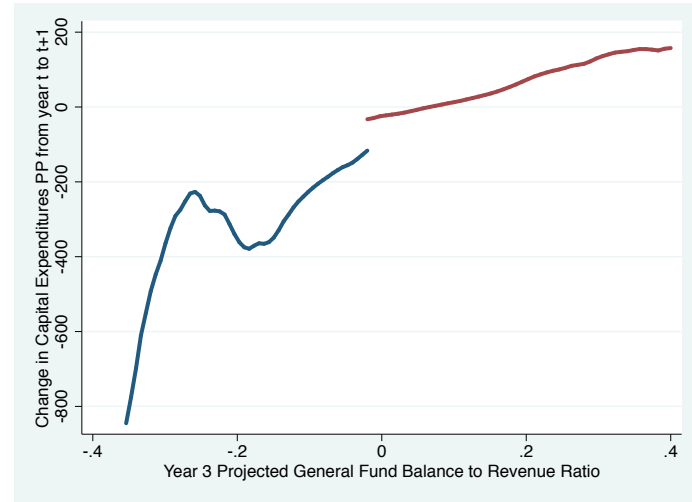


(c) Year 3 Projected Ratio

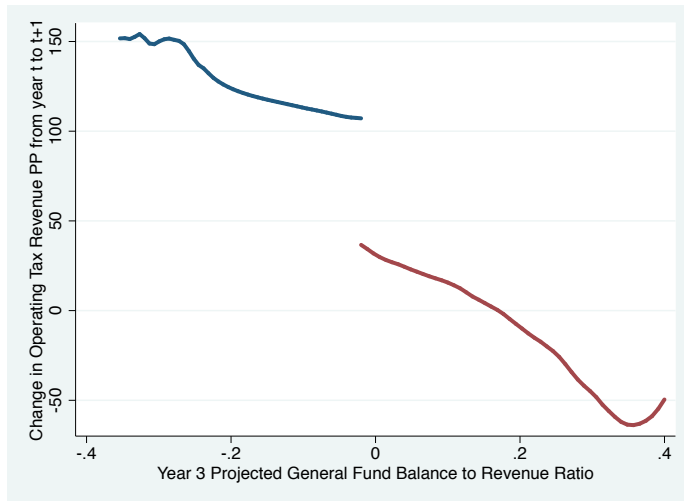
Figure A.10 Change in District Finances and Housing Prices, by Year 3 Projected Ratio



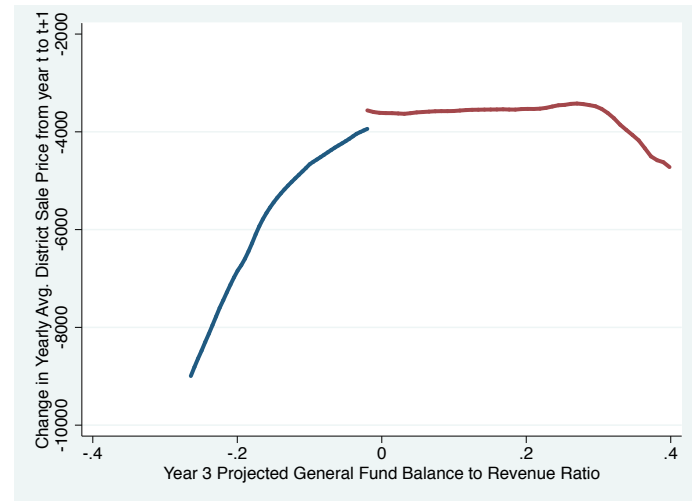
(a) Operating Expenditures per Pupil



(b) Capital Expenditures per Pupil



(c) Operating Property Tax Revenue per Pupil



(d) Avg. Yearly Housing Price

Table A.9 Regression Discontinuity Results

Panel A: Operating Expenditures per Pupil		
	Year t-1 to Year t	Year t to Year t+1
1[Projected Year 3 Ratio < -0.02]	42.01 (69.25)	-195.15** (50.42)
Projected Year 3 Ratio	275.56*** (70.38)	372.71*** (92.32)
Observations	5,814	5,315
R-squared	0.07	0.14
Panel B: Capital Expenditures per Pupil		
	Year t-1 to Year t	Year t to Year t+1
1[Projected Year 3 Ratio < -0.02]	126.83 (119.29)	-83.13 (116.44)
Projected Year 3 Ratio	857.06*** (316.85)	621.75** (301.84)
Observations	5,814	5,315
R-squared	0.01	0.01
Panel C: Operating Property Tax Revenue per Pupil		
	Year t-1 to Year t	Year t to Year t+1
1[Projected Year 3 Ratio < -0.02]	-9.61 (18.06)	34.94 (22.91)
Projected Year 3 Ratio	-195.08*** (59.12)	-294.18*** (67.64)
Observations	5,813	5,314
R-squared	0.07	0.07
Panel D: Avg. Yearly District Sale Price		
	Year t-1 to Year t	Year t to Year t+1
1[Projected Year 3 Ratio < -0.02]	345.62 (1,220.84)	2,002.13* (1,206.59)
Projected Year 3 Ratio	-1,627.81 (3,530.90)	743.35 (3,941.61)
Projected Year 3 Ratio*	7,910.29	27,409.54*
1[Projected Year 3 Ratio < -0.02]	(13,840.46)	(14,429.34)
Observations	4,286	3,929
R-squared	0.24	0.25

Dependent variable is yearly change in the listed variable. Each specification contains controls for yearly changes in school district financial characteristics, parcel characteristics, and a flexible polynomial of the projected year 3 ratio. Robust standard errors, clustered at the school district level given in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A.10 Variable Names and Definitions

Variable Name	Description
Fiscal Stress Indicators†	
Fiscal Oversight	= 1 if school district is labeled in fiscal caution or fiscal watch at time t
Fiscal Emergency	= 1 if school district is labeled in fiscal emergency at time t
District Financials†	
GF Balance/Rev Ratio	The end of year balance in the general fund divided by general fund revenue
Ratio < -0.02	=1 if the deficit in the gen fund exceeds 2% of gen fund revenue
Total Expend	Total expenditures divided by total enrollment
Operating Expend	Total operating expenditures divided by total enrollment
Total Salaries	Total expenditures on employee salaries divided by total enrollment
Employee Benefits	Total expenditures on employee benefits divided by total enrollment
Capital Expend	Total capital expenditures divided by total enrollment
New Construct	Total expenditures on new construction projects divided by total enrollment
Total Revenue	Total revenue divided by total enrollment
Federal Revenue	Total federal revenue divided by total enrollment
State Revenue	Total state revenue divided by total enrollment
Local Revenue	Total local revenue divided by total enrollment
Local Prop Tax Rev	Total local property tax revenue divided by total enrollment
Oper Prop Tax Rev	Total local property tax revenue for operating expenditure divided by total enrollment
Cap Prop Tax Rev	Total local property tax revenue for capital expenditure divided by total enrollment
Local Inc Tax Rev	Total local income tax revenue divided by total enrollment
District Demographics•	
Total Enrollment	Total district enrollment
Proportion Free/Reduced Lunch	Proportion of students that are eligible for free and reduced price lunch
Number of Schools in District	Total number of schools in the district
FTE Teachers	Total full-time equivalent teachers in the district
Poor School Report Card Grade	School Report Card Grade is below average
Urban District	= 1 if school district is in an urban area
Parcel Characteristics*	
ln(Sale Price)	The natural log of the sale price
Rooms	The total number of rooms in the house
Bedrooms	The total number of bedrooms in the house
Total Baths	The total number of bathrooms in the house (half bath = 0.5)
Living Area	The total square footage of the living area in the house
Year Home Built	The year when house was originally constructed

† Sources: Data from Ohio Auditor website and Ohio Department of Education website's list of fiscal caution, fiscal watch, and fiscal emergency districts.

‡ Sources: Data from Ohio Auditor, Ohio Department of Taxation, and Common Core of Data

• Sources: Common Core of Data and Ohio Department of Education

* Sources: Data from 62 of 88 counties in Ohio. Data was downloaded from individual county auditor websites using parcel search tools or data set downloads.

