8-2016

Essays in labor economics and panel data analysis

Evan S. Totty

Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Part of the Economic Theory Commons

Recommended Citation

https://docs.lib.purdue.edu/open_access_dissertations/869

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
This is to certify that the thesis/dissertation prepared

By Evan Totty

Entitled
Essays in Labor Economics and Panel Data Analysis

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Mohitosh Kejriwal
Kevin Mumford
Justin Tobias
Jack Barron

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University’s “Policy of Integrity in Research” and the use of copyright material.

Approved by Major Professor(s): Mohitosh Kejriwal

Approved by: Justin Tobias 7/21/2016

Head of the Departmental Graduate Program Date
To my grandparents.
ACKNOWLEDGMENTS

I would like to thank my co-chairs, Mohitosh Kejriwal and Kevin Mumford, for their advice and guidance on my research projects and the job market. I would like to thank Justin Tobias, Jack Barron, and the rest of the faculty for their helpful comments and suggestions. I would like to thank my classmates for their companionship. I would like to thank my co-authors from the third chapter of this dissertation: Hope Gerlach, Anu Subramanian, and Tricia Zebrowski. I would like to thank my parents and brother for their constant belief in and support of me. Finally, I would like to thank my friends Matthew, Shae, Neil, Hope, Anu, and Bill for their encouragement and many hours spent listening to presentations about minimum wages.
# TABLE OF CONTENTS

