University of Tennessee, Knoxville

TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations

Graduate School

8-2020

Three Essays in Public Finance

Md Sabbirul Haque
The University of Tennessee, Knoxville, mhaque3@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Part of the Public Economics Commons, and the Taxation Commons

Recommended Citation

Haque, Md Sabbirul, "Three Essays in Public Finance. " PhD diss., University of Tennessee, 2020.
https://trace.tennessee.edu/utk_graddiss/6876

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Md Sabbirul Haque entitled "Three Essays in Public Finance." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Economics.

Donald J. Bruce, Major Professor

We have read this dissertation and recommend its acceptance:

Celeste K. Carruthers, Enda P. Hargaden, LeAnn Luna

Accepted for the Council:

<u>Dixie L. Thompson</u>

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Three Essays in Public Finance

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Md Sabbirul Haque
August 2020

Copyright © by Md Sabbirul Haque, 2020
All rights reserved.

I dedicate this dissertation to my mother late Ferdousi Sultana.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Donald Bruce for his support and guidance throughout the entire journey. This achievement would not be possible without his support. I would like to thank Dr. Celeste Carruthers for her guidance as well as for sharing her ideas and experiences. I thankfully recognize Dr. Enda Hargaden's effort to help me improve my dissertation by sharing his ideas and experience. I am thankful to Dr. LeAnn Luna for sharing her invaluable ideas and suggestions. I would like to express my gratitude to all the faculty members for sharing their knowledge. I would like to thank Dr. Caroline Weber for sharing her code and providing me suggestions.

I am indebted to my wife Nilufer Sultana for sacrificing her own career and comfort and for providing me support during my stressful time. I am also indebted to my father Md Nurul Hoque and mother late Ferdousi Sultana for providing me every support.

ABSTRACT

The first chapter examines the long-run and short-run elasticity of income with respect to changes in tax rates. The Elasticity of Taxable Income (ETI) is a largely-debated parameter in both research and policy. Despite the growing importance of ETI, the literature has not fully considered the intertemporal impacts of taxation. I expand the literature by estimating short-run and long-run impacts of tax rate changes relying on the most recent estimation method and using appropriate lagged values of income when constructing the predicted net-of-tax rate instruments. The short-run ETI in the baseline specification is 0.69 whereas estimates for the Elasticity of Broad Income (EBI) are much smaller and imprecise. The second chapter studies the impact of tax base on the elasticity of income. Most of the existing literature has appropriately used a constant definition of taxable income to focus on the effects of tax rate changes. It is important to recognize that a decrease in the tax base (in the form of a new deduction, exemption, or credit, for example) can create new opportunities for legal tax avoidance without altering real behavior. Using the most recent estimation method, I estimate the impact of tax base on the behavioral responses to taxation. Estimated results for the impact of tax base are much smaller than those in the existing literature. The third chapter examines the possible linkages between school choice and home values. I use home prices to draw inferences about households' value for school choice, and a Herfindahl-Hirschman Index (HHI) for enrollment among four different types of schools as a proxy measure of school choice. I empirically test two hypotheses: 1) less concentrated counties will have less variability in home prices, 2) less concentrated counties will have higher median home prices. Based on county-level data, I find evidence that an increase in competition for enrollment is associated with a decrease in inequality of home prices within the county. Moreover, I find evidence of an overall increase in home prices within the counties following increased competition for enrollment among schools.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

DYNAMIC INCOME RESPONSES TO TAX REFORM: NEW ESTIMATES

1.1 Introduction

Taxation creates a burden on individuals, causing behavioral changes and implying inefficiencies and distortions in the free market. Studying this behavioral response is vital in understanding the deadweight loss derived from taxation as well as in formulating optimal taxation. The current literature on the elasticity of taxable income (ETI) considers mostly short-run responses of changes in tax rates, and despite the growing importance of ETI for both research and policy, the literature has not fully considered the intertemporal impacts of taxation. The existing literature on the intertemporal impacts for changes in tax rates uses instruments for marginal net-of-tax rates[1] constructed based on initial-year income. However, Weber (2014) shows that such instruments are endogenous. She addresses this concern by using appropriate lagged values of income when constructing the predicted marginal net-of-tax rate instruments. I apply the estimation method developed by Weber (2014) to estimate the intertemporal impacts of taxation.

The ETI literature is vast. However, the recent literature on ETI mostly examines short-run responses. The specifications used to examine such short-run responses assume that taxpayers' responses to taxation are immediate, which is a very strong assumption. The inherent nature of tax reform provides an opportunity for tax avoidance and adjustment of income across periods. In most cases there are large lags among the formulation of the draft proposal, enacting the tax reform and implementing it. For example, the Tax Reform Act of 1986 (TRA86) was formulated in 1985, approved and signed into law in 1986 and finally implemented in 1987. Furthermore, phase-in and phase-out mechanisms also contribute to these lags, which enable the taxpayers to be aware of future changes in tax law well before they go into effect. Because future tax rates are predictable,

[1] Marginal net-of-tax rate is one minus marginal tax rate.

tax filers have an incentive to shift income between adjacent years to avoid taxes without altering longer-term real behavior. In such case, the magnitude of the short-run responses can be larger than that of the long-run responses. However, responses to tax changes can also take time. For instance, one might change his/her long-term investment decision in response to tax changes, or one might consider changing educational or career plan in response to tax changes. In such cases, short-run responses can be smaller than long-run responses. Therefore, the relative magnitude of short- and long-run responses can be ambiguous and is subject to empirical investigation.

Giertz (2010) is one of the studies in the literature that examines intertemporal responses to tax changes. He follows the conventional literature by using instruments for marginal net-of-tax rates constructed based on base-year income. However, the recent literature has shown that the use of conventional instruments constructed using base-year income does not guarantee the exogeneity of the instruments, and therefore it is very unlikely that the estimates presented by Giertz (2010) are unbiased estimates of ETI. Utilizing the variation created by TRA86, I extend the literature by investigating the intertemporal behavioral responses and thereby separating the transient responses from the permanent responses to marginal tax rate changes using a more recently developed estimation method that addresses the endogeneity of the instruments due to mean reversion. I discuss mean reversion in the following section. Results are sensitive to specifications and provide some weak evidence for intertemporal adjustment of income although this evidence is suspect. For the baseline specification, the short-run ETI is 0.685 and there is no evidence for intertemporal adjustment. Estimates for the elasticity of broad income[2] (EBI) are smaller compared to ETI estimates. My estimates are larger in magnitude than those in the existing literature for intertemporal adjustments.

1.2 Literature Review

Estimating the sensitivity of income to changes in tax rates requires exogenous changes in tax rates. However, because the marginal tax rate is a function of income which is the dependent variable of the estimation equation, the observed tax rates are clearly endogenous. This concern has been widely recognized in the literature. The literature has addressed this issue by mostly utilizing a difference form specification and using marginal tax rates defined for some base-year income in the pair of observations as instruments. However, concerns of endogeneity still

[2] Broad income is an extensive definition of income which does not consider adjustments and deductions. I provide the definition of broad income in Section 1.4.2.

remain because the marginal tax rates constructed from base-year income are still a function of the dependent variable. As documented by Weber (2014), mean reversion and changing distribution of income can also lead to endogeneity in tax rates. For example, mean reversion causes the high income of top earners in the base year to fall in the subsequent year because the high income in the base year is likely to be due to a positive transitory income which is unrelated to tax changes, thereby producing a negative correlation between the base-year income and the error term in difference form. On the other hand, a widening distribution of income can produce a positive correlation between the base-year income and the error term. Because base-year income and the error term are correlated; and the tax rate is a direct function of base year income, the instrument for the net-of-tax rate will also be correlated with the error term.

Feldstein (1995) is one of the pioneer studies that estimates ETI utilizing panel data. Based on different specifications and using the difference-in-differences method, he reports ETI estimates that range from 1 to 3. Following Feldstein (1995), a large number of panel-based studies emerged addressing different econometric issues discussed in the literature to obtain consistent and unbiased estimates. Auten and Carroll (1999) address mean reversion and diverging distribution of income by controlling for lagged income in the estimating equation. They report an elasticity of 0.55. Moffitt and Wilhelm (2000) use panel data from the Survey of Consumer Finances instead of tax return data and examine the effect of the variation in marginal tax rates arising from TRA86. They report elasticity estimates ranging from 0.35 to 0.97. They also estimate the sensitivity of labor supply with respect to tax rate changes and conclude that the rise of taxable income of high-income individuals is not accompanied by an increase in reported labor supply. These conclusions are consistent with the notion of possible retiming of income. Gruber and Saez (2002) introduce the use of income splines to control for mean reversion and heterogeneous trends; and separate income and substitution effects of tax changes. They rely on the public use version of panel tax data and use the variation caused by TRA86. They find an elasticity of taxable income of 0.40 and an elasticity of broad income (EBI) of 0.12.

Weber (2014) introduces a new instrument to address the mean reversion issue widely discussed in the literature. She uses lagged income instead of base-year income to construct instruments for net-of-tax rates. She also verifies that as the number of lags increases, the lagged income and hence the instruments constructed from that lagged income become more orthogonal to the error term. She identifies the instruments based on lagged income using a testable assumption regarding the degree of serial correlation in the error term. Based on panel tax data for years 1979-1990 and addressing mean reversion issues, she reports an ETI estimate of 0.858 and EBI estimate of 0.475. Her reported ETI estimate is twice as large as that reported in Gruber and Saez (2002). Kawano et al. (2016) use the variation from the American Taxpayer Relief Act of 2012 to estimate EBI. They use an inverse probability weighting (IPW)[3] in the context of a difference in differences approach and find significantly lower estimates of EBI ranging from 0.013 to 0.034.

The notion of timing of income receipt was rooted in the hierarchy of behavioral responses to taxation by Slemrod (1990). He suggests a hierarchy of three tiers of behavioral response to taxation. First two tiers of his hierarchy include timing and avoidance whereas the third tier refers to behavioral responses. Using panel data and focusing on realized capital gains, Burman and Randolph (1994) provide evidence for substantially smaller permanent effects than transitory effects. Bakija and Heim (2008), on the other hand, investigate the impact of taxation on charitable giving. They use an instrumental variable approach for estimating the elasticity of charitable giving with respect to its current and future prices. They find evidence for re-timing of giving in response to predictable future changes in federal tax rates, however, this finding is sensitive to source of identification.

The ETI literature, however, contains a few companion studies investigating the intertemporal impacts of tax changes on taxable income; and these studies are most relevant for my study. Kreiner et al. (2016) examine the ETI using monthly payroll data and provide evidence for intertemporal shifting of wage income with respect to tax rate changes. However, removing the data for a few months around the point of time when the tax change takes place, they find the elasticity close to zero. Goolsbee (2000) examines the responsiveness of income of high-income executives with respect to changes in tax rates and provides evidence for a short-run shift in the timing of realization of compensation rather than a permanent reduction in income. Sammartino and Weiner (1997) demonstrate that income shifted backward in time from 1993 into 1992 (even when excluding capital gains) in response to President-elect Bill Clinton's promised tax increase on high-income taxpayers. One of the most relevant studies for my present study is Giertz (2010). He investigates intertemporal responses of taxable income by using the variation from the Omnibus Budget and Reconciliation Acts (OBRAs) of 1990 and 1993. Based on panels of U.S. tax returns, he estimates short-run and long-run responses of taxable income to changes in tax rates. Giertz (2010) addresses the endogeneity in the tax rates by using predicted marginal net-of-tax rates based on base-year income. However, as explained previously, the use of base-year income in constructing the instruments does not guarantee that endogeneity in marginal tax rates will be resolved. Giertz (2010) reports larger long-run estimates than short-run estimates. Another companion study examining intertemporal responses is Holmlund and Soderstrom (2011). They investigate the intertemporal responses using panel data on Swedish tax reform for years 1991-2002 and report the long-

[3] IPW approach removes the confounding by using inverse of probability as weights.

run estimates between 0.10 to 0.30. The intertemporal aspects considered in their study capture the habit persistence from past tax rates, however, the responses to predictable future tax rates are not considered. To the extent that taxpayers are aware of future tax changes ahead of time and that they intertemporally optimize by altering their current behavior in response to predictable future tax changes, the specification used in Holmlund and Soderstrom (2011) is also mis-specified.

All of these concerns involving current knowledge on intertemporal responses to tax changes call for a further extension of the literature. My study uses the instruments developed by Weber (2014) and extends the literature by addressing the endogeneity due to mean reversion in estimating short-run and long-run impacts of tax rates. Specifically, I use lagged values of income when constructing predicted net-of-tax rates, and test various lags of income for robustness. My baseline specification includes instruments based on 1-year lagged income and examines the endogeneity of this instrument assuming that instruments constructed from 2-year and 3-year lagged income are exogenous. I discuss the rationale behind such lag structure for constructing instruments at the end of the next section.

1.3 Model

Changes in tax rates can affect taxpayers' behavior in several ways. If tax rates increase, labor becomes more expensive relative to leisure. Therefore, an increase in tax rates can cause a decrease in labor supply thereby a decrease in income because of the substitution effect. An income effect can lead to an opposite result. Although in general, the magnitude of the elasticity of primary earners' labor supply varies across studies, the magnitude for the secondary earner has been found to be large (Eissa, 1995). Second, different forms of income are taxed at different rates, and some forms are exempt from taxation. For example, in 1990 taxpayers could claim tax exemption for a certain type of interest income as well as for a portion of pensions and annuities and IRA distributions. More interestingly, high-income taxpayers, in most cases, can shift their income towards tax-favored forms of income, for instance, fringe benefits, stock options, etc. An increase (or decrease) in tax rates encourages taxpayers to take advantage of such tax treatment by shifting income towards tax-favored components. Not only do the taxpayers shift their income across components of income, but they can also shift expenditure towards tax-favored components to reduce taxable income. For example, an increase in tax rates can motivate taxpayers to increase their tax-deductible expenditures on home mortgages, medical and dental treatments, or charitable contributions. A change in tax rates can also affect the extent of compliance with the tax laws regarding the accuracy of the reported income. An increased tax rate raises the opportunity cost of compliance and, therefore, may encourage tax evasion. Third, individuals can shift income across time to take advantage of a favorable tax treatment. For example, if a taxpayer expects the tax rate to fall in the future period, he/she may consider deferring some compensation or pension income from the current to the future period, and/or he/she may consider shifting income from taxable bond to municipal bond in the current period to avoid taxes. Kreiner et al. (2016) provide evidence that employees shift their wage income intertemporally in an anticipation of tax changes.

To explain intertemporal responses, I use the following intertemporal model that is employed by Goolsbee (2000). Giertz (2010) and Bakija and Heim (2008) have also employed similar specification.

$$\ln\left(\frac{Y_{i\,t+1}}{Y_{i\,t}}\right) = \alpha + \gamma_t + \boldsymbol{X'}_{it}\boldsymbol{\beta_1} + \beta_2 ln\frac{(1-\tau_{i\,t+1})}{(1-\tau_{i\,t})} + \beta_3 ln\frac{(1-\tau_{i\,t+2})}{(1-\tau_{i\,t+1})} + \nu_{i\,t} \qquad (1.1)$$

Equation (1.1) can also be written as follows.

$$\Delta lnY_{it} = \alpha + \gamma_t + \boldsymbol{X'}_{it}\boldsymbol{\beta_1} + \beta_2\Delta ln(1-\tau_{i\,t}) + \beta_3\Delta ln(1-\tau_{i\,t+1}) + \nu_{i\,t} \qquad (1.2)$$

In equation (1.1), $\tau_{i\,t}$ is current marginal tax rate[4] of individual i at time period t, $\tau_{i\,t+1}$ is the marginal tax rate of individual i at time period $t+1$, and Y_{it} is income. X_{it} represents a set of control variables used in the estimation, γ_t represent year fixed effects and v_{it} represents unobserved error. Here, $ln\frac{(1-\tau_{i\,t+1})}{(1-\tau_{i\,t})}$ represents change in current year's marginal tax rates in log form whereas β_2 represents the percentage change in taxable income for a one percent change in current period's net-of-tax rates. Moreover, $ln\frac{(1-\tau_{i\,t+2})}{(1-\tau_{i\,t+1})}$ represents the change in anticipatory future tax rates whereas the coefficient β_3 represents the responses to anticipatory future net-of-tax rates on income. Even though this model has been employed in the literature, one may cast doubt regarding the validity of the specification as the numerator of $ln\frac{(1-\tau_{i\,t+1})}{(1-\tau_{i\,t})}$ and the denominator of $ln\frac{(1-\tau_{i\,t+2})}{(1-\tau_{i\,t+1})}$ may seem to cancel out each other from the right had side of the equation. However, they may cancel out only when β_2 and β_3 are equal which is very unlikely. Furthermore, intuitively, the denominator of $ln\frac{(1-\tau_{i\,t+2})}{(1-\tau_{i\,t+1})}$ is an anticipated measure of the future year's marginal net-of-tax rates whereas the numerator of $ln\frac{(1-\tau_{i\,t+1})}{(1-\tau_{i\,t})}$ is the marginal net-of-tax rate that taxpayers are aware of. Because there can be error involved in the anticipation process, these two terms may not necessarily be equal and cancel out each other in the right-hand side of equation (1.1). Moreover, as I have explained in previous sections, tax rates are endogenous and therefore I use instrumental variables for

[4] This tax rate refers to the sum of federal and state tax rates.