Table A.11 Comparison of Full and Analytic Samples

	Full Sample	Analytic Sample
Fiscal Oversight	0.05 (0.21)	0.05 (0.21)
Fiscal Emergency	0.02 (0.13)	0.02 (0.13)
Below Avg Report Card Grade	0.19 (0.40)	0.19 (0.39)
Urban	0.19 (0.39)	0.19 (0.39)
GF Balance/Rev Ratio	0.18 (0.19)	0.18 (0.19)
Ratio \leq -0.02	0.10 (0.30)	0.09 (0.29)
Revenue per Pupil	11,129 (4,402)	11,313 (4,756)
Expenditures per Pupil	10,973 (4,591)	11,159 (4,889)
Total Class I Millage Rate	29.60 (6.56)	30.04 (6.93)
Proportion of Free/Reduced Lunch	0.24 (0.18)	0.23 (0.17)
Proportion Black	0.05 (0.14)	0.06 (0.15)
Total Enrollment	2,888 (4,684)	3,075 (5,170)
Number of Schools in District	5.86 (9.08)	6.22 (10.14)
Pupil Teacher Ratio	17.28 (2.47)	17.27 (2.60)
Proportion Above Math Proficiency	75.24 (11.37)	75.34 (11.51)
Proportion Above Reading Proficiency	0.82 (0.08)	0.82 (0.08)
Observations	6,749	5,209

APPENDIX B CHAPTER 2 TABLES AND FIGURES

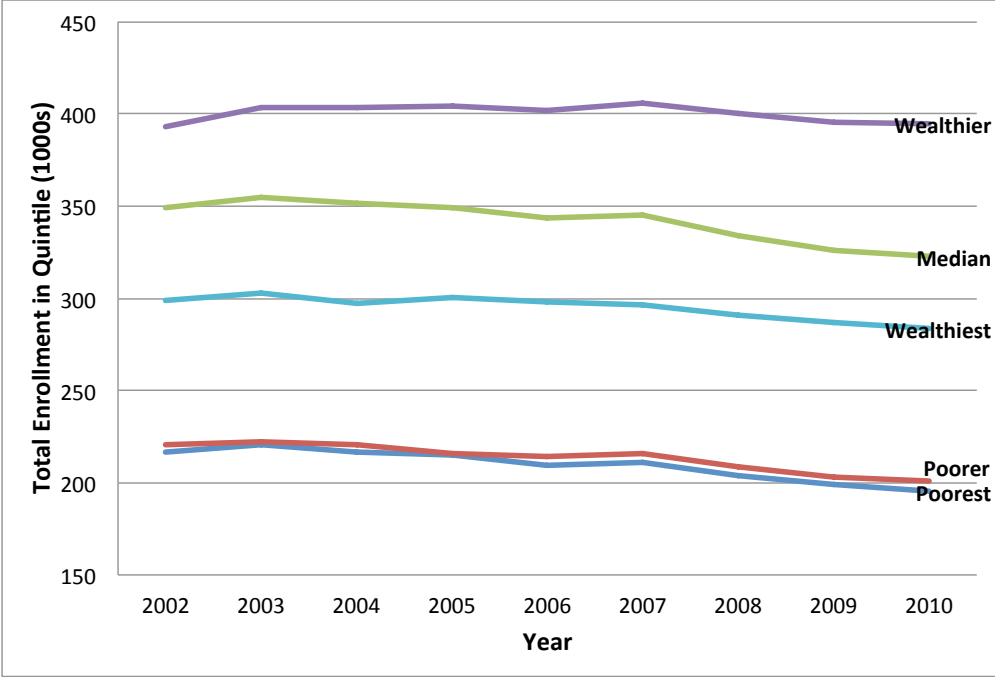
Table B.12 Property Tax Rates and Taxable Values 2002-2010

	Michigan	Ohio	Class I	Class II
School District Millage		School District Millage		
Debt	4.20 (2.85)	Debt	3.19 (2.61)	3.19 (2.61)
Hold-Harmless	0.33 (1.85)	Current Expenses	23.06 (5.96)	26.69 (9.42)
Sinking Fund	0.33 (0.75)	Emergency Operating	2.48 (3.97)	2.48 (3.97)
Recreational	0.02 (0.21)	Permanent Improvements	0.88 (1.03)	1.00 (1.13)
Non-Homestead	17.62 (1.72)	Classroom Facilities	0.14 (0.21)	0.14 (0.22)
Intermediate SD Millage				
Debt	0.003 (0.03)			
Vocational Education	0.76 (0.86)			
Special Education	2.67 (1.03)			
Enhancement	0.03 (0.19)			
Operating	0.21 (0.13)			
School District Taxable Values (in Millions)				
Homestead	349 (578)	Class I		269 (402)
Non-Homestead	218 (391)	Class II		79 (219)
Industrial Personal Property	16.2 (110)	Tangible Personal Property		28 (61)
Commercial Personal Property	2.00 (16.2)	Tangible Public Utility Property		15 (28)

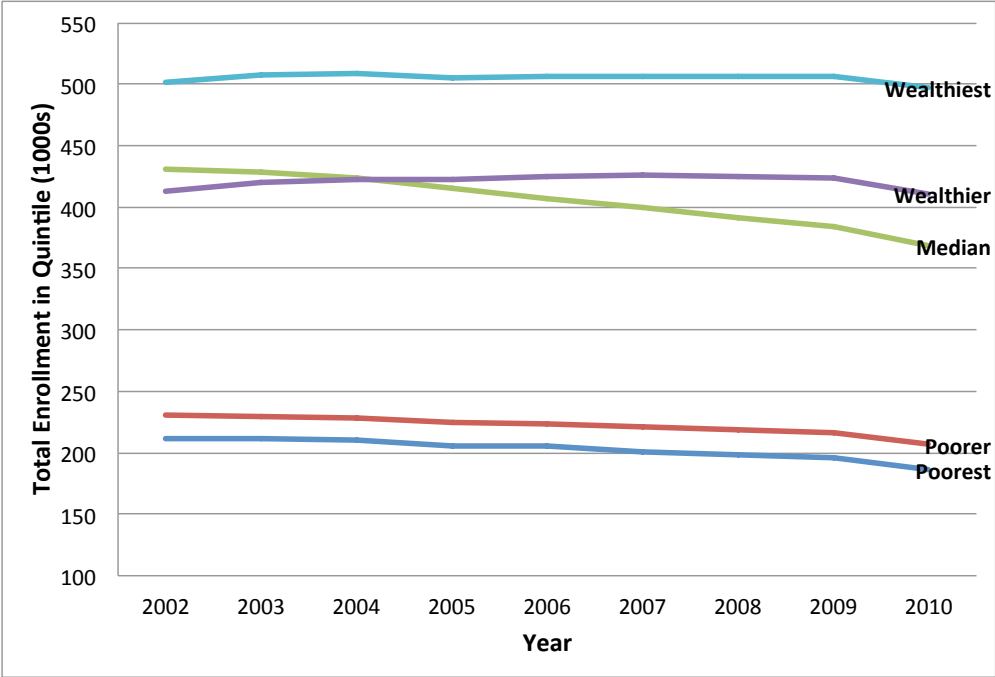
Table B.13 Revenue and Demographic Characteristics

	Michigan	Ohio
Students	2774 (4714)	2870 (4552)
Total Expenditures (in Millions)	32.6 (63.40)	33.8 (67.90)
Total Revenue (in Millions)	31.2 (59.80)	34 (67.40)
State Revenue (in Millions)	23.1 (41.50)	14.8 (32.30)
Local Revenue (in Millions)	6.22 (12.20)	16.8 (30.60)
Taxable Value per Student (in Thousands)	307 (908)	133 (159)
Total Local Operating Revenue (in Millions)	0.953 (3.61)	13.9 (27.70)
Total Local Capital Revenue (in Millions)	2.76 (5.57)	1.83 (3.78)
Total Local Income Tax Revenue (in Millions)		0.4 (1.07)
Special Education Students	117 (323)	396 (714)
White Students	2134 (2580)	2251 (2112)
Free or Reduced Lunch Students	939 (2930)	824 (2749)
Population (in Thousands)	17.5 (36.20)	18.7 (36.20)
Population \geq Ages 5-17	3165 (7423)	3314 (6301)
Population in Poverty \geq Ages 5-17	462 (2583)	489 (1851)
Number of Schools	6.11 (8.92)	5.81 (8.98)
Number of Title I Schools	2.97 (7.64)	4.09 (8.20)
Number of Observations	4953	5523