<p>| LIST OF TABLES | vii |
| LIST OF FIGURES | ix |
| ABSTRACT | x |
| 1 | The Effect of Minimum Wages on Employment: A Factor Model Approach | 1 |
| 1.1 | Introduction | 1 |
| 1.2 | Empirical approach | 7 |
| 1.2.1 | Multi-factor error structure | 7 |
| 1.2.2 | Factor model estimators | 10 |
| 1.2.3 | Factor model approach versus other approaches in literature | 13 |
| 1.2.4 | Limitations of the factor model approach | 19 |
| 1.3 | Data | 20 |
| 1.3.1 | Data sources | 20 |
| 1.3.2 | Cross section dependence and data size | 23 |
| 1.4 | Results | 24 |
| 1.4.1 | Minimum wage-employment elasticity | 24 |
| 1.4.2 | Alternative teenage employment data sources | 27 |
| 1.4.3 | Number of common factors for IFE estimation | 28 |
| 1.4.4 | Accounting for the difference between OLS and factor model estimators | 29 |
| 1.4.5 | Throwing out the baby with the bathwater? | 31 |
| 1.4.6 | Robustness checks | 32 |
| 1.5 | Simulation | 38 |
| 1.6 | Discussion and conclusion | 40 |
| 2 | The Lasting Impact of High School Value-Added | 70 |</p>
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>70</td>
</tr>
<tr>
<td>2.2</td>
<td>Model background</td>
<td>73</td>
</tr>
<tr>
<td>2.3</td>
<td>Methodology</td>
<td>77</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Estimating the impact of high schools on test scores</td>
<td>77</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Estimating the impact of high schools on college performance</td>
<td>81</td>
</tr>
<tr>
<td>2.4</td>
<td>Data</td>
<td>83</td>
</tr>
<tr>
<td>2.4.1</td>
<td>K-12 test score data</td>
<td>83</td>
</tr>
<tr>
<td>2.4.2</td>
<td>College performance data</td>
<td>84</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Dataset of analysis</td>
<td>85</td>
</tr>
<tr>
<td>2.5</td>
<td>Testing the assumptions for the analysis</td>
<td>87</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Correlation between high school value-added and unobservables</td>
<td>87</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Selection into or out of universities in the sample</td>
<td>89</td>
</tr>
<tr>
<td>2.6</td>
<td>Results</td>
<td>91</td>
</tr>
<tr>
<td>2.6.1</td>
<td>GPA and Graduation</td>
<td>93</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Gender and racial heterogeneity</td>
<td>95</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Quantile regressions</td>
<td>96</td>
</tr>
<tr>
<td>2.7</td>
<td>Conclusion</td>
<td>97</td>
</tr>
<tr>
<td>3</td>
<td>The Impact of Stuttering on Adult Labor Market Outcomes</td>
<td>110</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>110</td>
</tr>
<tr>
<td>3.2</td>
<td>Background</td>
<td>113</td>
</tr>
<tr>
<td>3.2.1</td>
<td>What is stuttering and who stutters?</td>
<td>113</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Stuttering, employment, and stigma</td>
<td>114</td>
</tr>
<tr>
<td>3.3</td>
<td>Data</td>
<td>118</td>
</tr>
<tr>
<td>3.3.1</td>
<td>The measure of stuttering</td>
<td>119</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Outcome variables</td>
<td>120</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Labor market characteristics, individual characteristics, and family characteristics</td>
<td>121</td>
</tr>
<tr>
<td>3.4</td>
<td>Methodology</td>
<td>123</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>3.4.1 Regression analysis and propensity score matching</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>3.4.2 Labor market characteristics and Blinder-Oaxaca</td>
<td>126</td>
<td></td>
</tr>
<tr>
<td>decomposition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5 Results</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>3.5.1 Regression analysis</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>3.5.2 Controls for individual heterogeneity</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>3.5.3 Propensity score matching</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>3.5.4 Labor market characteristics and Blinder-Oaxaca</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>decomposition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.6 Conclusion</td>
<td>136</td>
<td></td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td>APPENDICES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A The Effect of Minimum Wages on Employment: A Factor Model</td>
<td>158</td>
<td></td>
</tr>
<tr>
<td>Approach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A.1 Common correlated effects</td>
<td>158</td>