$ln\frac{(1-\tau_{i\,t+1})}{(1-\tau_{i\,t})}$ and $ln\frac{(1-\tau_{i\,t+2})}{(1-\tau_{i\,t+1})}$. The estimation of equation (1.1) is based on the fitted values of $ln\frac{(1-\tau_{i\,t+1})}{(1-\tau_{i\,t})}$ and $ln\frac{(1-\tau_{i\,t+2})}{(1-\tau_{i\,t+1})}$ from the first stage regressions instead of observed values for these two ratios. Since I am using fitted values instead of observed values, these two variables should not cancel out each other. Following previous studies in the literature, β_2 represents the short-run elasticity and $\beta_2 + \beta_3$ represents the long-run elasticity.

The literature mostly uses difference-in-differences methods to estimate this type of model. This approach yields consistent estimates if tax rate changes are the only sources of income shocks. However, other non-tax factors (e.g. heterogeneous income growth and mean reversion) can lead to inconsistent estimates. Furthermore, because of the graduated tax rate schedule, tax rates are endogenous and can lead to biased estimates. The existing literature uses the instrumental variable approach and constructs instruments for marginal net-of-tax rates based on initial-year income to address the endogeneity issue. However, Weber (2014) shows that such instruments are endogenous. She uses lagged income to construct the instrument and shows that increasing each additional lag of income used to construct the instruments will make the instruments more exogenous. I use a difference-in-differences estimation method with one-year differences and apply the instruments developed by Weber (2014) to estimate the intertemporal model (1). For one-year differences specification, I make pairs of observations one year apart and regress the changes in income between pairs of observations on changes in net-of-tax rates along with other control variables. The predicted marginal tax rate for year t based on k-year lagged income is defined as follows.

$$\tau_{i\,t}^{P\,\mathrm{k}\,lag} = \boldsymbol{\tau}(\boldsymbol{Y_{i\,t-k}}, \boldsymbol{c_t})$$

Here $\boldsymbol{c_t}$ is the tax law of year t. Therefore, the instrument, for example, for $\Delta ln(1-\tau_{i\,t})$ based on one-year lagged income in equation (1.2) is defined as $ln\frac{(1-\tau_{i\,t+1}^{P\,1\,lag})}{(1-\tau_{i\,t}^{P\,1\,lag})}$ and the instrument, for example, for $\Delta ln(1-\tau_{i\,t+1})$ based on one-year lagged income in equation (1.2) is defined as $ln\frac{(1-\tau_{i\,t+2}^{P\,1\,lag})}{(1-\tau_{i\,t+1}^{P\,1\,lag})}$. This simply means that the two predicted marginal tax rates in the difference form differ only by the tax law and all other information for constructing the two predicted tax rates are same.[5]

[5] Net of marginal tax rate instruments are computed by running income with the same lags through TAXSIM for the current year and the base year, and then taking the difference between the two

Following Weber (2014), I use a Difference-in-Sargan test to assess the endogeneity of the current- and lead-tax rate[6] instruments. The Difference-in-Sargan test is a test of overidentifying restrictions that assumes a subset of instruments as exogenous and tests the validity of suspect instruments. It checks the validity of suspect instruments by computing the increase in Sargan's J statistic when such suspect instruments are added to the estimation. To test the validity of two suspect instruments for two endogenous variables, at least three instruments are needed that are assumed to be exogenous. This means that at least five instruments are needed to implement Difference-in-Sargan test in a specification with two endogenous variables. This requirement for the minimum number of instruments is satisfied when I test the validity of instruments for current- and lead-tax rates constructed from lagged income of year $t-i$ assuming that instruments constructed from income of years $t-j$ and $t-k$ are exogenous where $i<j<k$. Here, I am using Weber's (2014) approach to test validity of the suspect instrument.

1.4 Data and Estimation Approach

1.4.1 Identification

Most studies in the ETI literature have employed a difference-in-differences or a similar approach. This study follows the spirit of the conventional approach and uses a difference-in-differences estimation method. For identification of the effect of tax changes in an ideal scenario, the control group would not experience a tax change, and the treatment group would experience a tax change. Unfortunately, such a control group does not exist because the groups of taxpayers that face differential tax treatments are also different in terms of demographic characteristics. Therefore, the next best solution is to find two similar groups that experience differential tax treatments. TRA86 creates an opportunity for identification of the effect of tax changes because high-income taxpayers experience larger reductions in tax rates than lower-income individuals. In an ideal scenario, the two comparison groups are similar in every aspect except for

marginal tax rates. For instance, for constructing a tax rate instrument as a function of income of two-year lag, I compute synthetic tax rates using income of year t-3 and tax codes for year t-1 (base year); and similarly I compute synthetic tax rates using income of year t-3 and tax codes for year t; and then I take the difference of the log of the two synthetic tax rates.

[6] For simplicity, I refer to the marginal net-of-tax rate for the current period as current tax rate; and marginal net-of-tax rate for the future period as lead tax rate.

treatment status. However, in reality, the high- and low-income groups may not necessarily be similar. Specifically, higher-income taxpayers generally have higher income growth rates than lower-income individuals. Bound and Johnson (1992), Kutz and Murphy (1992), Levy and Michel (1991), and Murphy and Welch (1992) provide evidence that labor earning became more unequal during 1980s. Moreover, Krugman (1992) highlights that very high-income households received a disproportionately large share of income growth during 1980s. One obvious approach to address these issues is to control for these heterogeneous characteristics. However, the US-based studies generally utilize tax panel data from the IRS, and these datasets generally include minimal socio-demographic information. These identification concerns have drawn considerable attention in the literature, and addressing these issues has been a difficult challenge because such heterogeneous effects are not well understood. Several non-US studies claim to be able to mitigate these issues by using rich socio-demographic information. On the other hand, given the unavailability of such rich data, the US studies commonly use income controls to mitigate these effects. I will discuss the treatments to address these concerns in detail in later sections.

1.4.2 Data

I obtain the data used in this study from Statistics of Income (SOI) tax files for U.S. tax returns for the years 1979-1990. The individual SOI tax files are combined to construct a panel of years 1979-1990. Because the instruments are functions of lagged income of up to four years, I lose observations of years 1979-1982 in the estimation. Moreover, I also lose observations of year 1990 because the model includes one-year lead tax rates and year 1989 because of the use of one-year difference form. Therefore, the estimation is restricted to between 1983 and 1988. The study primarily uses taxable income for the definition of income as a dependent variable. An extension of this analysis to broad income is also presented in a later section. Broad income is total income from all income sources (except capital gains) that can be computed from the data for all years 1979-1990. Most previous studies have excluded capital gains from the analysis. As is common in the literature, I exclude capital gains from the analysis. Social security benefits are also excluded from the definition of broad income because they are not included in the data before 1984. Taxable income is defined as the broad income minus above-the-line deductions (i.e., adjustments) minus the larger of below-the-line deductions (i.e., the larger of standard deductions and itemized deductions), minus personal exemptions that are available in each year.

Tax reforms generally involve changes in the tax base along with tax rates and other changes. If the tax base varies systematically, then this variation in the tax base can potentially bias the estimates. The literature has addressed this issue by using a constant-law definition of taxable income. Consistent with the literature, I use the definition of taxable income that is constant across reforms whereas the constant-law definition of taxable income is close to the 1990 definition of taxable income. All income, and deduction components are converted into 1992 dollars using inflation indices used in Gruber and Saez (2002).

Marginal tax rates are computed using TAXSIM (Feenberg and Coutts, 1993), a freely available internet version of the tax rate calculator developed by NBER.[7] TAXSIM computes federal and state marginal tax rates separately. The effective marginal tax rate is a summation of state and federal marginal tax rates. Only observations for which a taxpayer's marital status do not change between the base and the current years are included in the analysis because a change in marital status can cause changes in income unrelated to changes in tax rates. Finally, as has become standard in the literature, I exclude taxpayers whose taxable income is less than $10,000 in the base year to reduce the impact of mean reversion. It is evident in the literature that mean reversion is severe in the extreme low-income range and avoiding this income group can help reduce the impact from mean reversion. Following Weber (2014), I include several control variables in my estimation, for example, indicators for marital status, and the number of dependent children in the household, a full set of state and year dummy indicators. Table A.1 presents the descriptive statistics for the variables used in the analysis.

1.4.3 Identification Issues

As explained in the previous section, the existence of heterogeneous income trends and mean reversion poses a challenge in estimating ETI. These issues have drawn substantial attention in the literature. Some studies were able to partially control for heterogeneous trends by using demographic information. Other studies have employed income controls to control for heterogeneous trends. Early studies include income as a control in a linear fashion (Carroll, 1998). However, a linear relationship between the income control variable and the error term may not exist. More recent studies employ non-linear income splines to address the issue. Non-linear income splines allow a smooth and non-linear relationship between the income control variable and the error term. Gruber and Saez (2002) introduce the use of income splines in the ETI literature, and the subsequent studies followed Gruber and Saez (2002) by including income splines in the estimations to control for mean reversion and heterogeneous trends. Following the literature I also use income splines and include five-piece quintile splines based on income in the estimating equation to control for heterogeneous trends. I also examine the stability of the estimates by using non-linear income splines based on different lags of income.

[7] The internet version of TAXSIM is available at https://users.nber.org/~taxsim/taxsim27/

1.4.4 Instrument Selection

This section explores the suitable instruments for the two endogenous variables of interest: current- and lead-tax rates. Column 1 of Table 1.1 presents replication of Weber's (2014) ETI estimate using my data whereas columns 2-4 in Table 1.1 present estimates for the second stage regression of equation (1.1) using different lagged instruments for current- and lead-tax rates. The dependent variable in all four columns is ΔlnY_{it}.The first stage regression estimates for this table and all subsequent tables are included in Appendix A. The primary purpose of Table 1.1 is to empirically examine and find the instruments that are exogenous as well as correlated with the two endogenous regressors. Both first-stage and second-stage regressions presented in Table 1.1 and all other subsequent tables in this essay adjust for heteroscedasticity clustered by the individual level. Pflueger et al. (2013) raises concern that when first-stage F-statistics do not adjust for heteroscedasticity, the values of F-statistics can be large leading to rejection of weak instruments. They suggest that F-statistics need to be adjusted for heteroscedasticity to implement test for weakness of instruments. My results address this concern by adjusting heteroscedasticity-robust standard errors clustered at the individual level in all first-stage regressions. A Difference-in-Sargan test is employed to test the exogeneity of the instruments. The null hypothesis of this test is that the suspect instrument(s) is(are) exogenous. The weakness of the instruments is examined by comparing the F-statistics from the first stage regressions with the minimum F-statistic suggested by Stock and Yogo (2002). My estimated elasticity for Weber's (2014) baseline specification is 0.722 whereas her reported estimate[8] is 0.858. First-stage regression estimates for Weber's (2014) specification is presented in Table A.2 in Appendix A.

---

[8] I have some limitations while replicating Weber's (2014) estimates. Weber's estimates are based on the full version of TAXSIM which utilizes all the variables in SOI and which is available exclusively on NBER server. On the other hand, I use the publicly available version of TAXSIM which computes tax rates based on only 27 input variables and does not take into account all the available variables in SOI. For example, publicly available TAXSIM version does not have direct input variables for standard deduction and personal exemption, rather it computes them from the variables provided as input. However, it does not have any input for primary and secondary taxpayer's blindness status either and computes the personal exemption and standard deduction without taking into account blindness status. Similarly, there are other variables in SOI that are not used as input in the publicly available TAXSIM version. In addition to this limitation, I also don't have the exact information of how Weber cleaned and prepared the individual SOI files and combined them into a single file.

Table 1. 1: Second-stage regression estimates for instrument selection

	(1)	(2)	(3)	(4)
	Replication of Weber' estimates			
$\Delta ln(1-\tau_t)$	0.722**	-0.299***	0.685**	1.209***
	(0.294)	(0.115)	(0.293)	(0.375)
$\Delta ln(1-\tau_{t+1})$		0.147	0.151	0.0493
		(0.259)	(0.340)	(0.389)
Observations	23,438	24,731	24,731	24,731
R-squared	-0.140	0.071	-0.125	-0.262
Instruments	2 3 4 lags	0 2 3 lags	1 2 3 lags	2 3 4 lags
Diff-in-Sargan p-val	0.810	7.04e-05	0.245	0.217
First stage F-statistic	121.75	276.21, 46.05	52.40, 39.09	37.43, 39.71

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table (and in all other tables in this essay) adjust for heteroscedasticity clustered at the individual level. R-squared value being between 0-1 is the property of OLS. However, unlike in case of OLS, R-squared from IV regression can be negative because Residual Sum of Squared (RSS) can be larger than Total Sum of Squares (SST) in case of IV regressions. R-squared value in case of IV estimation is not useful as it is in case of OLS. *** p<0.01, ** p<0.05, * p<0.1.

Columns 2-4 of Table 1.1 examine the exogeneity and weakness of the instruments in the basic specification. The first stage regression estimates for these columns of Table 1.1 are included in Table A.3 in Appendix A. Column 2 of Table 1.1 includes instruments for current tax rate and lead tax rate constructed based on base-year income, two-year lagged income and three-year lagged income. First stage F-statistics from the regressions of current and lead-tax rates are 276.21 and 46.05 respectively which are much larger than the minimum value suggested by Stock and Yogo (2002).[9] The Difference-in-Sargan test for this column examines whether the two instruments constructed based on base-year income are exogenous, assuming that the instruments based on two- and three-year lagged income are exogenous. Although the instruments are not weak, the

[9] For two endogenous variables and six excluded instruments, if one wants to limit the bias of IV estimator to 5% of OLS bias, the minimum F-statistic suggested by Stock and Yogo (2002) is 15.72.

Difference-in-Sargan p-value is close to zero; therefore, we can strongly reject the null hypothesis that the instruments are exogenous at any conventional level of significance. Column 3 includes instruments constructed based on one-year lagged income, two-year lagged income and three-year lagged income. The Difference-in-Sargan p-value is 0.245, so the instruments are more plausibly exogenous compared to those in column 2. First stage F-statistics from the regressions of current- and lead-tax rates are 52.40 and 39.09 respectively, suggesting that the instruments are strong. Column 4 includes instruments constructed from two-, three- and four-year lagged income. The Difference-in-Sargan p-value now slightly decreases to 0.217. First stage F-statistics from the regressions of current- and lead-tax rates are 37.43 and 39.71 respectively which are also much larger than the minimum value suggested by Stock and Yogo (2002). I choose the specification in column (3) (instruments constructed based on one-, two- and three-year lagged income) as my preferred specification based on the Difference-in-Sargan p-value and values of first-stage F-statistics. Column (3) is my preferred specification because Difference-in-Sargan p-value is larger (i.e., instruments are more exogenous) and first-stage F-statistics are also larger compared to other specifications. For the baseline specification, the coefficient on current-tax rate is 0.69 and is significant at the 5% level whereas the coefficient on the lead tax rate is 0.15 but this coefficient is not significant at the 10% level of significance with large p value (p value is 0.658). These baseline results do not provide any evidence for intertemporal adjustments.[10]

1.4.5 Stability of Income Controls

As is discussed previously, heterogeneous income growth poses a challenge in estimating ETI because such heterogeneous effects are not well understood. The recent literature addresses this issue by controlling for income as an alternative to controlling for socio-demographic characteristics. Table 1.2 examines the stability of the estimates for various lags needed to construct these splines. The estimates from the first stage regressions for this table are included in Table A.4 in Appendix A.

[10] I do not adjust my estimates for sample weights as my data do not have high-income oversample. This means that all taxpayers have the same probability of being selected in the sample regardless of the income level. Weber (2014) also doesn’t adjust for sample weights as her data do not have high income oversample. My dataset includes 177 high-income taxpayers with missing state information. The estimated results for the baseline specification remain unchanged when I estimate ETI excluding those high-income taxpayers.

Table 1. 2: Second-stage regression estimates with income controls

	(1)	(2)	(3)	(4)
	Baseline of Table 1.1 No Splines	Base-year income splines	1-year lagged income splines	2-year lagged income splines
$\Delta ln(1-\tau_t)$	0.661**	1.810***	0.748**	0.537
	(0.282)	(0.458)	(0.347)	(0.329)
$\Delta ln(1-\tau_{t+1})$	0.161	1.160**	0.230	0.00771
	(0.336)	(0.454)	(0.399)	(0.380)
Observations	24,007	24,007	24,007	24,007
R-squared	-0.119	-0.377	-0.136	-0.097
Spline lags	No Spline	0	1	2
Lags of income for constructing instruments	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
Diff-in-Sargan p-val	0.454	0.282	0.435	0.500
First stage F-statistic	53.19, 39.92	23.93, 27.02	31.90, 26.55	33.17, 27.37

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table adjust for heteroscedasticity clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.2 includes instruments based on income lagged one, two and three years prior to the base year. However, columns 1-4 in Table 1.2 varies by the number of lags used to construct income splines. Column 1 repeats the baseline column 3 of Table 1.1 and does not include any splines. The coefficient on the current-year tax rates is 0.661 and is significant at the 5% level whereas the coefficient on the anticipated tax rates in imprecise. Column 2 includes income splines based on base-year income. The coefficient on the current-year tax rates in 1.81 and that on the future tax rates is 1.16. These estimates are much larger than those in the current literature. However, because these splines used in column 2 are a function of base-year income, it is unlikely that these splines will be able to absorb heterogeneous trends. For this reason, Kopczuk (2005) uses lagged income instead of initial year income to construct the income splines. Weber (2014) also finds very large elasticity (2.40) for this specification. For all these reasons, the results in column 2 are suspect. When I include income splines constructed based on one-year lagged income (column 3), the coefficient on the current tax rates becomes 0.75 and is significant at the 5% level, however, the coefficient on the anticipated future tax rates is again imprecise. Column 4 includes two-year lagged income splines. Both coefficients in column 4 are imprecise.