Figure B.11 Total Enrollment of Quintiles

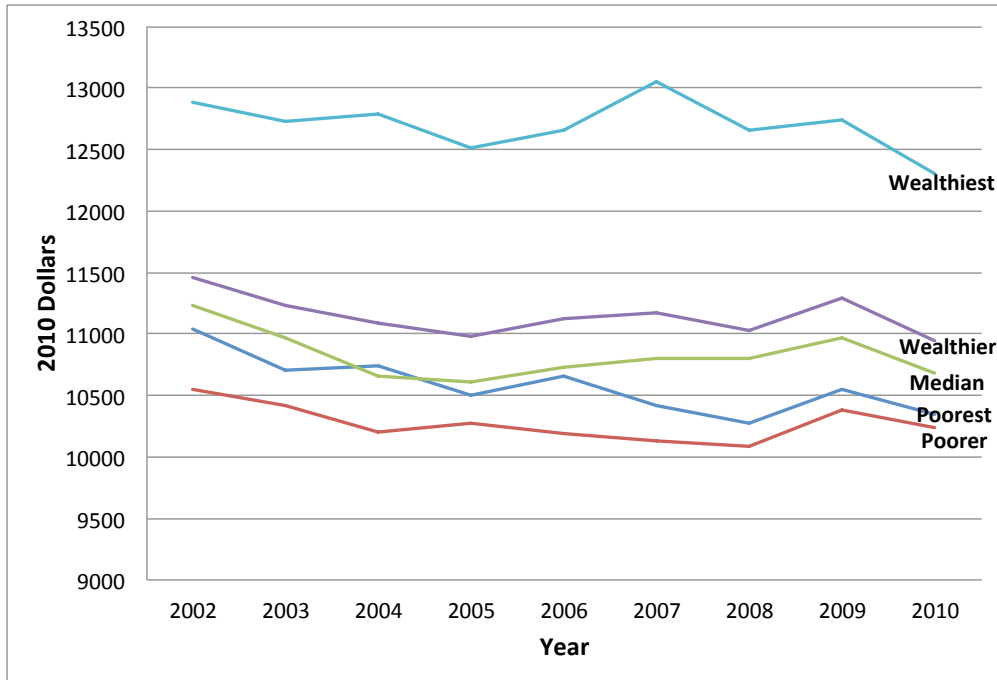


(a) Michigan

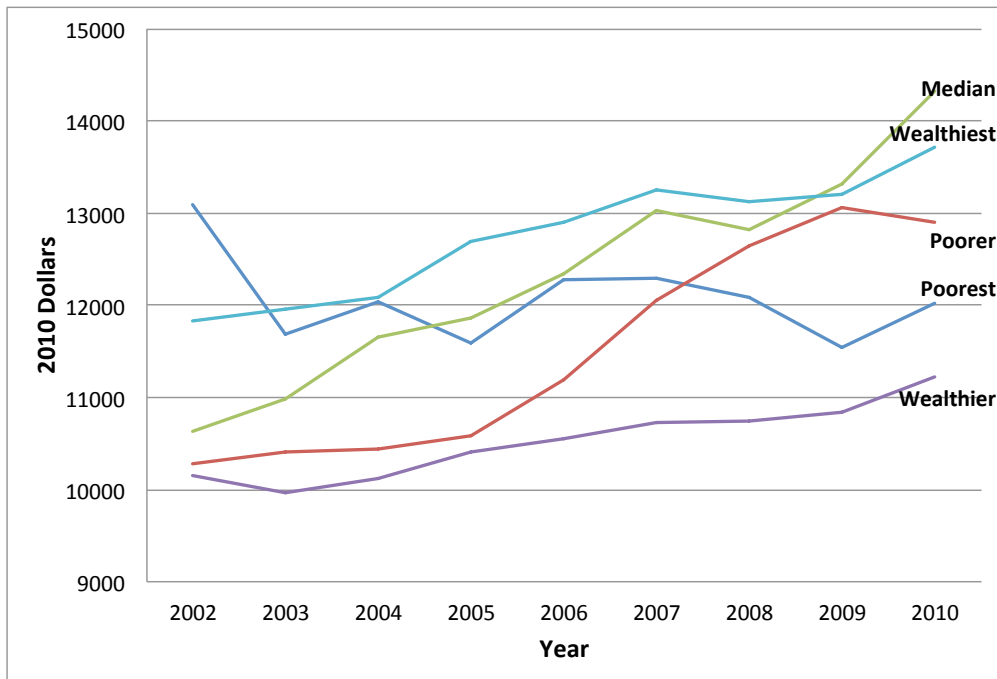


(b) Ohio

Figure B.12 Total Revenue Per Pupil

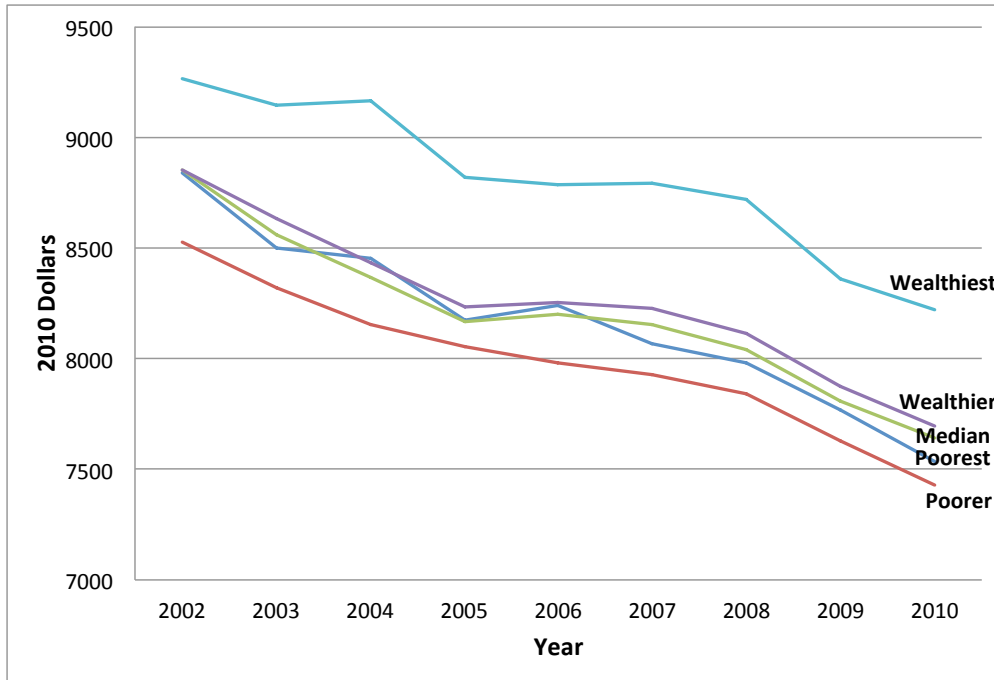


(a) Michigan

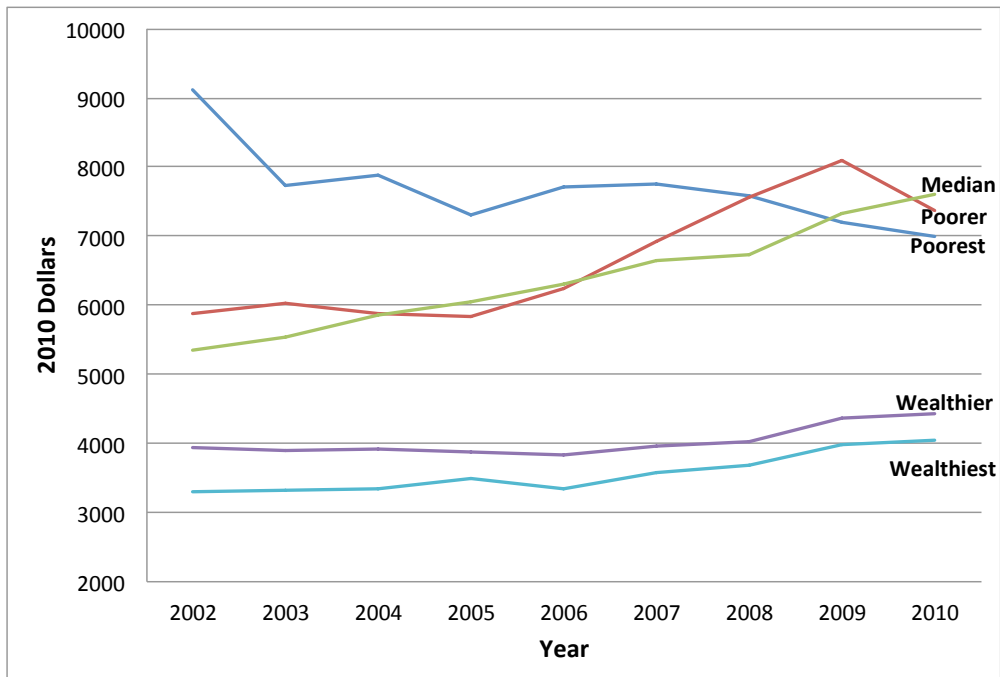


(b) Ohio

Figure B.13 State Revenue Per Pupil

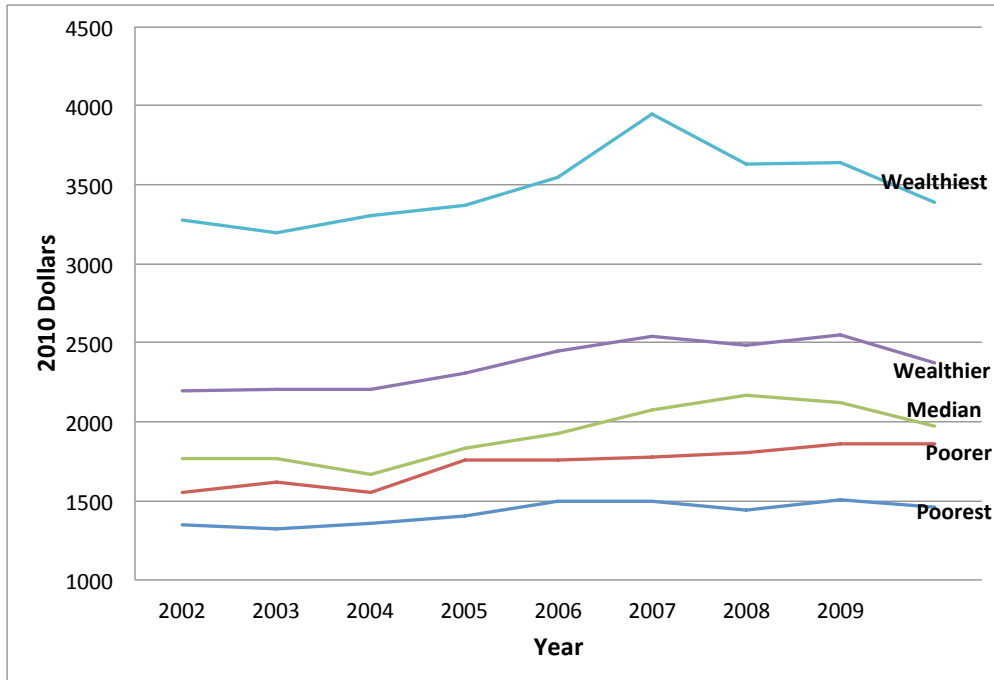


(a) Michigan

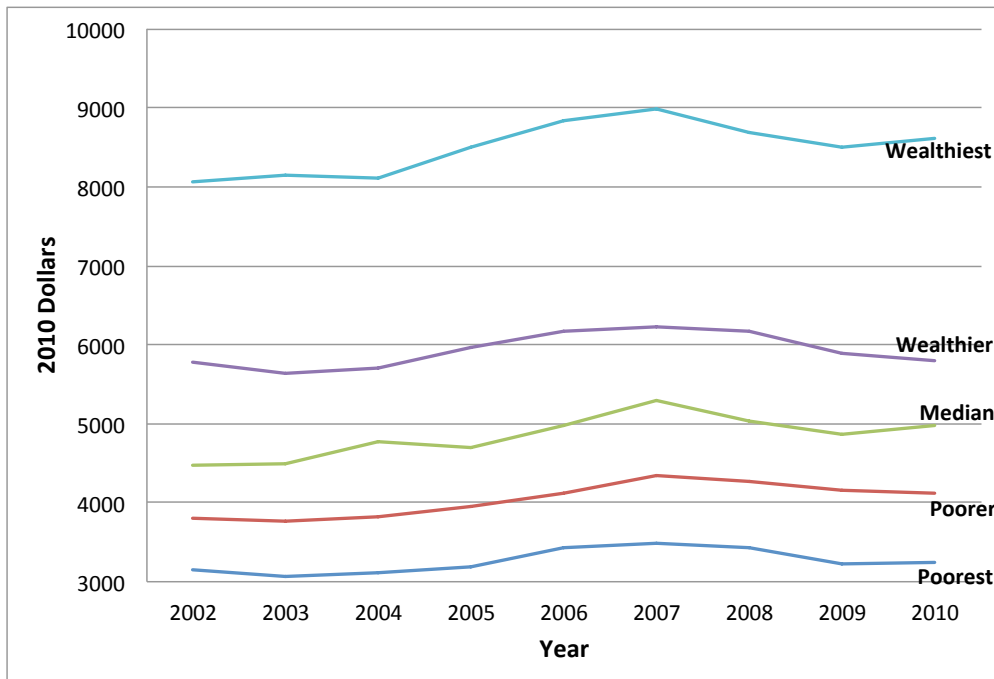