<td></td>
</tr>
<tr>
<td>A.2 Interactive fixed effects</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>B The Impact of Stuttering on Adult Labor Market Outcomes</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>B.1 Construction of the composite self-stigma index</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>B.2 Propensity score matching</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>VITA</td>
<td>166</td>
<td></td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>53</td>
</tr>
<tr>
<td>1.2</td>
<td>54</td>
</tr>
<tr>
<td>1.3</td>
<td>55</td>
</tr>
<tr>
<td>1.4</td>
<td>56</td>
</tr>
<tr>
<td>1.5</td>
<td>57</td>
</tr>
<tr>
<td>1.6</td>
<td>58</td>
</tr>
<tr>
<td>1.7</td>
<td>59</td>
</tr>
<tr>
<td>1.8</td>
<td>60</td>
</tr>
<tr>
<td>1.9</td>
<td>61</td>
</tr>
<tr>
<td>1.10</td>
<td>62</td>
</tr>
<tr>
<td>1.11</td>
<td>63</td>
</tr>
<tr>
<td>1.12</td>
<td>64</td>
</tr>
<tr>
<td>1.13</td>
<td>65</td>
</tr>
<tr>
<td>1.14</td>
<td>66</td>
</tr>
<tr>
<td>1.15</td>
<td>67</td>
</tr>
<tr>
<td>1.16</td>
<td>68</td>
</tr>
<tr>
<td>1.17</td>
<td>69</td>
</tr>
<tr>
<td>2.1</td>
<td>100</td>
</tr>
<tr>
<td>2.2</td>
<td>101</td>
</tr>
<tr>
<td>2.3</td>
<td>102</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2.4 Selection across universities within sample and selection out of</td>
<td>103</td>
</tr>
<tr>
<td>sample</td>
<td></td>
</tr>
<tr>
<td>2.5 The impact of attending a higher value-added high school on graduation and GPA</td>
<td>104</td>
</tr>
<tr>
<td>2.6 The impact of attending a higher value-added high school on graduation and GPA: students that graduated</td>
<td>105</td>
</tr>
<tr>
<td>2.7 The impact of attending a higher value-added high school on graduation and GPA: students that did not graduate</td>
<td>106</td>
</tr>
<tr>
<td>2.8 The impact of attending a higher value-added high school on graduation and GPA: gender and racial heterogeneity</td>
<td>107</td>
</tr>
<tr>
<td>2.9 Quantile regression impact of attending a higher value-added high school on GPA by semester</td>
<td>108</td>
</tr>
<tr>
<td>2.10 Quantile regression impact of attending a higher value-added high school on GPA by semester, students that graduated</td>
<td>109</td>
</tr>
<tr>
<td>3.1 Summary statistics</td>
<td>139</td>
</tr>
<tr>
<td>3.2 Summary statistics (continued)</td>
<td>140</td>
</tr>
<tr>
<td>3.3 OLS and probit estimates of the effect of stuttering on labor market outcomes</td>
<td>141</td>
</tr>
<tr>
<td>3.4 Robustness of results to controls for ADHD and learning disabilities</td>
<td>142</td>
</tr>
<tr>
<td>3.5 Propensity score matching estimates of the effect of stuttering on labor market outcomes</td>
<td>143</td>
</tr>
<tr>
<td>3.6 Remaining difference after controlling for labor market characteristics</td>
<td>144</td>
</tr>
<tr>
<td>3.7 Blinder-Oaxaca decomposition of the effect of stuttering on log hourly earnings</td>
<td>145</td>
</tr>
<tr>
<td>3.8 Robustness of Blinder-Oaxaca decomposition to removing self-stigma</td>
<td>146</td>
</tr>
<tr>
<td>B.1 Questions used to construct the composite self-stigma index</td>
<td>163</td>
</tr>
<tr>
<td>B.2 Difference in covariates before and after propensity score matching</td>
<td>164</td>
</tr>
<tr>
<td>B.3 Difference in covariates before and after propensity score matching (continued)</td>
<td>165</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 State minimum wages</td>
<td>43</td>
</tr>
<tr>
<td>1.2 IFE minimum wage-employment elasticity estimates, different numbers of factors</td>
<td>44</td>
</tr>
<tr>
<td>1.3 IFE minimum wage-earnings elasticity estimates, different numbers of factors</td>
<td>44</td>
</tr>
<tr>
<td>1.4 First common factor - restaurant employment</td>
<td>45</td>
</tr>
<tr>
<td>1.5 First common factor - teenage employment (QWI)</td>
<td>46</td>
</tr>
<tr>
<td>1.6 Effect of common factors on restaurant employment</td>
<td>47</td>
</tr>
<tr>
<td>1.7 Effect of common factors on teenage (CPS ORG) employment</td>
<td>48</td>
</tr>
<tr>
<td>1.8 Effect of common factors on teenage (CPS basic monthly) employment</td>
<td>49</td>
</tr>
<tr>
<td>1.9 Effect of common factors on teenage (QWI) employment</td>
<td>50</td>
</tr>
<tr>
<td>1.10 Pre-existing trends - restaurant employment</td>
<td>51</td>
</tr>
<tr>
<td>1.11 Pre-existing trends - teenage employment</td>
<td>52</td>
</tr>
<tr>
<td>3.1 Summary of Blinder-Oaxaca decomposition results</td>
<td>147</td>
</tr>
</tbody>
</table>
ABSTRACT