1.4.6 Heterogeneity

Tax treatments generally apply to different subgroups of taxpayers differentially. In this section, I examine heterogeneity among taxpayers based on income and marital status in response to tax changes. Columns 1 and 2 in Table 1.3 present ETI estimates for taxpayers with broad income less than $40,000 and larger than $40,000 respectively. Both columns include instruments based on one-, two- and three-year lagged income and does not include any income splines (my preferred specification in Table 1.1). For taxpayers with broad income less than $40,000 (column 1), the coefficients on both current and lead tax rates are imprecise at the 10% level of significance. Moreover, both first stage F-statistics suggest that instruments are weak. For taxpayers with broad income larger than $40,000 (Column 2), the coefficients on the current tax rates is 0.702 and is significant at the 5% level. Both first stage F-statistics from this specification are larger than suggested minimum value. These results provide evidence that higher-income taxpayers are more responsive to changes in tax rates compared to lower income group. However, I don't find evidence for intertemporal adjustment as the coefficient on the lead tax rates in insignificant.

Table 1. 3: Second-stage regression estimates for heterogeneity based on income

	(1)	(2)
	<40k BI	>40k BI
$\Delta ln(1-\tau_t)$	0.628	0.702**
	(1.083)	(0.312)
$\Delta ln(1-\tau_{t+1})$	-0.479	0.509
	(1.158)	(0.373)
Observations	10,984	13,747
R-squared	-0.121	-0.155
Diff-in-Sargan p-val	0.164	0.675
Spline lags	No splines	No splines
Lags of income for constructing the instruments	1, 2, 3	1, 2, 3
First stage F-statistic	4.37, 3.05	39.27, 32.07

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table adjust for heteroscedasticity clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.4 shows the estimates for two groups based on marital status using the preferred specification in Table 1.1. Column 1 presents estimates for single taxpayers whereas column 2 presents estimates for married taxpayers. For single filers, the coefficients on both current and lead tax rates are not significant whereas the instruments are weak. For married taxpayers, the first stage F-statistics are much larger than the minimum value suggested by Stock and Yogo (2002). The results are in agreement with those from all previous specifications suggesting that married taxpayers are responsive to tax changes. The results further indicate that married tax filers are responsive to current tax changes only, but they do not adjust income intertemporally.

Table 1. 4: Second-stage regression estimates for heterogeneity based on marital status

	(1)	(2)
	Single	Married
$\Delta ln(1-\tau_t)$	-0.256	0.869**
	(0.503)	(0.437)
$\Delta ln(1-\tau_{t+1})$	-0.707	0.430
	(0.521)	(0.486)
Observations	6,198	17,202
R-squared	0.021	-0.150
Diff-in-Sargan p-val	0.175	0.688
Spline lags	No splines	No splines
Lags of income for constructing instruments	1, 2, 3	1, 2, 3
First stage F-statistic	16.52, 7.36	32.14, 30.54

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table adjust for heteroscedasticity clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

1.4.7 Elasticity of Broad Income

As highlighted in Section 1.3, an individual taxpayer can respond to changes in tax rates through several margins including adjustment of the amount of labor supply, tax avoidance by changing the form and timing of compensation, tax evasion, etc. ETI captures all these responses. For instance, a taxpayer facing an increased tax rates may shift his/her incomes and expenditures towards tax-favored components thereby reducing overall taxable income without altering real behavior. Such a behavioral response affects ETI although the real behavior is unaltered. ETI is considered to be a sufficient statistic for welfare analysis when there are no classic and fiscal externalities. However, in case of externalities, an elasticity of a broader definition of income is needed along with ETI. The literature argues that some below-the-line deductions, for example, deductions for charitable contributions, state and local taxes, home mortgage interests, etc., can create externalities. In that light, I extend my analysis to EBI as well.

Table 1. 5: Second-stage regression estimates for elasticity of broad income

	(1)	(2)	(3)
	No Splines	Base-year income splines	1-year lagged income splines
$\Delta ln(1-\tau_t)$	0.186	0.739**	0.304
	(0.192)	(0.289)	(0.238)
$\Delta ln(1-\tau_{t+1})$	0.122	0.604**	0.222
	(0.201)	(0.270)	(0.238)
Observations	24,937	24,937	24,937
R-squared	-0.040	-0.212	-0.073
Spline lags	No Spline	0	1
Instrument lags	1, 2, 3	1, 2, 3	1, 2, 3
Diff-in-Sargan p-val	0.872	0.572	0.857
First stage F-statistic	48.89, 38.81	24.04, 26.55	29.40, 26.93

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table adjust for heteroscedasticity clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

The primary difference between taxable income and broad income is that broad income does not include deductions whereas taxable income does. Table 1.5 shows the estimates for EBI using different specifications. Each column in Table 1.5 includes instruments based on income lagged one, two and three years prior to the base year. Column 1 does not include any income splines whereas columns 2 and 3 include five-piece splines constructed from base-year and one-year lagged income respectively. The results suggest that estimates are sensitive to income splines used in the specification. Both coefficients are significant at the 5% level when income splines are constructed based on base-year income (Column 2). However, for reasons explained in Section 1.4.5, these results are suspect. For the remaining two specifications, first-stage F-statistics are larger than the minimum values but both the coefficients are imprecise at any conventional level of significance. Therefore, we cannot reject the null hypothesis that the coefficients are equal to zero. The EBI estimates from all three specifications are smaller than those in the corresponding specifications for ETI (Table 1.2).

1.5 Conclusion

This study revisits the intertemporal responses to changes in tax rates. The results from this study provide some evidence that the estimates for intertemporal responses to taxation in the existing literature are likely biased because the instruments are endogenous. This study also identifies the instruments that should yield estimates that are, at least, closer to consistent estimates compared to those in the prior literature. For most of the specifications, estimated results do not provide evidence for intertemporal adjustment. . The results from the specification controlling for base-year income splines provide some weak evidence of intertemporal adjustment, however, these estimates are suspect. The short-run ETI in my preferred specification is 0.69 whereas estimates for EBI are much smaller and imprecise. My short-run elasticity is larger than that found in Giertz (2010) but smaller than the estimate found in Goolsbee (2000). Giertz (2010) and Goolsbee (2000) provide evidence for existence of intertemporal adjustment. However, using more exogenous instruments, my results, in general, do not support the existence of intertemporal adjustment. Kreiner et al. (2016) provide evidence for intertemporal shifting using Danish payroll data. However, they do not find evidence for shifting income earned over entire year when they remove last two months' income before the implementation of tax reform and one month's income after the tax changes take place. Perhaps my results are different from Kreiner et al. (2016) because of the differences in the nature of tax laws in two countries. In Denmark, it is possible to report income earned from one period as income earned in another period without violating the tax laws whereas it is not possible in the USA. Because ETI estimates are larger than those for EBI, the results in my study are also indicative that most of the responses of taxable income take place through deductions. ETI may be more relevant for policymakers more concerned about revenue-maximizing tax rates. On the other hand, EBI may be more useful for analyzing real behavior because EBI does not include tax avoidance behavior. Results found in this study have policy consequences. Since, in general, I don't find evidence for intertemporal adjustment, these results are indicative that the actual efficiency cost of taxation may not necessarily be larger in the long run compared to that in the short run. Therefore, policymakers may have more flexibility in raising revenue without affecting much efficiency cost in the long run. However, estimating the exact magnitude of optimal tax rates considering the intertemporal responses is beyond the scope of this study and remains an area for future research.

CHAPTER II

DOES TAX BASE AFFECT THE BEHAVIOR OF TAXPAYERS?

2.1 Introduction

The Elasticity of Taxable Income (ETI) and the Elasticity of Broad Income (EBI) are two central parameters of interest in tax policy analysis. Although there has been a large body of literature addressing these policy parameters, there has been little consensus regarding the magnitude of these parameters. The literature recognizes that these parameters cannot be thought of independent behavioral responses but, rather these parameters can be affected by tax policy itself. Slemrod and Kopczuk (2002) and Kopczuk (2005) provide evidence that elasticities are larger for a tax system with a larger amount of deductions and therefore, any changes in tax rates have a direct impact on income as well as an indirect impact caused by the interaction between tax rates and the amount of deductions. This implies that the behavioral responses and, therefore, the efficiency cost of taxation can be controlled by controlling the amount of allowable deductions. Kopczuk (2005) estimates this indirect effect of changes in tax rates and his estimated parameter is in the order of 0.7 to 0.8. In my study, I revisit this indirect effect of changes in tax rates using a recent estimation method developed by Weber (2014). My estimated results support this hypothesis. The findings from my study demonstrate that EBI is not an exogenous parameter, rather it depends on the tax base. However, the magnitude of the tax base effect is smaller when compared to the magnitudes in the prior literature. This finding has significant implications on the efficiency cost of taxation. This finding implies that the tax base has an impact on the efficiency of taxation and, therefore, any analysis of the efficiency cost of taxation ignoring the tax base effect is incomplete. We need to take into account both tax rates and tax base to better understand the efficiency cost of taxation.

2.2 Literature Review

Prior research recognizes that that the behavioral responses of tax rate changes are not independent, and external factors can influence this parameter thereby affecting the efficiency cost of taxation. For instance, Harju and Matikka (2016) argue that external factors such as third-party reporting can affect the behavioral responses of taxpayers. Keiner et al. (2014, 2016), Kleven et al. (2011) as well as Kleven et al. (2016) also provide similar arguments. The literature has also documented that behavioral responses can be affected by economic conditions. Hargaden (2020) demonstrates that the behavioral responses are smaller during the economic recession and therefore the behavioral responses are affected by economic fluctuations. These pieces of evidence demonstrate that behavioral responses are not exogenous parameters and, therefore, can be affected by other external factors.

The conceptual foundation of how the tax base can affect taxable income has been discussed in Slemrod (1995) and Slemrod and Kopczuk (2002). They argue that a larger amount of deductions is associated with larger elasticities. This implies that the behavioral responses can depend on allowable deductions and therefore the efficiency cost of taxation also depends on the prevailing tax structure. Kopczuk (2005) provides extensive empirical evidence supporting this hypothesis. He uses tax return data to identify the direct and indirect impact of taxation. Using an instrumental variables approach and a difference-in-differences estimation method, he estimates the indirect elasticity of broad income with respect to changes in net-of-tax rates in the order of 0.7 to 0.8. As discussed in the literature, tax rates are endogenous as they are direct functions of income because of graduated tax rates. Moreover, he also points out that the tax base can also depend on income indirectly. This turns out that both net-of-tax-tax rates and net-of-tax-tax base[11] are endogenous. He addresses this endogeneity by using an instrumental variable approach. He constructs the instruments as functions of base year income.

As discussed in the first essay, Weber (2014) introduces a new instrument to address the mean reversion issue widely discussed in the literature. She uses lagged income instead of base-year income to construct the predicted net-of-tax rate instruments. I estimate the elasticity of broad income with respect to net-of-tax rates and the indirect impact of net-of-tax rates using the estimation method developed by Weber (2014) and address the endogeneity due to the mean reversion using instruments constructed based on lagged income. My baseline specification includes instruments based on one-year lagged income and examines the endogeneity of these instruments assuming that instruments constructed from two-year and three-year lagged income are exogenous. Using instruments based on lagged income, I find an estimate for the indirect impact of

[11] Net-of-tax base is one minus the share of taxable income with respect to broad income. Net-of-tax base also refers to the share of deductions with respect to broad income. Net-of-tax-base and the share of deductions are used interchangeably in this study.

changes in tax rates in the range of 0.07-0.12 which is significantly smaller than those reported by Kopczuk (2005).

2.3 Data

Data used in this essay are mostly similar to those used in the first essay. Similar to the first essay, I combine the individual SOI tax files into a panel of years 1979-1990. The estimates are primarily based on two-year differences. Moreover, a sensitivity analysis has also been presented for one- and three-year differences. I lose six years of observations for using two-year differences and for using instruments constructed based on up to four-year lagged income. Therefore, the estimation is restricted to between 1983 and 1988. The study uses broad income for the definition of income as the dependent variable.[12]

In my study, I follow the definition of the net-of-tax base used by Kopczuk (2005). Specifically, I consider all the itemized deductions, adjustments for AGI, and reported income that are not included in AGI (Adjusted Gross Income).[13] Inelastic standard deductions and personal exemptions are not included in the definition of net of tax base. The net of tax base is then computed as the ratio of the total of such itemized deductions, adjustments, and non-taxable reported income with respect to broad income. All income and, deduction components are converted into 1992 dollars using inflation indices used in Gruber and Saez (2002).

Similar to the first essay, I compute the marginal tax rates for this study using TAXSIM (Feenberg and Coutts, 1993) developed by NBER. Data used in this study are limited to those observations for which a taxpayer's marital status does not change between two years for the paired observations because a change in marital status can cause changes in income unrelated to changes in tax rates or tax bases. Finally, as I have done in the first essay, I exclude taxpayers whose broad income is less than $10,000 in the base year to avoid extreme mean reversion. I include several control variables in my estimation including indicators for marital status and the number of dependent children in the household, a full set of state and year dummy indicators. After imposing these restrictions, I obtain 16,184 observation for the baseline specification. The number of observations is smaller than that in the first essay primarily because there are observations with missing itemized deductions. Those observations with missing deductions can still be used in the estimations in the first essay but cannot be used in the second essay. Table B.1 presents the descriptive statistics for the variables used in the analysis.

---

[12] Broad income is defined in Section 1.4.2.
[13] Reported income that are not included in AGI are also not included in the taxable income.

2.4 Identification and Estimation

The Tax Reform Act 1986 (TRA86) was the most significant tax reform within a thirty-year window of that decade. TRA86 not only changed the marginal tax rates but also broadened the tax base. For instance, TRA86 reduced the deductible portion for certain business meals and entertainment. TRA86 eliminated the adjustment for married couples when both work. It also eliminated the deduction for personal interest (for itemizers) and eliminated the deductions for charitable contributions made by a non-itemizer. Moreover, TRA86 changed the treatment of moving expenses from an above-the-line adjustment to an itemized deduction. In addition to these changes, TRA86 also changed IRA limits and deductibility for medical and miscellaneous expenses. I use this variation caused by TRA86 to identify the impact of changes in tax base on the behavioral responses.

To estimate the impact of changes in tax base, I use a difference-in-differences estimation method. My baseline specification uses two-year differences where I make pairs of observations two years apart and regress the changes in income between pairs of observations on changes in net-of-tax rates and changes in the net-of-tax base along with other control variables. In the basic specification, I investigate the impact of tax base only and use only those observations whose marginal tax rates are unchanged across two years of the differences. For this analysis, I estimate the following equation.

$$\Delta ln Y_{it} = \alpha + \boldsymbol{X'}_{it}\boldsymbol{\beta_1} + \beta_3 \Delta ln \gamma_{i\,t} + \nu_{i\,t} \tag{2.1}$$

In equation (2.1), $\gamma_{i\,t}$ is the share of deductions with respect to broad income for individual i at time period t, Y_{it} is income of individual i at time period t and X_{it} represents a set of control variables used in the estimation. Then, I allow net-of-tax rates to vary and extend the analysis by estimating the following two equations.

$$\Delta ln Y_{it} = \alpha + \boldsymbol{X'}_{it}\boldsymbol{\beta_1} + \beta_2 \Delta ln(1 - \tau_{i\,t}) + \beta_3 \Delta ln \gamma_{i\,t} + \nu_{i\,t} \tag{2.2}$$

$$\Delta ln Y_{it} = \alpha + \boldsymbol{X'}_{it}\boldsymbol{\beta_1} + \beta_2 \Delta ln(1 - \tau_{i\,t}) + \beta_3 \Delta\, ln(\gamma_{i\,t} * (1 - \tau_{i\,t})) + \nu_{i\,t} \tag{2.3}$$

Here, τ_t is the marginal tax rate. β_3 in equation (2.2) represents the impact of changes in net-of-tax base controlling for marginal net-of-tax rates and other variables whereas β_3 in equation (2.3) represents the impact of tax base on EBI or the indirect impact of changes in tax rates. Kopczuk (2005) estimates equation (2.3) using instruments constructed from base-year income and I estimate the same equation using a method developed by Weber (2014).