(b) Ohio

Figure B.14 Local Revenue Per Pupil

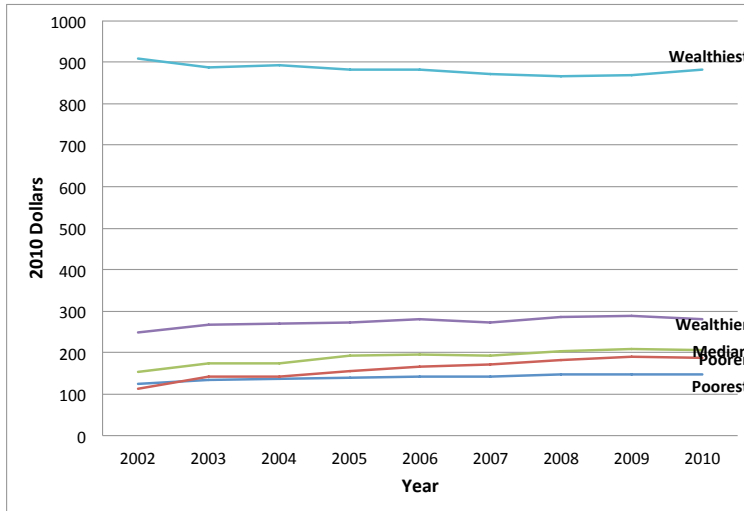


(a) Michigan

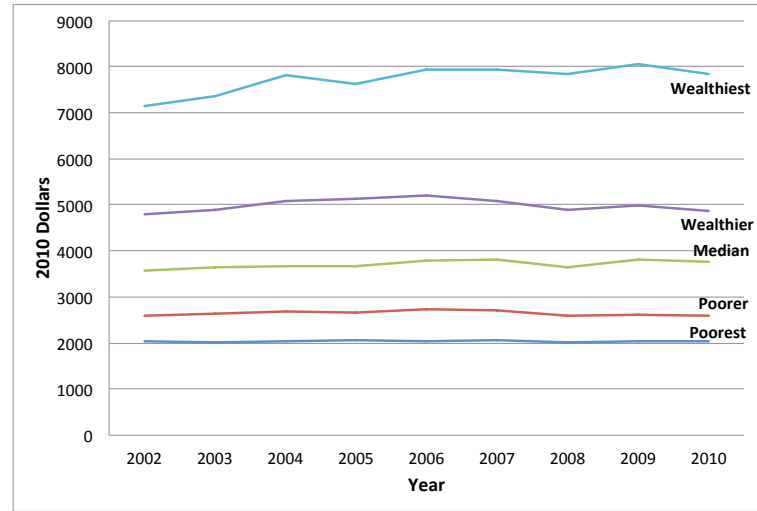


(b) Ohio

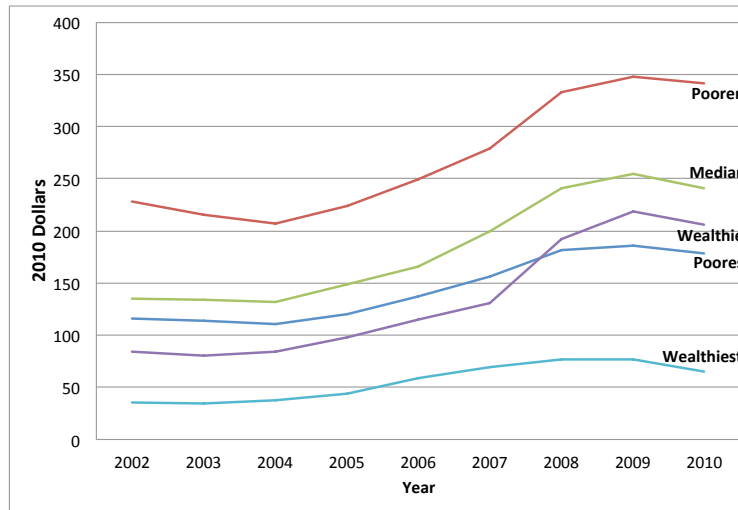
Figure B.15 Local Operating Tax Revenue Per Pupil



(a) Michigan Property Tax



(b) Ohio Property Tax



(c) Ohio Income Tax

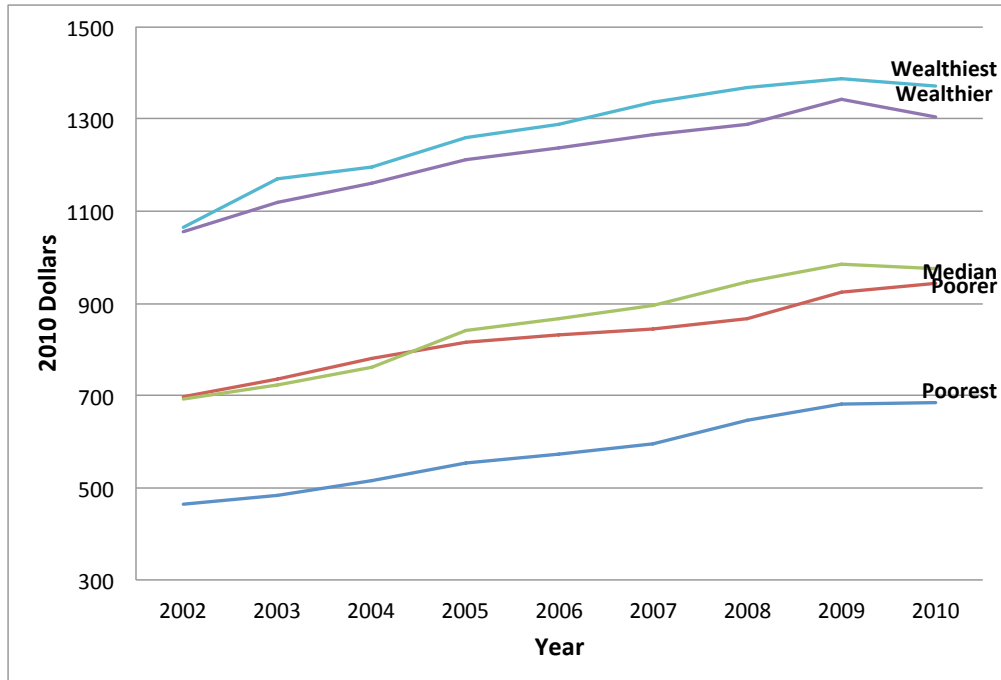
Table B.14 Revenue Regression Results Without District Fixed Effects