Totty, Evan S. PhD, Purdue University, August 2016. Essays in Labor Economics and Panel Data Analysis. Major Professors: Mohitosh Kejriwal and Kevin Mumford.

This dissertation is composed of three independent chapters. The first chapter studies the impact of minimum wage hikes on employment of low-skill workers in the United States. The second chapter studies the lasting impact of attending a higher value-added high school on college performance. The third chapter studies the impact of stuttering on labor market outcomes.

The first chapter resolves issues in the minimum wage-employment debate by using factor model econometric methods to address concerns related to unobserved heterogeneity. Recent work has shown that the negative effects of minimum wages on employment found using traditional methods are sensitive to the inclusion of controls for regional heterogeneity and selection of states that experience minimum wage hikes, leaving the two sides of the debate in disagreement about the appropriate approach. Factor model methods are an ideal solution for this disagreement, as they allow for the presence of multiple unobserved common factors, which can be correlated with the regressors. These methods provide a more flexible way of addressing concerns related to unobserved heterogeneity and are robust to critiques from either side of the debate. The factor model estimators produce minimum wage-employment elasticity estimates that are much smaller than the traditional OLS results and are not statistically different from zero. These results hold for many specifications and two datasets that have been used in the minimum wage-employment literature. A simulation shows that unobserved common factors can explain the different estimates seen across approaches in the literature.
The second chapter studies whether there is any lasting benefit to attending a better "value-added" high school. Many states and school districts have tied teacher and school evaluation to value-added scores, which attempt to assess the impact that teachers and schools have on student test scores. Much work has been done to develop value-added models that can provide unbiased estimates of the impact that teachers and schools have on test scores. Much less work has been done to determine whether these higher value-added teachers and schools have a lasting impact on their students. This chapter address this issue by estimating value-added scores for high schools in Florida and North Carolina and then linking these value-added scores to a student-level dataset on college performance that includes the high school that the student attended. Results suggest that attending a higher value-added high school has a positive and statistically significant impact on college GPA, although the impact is small; a one standard deviation increase in high school value-added increases first semester GPA by 0.022 points and final GPA by 0.017 points. The effect on graduation is not statistically significant. Results are consistent across race and gender, although the effect is slightly larger for White students.

The third chapter quantifies the difference in labor market outcomes between people who stutter and people who do not. Stuttering is a neurodevelopmental disorder of speech which is characterized by repetitions and prolongations of sounds, syllables, and words. It is estimated that 70 million people stutter worldwide and 3 million in the United States. Despite this prevalence, no study has adequately quantified the impact of stuttering on adult labor market outcomes; considering the ubiquitous nature of communication, it is likely that communication differences and difficulties would affect labor market outcomes. Results suggest that stuttering impacts employment status, labor force participation, receipt of public assistance, and hourly earnings. Additionally, stuttering seems to impact males and females differently.
CHAPTER 1. THE EFFECT OF MINIMUM WAGES ON EMPLOYMENT: A FACTOR MODEL APPROACH

1.1 Introduction

Understanding the effect of minimum wages on employment has long been of interest to economists, with empirical work on the subject dating back approximately 100 years [Obenauer and von der Nienburg, 1915]. Despite this long history of attention, economists are still very much divided on the effect of minimum wages. The last two decades, in particular, have produced an abundance of work on the subject, without providing a consensus. The empirical evidence in these studies differs depending on both the datasets used and the methodology\(^1\). The goal of this paper is to resolve the issues in the minimum wage-employment literature by using panel data econometric methods that are robust to critiques from either side of the debate. Specifically, this study uses the common correlated effects estimators developed by [Pesaran, 2006] and the interactive fixed effects estimator developed by [Bai, 2009]. These estimators are applied to many datasets and specifications that have recently been used in the literature. The factor model methods used in this paper are well suited for a wide variety of empirical studies, although they have not yet received much use.

Minimum wage hikes are very common in the United States. There have been three instances of federal minimum wage hikes since 1990, each of which involved phasing in a higher minimum wage through two separate hikes in consecutive years.

\(^1\)The theory is also ambiguous. While the simple competitive model predicts a decrease in employment in response to a minimum wage hike, the monopsony model can predict no effect or even a small positive effect. Additionally, there are many other channels through which minimum wages could impact firms and employees that would mitigate the employment effect, such as labor-labor substitution, decreased costly labor turnover, reductions in non-wage benefits, improved organizational efficiency, increased worker effort, price increases, or reductions in profit [Schmitt, 2013]. Evidence on these other channels is sparse and mixed, but the most convincing evidence may support decreased labor turnover [Dube et al., 2014] and price increases [Aaronson et al., 2008, Lemos, 2008, Harasztosi and Lindner, 2015].
However, much of the minimum wage policy variation takes place at the state level: there has been at least one state-level minimum wage hike in every year since 1990. Additionally, state-level minimum wage hikes have become very common. In 2016, for example, 17 states will increase their minimum wage. Several states have even begun indexing their minimum wage to adjust annually based on inflation. In total, 29 states currently have a minimum wage higher than the federal minimum wage. Recent campaigns among US workers to increase the federal minimum wage from $7.25/hour to $15.00/hour have received a lot of attention in the media. Not surprisingly, the merits of minimum wage hikes have been heavily debated by media members and politicians. However, minimum wages have also been heavily debated among economists in recent years. A 2015 poll by the Institute for Research on Global Markets asked 37 prominent economists if they agreed with the following statement: "If the federal minimum wage is raised gradually to $15-per-hour by 2020, the employment rate for low-wage US workers will be substantially lower than it would be under the status quo." Nine-percent strongly agreed, twenty-five percent agreed, thirty-seven percent were uncertain, and twenty-nine percent disagreed.