The primary focus of this study is to estimate equation (2.3). However, to understand how the estimates vary across specifications, I also estimate equations (2.1) and (2.2) along with (2.3). I include the interaction between net-of-tax tax rates and net-of-tax base as well as net-of-tax rates but not the net-of-tax base separately. The rationale for such specification has been explained in Kopczuk (2005). Tax base matters to the taxpayers only through tax saving and, therefore, we may not expect an independent tax base effect.

The challenge of estimating the above equation is the endogeneity of the two explanatory variables: net-of-tax-rate and share of deductible expenses (i.e., net-of-tax base). Net-of-tax-rate is a direct function of income and therefore the exogeneity condition for identification is violated. On the other hand, some deductions, e.g. deductions for medical and miscellaneous expenditures are also direct functions of income and therefore subject to additional limitations. Moreover, the net-of-tax base can indirectly depend on income. This implies that the net-of-tax base is also correlated with the unobserved error term in the estimating equation and therefore is endogenous. Kopczuk (2005) addresses this endogeneity concern and uses an instrumental variable approach to consistently estimate the parameter. He constructs the instrumental variables for net-of-tax base and net-of-tax rates as functions of base-year income. However, as evident in Weber (2014), such instruments may not guarantee the exogeneity of instruments because these instruments are still functions of the base-year income, which is the dependent variable of the estimating equation.

I use lagged values of income to construct the instruments for net-of-tax rates and the share of deductible expenses. Synthetic tax rates are computed using the TAXSIM program provided by NBER. Specifically, I compute the predicted tax rates by running income components and other information for a lagged period through TAXSIM for the year of interest. Similarly, I compute the predicted net-of-tax base using lagged incomes.

I use a Difference-in-Sargan test to assess the endogeneity of the instruments used in the estimation. The Difference-in-Sargan test is a test of overidentifying restrictions that assumes a subset of instruments as exogenous and tests the validity of suspect instruments. It checks the validity of suspect instruments by computing the increase in Sargan’s J statistic when such suspect instruments are added to the estimation. To test the validity of two suspect instruments for two endogenous variables, at least three instruments are needed that are assumed to be exogenous. This means that at least five instruments are needed to implement the Difference-in-Sargan test in a specification with two endogenous variables. I use three instruments for each endogenous variable with a total of six instruments.

2.5 Empirical Results

I begin my analysis by first investigating the impact of the net-of-tax base only while keeping tax rates constant across two years of the paired observations. For that purpose, I estimate equation (2.1) using a limited sample. As opposed to the full sample, this limited sample includes only those pair of observations for which net-of-tax rates are equal across two years of the paired observations whereas the full sample does not have such restriction and the net-of-tax rates across two years of the paired observations in the full sample can be equal or unequal. The estimates using the limited sample are presented in Table 2.1.

Table 2. 1: Second-stage IV regression estimates using net-of-tax base only

	(1)	(2)	(3)
$\Delta ln(\boldsymbol{\gamma})$	0.0709***	0.0699***	0.0683***
	(0.0121)	(0.0119)	(0.0118)
Observations	2,226	2,226	2,226
R-squared	-0.065	-0.063	-0.060
Instruments	0 2 3 lags	1 2 3 lags	2 3 4 lags
First stage F-statistic	2391	2217	2289

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table (and in all other tables in this essay) adjust for heteroscedasticity clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1

Columns 1-3 of Table 2.1 present estimates using instruments of different lagged values. The first stage regression estimates for these columns of Table 2.1 are included in Table B.2 in Appendix B. The weakness of the instruments is examined by comparing the F-statistics from the first stage regressions with the minimum F-statistic suggested by Stock and Yogo (2002). The endogeneity of the instruments is examined by observing the p-value of the Difference-in-Sargan test. The null hypothesis of this test is that the suspect instrument(s) is(are) exogenous. The first stage F-statistic is large for each column implying that the instruments are strong. However, the Difference-in-Sargan p-value is not reported for all three columns because estimates in these three columns are based on only a small number of observations in the limited sample which includes a restriction that marginal net-of-tax rates are equal across two years of the paired observations. Hence, I do not have any information regarding the endogeneity of the instruments used in this table. However, in the subsequent tables, I remove this constraint, and as a result, I end up with a larger number of observations which are sufficient to compute the Different-in-Sargan p-value. Column (1) in Table 2.1 presents the estimates using the instruments for net-of-tax base constructed as a function of base-year income as well as the income of two- and three-year lags. Column (2) presents estimates using instruments with one-, two- and three-year lagged income. Column (3) presents estimates using instruments with two-, three- and four-year lagged income. In all three specifications, the coefficient on the net of tax base is positive and significant at the 1% level. The magnitudes are similar and in the order of 0.07 in all three columns.

Estimates in Table 2.1 are based on a limited sample where tax rates are unchanged across years of the paired observations. Table 2.2 eliminates this constraint and utilizes the full sample to estimate equation (2.2) which includes net-of-tax base and net-of-tax rates in the estimating equation.

Table 2. 2: Second stage regression estimates for the IV regression with net-of-tax rates and net-of-tax base

	(1)	(2)	(3)	(4)
$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$		-0.214*	0.203	-0.0320
		(0.112)	(0.201)	(0.219)
$\Delta ln(\boldsymbol{\gamma})$	0.106***	0.105***	0.106***	0.105***
	(0.00685)	(0.00666)	(0.00705)	(0.00682)
Observations	16,184	16,184	16,184	16,184
R-squared	-0.124	-0.052	-0.198	-0.112
Instruments	0 2 3 lags	0 2 3 lags	1 2 3 lags	2 3 4 lags
Diff-in-Sargan p-val	0.140	0.0485	0.862	0.366
First stage F-statistic		231, 4253	61, 4169	56, 4158

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table adjust for heteroscedasticity clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

I present column 1 in Table 2.2 to compare the results with those in Table 2.1. Column 1 in Table 2.1 and column 1 in Table 2.2 are based on the same specification but use different samples. Columns 2-4 of Table 2.2 examine the endogeneity and weakness of the instruments. The first stage regression estimates for this Table 2.2 are included in Table B.3 in Appendix B. Column 2 of Table 2.2 includes instruments for net-of-tax rates and net-of-tax base constructed based on base-year income, two-year lagged income and three-year lagged income. First stage F-statistics from the regressions of net-of-tax rates and net-of-tax base are 231 and 4253 respectively which are much larger than the minimum value suggested by Stock and Yogo[14] (2002). The Difference-in-Sargan test for this column examines whether the instruments for the two endogenous variables constructed based on base-year income are exogenous, assuming that the instruments based on two-year and three-year lagged income are exogenous. Although the instruments are not weak, the Difference-in-Sargan p-value is close to zero; therefore, we can strongly reject the null hypothesis that the instruments are exogenous. Column 3 includes instruments constructed based on one-year lagged income, two-year lagged income, and three-year lagged income. The Difference-in-Sargan p-value is 0.86, and therefore the instruments are more exogenous when constructed from one-, two- and three-year lagged income compared to those when constructed from base-year income, two-year lagged income and three-year lagged income. First stage F-statistics from the regressions of net-of-tax rates and net-of-tax base are 61 and 4169 respectively, suggesting that the endogeneity of the instruments is not a concern. Column 4 includes instruments constructed from two-year, three-year and four-year lagged income. While the instruments are strong according to the guidelines suggested by Stock and Yogo (2002), the Difference-in-Sargan p-value now decreases to 0.37. My preferred specification in Table 2.2 is that of column 3 for which the Diff-in-Sargan p-value is larger implying that instruments are more exogenous compared to those in the other two columns, and the first stage F-statistic is also large implying the instruments are strong. The coefficient on the net-of-tax rates is 0.20 and is not significant whereas the coefficient on the net-of-tax base is 0.11 and is significant at the 1% level. These results support the finding from Table 2.1, however, the magnitude of the coefficient on the net-of-tax base is slightly larger.

Estimated results from the IV regression with net-of-tax rates and the interaction are presented in Table 2.3. The first stage regression estimates for this table are included in Table B.4 in Appendix B. Columns 1-3 of Table 2.3 examine the endogeneity and weakness of the instruments. First stage F-statistics from all three columns imply strong instruments. A comparison of the Diff-in-Sargan p-values provides findings similar to the ones in Table 2.1. Instruments are more exogenous when they are constructed based on one-, two- and three-year lagged income (column 2). Column 2 is my preferred specification and the estimated results for this specification suggest that the coefficient on the net-of-tax rates is positive but insignificant. The coefficient on the interaction term is positive and significant at any conventional level of significance and its magnitude is 0.11. This

[14] For two endogenous variables and six excluded instruments, if one wants to limit the bias of IV estimator to 5% of OLS bias, the minimum F-statistic suggested by Stock and Yogo (2002) is 15.72.

simply means that for individuals with share of deductions equal to one, the indirect impact of broad income with respect to changes in net-of-tax rates is 0.11.

Tax return data include rich information on incomes and deductions but limited socio-demographic information. As highlighted in the literature, the existence of heterogeneous income growth can cast doubt on the validity of the estimated results because such heterogeneous growth is not well understood. Controlling for income as an alternative to controlling for socio-demographic information is a common approach in the literature. Table B.5 presents results for specifications including five-piece income splines constructed from different lagged values of income. The results are similar to the ones obtained from the preferred specifications in Table 2.3. In all columns, the first stage F-statistic is large. Except for column 2, estimate for the direct impact of changes in net-of-tax rates are small, positive, and insignificant in all specifications. The estimate for the indirect effect of changes in net-of-tax rate is strongly significant in all specifications and is robust to the inclusion of different income splines.

Table 2. 3: Second stage regression estimates from the IV regression with net-of-tax rates and the interaction

	(1)	(2)	(3)
$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	-0.319***	0.0972	-0.137
	(0.112)	(0.201)	(0.219)
$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))$	0.105***	0.106***	0.105***
	(0.00666)	(0.00705)	(0.00682)
Observations	16,184	16,184	16,184
R-squared	-0.052	-0.198	-0.112
Instruments	0 2 3 lags	1 2 3 lags	2 3 4 lags
Diff-in-Sargan p-val	0.0485	0.862	0.366
First stage F-statistic	231, 3257	62, 3178	56, 3178

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table adjust for heteroscedasticity clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

My last step of analysis includes examining the estimates with different difference lengths. Table B.7 presents such results and considers one-, two- and three-year difference lengths.[15] For the one-year differences specification, the estimate for the direct impact of changes in net-of-tax rates is 0.374 and is significant at the 10% level whereas the estimate for the indirect impact is strongly significant and is smaller (the magnitude is 0.07) compared to the ones from the previously discussed two-year differences specifications. In the case of three-year differences specification, the estimate for the direct impact is now negative (-0.997) and is significant at the 10% level of significance. The coefficient on the interaction term is significant at any conventional level and is slightly larger than the ones from the two-year differences specifications. The results are, in general, consistent with the previously obtained results and suggest much a smaller indirect impact as compared to the one in the prior literature.

The estimated results from all the specifications support the claim that tax elasticity is not an exogenous parameter, rather it depends on the net of tax base. To better understand the behavioral responses, I run a join significance test for the coefficients on the net of tax rates and the interaction term. I find both the coefficients are jointly significant at 1% level of significance. I compute the total tax elasticity by plugging the values of $\boldsymbol{\gamma}$ in estimated coefficients. I find the tax elasticity at the average net of tax base as 0.12. Moreover, depending on the value of $\boldsymbol{\gamma}$, tax elasticity varies between 0.09 to 0.26.

The estimated coefficient on the interaction term is largely different from that in Kopczuk (2005). One of the differences between Kopczuk's model and the model used in my study is how the interaction term between the net-of-tax rates and the tax base is defined. I define the interaction as $\Delta\, ln(\gamma_{i\,t} * (1 - \tau_{i\,t}))$ whereas Kopczuk (2005) defines it as $\Delta(\gamma_{i\,t} *\ ln(1 - \tau_{i\,t}))$. This means that there is a difference because of the scale change. However, to understand the other sources of differences, I estimate equation (2.3) using Kopczuk's (2005) baseline specification and using his definition. Kopczuk (2005) uses three-years differences and his baseline specification includes instruments for net-of-tax rates and the interaction term based on base-year information. I start with his baseline specification and make changes to specifications to arrive at my specification with three-year differences (column 3 of Table B.7). Table 2.4 presents the estimated results for such an analysis. Column (1) presents estimates using Kopczuk's (2005) specification and his definition of the interaction. This column includes instruments based on base-year income, two-year lagged income and three-year lagged income as well as income splines based on one-year lagged income. The estimate reported in his study for this specification is 0.79 whereas my replicated estimate is 0.30. My replicated estimate is largely different from his reported

[15] The number of observations is different for different columns in this table because I lose one additional year of observations for each additional year in the difference length.

estimate. However, it is noteworthy that the coefficient on the interaction term changes significantly as I increase lags of income for constructing the instruments.

Table 2. 4: Understanding the sources of differences from Kopczuk (2005)

	(1)	(2)	(3)	(4)	(5)
Net-of-tax rates	-0.909***	-1.484	-4.665**	8.488	-1.205
	(0.232)	(0.910)	(1.960)	(9.374)	(1.127)
Interaction between tax rate and tax base	0.295***	0.623*	0.293*	0.092***	0.087***
	(0.055)	(0.349)	(0.164)	(0.020)	(0.008)
Observations	13,086	13,086	13,086	3,279	4,823
R-squared	0.028	-0.083	-0.285	-4.343	-0.038
Lags of income for constructing instruments	0 2 3 lags	1 2 3 lags	2 3 4 lags	1 2 3 lags	1 2 3 lags
Lags of income for constructing splines	1 lag	1 lag	1 lag	1 lag	No spline
First stage F statistics	101, 35	10, 19	7, 2	0.6, 2539	11, 3106
Diff-in-Sargan p-val	0.324	0.493	0.0172	0.878	0.344

Note: Heteroscedasticity-robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Column (2) includes instruments based on one-, two-, three-year lagged income while keeping splines and other information unchanged from those of columns (1). Instruments are now more exogenous while first-stage F-statistics are still larger than the minimum value for 10% of OLS bias suggested by Stock and Yogo (2002). For this specification, the coefficient on the interaction term now increases to 0.623 and is significant at the 10% level of significance. Column (3) includes instruments based on two-, three- and four-year lagged income, as well as income splines based on one-year lagged income. However, the instruments are now weak. Column (4) uses the same specification as that in column (2) but uses the definition of the interaction term used in my study instead of the one used in Kopczuk (2005). Therefore, column (4) uses the same specification as in column (2) but a different definition of the interaction term and different sample. The coefficient is now much smaller. This suggests that much of the difference can be explained by the difference in the definition and the sample. The instruments are now weak. Column (5) uses the instruments based on one-, two-, and three-year lagged income as well as the same definition of the interaction term as that in column (4) but does not include any income splines. The specification in this column is similar to that used in column (3) of Table B.7. The first stage F-statistics are now larger than the suggested minimum value. A comparison between columns (2) and (5) suggests that much of the difference is caused by the difference in the definition of the interaction term. However, looking at columns 1, 2 and 3, it is also suggestive that the number of lags of income used to construct the instruments is also an important factor in determining the magnitude of the estimate. The coefficient largely varies as the number of lags to construct the instruments changes. Although the definitions of the interaction term are different, this difference should not matter when it comes to computing the overall elasticity. Having said that, the use of suitable instrument is needed for consistently estimating the elasticities. This implies that, because my estimation is based on plausibly more exogenous instruments, the results in my study provide some improvement over those in the prior literature.

2.6 Conclusion

The present study revisits the direct and indirect impact of changes in tax rates using a recent estimation method. In most of the specifications, the estimate for the direct impact is of the order of 0 to 0.37 and is insignificant. On the other hand, the indirect impact is strongly significant (at the 1% level) for all the specifications considered. Its value is also very much consistent across specifications with the value ranging between 0.102-0.106 for the specifications with two-year differences. The estimated indirect impact is strongly significant in the case of specifications with one-year differences as well, however, its magnitude is smaller (0.07). Similar to all other specifications, the estimated indirect impact is significant at the 1% level in the case of three-year differences, but its magnitude (0.123) is slightly larger than that in two-year differences. The estimated direct impact is, in general, in line with the prior literature. However, the estimated indirect impact is largely different from that of the prior literature. Kopczuk (2005) reports a large indirect impact of the order of 0.7 to 0.8 implying large efficiency cost of taxation through tax base. However, my study provides strong evidence that the indirect efficiency cost of taxation may not necessarily be as large as suggested by the present literature. These findings imply that EBI is not an exogenous parameter, rather it depends on the tax base. These findings have significant implications on the efficiency cost of taxation. These findings imply that the tax base has an impact on the efficiency of taxation and, therefore, any analysis of the efficiency cost of taxation ignoring the tax base effect is incomplete. We need to take into account both tax rates and tax base to better understand the efficiency cost of taxation.