	Total Revenue		State Revenue		Local Revenue	
	Michigan	Ohio	Michigan	Ohio	Michigan	Ohio
ln(Total Taxable Value)	0.146** (0.017)	0.139** (0.029)	0.103** (0.015)	-0.647** (0.047)	0.386** (0.050)	0.849** (0.025)
ln(Enrollment)	0.730** (0.072)	0.751** (0.055)	0.733** (0.073)	1.151** (0.081)	0.763** (0.133)	0.353** (0.060)
ln(Spec Ed Enrollment)	0.038** (0.012)	0.153** (0.025)	0.023* (0.009)	0.143** (0.036)	0.109** (0.039)	0.081** (0.026)
ln(White Enrollment)	-0.093** (0.018)	-0.123** (0.026)	-0.052** (0.009)	-0.070** (0.018)	-0.130* (0.051)	-0.148** (0.033)
ln(Free/Reduced Lunch Enroll)	-0.007 (0.013)	-0.036* (0.016)	0.012 (0.007)	0.011 (0.021)	-0.106* (0.043)	-0.023 (0.016)
Number of Schools	0.005** (0.001)	0.002 (0.003)	0.004** (0.001)	-0.009* (0.004)	0.004 (0.005)	0.020** (0.003)
Number of Title I Schools	-0.003* (0.001)	0.001 (0.003)	-0.003** (0.001)	0.013** (0.004)	-0.003 (0.004)	-0.017** (0.003)
ln(Population)	0.087* (0.039)	-0.046 (0.053)	0.053 (0.028)	0.027 (0.079)	0.113 (0.161)	-0.063 (0.058)
ln(Population under age 18)	0.102 (0.070)	0.078 (0.055)	0.112 (0.071)	0.350** (0.093)	0.166 (0.177)	-0.070 (0.064)
ln(Pop. under age 18 in poverty)	-0.029* (0.013)	0.045** (0.013)	-0.014 (0.011)	0.020 (0.018)	-0.260** (0.042)	-0.058 (0.015)
Constant	7.57** (0.169)	8.34** (0.249)	7.98** (0.148)	16.42** (0.402)	1.70** (0.598)	-0.846** (0.247)
R-squared	0.99	0.95	0.99	0.9	0.87	0.97
Observations	4726	4836	4726	4836	4723	4836

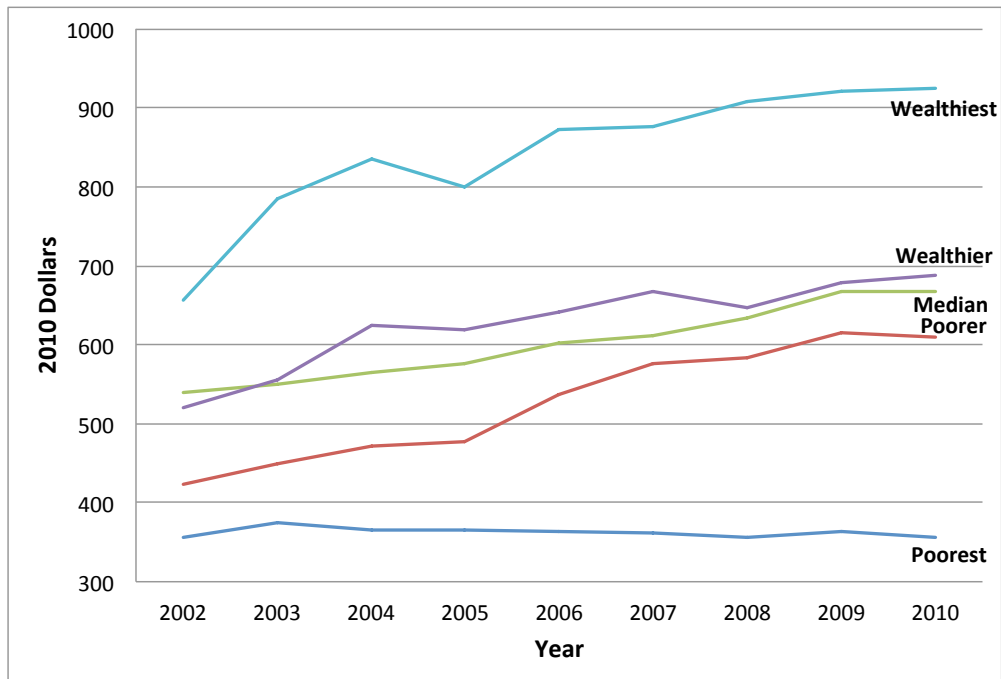
Dependent variables are natural logs of the listed revenue variables. Each specification contains year fixed effects. Robust standard errors, clustered at the school district level given in parentheses.

** p<0.01, * p<0.05

Figure B.16 Local Capital Property Tax Revenue Per Pupil



(a) Michigan



(b) Ohio

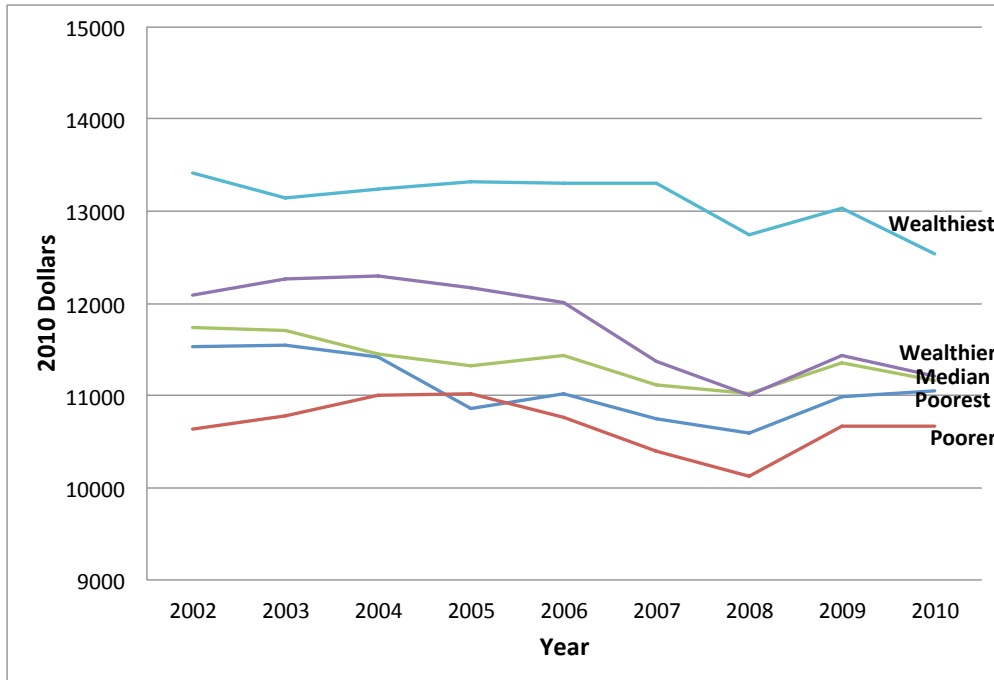
Table B.15 Revenue Regression Results With District Fixed Effects