CHAPTER III

THE CAPITALIZATION EFFECT OF SCHOOL COMPETITION ON HOME PRICES

3.1 Introduction

Theoretical residential sorting models argue that individuals value local public services and therefore choose their residential location according to the public services available in geographical regions. Their choice reflects how they value public services available in those geographical regions. One such public service is school quality. Parents care about the education quality available in regions and this concern translates into higher values of residential property in those regions. However, increased competition for enrollment provides students with opportunities to transfer to the school of their choice without having to change their residence. Therefore, school choice weakens the link between school quality and residential property value. My study examines the connection between increased school competition and property values. Specifically, this study empirically examines the possible linkages between (1) school competition and the distribution of home prices within the county, and (2) school competition and median home prices. This study provides evidence that an increase in school competition is associated with an overall increase in home value within the county. Moreover, this study also provides evidence that an increase in school competition is associated with a decrease in variability in home values within the county, however, these findings are sensitive to fixed effects and time trends included in the specifications.

This study proceeds as follows: Section 3.2 presents a conceptual framework for linkages between (1) school competition and the distribution of home prices within the county, and (2) school competition and median home prices; Section 3.3 describes the existing literature; Section 3.4 describes the data used in the study; Section 3.5 presents the empirical methods; Section 3.6 presents the results obtained in the study; and Section 3.7 presents the conclusion of the study.

3.2 Conceptual Framework

Consider two school districts: a good school district and a bad one.[16] The services from the good school district are more valuable to parents than those from the bad one. Initially, the state does not adopt open enrollment and thereby one's school attendance zone is a hard default. In such a case, it is likely that home prices are higher in good school districts and lower in bad ones (Black 1999,; and Figlio and Lucas 2004). After the adoption of open enrollment, attendance zones are no longer relevant, and parents are no longer restricted to attendance zones. Parents now have the option to choose any school from the good and the bad school districts regardless of their residential locations. Because houses are more expensive and schools are better in good school districts, parents in a bad school district will maximize their utility by sending their children to the schools in the good district where services are better while continuing to live in the bad school district where houses are cheaper. The empirical analyses in this paper are based on the following assumptions:

- Transportation costs to/from schools are negligible;
- Families are well-informed about the exact services provided by the schools, recent school choice reform initiatives by the states and the home prices;
- In the absence of school choice, the attendance zone is a hard default and difficult to opt out of; and
- Alternatives to traditional public schools (magnet, charter and private schools) are as good as "good" public schools.

The adoption of open enrollment is associated with the change in residential property value in the two districts. Residents in the bad school district are now enjoying better services provided by the schools in the good school district while maintaining the same residential area with cheaper house prices. This, in turn, implies that residences in the bad school district are now more valuable after the adoption of open enrollment than they were before the adoption of open enrollment. On the other hand, demand for homes in the good school district decreases after adoption of open enrollment. The changes in the relative valuation for homes between the two school districts will lead to changes in market prices for homes in those districts. One should expect home prices in bad school districts

[16] Good and bad school districts are defined based on the aspects that the parents care about and expect from schools.

to increase weakly[17] whereas home prices in the good school district should decrease weakly. This relative change in home prices influences the distribution of home prices across two districts. In addition to affecting the relative home prices of the two districts, the adoption of open enrollment also increases the competition for enrollment among schools as parents now have the freedom to choose a school from neighboring districts as well. The adoption of open enrollment should, therefore, reduce the inequality in home prices across two districts and increase school competition. With adoption of open enrollment, the residents have more freedom of sending their children to the schools of their choice. Thus, overall satisfaction of the residents of the combined region of the two districts should increase, leading to an increase in overall home prices.

This story extends to school choice among traditional public, charter, magnet, and private schools as well. The introduction of alternatives to traditional public school allows parents to choose a school among four different types of schools anywhere within the county and the relative home prices will be affected among the areas with varying school quality within the county in a similar manner as in the case of open enrollment. I use the Herfindahl-Hirschman Index[18] (HHI) for enrollment among four different types of schools to represent school competition. A lower value of HHI is associated with less concentrated enrollment which in turn represents a greater choice for parents in selecting a school for their children and vice versa. I provide a precise definition of HHI in a later section. It might be possible that there is only one type of school within the county when the value of HHI is at its maximum of HHI=10000. The way HHI for enrollment is defined, the minimum concentration and thereby maximum choice might be possible when there is equal enrollment among all four types of schools and this case will be represented by HHI=4*25^2=2500. Based on the assumptions and conceptual analyses described above, the following hypotheses are drawn:

Hypothesis 1: Less concentrated counties will have less variability in home prices.

Hypothesis 2: If households value school choice, less concentrated counties will have higher median home prices.

---

[17] Quality of services at schools in a good school district as perceived by residents in a bad school district may be higher or equal to the perceived quality of services at schools in bad school districts.
[18] Herfindahl-Hirschman Index is defined as the sum of squares of enrollment shares for four different types of schools within the county.

3.3 Literature Review

Considering its importance, there has been a substantial discussion in the literature regarding the linkage between school quality and property values as well as other welfare measures. Black (1999) investigates parental valuation for good schools using a school attendance zone boundary approach. She uses a regression discontinuity design based on the data on both sides of school attendance zones in Massachusetts to estimate the impact of test scores on home prices. This study assumes that residential houses near the boundaries of school attendance zones are expected to have similar neighborhood characteristics thereby reducing the possibility of bias resulting from omitted neighborhood characteristics. Her results show that a one percent increase in test scores results in a 0.5 percent increase in home prices. Figlio and Lucas (2004) address the question of whether the state-administered grades assigned to schools have any influence on house prices as well as on residential location. They use data on repeated sales on individual residential properties in Florida and found that school grades impact house prices and residential locations, however, they also found that these estimated effects diminish over time. Dhar and Ross (2012) address a similar question by examining differences across school district boundaries rather than those across attendance zones. Based on the data from Connecticut, they found a significant positive effect of test scores on property values. Reback (2005) examines the effect of public-school choice reforms on house prices. Using data on inter-district choice from Minnesota, he finds that school districts with the net exit of students to neighboring districts experience home price increases whereas net student inflow is associated with a decrease in home prices. His estimation results show that the magnitude of the effect for student outflow is larger than the magnitude of the effect for student inflow. Based on this comparison he concludes that the net welfare impact of the expansion of school choice is positive, even though the difference between the magnitudes of effects for incoming and outgoing transfer rates is not statistically significant. Other related works include those by Brunner et al., (2012), Schwartz et al., (2014) and Chung (2015). These related papers, in general, investigate the effects of outflow and inflow of students, as a result of school choice reforms, on the residential property value as well as on the residential location. They find that the effect of the expanded school choice on low performing geographical regions (as defined by school districts or school zones) is positive and vice versa.

There are a few companion studies in the literature that are most relevant for my study. For example, Reback (2005) addresses the question of school choice reform and evaluates the expanded school choice in terms of effects for inflow and outflow of students. Other related works by Brunner et al., (2012), Schwartz et al., (2014) and Chung (2015) also follow Reback (2005) by evaluating school choice in terms of the effects of inflow and outflow of students on home value. Reback (2005) and Brunner et al., (2012) find that initially "low-quality" districts experience a net outflow of students and an increase in home value whereas initially "high-quality" districts experience a net inflow of students and a decrease in home value. These patterns of student mobility, in turn, reduce the inequality of home values. None of these studies explicitly investigate and empirically test a direct relation between school choice and the inequality of home prices. Moreover, the findings from these studies are based on smaller geographical regions. There is no evidence that the findings from these studies based on smaller geographic regions generalize to the entire country. My study adds to the literature by examining the relationships between (1) school competition and the distribution of home prices within the county, and (2) school competition and median home prices. In addition, in contrast to the previous literature, I utilize data representing the entire country. A strong set of control variables on demographic as well as housing characteristics is utilized to estimate the model. Utilizing county-level data, this study provides evidence that an expanded school competition is associated with an increase in home value. Moreover, this study also provides evidence that an expanded school competition is associated with a decrease in variability in home value within a county, however, this finding is sensitive to sample size and to controls for unobserved heterogeneity included in the specifications.

3.4 Data

The dataset for this study includes county-level panel data for four years (2010, 2012, 2014 and 2016) with 819 counties from all over the US. Data are collected from two different sources: the National Center for Education Statistics (NCES) and American Community Survey (ACS) by the U.S. Census Bureau.[19] Data collected from NCES are school-level data on enrollment into all traditional public, charter, magnet and private schools. These school-level data on enrollment are then used to compute the Herfindahl-Hirschman Index (HHI) for enrollment among four different types of schools within each county, therefore, school level

[19] Data are collected from the following tables:
U.S. Census Bureau; American Community Survey, 2010 American Community Survey 1-Year Estimates, Tables s2506, CP04 and DP05; generated using American FactFinder; <http://factfinder2.census.gov>; (11 May 2019)
U.S. Census Bureau; American Community Survey, 2012 American Community Survey 1-Year Estimates, Tables s2506, CP04 and DP05; generated using American FactFinder; <http://factfinder2.census.gov>; (11 May 2019)
U.S. Census Bureau; American Community Survey, 2014 American Community Survey 1-Year Estimates, Tables s2506, CP04 and DP05; generated using American FactFinder; <http://factfinder2.census.gov>; (11 May 2019)
U.S. Census Bureau; American Community Survey, 2016 American Community Survey 1-Year Estimates, Tables s2506, CP04 and DP05; generated using American FactFinder; <http://factfinder2.census.gov>; (11 May 2019)

observations are converted into county-level data. The HHI measure serves as a proxy for school competition within each county. HHI for school enrollment is defined as:

$$HHI = \sum_i s^2 \tag{3.1}$$

Here i represents the type of school and s represents the share of enrollment for school type. Based on the definition of HHI for enrollment, HHI is bounded between a minimum of 2500 and a maximum of 10000. A lower value of HHI represents less concentrated enrollment which further implies a greater choice for parents in selecting a school for their children and vice versa.

Data on home value and other socio-demographic information are collected from three different tables (s2506, CP04, and DP05) of ACS 1-year estimates by U.S. Census Bureau. The unit of observation for each of these data tables is the county. Table DP04 contains the distributional data on home prices, among other information. These distributional data have been used to construct the Gini coefficient for home values. Gini coefficient for home value is a measure of dispersion for home value within the county and is one of the two dependent variables of interest. The definition of the Gini coefficient is given in equation (3.2).

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2\sum_{i=1}^{n}\sum_{j=1}^{n}x_j} \tag{3.2}$$

In the above definition, x_i and x_j are individual home values. There are certain advantages of Gini coefficient that makes it a preferred choice over other measures (standard deviation or variance) to represent dispersion. First, Gini coefficient is a more generalized form of measure for dispersion. Moreover, it is not only invariant to the scale of observations but also bounded. One can analyze extreme cases (Gini=0 representing perfect equality and Gini=10,000 representing perfect inequality) with the help of Gini coefficient which is not possible using other measures. In this study, Gini coefficient for home value is bounded[20] between a minimum of 0 and a maximum of 10,000. Table for data on HHI from NCES and all three tables from ACS contain county names and state names. The combination of county and state names is unique in each dataset. This unique combination of county and state names is used to combine all the individual data sets to obtain a

[20] To maintain consistency in scale with other variables expressed in percentages, Gini Coefficient is measured in percentage squares and bounded between a minimum of 0 and a maximum of 10,000. This scale of the Gini coefficient used in this paper contrasts the conventional value of Gini coefficient bounded between 0 and 1.

single merged data set. The summary statistics for the combined data set are shown in Table C.1.

3.5 Estimation Method

Figures 1-4 provide a visualization of the linkages examined in this study. Figure 1 and 2 illustrate the unconditional relation between HHI and the median home price for 2010 and 2016 respectively. These two figures display a weakly discernable negative correlation between the two variables. Figure 3 and 4 show scatter plots for the relation between school competition and the Gini coefficient of home prices for 2010 and 2016 respectively. These two scatter plots display a very subtle positive relation between school competition and the inequality of home prices.[21]

To provide empirical evidence concerning whether these relations hold, I estimate the following equation (3.3) with Pooled OLS regression:

$$Y_i = \alpha + \beta_1 * hhi_i + \beta_2 * X_i + \varepsilon_i \quad (3.3)$$

In the equation (3.3), Y_i represents two dependent variables of interest: median home prices and variability of home prices as measured by the Gini coefficient for home prices. The independent variable of interest is HHI and the coefficient of interest is β_1. X represents a vector of control variables. This set of controls include data on demographic as well as housing characteristics.

[21] Years 2012 and 2014 are not displayed for brevity. However, the scatter plots for those two years also have similar patterns.

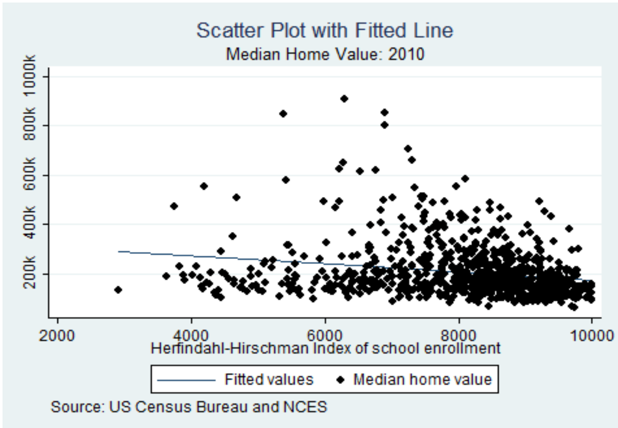

Figure 1: Home value vs HHI: 2010

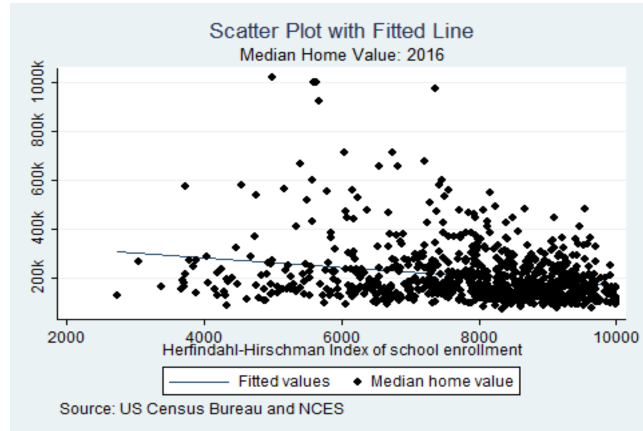

Figure 2: Home value vs HHI: 2016

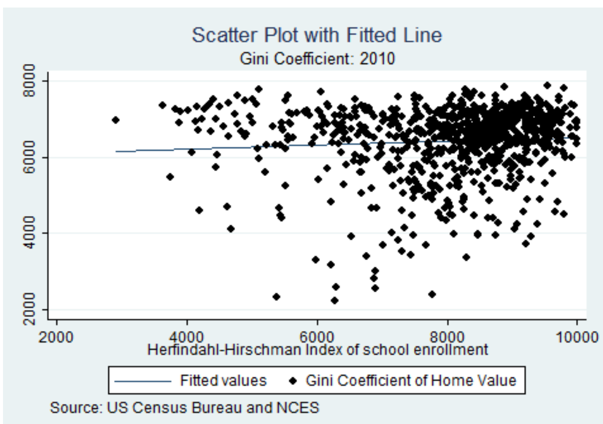

Figure 3: Gini-coefficient vs HHI: 2010

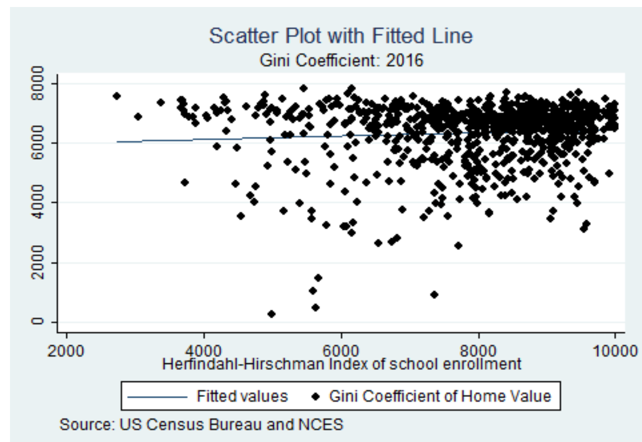

Figure 4: Gini-coefficient vs HHI: 2016

As reported in the literature, one of the challenges in estimating the capitalization effect of school competition on home value arises because the school competition may be correlated with unobserved neighborhood characteristics such as crime rate, other public services available within the geographical unit, demographic compositions, etc. Although the dataset includes a rich set of sociodemographic control variables, regressing the property values on other independent and control variables can result in biased estimates because of such correlation between the school competition and unobserved neighborhood characteristics. Such unobserved neighborhood characteristics that are invariant with respect to time for a geographical unit can be removed for by adding geographical unit-specific fixed effects. The estimation equation including such fixed effects is as below:

$$Y_{i\,t} = \theta_i + \beta_1 * hhi_{i\,t} + \beta_2 * X_{i\,t} + \varepsilon_{i\,t} \qquad (3.4)$$

In equation (3.4), θ_i represents county-specific time-invariant effects. In the following sections, I estimate equation (3.4) using both fixed effects and random effects model and compare the results.

Estimates including county-specific fixed effects rely on the assumption that there are no inter-county movements for attending schools. However, it is reasonably likely that parents living near the border of two counties may want to send their children to a school located outside of their own county of residence in order for their children to obtain better schooling. In this case, the identification assumption is violated. To address this concern, I extend the analysis by including metropolitan area-specific fixed effects instead of county-specific fixed effects. For such analysis, equation (3.4) is estimated, however, in this case, θ_i represents metropolitan area-specific time-invariant fixed effects. These estimates considering metropolitan area fixed effects remain valid if students do not attend a school located outside of their own metropolitan area of residence.

Including county-specific fixed effects controls for time-invariant unobserved characteristics of counties. However, concerns remain about the unobserved characteristics that change over time, where county-specific fixed effects are unable to absorb such unobserved characteristics. These unobserved time-varying characteristics can also lead to bias in the estimates. To address such concern, I test the robustness of estimates by relaxing the assumption of time invariance in the characteristics of geographical units and by absorbing any linear time-varying unobserved characteristics by including county-specific linear time trends in the estimation. With county-specific linear time trends, the estimation equation takes the following form:

$$Y_{i\,t} = \alpha + \gamma_i t + \beta_1 * hhi_{i\,t} + \beta_2 * X_{i\,t} + \varepsilon_{i\,t} \qquad (3.5)$$

3.6 Results

3.6.1 The Effect of School Competition on Median Home Value

Table 3.1 presents the estimates for the effect of school competition on median home prices. Column 1 includes estimates from Pooled OLS regression without any control variables. The coefficient on HHI is -16.06 and is significant at the 1% level. The estimated results suggest that for one unit decrease in concentration (ΔHHI = -1), home price increases by 16.06 dollars which is 0.01 % at the median. The estimated coefficient on HHI in equation (3.3) using Pooled OLS including control variables is -5.251 (column 2). The point estimate is significant at the 1% level. This result suggests that one unit decrease in HHI is associated with an increase in home value by $5.25 which is a 0.003% at the median. To interpret this result, consider the case when a county goes from fully concentrated enrollment (HHI=10000 with no competition for school enrollment) to fully unconcentrated enrollment (HHI=0 with maximum school competition). This means that initially the county does not adopt open enrollment in which case the attendance zone is a hard default and difficult to opt-out of and there is no school competition, and then the county adopts an open enrollment in which case the student is fully flexible to choose any school within that county and the school competition is maximum. Such a change in enrollment is associated with an increase in home value by $52,510. This increase in home value represents a 26.50% increase at the median.

The OLS estimate in column 2 relies on the assumption that the unobserved error is uncorrelated with school competition which is very strong assumption. Columns 3 and 4 show the estimates using fixed effects and random effects models respectively recognizing county-specific fixed effects. My preferred specification is fixed effects model. The underlying assumption for the random effects model is that county specific unobserved characteristics are uncorrelated with HHI which is very strong assumption. Although the dataset contains socio-demographic information, variables included in the dataset may not be sufficient to absorb the unobserved county specific characteristics. If the county specific unobserved heterogeneity is correlated with HHI, the random effects model will produce inconsistent estimates. Fixed effects model relaxes such assumption and can produce consistent estimate when HHI is correlated with unobserved heterogeneity. Hausman test results also suggest that the fixed effects model is appropriate (p-value from the Hausman test is 0). Column 3 presents the estimated results from fixed effects model. The coefficient on HHI of school enrollment is -2.999 and is significant at the 1% level. These estimates support the hypothesis that a decrease in concentration is associated with an increase in home value where a one unit decrease in concentration of enrollment is associated with an increase in home value by 0.002% at the median.

Table 3. 1: Estimates for regression of home value using full sample

	(1) OLS - No Control	(2) OLS - Controlled	(3) FE	(4) RE
HHI	-16.06***	-5.251***	-2.999***	-4.924***
	(1.304)	(0.580)	(0.676)	(0.629)
Observations	3,250	3,165	3,165	3,165
R-squared	0.050	0.891	0.433	
Number of counties			819	819

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The coefficients on control variables are not listed in this (and subsequent) table (tables) for brevity.

A careful examination of the values of HHI for enrollments shows that not every panel unit has enough within-county variation in HHI across years. To avoid the fact that panel units with low within-county variation may affect the estimates and to confirm that the results are based on enough variation in HHI across years, I identify the counties for the largest 50 percent within-county variation in absolute value in HHI across years and create a new sample of observations which experience the largest 50 percent swing in within-county variation in HHI across years. This limited sample consists of a balanced panel data set for 669 counties. The results based on this limited sample are shown in Table 3.2. With the new limited sample, the sign of the coefficients on HHI from Pooled OLS, fixed effects as well as random effects models remains unchanged (negative) and are, therefore, in agreement with the hypothesis. The estimates from all the specifications are precise at the 5% or lower level of significance.[22]

The estimates presented so far rely on the assumption that the observations are homoscedastic. However, Breusch-Pagan as well as White test for heteroskedasticity suggests a strong rejection of no heteroskedasticity. To examine the direction and magnitude of the coefficients and the magnitude of standard errors recognizing heteroskedasticity, heteroskedasticity-robust

[22] The data fail to meet the asymptotic assumption for Hausman Test, so it is not possible to determine which model is appropriate.

estimates are obtained considering clusters at the county level. Columns 1-3 of Table 3.3 present heteroskedasticity-robust estimates. Consistent with previously estimated coefficients, the sign of the coefficients is unchanged suggesting that a decrease in the concentration of enrollments is associated with an increase in home value. The coefficients from Pooled OLS, fixed effects, and random effects specifications are significant at the 1% level.

Table 3. 2: Estimates for regression of home value using limited sample

	(1) No Control	(2) OLS - Controlled	(3) FE	(4) RE
HHI	-15.81***	-4.838***	-2.287**	-4.555***
	(1.996)	(0.839)	(0.896)	(0.797)
Observations	1,627	1,591	1,591	1,591
R-squared	0.046	0.898	0.437	
Number of counties			669	669

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Heteroskedasticity-robust estimates with clusters at the county levels are also obtained with the limited sample of 669 counties that have the largest 50 percent swing in within-county variation in HHI across years. The results based on this limited sample are shown in Table C.2. With the new limited sample, coefficients from Pooled OLS and fixed effects models are -4.838 and -2.287 respectively and are both significant at the 1% level.

Table 3. 3: Estimates for regression of home value using full sample with cluster at county level

	(1) OLS	(2) FE	(3) RE
HHI	-5.251***	-2.999***	-4.924***
	(0.919)	(0.700)	(0.765)
Observations	3,165	3,165	3,165
R-squared	0.891	0.433	
Number of counties		819	819

Note: Heteroscedasticity-robust standard errors clustered by the county are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The estimates presented so far include fixed effects at the county level. The estimates considering county-specific fixed effects rely on the assumption that students attend schools within the county of their residence. However, it is possible that parents near the border of two counties send their children to a school on the other side of the border in order for their children to obtain better schooling. In such a case, the identification assumption is violated. I address this concern by re-estimating equation (3.4) including metropolitan area-specific fixed effects instead of county-specific fixed effects. The estimates are shown in Table 3.4. The results reiterate the negative relation between the concentration of enrollments and home value. Specifically, the coefficient on HHI of enrollments from OLS and random effects are -2.581 and -2.934 respectively and are significant at the 5% and 1% levels respectively.

The final step of analysis involves including county-specific linear time trends. Fixed effects included in the previous estimates are time-invariant and are, therefore, unable to absorb any characteristics that change over time. County-specific linear time trend can absorb unobserved county-specific characteristics that vary linearly over time. Such estimated coefficients of equation (3.5) are presented in Table C.3. Consistent with the hypothesis, the coefficient on HHI of enrollment is again negative. However, the coefficient is now imprecise at the 10% level. The fact that the estimates are imprecise when I absorb county specific linear time trends are suggestive that HHI may be correlated with the unobserved linear time trends.

Table 3. 4: Estimates for regression of home value with metropolitan area level data (clustered at metro level)

	(1) OLS	(2) FE	(3) RE
HHI of school enrollment	-2.581**	-0.916	-2.934***
	(1.221)	(1.070)	(1.041)
Constant	309,039**	77,235	6,793
	(132,142)	(133,374)	(98,197)
Observations	1,327	1,327	1,327
R-squared	0.881	0.328	
Number of Metro Area		383	383

Note: Heteroscedasticity-robust standard errors clustered by the metropolitan area are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

3.6.2 Effect of School Competition on Gini Coefficients of Home Prices

The estimated coefficient for the effect of school competition on the inequality of home prices are shown in Table 3.5. The results in this table assume homogeneity in the unobserved errors. Column 1 represents the estimates from Pooled OLS regression without any control variables. The estimated coefficient on HHI of school enrollment in this specification is 0.0330 and is significant at the 1% level. Column 2 presents coefficients in equation (3.3) using Pooled OLS regression including the full set of control variables. The estimate of interest is -0.0179 which is significant at the 5% level of significance. Results shown in columns 3 and 4 recognize county-specific fixed effects. Although I report estimates from random effects model, my preferred specification is fixed effects model. Hausman test also suggest that the fixed effects model is appropriate. The coefficient on the HHI of school enrollment in the fixed effects model is 0.0392 and is significant at the 1% level. The estimated coefficient from the fixed effects model supports the hypothesis and suggests that going from fully unrestricted enrollment (HHI=0) to fully restricted enrollment (HHI=10000) increases the variability by 6.09% at the median.

As in the case of analysis for median home value, I identify the counties with enough variation across years. This new sample includes 669 counties with the largest mere 50 percent within-county variation in absolute value in HHI across years. The results based on this limited sample are shown in Table C.4. Although the sign of coefficient from Pooled OLS now changes to negative, the sign of the coefficient from the preferred fixed effects model remains unchanged (positive), thereby, supporting the hypothesis that an increase in competition is associated with a decrease in inequality of home prices.

The specifications in Table C.5 relax the assumption of homoskedasticity and present heteroskedasticity-robust estimates considering clusters at the county levels. The sign of the coefficient on HHI in the preferred fixed effects model is positive which is in line with the findings in the previous tables. The coefficient is significant at the 1% level.

Heteroskedasticity-robust estimates based on the limited sample with clusters at the county levels are also obtained. These estimates are presented in Table C.6. Similar to Table C.5, the Pooled OLS estimates contrast the hypothesis whereas estimates from the preferred fixed effects model are in agreement with the hypothesis. The coefficient is significant the 1% level of significance.

Table 3. 5: Estimates for regression of Gini-coefficient of home value using full sample

	(1) No Control	(2) OLS	(3) FE	(4) RE
HHI	0.0330***	-0.0179**	0.0392***	0.0168*
	(0.0112)	(0.00805)	(0.0129)	(0.00970)
Observations	3,250	3,165	3,165	3,165
R-squared	0.009	0.705	0.150	
Number of conuties			819	819

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

To address the concern discussed in a previous section, estimates are obtained by using metropolitan area level observations and including metropolitan area fixed effects in the estimating equation. Table 3.6 includes such estimates with metropolitan areas as units of observations. The coefficients from Pooled OLS, fixed effects as well as random effects specifications are now positive suggesting that a decrease in concentration is associated with a decrease in variability in home prices within the metropolitan area. The coefficients from fixed effects and random effects models are significant at the 10% level of significance whereas the coefficient from the Pooled OLS is not.

Table 3. 6: Estimates for regression of Gini-coefficient of home value with metropolitan level data (clustered at metro level)

	(1) OLS	(2) FE	(3) RE
HHI of school enrollment	0.000723	0.0289*	0.0234*
	(0.0145)	(0.0171)	(0.0133)
Constant	11,427***	8,908***	11,660***
	(1,637)	(2,366)	(1,552)
Observations	1,327	1,327	1,327
R-squared	0.669	0.171	
Number of metro areas		383	383

Note: Heteroscedasticity-robust standard errors clustered by the metropolitan area are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

As for the final step of the robustness check, the estimates considering county-specific linear time trend are obtained and shown in Table C.7. Consistent with the hypothesis and with findings from previously presented specifications, the coefficient is still positive suggesting that as school competition increases, home prices become less dispersed. However, the coefficient is now imprecise at any conventional level of significance.

3.6.3 Understanding the Driving Force in the Composition of HHI

So far, I have examined the response for a change in HHI as a measure of school competition on home values and the inequality of home value. In general, I find that an increase in school competition increases home values and decreases the inequality of home values. However, this analysis does not provide any information regarding the relative magnitude of driving forces within the composition of HHI. HHI has been constructed based on four different types of schools: public, charter, magnet and private. Each of the four types of school is different in terms of quality of services provided, and therefore, each type of school may have different driving forces on the overall response. To better understand how each type of school affects home values and the inequality of home values, I regress home values and the inequality of home values on the share of total enrolment for private, charter and magnet school categories along with other independent variables using fixed effects model. The estimates are presented in Table 3.7. Column (1) presents the estimates for the regression of median home values on the share of enrolment in each category. The coefficients on all three shares are positive and significant. The coefficient on the share of enrolments of the charter school is more prominent than the coefficients for the other categories. This is suggestive that charter school is the most prominent in explaining the variation in home values as compared to other categories. Column (2) presents the estimates for the regression of the Gini-coefficient of home values on the share of enrolments in each category. The coefficients on all three shares are negative, however, the coefficient on the share of enrolment of the magnet school is significant at the 10% level whereas the other two coefficients are insignificant. These results are indicative that magnet school has an impact in reducing the inequality of home prices whereas the other two types of schools are not prominent.

Table 3. 7: Understanding the variation in the composition of enrolment in HHI

	(1)	(2)
	Median home value	Gini coefficient of home values
Share of enrolment of charter school	787.7**	-2.227
	(316.7)	(4.987)
Share of enrolment of magnet school	239.2***	-4.870***
	(84.06)	(1.659)
Share of enrolment of private school	599.5**	-6.130
	(247.6)	(4.988)
Observations	3,165	3,165
R-squared	0.434	0.151
Number of counties	819	819

Note: Heteroscedasticity-robust standard errors clustered by county in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

3.7 Conclusion

My study examines the connection between the expansion of school competition and property values. Specifically, this study empirically examines the possible linkages between (1) school competition and the distribution of home prices, and (2) school competition and median home prices. Utilizing county-level data from all 50 states across the United States, this study provides evidence that an increase in school competition is associated with an increase in home value. Moreover, this study also provides evidence that an expansion of school competition is associated with a decrease in variability in home value, however, both findings are sensitive to fixed effects and time trends included in the specifications. My findings add to the prediction from the existing literature. Reback (2005) and Brunner et al., (2012) establish a loose connection between the variability of home prices and school competition using inflow and outflow of students. Predictions from both studies suggest a reduction in inequality of home prices for a greater school competition. The finding from my study adds to the findings from these studies by providing evidence that those predictions hold in the context of nationally representative data. Moreover, the findings from this study support the hypothesis that an increase in school competition is associated with an increase in home prices and vice versa. Furthermore, I find the relative magnitudes of the driving forces among different types of schools. My findings provide evidence that the share of enrolment of the charter school is most prominent in determining the home values whereas the share of enrolment of the magnet school is the only statistically significant factor in explaining the inequality in home values. These results provide some evidence that the market as a whole values expanded school competition. Whereas the results from my study support a linkage between school competition and home values, these results should be interpreted with caution because these results are inconclusive and sensitive to specifications. Results are significant when I include time-invariant county fixed effects. However, the estimates are imprecise when I absorb county-specific linear time trends. These findings are suggestive that school competition may be correlated with the unobserved linear time trend. Perhaps the growth of wealth of a county, which can be a determinant of home prices, may increase the demand for better schooling thereby affecting the school competition, or maybe, other time-varying determinants of home prices can be correlated with the school competition.

BIBLIOGRAPHY

Anderson, T.W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. Journal of Econometrics, 18(1): 47–82.

Auten, G. E. and Carroll, R. (1999). The effect of income taxes on household income. Review of Economics and Statistics, 81 (4): 681–93.

Bakija, J. and Heim, B. (2008). How does charitable giving respond to incentives and income? Dynamic panel estimates accounting for predictable changes. NBER Working Paper Series, No. 14237.

Black, S.E. (1999). Do better schools matter? Parental valuation of elementary education. Quarterly Journal of Economics, 114 (2): 577–600.

Bound, J. and Johnson, G. (1992). Changes in the structure of wages during the 1980s: An evaluation of alternative explanations. American Economic Review, 82: 371–392.

Brunner, E. J., Cho, S. W. and Reback, R. (2012). Mobility, housing markets, and schools: Estimating the effects of inter-district choice programs. Journal of Public Economics, 96(7- 8): 604–614.

Brunner, E. and Imazeki, J. (2008). Tiebout choice and universal school vouchers. Journal of Urban Economics, 63 (1): 253–279.

Brunner, E., Sonstelie, J. and Thayer, M., 2001. Capitalization and the voucher. Journal of Urban Economics, 50: 517–536.

Burman, L. E. and Randolph, W. C. (1994). Measuring permanent responses to capital-gains tax changes in panel data. The American Economic Review, 84 (4): 794–809.

Calabrese, S., Epple, D., Romer, T. and Sieg, H. (2006). Local public good provision: voting, peer effects, and mobility. Journal of Public Economics, 90 (6–7): 959–981.

Carlson, D., Lavery, L. and Witte, J. F. (2011). The determinants of inter-district open enrollment flows: evidence from two states. Educational Evaluation and Policy Analysis, 33 (1): 76–94.