	Total Revenue		State Revenue		Local Revenue	
	Michigan	Ohio	Michigan	Ohio	Michigan	Ohio
ln(Total Value)	0.245** (0.035)	0.097** (0.037)	0.238** (0.040)	-0.574** (0.060)	0.405** (0.112)	0.671** (0.037)
ln(Enrollment)	0.335** (0.120)	0.414** (0.081)	0.381** (0.136)	0.789** (0.119)	0.103 (0.068)	0.165 (0.089)
ln(Spec Ed Enrollment)	0.027* (0.012)	0.041 (0.023)	0.030* (0.013)	0.048 (0.035)	0.021 (0.017)	0.037* (0.017)
ln(White Enrollment)	-0.005 (0.007)	-0.169** (0.056)	-0.002 (0.007)	-0.347** (0.070)	-0.006 (0.015)	0.001 (0.072)
ln(Free/Reduced Lunch Enroll)	0.037** (0.013)	0.018 (0.015)	0.037* (0.015)	0.054* (0.022)	0.041** (0.014)	0.001 (0.012)
Number of Schools	0.003* (0.002)	0.004 (0.003)	0.005* (0.002)	0.006 (0.004)	-0.004 (0.002)	0.005* (0.002)
Number of Title I Schools	-0.001** (0.000)	0.000 (0.002)	-0.001 (0.000)	0.001 (0.003)	0.001 (0.001)	-0.003 (0.002)
ln(Population)	0.356** (0.092)	-0.116 (0.202)	0.141 (0.106)	-0.063 (0.305)	1.505** (0.343)	0.205 (0.164)
ln(Population under age 18)	0.015 (0.091)	0.515** (0.161)	-0.070 (0.101)	0.551* (0.240)	0.056 (0.241)	0.017 (0.119)
ln(Pop. under age 18 in poverty)	0.025** (0.007)	0.002 (0.010)	0.036** (0.007)	-0.002 (0.014)	-0.011 (0.021)	-0.006 (0.008)
Constant	5.63** (0.690)	9.86** (1.098)	7.73** (0.747)	19.44** (1.687)	-8.42** (2.701)	-0.413 (0.931)
R-squared	0.99	0.97	0.99	0.94	0.97	0.99
Observations	4726	4836	4726	4836	4723	4836

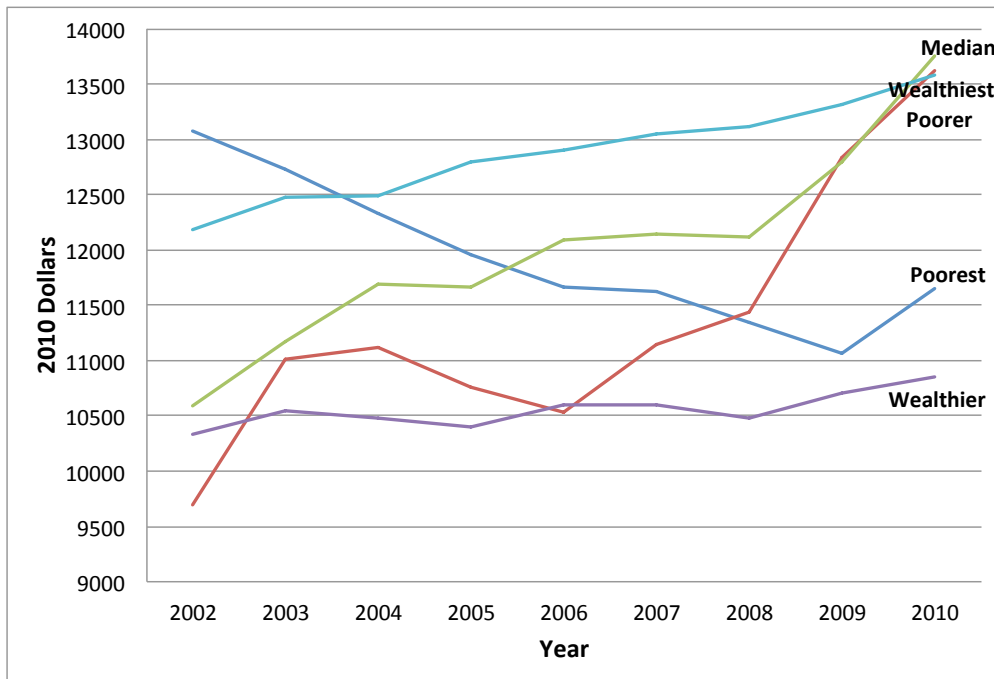
Dependent variables are natural logs of the listed revenue variables. Each specification contains school district fixed effects and year fixed effects. Robust standard errors, clustered at the school district level given in parentheses.

** p<0.01, * p<0.05

Figure B.17 Total Expenditures Per Pupil



(a) Michigan



(b) Ohio

APPENDIX C CHAPTER 3 TABLES AND FIGURES

Table C.16 Simulation Results: Comparing Fixed and Random Teacher Effects Estimators

$\lambda = 0.5$		Four Cohorts			One Cohort		
G-A Mechanism	Evaluation Type	DOLS	AR	EB LAG	DOLS	AR	EB LAG
RA	Rank Correlation	0.85	0.85	0.86	0.65	0.65	0.67
	Misclassification	0.15	0.15	0.15	0.25	0.25	0.26
	Avg. Theta	1.01	1.01	0.78	1.03	1.03	0.49
	Avg. Std. Dev.	0.13	0.14	0.12	0.28	0.27	0.18
	MSE	0.02	0.02	0.01	0.08	0.08	0.03
DG-RA	Rank Correlation	0.85	0.85	0.86	0.64	0.64	0.65
	Misclassification	0.15	0.16	0.16	0.26	0.25	0.25
	Avg. Theta	1.01	0.99	0.77	1.00	0.98	0.45
	Avg. Std. Dev.	0.14	0.14	0.12	0.27	0.27	0.19
	MSE	0.02	0.02	0.01	0.07	0.07	0.03
DG-PA	Rank Correlation	0.86	0.60	0.76	0.63	0.38	0.45
	Misclassification	0.15	0.28	0.22	0.26	0.35	0.48
	Avg. Theta	0.99	0.53	0.49	0.98	0.52	0.16
	Avg. Std. Dev.	0.13	0.19	0.16	0.27	0.30	0.22
	MSE	0.02	0.04	0.02	0.07	0.09	0.05
DG-NA	Rank Correlation	0.85	0.62	0.78	0.67	0.41	0.48
	Misclassification	0.14	0.26	0.20	0.25	0.34	0.47
	Avg. Theta	1.01	0.54	0.53	1.03	0.54	0.17
	Avg. Std. Dev.	0.14	0.19	0.15	0.27	0.29	0.22
	MSE	0.02	0.03	0.02	0.07	0.09	0.05
HG-RA	Rank Correlation	0.72	0.73	0.73	0.58	0.59	0.60
	Misclassification	0.23	0.22	0.23	0.29	0.29	0.30
	Avg. Theta	1.02	1.02	0.86	1.00	0.99	0.54
	Avg. Std. Dev.	0.21	0.21	0.18	0.32	0.31	0.21
	MSE	0.05	0.04	0.03	0.10	0.10	0.04
HG-PA	Rank Correlation	0.94	0.93	0.94	0.81	0.79	0.81
	Misclassification	0.09	0.10	0.10	0.17	0.18	0.19
	Avg. Theta	1.61	1.52	1.43	1.60	1.51	1.06
	Avg. Std. Dev.	0.20	0.19	0.16	0.31	0.30	0.19
	MSE	0.04	0.04	0.03	0.10	0.09	0.04
HG-NA	Rank Correlation	0.39	0.38	0.41	0.26	0.25	0.28
	Misclassification	0.35	0.35	0.38	0.40	0.41	0.55
	Avg. Theta	0.33	0.32	0.15	0.34	0.33	0.06
	Avg. Std. Dev.	0.22	0.23	0.22	0.32	0.32	0.24
	MSE	0.05	0.05	0.05	0.10	0.10	0.06

Note: Rows of each scenario represent the following:
 First - Rank corr. of estimated effects and true effects
 Second - Fraction of above average teachers misclassified as below average
 Third - Average value of $\hat{\theta}$
 Fourth - Average standard deviation of estimated teacher effects across 100 reps
 Fifth - MSE measure