Carroll, R. (1998). Tax rates, taxpayer behavior, and the 1993 Tax Act. Office of Tax Analysis, US Department of the Treasury, mimeograph.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced form methods, Annual Review of Economics 1: 878–889.

Chung, I. H. (2015). School choice, housing prices, and residential sorting: empirical evidence from inter-and intra-district choice. Regional Science and Urban Economics, 52: 39–49.

Clapp, J., Nanda, A. and Ross, S. (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.

Dee, T.S. (2000). The capitalization of education finance reforms. Journal of Law and Economics, 43 (1): 185–214.

Dhar, P. and Ross, S. L. (2012). School district quality and property values: examining differences along school district boundaries. Journal of Urban Economics, 71: 18–25.

Downes, T. and Zabel, J. (2002). The impact of school characteristics on housing prices: Chicago 1987–1991. Journal of Urban Economics, 52: 1–25.

Eissa, N. (1995). Taxation and labor supply of married women: The Tax Reform Act of 1986 as a natural experiment. NBER Working Paper 5023.

Eissa, N. and Giertz, S. (2008). Trends in high-income and behavioral responses to taxation: evidence from executive compensation and statistics of income data. Working Paper 2006-14, Congressional Budget Office.

Epple, D. and Romano, R. E. (1998). Competition between private and public schools, vouchers, and peer-group effects. American Economic Review, 88 (1): 33–62.

Feenberg, D. and Coutts, E. (1993). An introduction to the TAXSIM model. Journal of Policy Analysis and Management, 12 (1): 189–194.

Feldstein, M. (1995). The effect of marginal tax rates on taxable income: A panel study of the 1986 Tax Reform Act. Journal of Political Economy, 103 (3): 551–572.

Ferreyra, M. (2007). Estimating the effects of private school vouchers in multidistrict economies. American Economic Review, 97: 789–817.

Figlio, D., Husted, T. and Kenny, L. (2004). Political economy of the inequality in school spending. Journal of Urban Economics, 55: 338–349.

Figlio, D. and Lucas, M. (2004). What's in a grade? School report cards and the housing market. American Economic Review, 94(3): 591–604.

Giertz, S. (2010). The elasticity of taxable income during the 1990s: new estimates and sensitivity analysis. Southern Economic Journal, 77(2): 406–433.

Goolsbee, A. (2000). What happens when you tax the rich? Evidence from executive compensation. Journal of Political Economy, 108: 352–78.

Gruber, J. and Saez, E. (2002). The elasticity of taxable income: evidence and implications. Journal of Public Economics, 84: 1–32.

Hall, B. and Liebman, J. (2000). The taxation of executive compensation. In Poterba, J., editor, Tax policy and the Economy 10. Cambridge, MA: MIT Press. 1–44.

Hargaden, E. P. (2020). Taxpayer responses in good times and bad. Journal of Economic Behavior and Organization, forthcoming.

Harju, J. and Matikka, T. (2016). The elasticity of taxable income and income-shifting: what is “real“ and what is not? International Tax and Public Finance, 23(4): 640–669.

Holmlund, B. and Soderstrom, M. (2011). Estimating dynamic income responses to tax reform. The B.E. Journal of Economic Analysis & Policy, 11(1): 1–38.

Kawano, L., Weber, C. and Whitten, A. (2016). Estimating the elasticity of broad income for high-income taxpayers, Working Paper.

Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. Journal of Public Economics, 89: 2093–2119.

Kreiner, C. T., Leth-Petersen, S. and Skov, P. E. (2014). Year-end tax planning of top management: evidence from high-frequency payroll data. American Economic Review, 104(5):154–158.

Kreiner, C. T., Leth-Petersen, S. and Skov, P. E. (2016). Tax reforms and intertemporal shifting of wage income: evidence from Danish monthly payroll records. American Economic Journal: Economic Policy, 8(3): 233–257

Krugman, P. (1992). The right, the rich, and the facts: deconstructing the income distribution debate. The American Prospect (Fall 1992).

Levy, F. and Micheal, R. C. (1991). The economic future of American families: income and wealth trends. Washing, DC: The Urban Institute Press.

Moffitt, R. and Wilhelm, M. (2000). Taxation and the labor supply: decisions of the affluent. In Slemrod, J., editor, Does Atlast Shrug? The economic consequences of taxing the rich. New York: Harvard University Press and Russell Sage Foundation.

Murphy, K. M. and Welch, F. (1992). The structure of wages. Quarterly Journal of Economics, 107: 285–326.

Pflueger, C. E. and Wang, S. (2013). WEAKIVTEST: Stata module to perform weak instrument test for a single endogenous regressor in TSLS and LIML, Statistical Software Components S457732, Boston College Department of Economics, revised 09 Sep 2018.

Reback, R. (2005). House prices and the provision of local public services: capitalization under school choice programs. Journal of Urban Economics, 57: 275–301.

Sammartino, F. and Weiner, D. (1997). Recent evidence on taxpayers’ response to the rate increases in the 1990’s. National Tax Journal, 50 (3): 683–705.

Schwartz, A.E., Voicu, I. and Horn, K.M. (2014). Do choice schools break the link between public schools and property values? Evidence from house prices in New York City. Regional Science and Urban Economics, 49: 1–10.

Slemrod, J. (1990). The economic impact of Tax Reform Act of 1986. In Slemrod, J., editor, Do Taxes Matter? The Impact of The Tax Reform Act of 1986. Cambridge. MA: MIT Press, 1–12.

Slemrod, J. (1995). Income creation or income shifting? Behavioral responses to the Tax Reform Act of 1986. American Economic Review, 85(2):175–180.

Slemrod, J. (1996). High income families and the tax changes of the 1980s: the anatomy of behavioral responses. In Feldstein, M. and Poterba, J., editors, Empirical Foundation of Household Taxation. Chicago: University of Chicago Press, 169–189.

Slemrod, J. and Kopczuk, W. (2002). The optimal elasticity of taxable income. Journal of Public Economics, 84(1):91–112.

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear IV regression. NBER Working Paper No. 284.

Weber, C. (2014). Toward obtaining a consistent parameter of the elasticity of taxable income using Difference-in-Differences. Journal of Public Economics. 117, 90–103.

APPENDIX

A Appendix for Chapter 1

Table A. 1: Descriptive statistics

	(2)	(3)	(5)	(6)	(8)	(9)
	1983		1983-88		1988	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Broad income	51,861.23	45,917.04	52,783.29	53,953.10	55,278.36	79,150.78
Taxable income	36,278.10	37,490.10	36,901.19	46,063.13	39,679.14	71,014.75
Federal tax rate	24.83	7.27	23.49	7.45	20.79	6.65
State tax rate	4.65	3.25	4.49	3.09	4.34	2.84
Single	0.24	0.43	0.25	0.43	0.26	0.44
Married	0.71	0.45	0.70	0.46	0.68	0.47
No dependent	0.51	0.50	0.52	0.50	0.54	0.50
One dependent	0.18	0.38	0.19	0.39	0.19	0.39
Two dependents	0.21	0.40	0.20	0.40	0.19	0.39
Three dependents	0.07	0.26	0.06	0.25	0.06	0.23
Under age 65 & not blind	0.89	0.31	0.90	0.30	0.90	0.29
One person over age 65 or blind	0.07	0.25	0.06	0.24	0.06	0.23
Number of observations	4,002		24,731		4,247	

Table A. 2: First stage estimates for Weber's baseline specification in column 1 of Table 1.1

	(1) $\Delta ln(1-\tau_t)$
$\Delta\ln(1-\tau_t^{p\ 2\ lags})$	0.17***
	(0.02)
$\Delta\ln(1-\tau_t^{p\ 3\ lags})$	0.12***
	(0.02)
$\Delta\ln(1-\tau_t^{p\ 4\ lags})$	0.15***
	(0.02)
Observations	23,438
R-squared	0.10
First stage F-statistic	121.75

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A. 3: First stage estimates for columns 2-4 of Table 1.1

	(1)	(2)	(3)	(4)	(5)	(6)
	Column 2 of Table 1.1		Column 3 of Table 1.1		Column 4 of Table 1.1	
	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$
$\Delta\ln(1-\tau_t^{p\ 0\ lag})$	0.72***	-0.03				
	(0.02)	(0.02)				
$\Delta\ln(1-\tau_t^{p\ 2\ lags})$	0.01	0.03*	0.11***	0.02	0.15***	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta\ln(1-\tau_t^{p\ 3\ lags})$	0.03*	-0.02	0.11***	-0.03	0.12***	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta\ln(1-\tau_{t+1}^{p\ 0\ lag})$	0.36***	0.21***				
	(0.02)	(0.03)				
$\Delta\ln(1-\tau_{t+1}^{p\ 2\ lags})$	-0.06***	0.14***	-0.02	0.14***	0.02	0.16***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta\ln(1-\tau_{t+1}^{p\ 3\ lags})$	-0.03	0.12***	0.01	0.12***	0.02	0.12***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta\ln(1-\tau_t^{p\ 1\ lag})$			0.21***	0.02		
			(0.02)	(0.02)		
$\Delta\ln(1-\tau_{t+1}^{p\ 1\ lag})$			0.10***	0.15***		
			(0.02)	(0.02)		
$\Delta\ln(1-\tau_t^{p\ 4\ lag})$					0.10***	0.04**
					(0.02)	(0.02)
$\Delta\ln(1-\tau_{t+1}^{p\ 4\ lags})$					0.02	0.10***
					(0.02)	(0.02)
Observations	24,731	24,731	24,731	24,731	24,731	24,731
R-squared	0.15	0.05	0.06	0.05	0.06	0.05

Note: Heteroscedasticity-robust standard errors clustered by individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A. 4: First stage estimates for Table 1.2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Column 1 of Table 1.2		Column 2 of Table 1.2		Column 3 of Table 1.2		Column 4 of Table 1.2	
	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$
$\Delta\ln(1-\tau_t^{p\ 1\ lag})$	0.20***	0.02	0.16***	-0.00	0.18***	-0.01	0.19***	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta\ln(1-\tau_t^{p\ 2\ lags})$	0.12***	0.02	0.08***	-0.00	0.10***	0.00	0.09***	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta\ln(1-\tau_t^{p\ 3\ lags})$	0.11***	-0.02	0.08***	-0.04**	0.10***	-0.04*	0.10***	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta\ln(1-\tau_{t+1}^{p\ 1\ lag})$	0.09***	0.16***	0.05**	0.14***	0.07***	0.14***	0.07***	0.15***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
$\Delta\ln(1-\tau_{t+1}^{p\ 2\ lags})$	-0.01	0.13***	-0.05**	0.11***	-0.03	0.11***	-0.03	0.11***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
$\Delta\ln(1-\tau_{t+1}^{p\ 3\ lags})$	0.01	0.12***	-0.03	0.10***	-0.01	0.11***	-0.01	0.11***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	24,007	24,007	24,007	24,007	24,007	24,007	24,007	24,007
R-squared	0.06	0.05	0.09	0.06	0.07	0.05	0.07	0.05

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A. 5: First stage estimates for Table 1.3

	(1)	(2)	(3)	(4)
	Column 1 of Table 1.3		Column 2 of Table 1.3	
	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$
$\Delta \ln\,(1-\tau_t^{p\ 1\ lag})$	0.10***	-0.02	0.19***	0.00
	(0.04)	(0.03)	(0.03)	(0.02)
$\Delta \ln\,(1-\tau_t^{p\ 2\ lags})$	0.02	0.02	0.09***	-0.02
	(0.04)	(0.03)	(0.03)	(0.03)
$\Delta \ln\,(1-\tau_t^{p\ 3\ lags})$	0.05	-0.05	0.07***	-0.06**
	(0.04)	(0.03)	(0.03)	(0.03)
$\Delta \ln\,(1-\tau_{t+1}^{p\ 1\ lag})$	0.09**	0.04	0.04	0.16***
	(0.04)	(0.04)	(0.03)	(0.03)
$\Delta \ln\,(1-\tau_{t+1}^{p\ 2\ lags})$	-0.04	0.10**	-0.05	0.11***
	(0.03)	(0.04)	(0.03)	(0.03)
$\Delta \ln\,(1-\tau_{t+1}^{p\ 3\ lags})$	-0.01	0.04	-0.04	0.14***
	(0.03)	(0.03)	(0.03)	(0.03)
Observations	10,984	10,984	13,747	13,747
R-squared	0.06	0.04	0.11	0.07

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A. 6: First stage estimates for Table 1.4

	(1) Single	(2) Single	(3) Married	(4) Married
	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$
$\Delta \ln(1-\tau_t^{p\ 1\ lags})$	0.27***	-0.01	0.18***	0.04*
	(0.05)	(0.04)	(0.03)	(0.02)
$\Delta \ln(1-\tau_t^{p\ 2\ lags})$	0.04	0.01	0.12***	0.02
	(0.06)	(0.05)	(0.02)	(0.02)
$\Delta \ln(1-\tau_t^{p\ 3\ lags})$	0.13**	-0.05	0.09***	-0.02
	(0.06)	(0.04)	(0.02)	(0.02)
$\Delta \ln(1-\tau_{t+1}^{p\ 1\ lag})$	0.17***	0.10	0.08***	0.16***
	(0.05)	(0.06)	(0.03)	(0.03)
$\Delta \ln(1-\tau_{t+1}^{p\ 2\ lags})$	-0.05	0.16**	0.00	0.12***
	(0.05)	(0.06)	(0.03)	(0.03)
$\Delta \ln(1-\tau_{t+1}^{p\ 3\ lags})$	-0.02	0.14**	0.02	0.11***
	(0.05)	(0.05)	(0.02)	(0.03)
Observations	6,198	6,198	17,202	17,202
R-squared	0.05	0.03	0.07	0.06

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A. 7: First stage estimates for Table 1.5

	(1)	(2)	(3)	(4)	(5)	(6)
	Column 1 of Table 1.5	Column 1 of Table 1.5	Column 2 of Table 1.5	Column 2 of Table 1.5	Column 3 of Table 1.5	Column 3 of Table 1.5
	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$	$\Delta ln(1-\tau_t)$	$\Delta ln(1-\tau_{t+1})$
$\Delta \ln(1-\tau_t^{p\ 1\ lags})$	0.21***	0.02	0.17***	-0.01	0.18***	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta \ln(1-\tau_t^{p\ 2\ lags})$	0.11***	0.02	0.07***	0.00	0.09***	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta \ln(1-\tau_t^{p\ 3\ lags})$	0.11***	-0.03	0.08***	-0.05**	0.09***	-0.04**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta \ln(1-\tau_{t+1}^{p\ 1\ lag})$	0.09***	0.16***	0.05**	0.13***	0.06**	0.13***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta \ln(1-\tau_{t+1}^{p\ 2\ lags})$	-0.01	0.13***	-0.05*	0.11***	-0.03	0.12***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\Delta \ln(1-\tau_{t+1}^{p\ 3\ lags})$	0.00	0.12***	-0.03	0.10***	-0.02	0.11***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	24,937	24,663	24,937	24,663	24,937	24,663
R-squared	0.06	0.05	0.08	0.06	0.06	0.05

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

B Appendix for Chapter 2

Table B. 1: Descriptive statistics

	(2)	(3)	(5)	(6)	(8)	(9)
	1983		1983-88		1988	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Broad income	52,925.09	41,374.39	58,895.28	50,764.16	68,318.12	75,087.10
Taxable income	36,008.49	34,458.23	40,279.93	43,225.07	48,186.69	66,843.28
Federal tax rate	24.14	8.33	23.86	8.15	21.62	7.02
State tax rate	4.56	3.30	4.57	3.17	4.52	2.92
Net of tax base	22.83	15.28	24.23	14.15	22.40	12.12
Single	0.15	0.36	0.15	0.36	0.16	0.37
Married	0.81	0.39	0.81	0.40	0.80	0.40
No dependent	0.44	0.50	0.44	0.50	0.44	0.50
One dependent	0.19	0.39	0.20	0.40	0.21	0.41
Two dependents	0.24	0.42	0.24	0.43	0.24	0.43
Three dependents	0.09	0.28	0.08	0.28	0.07	0.26
Under age 65 & not blind	0.89	0.31	0.91	0.28	0.92	0.27
One person over age 65 or blind	0.07	0.26	0.05	0.23	0.05	0.21
Number of observations	3,348		16,184		2,214	

Table B. 2: First stage regression estimates for Table 2.1

	(1)	(2)	(3)
	$\Delta ln(\boldsymbol{\gamma})$	$\Delta ln(\boldsymbol{\gamma})$	$\Delta ln(\boldsymbol{\gamma})$
$\Delta \ln \gamma^{p\ 0\ lag}$	0.67***		
	(0.22)		
$\Delta \ln \gamma^{p\ 2\ lag}$	0.18	0.54	0.74**
	(0.38)	(0.48)	(0.35)
$\Delta \ln \gamma^{p\ 3\ lag}$	0.08	0.13	-0.53
	(0.34)	(0.36)	(0.35)
$\Delta \ln \gamma^{p\ 1\ lag}$		0.26	
		(0.37)	
$\Delta \ln \gamma^{p\ 4\ lag}$			0.72***
			(0.09)
Observations	2,226	2,226	2,226
R-squared	0.89	0.89	0.89

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B. 3: First stage regression estimates for Table 2.2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\Delta ln(\boldsymbol{\gamma})$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\boldsymbol{\gamma})$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\boldsymbol{\gamma})$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\boldsymbol{\gamma})$
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{p\ 0\ lag})$		0.68***	0.22***				
		(0.02)	(0.08)				
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{p\ 2\ lag})$		0.10***	0.04	0.20***	0.11	0.25***	0.09
		(0.02)	(0.08)	(0.02)	(0.08)	(0.02)	(0.08)
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{p\ 3\ lag})$		0.02	-0.26***	0.10***	-0.21**	0.10***	-0.23***
		(0.02)	(0.08)	(0.02)	(0.08)	(0.02)	(0.08)
$\Delta \ln \gamma^{p\ 0\ lag}$	0.99***	0.01	0.99***				
	(0.09)	(0.02)	(0.09)				
$\Delta \ln \gamma^{p\ 2\ lag}$	-0.04	0.01	-0.03	0.02	0.10	0.00	0.47**
	(0.18)	(0.04)	(0.18)	(0.05)	(0.25)	(0.04)	(0.19)
$\Delta \ln \gamma^{p\ 3\ lag}$	-0.05	-0.02	-0.06	-0.03	0.07	-0.09*	-0.47
	(0.18)	(0.04)	(0.17)	(0.04)	(0.19)	(0.05)	(0.30)
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{p\ 1\ lag})$				0.25***	-0.04		
				(0.02)	(0.08)		
$\Delta \ln \gamma^{p\ 1\ lag}$				0.01	0.73***		
				(0.05)	(0.23)		
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{p\ 4\ lag})$						0.17***	0.04
						(0.02)	(0.08)
$\Delta \ln \gamma^{p\ 4\ lag}$						0.09*	0.90***
						(0.05)	(0.27)
Observations	16,184	16,184	16,184	16,184	16,184	16,184	16,184
R-squared	0.85	0.17	0.85	0.10	0.85	0.09	0.85

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B. 4: First stage regression estimates for Table 2.3

	(1)	(2)	(3)	(4)	(5)	(6)
	Column 1 of Table 2.3		Column 2 of Table 2.3		Column 3 of Table 2.3	
	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1} - \boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1} - \boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1} - \boldsymbol{\tau}))$
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{0}\ \boldsymbol{lag}})$	0.66***	-0.10				
	(0.03)	(0.14)				
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{2}\ \boldsymbol{lag}})$	0.09**	0.16	0.18***	0.18	0.24***	-0.13
	(0.05)	(0.23)	(0.06)	(0.31)	(0.05)	(0.25)
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{3}\ \boldsymbol{lag}})$	0.04	-0.16	0.12**	-0.16	0.19***	0.42
	(0.05)	(0.23)	(0.05)	(0.24)	(0.06)	(0.35)
$\Delta ln(\gamma^{p\ 0\ lag} * (\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{0}\ \boldsymbol{lag}}))$	0.01	1.00***				
	(0.02)	(0.10)				
$\Delta ln(\gamma^{p\ 2\ lag} * (\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{2}\ \boldsymbol{lag}}))$	0.01	-0.02	0.02	0.12	0.00	0.47**
	(0.04)	(0.21)	(0.05)	(0.29)	(0.04)	(0.22)
$\Delta ln(\gamma^{p\ 3\ lag} * (\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{3}\ \boldsymbol{lag}}))$	-0.02	-0.08	-0.03	0.04	-0.09*	-0.56*
	(0.04)	(0.20)	(0.04)	(0.22)	(0.05)	(0.34)
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{1}\ \boldsymbol{lag}})$			0.24***	-0.53*		
			(0.05)	(0.28)		
$\Delta ln(\gamma^{p\ 1\ lag} * (\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{1}\ \boldsymbol{lag}}))$			0.01	0.74***		
			(0.05)	(0.27)		
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{4}\ \boldsymbol{lag}})$					0.08	-0.78**
					(0.05)	(0.33)
$\Delta ln(\gamma^{p\ 4\ lag} * (\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{4}\ \boldsymbol{lag}}))$					0.09*	0.99***
					(0.05)	(0.31)
Observations	16,184	16,184	16,184	16,184	16,184	16,184
R-squared	0.17	0.80	0.10	0.80	0.09	0.80

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B. 5: Second stage regression estimates for the sensitivity of income controls

	(1)	(2)	(3)	(4)
	No Spline	Base-Year Spline	1-Year Lagged Spline	2-Year Lagged Spline
$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	0.111	1.155***	0.253	0.152
	(0.201)	(0.365)	(0.265)	(0.250)
$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))$	0.106***	0.104***	0.102***	0.102***
	(0.00712)	(0.00783)	(0.00713)	(0.00710)
Observations	16,003	16,003	16,003	16,003
R-squared	-0.203	-0.630	-0.241	-0.206
Instruments	1 2 3 lags	1 2 3 lags	1 2 3 lags	1 2 3 lags
Diff-in-Sargan p-val	0.882	0.698	0.869	0.899
First stage F-statistic	61, 3239	24, 3354	32, 3221	35, 3171

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table adjust for heteroscedasticity clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

Table B. 6: First stage estimates for Table B.5

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Column 1		Column 2		Column 3		Column 4	
	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))$
$\Delta ln(\mathbf{1}-\boldsymbol{\tau^{p\ 1\ lag}})$	0.24***	-0.50*	0.16***	-0.76***	0.19***	-0.59**	0.22***	-0.53*
	(0.05)	(0.28)	(0.05)	(0.28)	(0.05)	(0.28)	(0.05)	(0.28)
$\Delta ln(\mathbf{1}-\boldsymbol{\tau^{p\ 2\ lag}})$	0.18***	0.17	0.11**	-0.04	0.14**	0.10	0.13**	0.10
	(0.06)	(0.31)	(0.06)	(0.30)	(0.06)	(0.30)	(0.06)	(0.31)
$\Delta ln(\mathbf{1}-\boldsymbol{\tau^{p\ 3\ lag}})$	0.11**	-0.19	0.07	-0.27	0.09*	-0.20	0.09*	-0.20
	(0.05)	(0.24)	(0.05)	(0.25)	(0.05)	(0.24)	(0.05)	(0.24)
$\Delta ln(\gamma^{p\ 1\ lag} * (\mathbf{1}-\boldsymbol{\tau^{p\ 1\ lag}}))$	0.02	0.75***	0.03	0.80***	0.02	0.77***	0.01	0.75***
	(0.05)	(0.27)	(0.04)	(0.26)	(0.04)	(0.27)	(0.05)	(0.27)
$\Delta ln(\gamma^{p\ 2\ lag} * (\mathbf{1}-\boldsymbol{\tau^{p\ 2\ lag}}))$	0.00	0.08	0.01	0.11	0.01	0.10	0.01	0.11
	(0.05)	(0.29)	(0.05)	(0.28)	(0.05)	(0.28)	(0.05)	(0.29)
$\Delta ln(\gamma^{p\ 3\ lag} * (\mathbf{1}-\boldsymbol{\tau^{p\ 3\ lag}}))$	-0.02	0.06	-0.03	-0.01	-0.03	0.03	-0.03	0.05
	(0.04)	(0.22)	(0.04)	(0.22)	(0.04)	(0.22)	(0.04)	(0.22)
Observations	16,003	16,003	16,003	16,003	16,003	16,003	16,003	16,003
R-squared	0.10	0.80	0.15	0.80	0.11	0.80	0.11	0.80

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B. 7: Second stage regression estimates for IV regressions with different difference lengths

	(1)	(2)	(3)
	1-year diff	2-year diff	3-year diff
$\Delta ln(\mathbf{1}-\boldsymbol{\tau})_{1-year\ diff}$	0.374*		
	(0.223)		
$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))_{1-year\ diff}$	0.0682***		
	(0.00605)		
$\Delta ln(\mathbf{1}-\boldsymbol{\tau})_{2-year\ diff}$		0.0972	
		(0.201)	
$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))_{2-year\ diff}$		0.106***	
		(0.00705)	
$\Delta ln(\mathbf{1}-\boldsymbol{\tau})_{3-year\ diff}$			-0.997*
			(0.570)
$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))_{3-year\ diff}$			0.123***
			(0.00811)
Observations	21,024	16,184	12,221
R-squared	-0.272	-0.198	-0.072
Instruments	1 2 3 lags	1 2 3 lags	1 2 3 lags
Diff-in-Sargan p-val	0.237	0.862	0.817
First stage F-statistic	40, 4392	62, 3278	23, 2275

Note: Heteroscedasticity-robust standard errors clustered by the individual are in parentheses. The first-stage F-statistics in this table adjust for heteroscedasticity clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

Table B. 8: First stage estimates for Table B.7

	(1)	(2)	(3)	(4)	(5)	(6)
	Column 1 of Table B.7		Column 2 of Table B.7		Column 3 of Table B.7	
	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1} - \boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1} - \boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1} - \boldsymbol{\tau}))$
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{1}\ \boldsymbol{lag}})_{1-year\ diff}$	0.20***	-0.54**				
	(0.05)	(0.23)				
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{2}\ \boldsymbol{lag}})_{1-year\ diff}$	0.14***	-0.04				
	(0.05)	(0.25)				
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{3}\ \boldsymbol{lag}})_{1-year\ diff}$	0.12***	-0.10				
	(0.04)	(0.25)				
$\Delta ln(\gamma^{p\ 1\ lag} * (\mathbf{1} - \boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{1}\ \boldsymbol{lag}}))_{1-year\ diff}$	-0.03	0.53**				
	(0.04)	(0.21)				
$\Delta ln(\gamma^{p\ 2\ lag} * (\mathbf{1} - \boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{2}\ \boldsymbol{lag}}))_{1-year\ diff}$	0.01	0.22				
	(0.04)	(0.23)				
$\Delta ln(\gamma^{p\ 3\ lag} * (\mathbf{1} - \boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{3}\ \boldsymbol{lag}}))_{1-year\ diff}$	0.03	0.18				
	(0.04)	(0.24)				
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p}\ \mathbf{1}\ \boldsymbol{lag}})_{2-year\ diff}$			0.24***	-0.29		
			(0.05)	(0.33)		

Table B.8: First stage estimates for Table B.7 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)
	Column 1 of Table B.7		Column 2 of Table B.7		Column 3 of Table B.7	
	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma * (\mathbf{1}-\boldsymbol{\tau}))$
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 2\ lag}})_{2-year\ diff}$			0.18***	-0.42		
			(0.06)	(0.41)		
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 3\ lag}})_{2-year\ diff}$			0.12**	0.33		
			(0.05)	(0.31)		
$\Delta ln(\gamma^{p\ 1\ lag} * (\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 1\ lag}}))_{2-year\ diff}$			0.01	0.22		
			(0.05)	(0.30)		
$\Delta ln(\gamma^{p\ 2\ lag} * (\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 2\ lag}}))_{2-year\ diff}$			0.02	0.62		
			(0.05)	(0.38)		
$\Delta ln(\gamma^{p\ 3\ lag} * (\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 3\ lag}}))_{2-year\ diff}$			-0.03	-0.41		
			(0.04)	(0.28)		
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 1\ lag}})_{3-year\ diff}$					0.12***	-0.01
					(0.04)	(0.39)

Table B.8: First stage estimates for Table B.7 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)
	Column 1 of Table B.7		Column 2 of Table B.7		Column 3 of Table B.7	
	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma *(\mathbf{1}-\boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma *(\mathbf{1}-\boldsymbol{\tau}))$	$\Delta ln(\mathbf{1}-\boldsymbol{\tau})$	$\Delta ln(\gamma *(\mathbf{1}-\boldsymbol{\tau}))$
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 2\ lag}})_{3-year\ diff}$					0.06	-0.46
					(0.05)	(0.54)
$\Delta ln(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 3\ lag}})_{3-year\ diff}$					0.03	0.34
					(0.04)	(0.32)
$\Delta ln(\gamma^{p\ 1\ lag} *(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 1\ lag}}))_{3-year\ diff}$					-0.01	0.12
					(0.04)	(0.39)
$\Delta ln(\gamma^{p\ 2\ lag} *(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 2\ lag}}))_{3-year\ diff}$					-0.01	0.50
					(0.05)	(0.51)
$\Delta ln(\gamma^{p\ 3\ lag} *(\mathbf{1}-\boldsymbol{\tau}^{\boldsymbol{p\ 3\ lag}}))_{3-year\ diff}$					0.02	-0.28
					(0.03)	(0.28)
Observations	21,024	21,024	16,184	16,057	12,233	11,769
R-squared	0.05	0.82	0.10	0.27	0.06	0.22

Note: Heteroscedasticity-robust standard errors clustered by individual are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

C Appendix for Chapter 3

Table C. 1: Summary statistics

	Mean	Std. Dev.
Median home value	198,156	108,921
Gini Coefficient of Home Value	6,438	919.7
Herfindahl-Hirschman Index of school enrollment	7,942	1,430
Total housing units	135,037	219,430
Percent of occupied units	88.09	6.551
Percent of structure with 1 unit detached	65.94	11.09
Percent of structures with 1 unit attached	5.109	5.210
Percent of structures with 2 units	3.565	3.121
Percent of structures with 3 or 4 units	3.982	2.409
Percent of structures with 5 to 9 units	4.342	2.178
Percent of structures with 10 to 19 units	3.782	2.575
Percent of structures with 20 or more units	5.676	5.907
Percent of units with 1 room	1.683	1.705
Percent of units with 2 rooms	2.065	1.482
Percent of units with 3 rooms	7.511	3.403
Percent of units with 4 rooms	15.65	3.958
Percent of units with 5 rooms	20.72	4.612
Percent of units with 6 rooms	18.76	3.388
Percent of units with 7 rooms	12.92	2.576
Percent of units with 8 rooms	9.072	2.773
Percent of units occupied by owner	67.00	8.998
Percent of units with no vehicles	7.075	5.176
Percent of units with one vehicle	32.72	5.488
Percent of units with two vehicles	38.80	4.779
Percent of units with no telephone	2.569	1.380
Percent of units with mortgage	64.52	8.797
Monthly cost for units with mortgage	1,472	439.6
Percent of structures built after 1980	47.68	16.99
Total population	327,870	575,999
Percent of male population	49.23	1.204
Median Age	38.20	4.599
Percent of white population	81.59	14.57
Percent of Hispanic or Latino	11.16	13.05
Median household income	80,995	18,278
Observations	3,250	3,250

Note: The data set includes county-level panel data for four years for 819 counties.

Table C. 2: Robust estimates for regression of home value using limited sample with cluster at county level

	(1) OLS	(2) FE	(3) RE
HHI	-4.838***	-2.287***	-4.555***
	(1.102)	(0.804)	(0.858)
Observations	1,591	1,591	1,591
R-squared	0.898	0.437	
Number of counties		669	669

Note: Heteroscedasticity-robust standard errors clustered by the county are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table C. 3: Estimates for regression of home value with county-specific time trends

	(1)
HHI of school enrollment	-0.107
	(0.677)
Constant	-274,233***
	(80,115)
Observations	3,156
R-squared	0.996

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C. 4: Estimates for regression of Gini coefficient of home value using limited sample

	(1) NO CONTRL	(2) OLS	(3) FE	(4) RE
HHI	0.0158	-0.0198*	0.0391**	0.0211*
	(0.0169)	(0.0114)	(0.0160)	(0.0117)
Observations	1,627	1,591	1,591	1,591
R-squared	0.013	0.725	0.218	
Number of counties			669	669

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C. 5: Heteroskedasticity-robust estimates for regression of Gini coefficient of home values clustered at county level using full sample

	(1) OLS	(4) FE	(5) RE
HHI	-0.0179	0.0392***	0.0168
	(0.0130)	(0.0135)	(0.0106)
Observations	3,165	3,165	3,165
R-squared	0.705	0.150	
Number of counties		819	819

Note: Heteroscedasticity-robust standard errors clustered by the county are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C. 6: Robust estimates for regression of Gini coefficient of home values clustered at county level using limited sample

	(1) OLS	(4) FE	(5) RE
HHI	-0.0198	0.0391***	0.0211*
	(0.0145)	(0.0137)	(0.0110)
Observations	1,591	1,591	1,591
R-squared	0.725	0.218	
Number of counties		669	669

Note: Heteroscedasticity-robust standard errors clustered by the county are in parentheses*** p<0.01, ** p<0.05, * p<0.1

Table C. 7: Estimates for regression of Gini-coefficient with county-specific time trend

	(1)
Herfindahl-Hirschman Index of school enrollment	0.00401
	(0.0145)
Constant	13,819***
	(1,719)
Observations	3,156
R-squared	0.971

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

VITA

Md Sabbirul Haque was born in Netrokona, Bangladesh on January 16, 1979. He received a Bachelor of Science in Applied Physics, Electronics and Communication Engineering from the University of Dhaka, Bangladesh in 2004 and a Master of Arts in Economics from the University of Tennessee, Knoxville in 2017. He earned his Doctor of Philosophy in Economics in August 2020.