Table C.17 Simulation Results: Comparing Shrunk and Unshrunk Estimators

$\lambda = 0.5$		Four Cohorts				One Cohort			
G-A Mechanism	Evaluation Type	DOLS	SDOLS	AR	SAR	DOLS	SDOLS	AR	SAR
RA	Rank Correlation	0.85	0.85	0.85	0.85	0.65	0.66	0.65	0.66
	Misclassification	0.15	0.15	0.15	0.15	0.25	0.25	0.25	0.25
	Avg. Theta	1.01	1.01	1.01	1.01	1.03	0.99	1.03	0.99
	Avg. Std. Dev.	0.13	0.13	0.14	0.14	0.28	0.26	0.27	0.26
	MSE	0.02	0.02	0.02	0.02	0.08	0.07	0.08	0.07
DG-RA	Rank Correlation	0.85	0.85	0.85	0.85	0.64	0.64	0.64	0.64
	Misclassification	0.15	0.15	0.16	0.16	0.26	0.25	0.25	0.25
	Avg. Theta	1.01	1.01	0.99	0.99	1.00	0.96	0.98	0.94
	Avg. Std. Dev.	0.14	0.14	0.14	0.14	0.27	0.26	0.27	0.25
	MSE	0.02	0.02	0.02	0.02	0.07	0.07	0.07	0.06
DG-PA	Rank Correlation	0.86	0.86	0.60	0.60	0.63	0.63	0.38	0.38
	Misclassification	0.15	0.15	0.28	0.28	0.26	0.27	0.35	0.36
	Avg. Theta	0.99	0.99	0.53	0.53	0.98	0.92	0.52	0.49
	Avg. Std. Dev.	0.13	0.13	0.19	0.19	0.27	0.25	0.30	0.29
	MSE	0.02	0.02	0.04	0.04	0.07	0.06	0.09	0.08
DG-NA	Rank Correlation	0.85	0.85	0.62	0.62	0.67	0.67	0.41	0.41
	Misclassification	0.14	0.14	0.26	0.26	0.25	0.25	0.34	0.34
	Avg. Theta	1.01	1.01	0.54	0.53	1.03	0.97	0.54	0.51
	Avg. Std. Dev.	0.14	0.14	0.19	0.19	0.27	0.25	0.29	0.28
	MSE	0.02	0.02	0.03	0.03	0.07	0.06	0.09	0.08
HG-RA	Rank Correlation	0.72	0.72	0.73	0.73	0.58	0.59	0.59	0.59
	Misclassification	0.23	0.23	0.22	0.22	0.29	0.29	0.29	0.29
	Avg. Theta	1.02	1.02	1.02	1.02	1.00	0.96	0.99	0.96
	Avg. Std. Dev.	0.21	0.21	0.21	0.21	0.32	0.30	0.31	0.30
	MSE	0.05	0.05	0.04	0.04	0.10	0.09	0.10	0.09
HG-PA	Rank Correlation	0.94	0.94	0.93	0.93	0.81	0.81	0.79	0.79
	Misclassification	0.09	0.09	0.10	0.10	0.17	0.17	0.18	0.18
	Avg. Theta	1.61	1.61	1.52	1.52	1.60	1.56	1.51	1.46
	Avg. Std. Dev.	0.20	0.20	0.19	0.19	0.31	0.30	0.30	0.29
	MSE	0.04	0.04	0.04	0.04	0.10	0.09	0.09	0.08
HG-NA	Rank Correlation	0.39	0.40	0.38	0.38	0.26	0.27	0.25	0.26
	Misclassification	0.35	0.35	0.35	0.35	0.40	0.41	0.41	0.41
	Avg. Theta	0.33	0.33	0.32	0.32	0.34	0.32	0.33	0.31
	Avg. Std. Dev.	0.22	0.22	0.23	0.23	0.32	0.30	0.32	0.30
	MSE	0.05	0.05	0.05	0.05	0.10	0.09	0.10	0.09

Note: Rows of each scenario represent the following:

First - Rank corr. of estimated effects and true effects

Second - Fraction of above average teachers misclassified as below average

Third - Average value of $\hat{\theta}$

Fourth - Average standard deviation of estimated teacher effects across 100 reps

Fifth - MSE measure

Figure C.18 Spearman Rank Correlations Across Different VAM Estimators

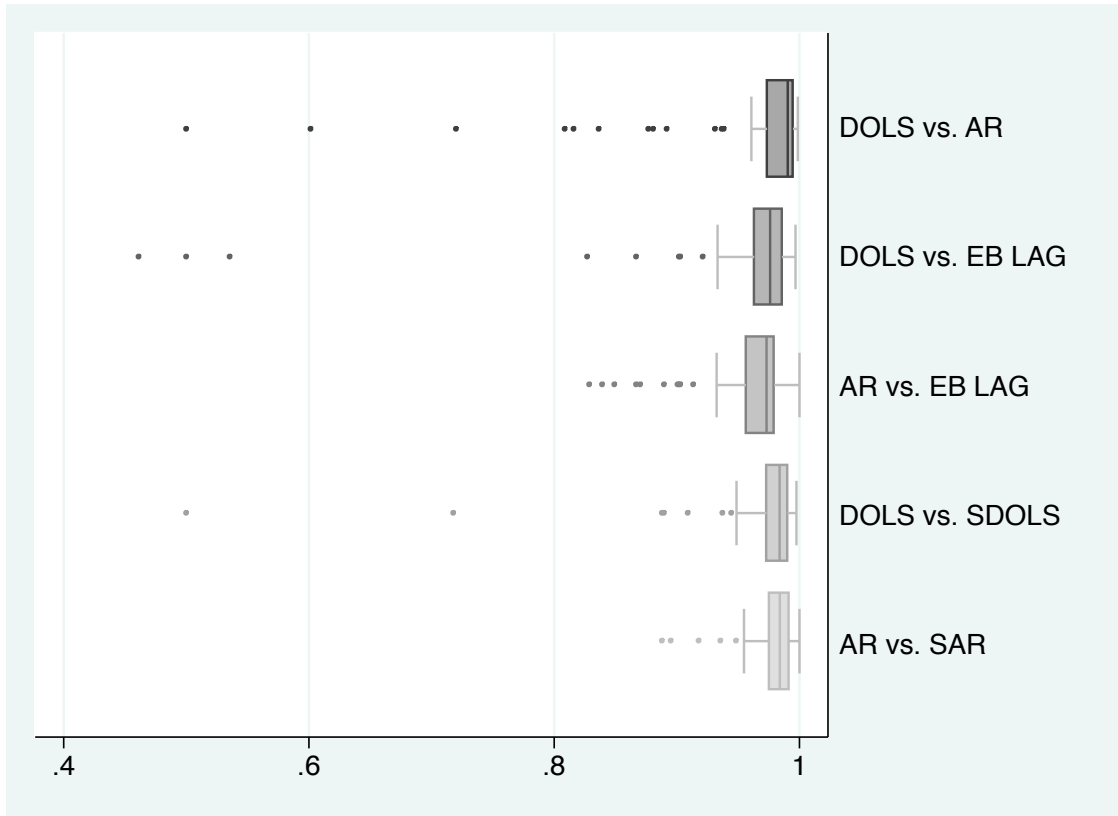


Table C.18 Fraction of Teachers Ranked in Same Quintile by Estimator Pairs

	DOLS	SDOLS	AR	SAR
Top Quintile				
SDOLS	0.91			
AR	0.94	0.89		
SAR	0.89	0.94	0.91	
EB LAG	0.87	0.95	0.86	0.93
Bottom Quintile				
SDOLS	0.89			
AR	0.96	0.88		
SAR	0.88	0.95	0.89	
EB LAG	0.87	0.98	0.86	0.96

Table C.19 Description of Value-Added Estimators

Estimator	Acronym	Description	Teacher Effects
Empirical Bayes'	EB LAG	Two-step approach: Estimate teacher effects using MLE on dynamic equation and then shrink estimates by shrinkage factor	Random
Average Residual	AR	Estimate dynamic equation by OLS and compute residuals for each student. Then compute the average of these residuals for each teacher to get estimated teacher effect	Random
Shrunken Avg. Residual	SAR	Two-step approach: Compute average residual for each teacher using residuals from OLS on dynamic equation. Then shrink average residual for each teacher by shrinkage factor	Random
Dynamic OLS	DOLS	Estimate teacher effects using ordinary least squares on dynamic equation	Fixed
Shrunken DOLS	SDOLS	Two-step approach: Estimate teacher effects using dynamic equation and then shrink estimates by shrinkage factor	Fixed

Table C.20 Definitions of Grouping-Assignment Mechanisms

Name of G-A Mechanism	Acronym	Grouping students in classrooms	Assigning students to teachers
Random Assignment	RA	Random	Random
Dynamic Grouping - Random Assignment	DG-RA	Dynamic (based on prior test scores)	Random
Dynamic Grouping - Positive Assignment	DG-PA	Dynamic (based on prior test scores)	Positive corr. between teacher effects and prior student scores
Dynamic Grouping - Negative Assignment	DG-NA	Dynamic (based on prior test scores)	Negative corr. between teacher effects and prior student scores
Heterogeneity Grouping - Random Assignment	HG-RA	Static (based on student heterogeneity)	Random
Heterogeneity Grouping - Positive Assignment	HG-PA	Static (based on student heterogeneity)	Positive corr. between teacher effects and student fixed effects
Heterogeneity Grouping - Negative Assignment	HG-NA	Static (based on student heterogeneity)	Negative corr. between teacher effects and student fixed effects

Table C.21 Description of Evaluation Measures of Value-Added Estimator Performance

Evaluation Measure	Description
Rank Correlation	Rank correlation between estimated teacher effect and true teacher effect
Misclassification	Fraction of above average teachers that are misclassified as below average
Average Theta	Average value of $\hat{\theta}$
Avg. Std. Dev.	Average standard deviation of estimated teacher effects across the 100 simulation reps
MSE	Average value of $\widehat{MSE} = (\beta_j - \hat{\beta}_j)^2$

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] Ballou, D., Sanders, W., & Wright, P. (2004), Controlling for Student Background in Value-Added Assessment of Teachers, *Journal of Educational and Behavioral Statistics*, 29(1): 37-65
- [2] Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4): 588-638
- [3] Berry, C. (2007). The Impact of School Finance Judgments on State Fiscal Policy. In *School Money Trials: The Legal Pursuit of Educational Adequacy*, edited by M. West and P. Peterson, pp. 213-242. Washington, D.C.: Brookings Institution Press
- [4] Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*, 114(2): 577-599
- [5] Black, S. E., & Machin, S. (2011). Housing valuations of school performance. *Handbook of the Economics of Education*, 3: 485-519
- [6] Brasington, D. M., & Haurin, D. R. (2006). Educational outcomes and house values: a test of the value-added approach. *Journal of Regional Science*, 46(2): 245-268
- [7] Brunner, E., & Sonstelie, J. (2003). School finance reform and voluntary fiscal federalism. *Journal of Public Economics*, 87(9): 2157-2185
- [8] Brunner, E., & Imazeki, J. (2004). Fiscal stress and voluntary contributions to public schools. In *Developments in school finance*, edited by W. J. Fowler, pp. 39-54. Washington, D.C.: National Center for Education Statistics
- [9] Cellini, S. R., Ferreira, F., & Rothstein, J. (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, 125(1): 215-261
- [10] Chetty, R., Freidman, J., and Rockoff, J. (2011), The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood. *NBER*, Working Paper 17699
- [11] Clapp, J. M., Nanda, A., & Ross, S. L. (2008). Which school attributes matter? The influence of school district performance and demographic composition on property values. *Journal of Urban Economics*, 63(2): 451-466
- [12] Corcoran, S. P., Jennings, J., and Beveridge, A. (2011). Teacher effectiveness on high- and low-stakes tests, *Unpublished Draft*
- [13] Corcoran, S. P., & Evans, W. N. (2008). Equity, adequacy, and the evolving state role in education finance. In *Handbook of Research in Education Finance and Policy*, edited H. F. Ladd and E. B. Fiske, pp. 332-356. New York: Routledge

- [14] Corcoran, S. P., Evans, W. N., Godwin, J., Murray, S. E., & Schwab, R. M. (2004). The changing distribution of education finance: 1972-1997. In *Social inequality*, edited by K. M. Neckerman, pp. 433-465. New York: Russell Sage Foundation
- [15] Davidoff, I., & Leigh, A. (2008). How much do public schools really cost? Estimating the relationship between house prices and school quality. *Economic Record*, 84(265): 185-214
- [16] Dougherty, J., Harrelson, J., Maloney L., Murphy, D., Smith, R., Snow, M., & Zannoni D. (2009). School choice in suburbia: test scores, race, and housing markets. *American Journal of Education*, 115(4): 523-548
- [17] Downes, T. A., & Zabel, J. E. (2002). The impact of school characteristics on house prices: Chicago 1987-1991. *Journal of Urban Economics*, 52(1): 1-25.
- [18] Dye, Richard & Reschovsky, A. (2008). Property tax responses to state aid cuts in the recent fiscal crisis. *Public Budgeting Finance*, 28(2): 87-111.
- [19] Fack, G. & Grenet, J.(2010). When do better schools raise housing prices? Evidence from Paris public and private schools. *Journal of Public Economics*, 94(1): 59-77
- [20] Fernandez, R., & Rogerson, R. (2003). Equity and resources: An analysis of education finance systems. *Journal of Political Economy*, 111(4): 858-897
- [21] Figlio, D. N., & Maurice, E. L. (2004). What's in a grade? School report cards and the housing market. *The American Economic Review*, 94(3): 591-604
- [22] Fiva, J. F., & Kirkeboen, L. J. (2010). Information shocks and the dynamics of the housing market. *Scandinavian Journal of Economics*, 113(3): 525-552
- [23] Gibbons, S., & Machin, S. (2003). Valuing English primary schools. *Journal of Urban Economics*, 53(2): 197-219
- [24] Gibbons, S., & Machin, S. (2006). Paying for primary schools: Admission constraints, school popularity or congestion? *Economic Journal*, 116(510): 77-92
- [25] Gibbons, S., Machin, S., & Silva, O. (2009). Valuing school quality, better transport, and lower crime: Evidence from house prices. *Oxford Review of Economic Policy*, 24(1): 99-199
- [26] Guarino, C. M., Reckase, M. D., and Wooldridge, J. M. (2012), Can Value-Added Measures of Teacher Performance Be Trusted?, *Education Policy Center at Michigan State University*, Working Paper 18
- [27] Hahn, J., Todd, P., & Van Der Klaauw, W. (2001). Regression discontinuity. *Econometrica*, 69(1): 201-209
- [28] Hall, J. C. & J. M. Ross. 2010. Tiebout Competition, Yardstick Competition and Tax Instrument Choice: Evidence from Ohio School Districts. *Public Finance Review*, 38(6): 710-737

- [29] Hansen, C. B. (2007), Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large. *Journal of Econometrics*, 141: 597-620
- [30] Hoene, C. (2004). Fiscal Structure and the Post-Proposition 13 Fiscal Regime in California's Cities. *Public Budgeting Finance*, 24(4): 51-72
- [31] Honadle, B. W. (2003). The states' role in U.S. local government fiscal crises: A theoretical model and results of a national survey. *International Journal of Public Administration*, 26(2): 1431-1472
- [32] Hoxby, C. M. (2001). All school finance equalizations are not created equal. *The Quarterly Journal of Economics*, 116(4): 1189-1231
- [33] Imbens, G., & Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 61(2): 467-476
- [34] Jacob, B. and Lefgren, L. (2005). Principals as Agents: Subjective Performance Measurement in Education. *NBER*, Working Paper 11463
- [35] Jacob, B. and Lefgren, L. (2008). Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education. *Journal of Labor Economics*, 26(1): 101-136
- [36] Kane, T. J., Riegg S. K., & Staiger, D. O. (2006) School quality, neighborhoods, and housing prices. *American Law and Economics Review*, 8(2): 183-212
- [37] Kane, T. and Staiger, D. (2008). Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. *NBER*, Working Paper 14607
- [38] Kane, T. J., Staiger, D. O., & Samms, G. (2003). School accountability ratings and housing values *Brookings-Wharton Papers on Urban Affairs*, 83-137
- [39] Kloha, P., Weissert, C. S., & Kleine, R. (2005). Someone to watch over me: State monitoring of local fiscal conditions. *The American Review of Public Administration*, 35(3): 236-255
- [40] Lockwood, J. R. and McCaffrey, D. (2007). Models for Value-Added Modeling of Teacher Effects. *Electronic Journal of Statistics*, 1: 223-252
- [41] Loeb, S. (2001). Estimating the effects of school finance reform: a framework for a federalist system. *Journal of Public Economics*, 80(2): 225-247
- [42] McCaffrey, D., Lockwood, J. R., Louis, T., and Hamilton, L. (2004). Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement. *Journal of Educational and Behavioral Statistics*, 29(1): 67-101
- [43] McCrary, J. (2008). Manipulation of the running variable in the Regression Discontinuity Design: A Density Test *Journal of Econometrics*, 142: 698-714

- [44] Morris, C. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association*, 78(381), 47-55
- [45] Murray, S. E., Evans, W. N., & Schwab, R. M. (1998). Education-finance reform and the distribution of education resources. *American Economic Review*, 88(4): 789-812
- [46] Nguyen-Hoang, P., & Yinger, J. (2011). The capitalization of school quality into house values: A review *Journal of Housing Economics*, 20: 30-48
- [47] Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata, 3e*. Stata Press: College Station, TX
- [48] Raudenbush, S. W. (2009). Adaptive centering with random effects: An alternative to the fixed effects model for studying time-varying treatments in school settings. *Education Finance and Policy*, 4(4), 468-491.
- [49] Reardon, S. F., Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, 4(4), 492-519.
- [50] Reschovsky, A. (2004). The impact of state government fiscal crises on local governments and schools. *State Local Government Review*, 36(2): 86-102.
- [51] Ross, Justin M., Phuong Nguyen-Hoang. 2013. School District Income Taxes: New Revenue or a Property Tax Substitute? *Public Budgeting Finance*, 33(2): 19-40
- [52] Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4), 537-571
- [53] Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1), 175-214
- [54] Roy, Joydeep. (2011). Impact of school finance reform on resource equalization and academic performance: Evidence from Michigan. *Education Finance and Policy*, 6(2): 137-167
- [55] Springer, M. G., Liu, K., Guthrie, J. W. (2009). The impact of school finance litigation on resource distribution: a comparison of court-mandated equity and adequacy reforms. *Education Economics*, 17(4): 421-444
- [56] Spry, J. A. (2005). The Effects of Fiscal Competition on Local Property and Income Tax Reliance. *Topics in Economic Analysis Policy*, 5(1): 1-19
- [57] Stone, J. (2014). Foundation-Style Funding and Capital Outlays in Primary and Secondary Schools in Michigan. working paper
- [58] Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2e. MIT Press: Cambridge, MA

- [59] Yinger, John. (Ed.). (2004). *Helping children left behind: State aid and the pursuit of educational equity*. Cambridge, MA: MIT Press
- [60] Zahirovic-Herbert, V., & Turnbull, G. (2009). Public school reform, expectations, and capitalization: What signals quality to homebuyers? *Southern Economic Journal*, 75 (4), 1094-1113