Much of the reason why there is still an ongoing debate regarding the employment effects of minimum wage hikes is that, while the abundance of state-level minimum wage variation is useful for empirical studies, minimum wage hikes are, of course, not randomly distributed across space or time. Figure 1.1 shows the average minimum wage in each state from 1980-2015 and the total number of minimum wage hikes in each state during the same time period. Minimum wages are generally higher and raised more frequently in the Northeast, parts of the Midwest, and the West Coast. Because regions of the US differ along many dimensions other than just minimum wage policy, such as employment patterns, demographics, education levels, industrial compositions, and any number of unobservable variables, estimating an unbiased employment effect of minimum wage hikes using aggregate state-level employment data is challenging. As evidence of this, note that case studies which compare adjoining local areas with different minimum wages around the time of a policy change have
tended to find small or no disemployment effects [Card and Krueger, 1994, Card and Krueger, 2000], while panel studies using data aggregated to the state or county level have tended to find larger disemployment effects [Neumark and Wascher, 1992, Neumark and Wascher, 2007]. One potential reason for the inconsistency across methods could be that the panel studies are producing spurious estimates due to unobserved confounders in the aggregate data which are correlated with both employment and minimum wages.

This challenge associated with minimum wage-employment studies using aggregate panel data has led to an abundance of new research on the topic in recent years; the recent debate in the minimum wage-employment literature has focused on how to generate credible estimates of employment effects when using aggregate state- or county-level panel data. An incomplete but representative summary of the recent literature is shown in Table 1.1. The traditional approach in early panel studies was to include two-way fixed effects (one for each time period and one for each state or county) in order to address the concerns about unobserved heterogeneity across areas and time periods [Neumark and Wascher, 1992, Neumark and Wascher, 2007]. This approach tends to find minimum wage-employment elasticity estimates in the range of -0.1 to -0.2, meaning that a 10% increase in the minimum wage causes a 1%-2% decrease in employment for low-skilled workers such as teenagers. The recent minimum wage debate began when [Dube et al., 2010] and [Allegretto et al., 2011] argued that two-way fixed effects are not sufficient to fully address concerns about unobserved heterogeneity. They propose the use of state-specific time trends and Census division-by-period fixed effects in order to allow for even more heterogeneity. [Dube et al., 2010] also uses a border discontinuity approach for county-level analysis, which uses policy discontinuity at state borders to identify the effect of minimum wage hikes. This essentially embeds the case study approach into the panel setting by restricting the sample to all contiguous counties along state borders and adding contiguous county-pair fixed effects to the two-way fixed effects. Each of these approaches pro-
duces minimum wage-employment elasticity estimates that are much smaller than the traditional two-way fixed effects approach and not statistically different from zero.

There are several potential issues associated with the approaches in [Dube et al., 2010] and [Allegretto et al., 2011]. [Neumark et al., 2014b] argue that the Census division-by-period fixed effects and border discontinuity approach throw out too much valid identifying information. This conclusion is reached based on the weights that a synthetic control approach places on same-division or border-county areas. They have also argued in a subsequent paper that Census division-by-period fixed effects and the border discontinuity approach may actually worsen policy endogeneity by changing the identifying variation from both state and federal variation, to only state variation, which is more likely to be endogenously determined [Neumark et al., 2014a]. [Neumark et al., 2014b] also argue that specifications with state-specific time trends may suffer from endpoint bias or the lack of flexible higher order state-specific time trends. They show that negative and statistically significant elasticity estimates return when they account for potential endpoint bias or allow for higher order state-specific time trends. Essentially, [Neumark et al., 2014b] argue in favor of the traditional two-way fixed effects specification over one that adds Census division-by-period fixed effects or state-specific time trends.

Synthetic controls have also been used in the literature as an alternative way to address the concerns about unobserved heterogeneity discussed above. Unsatisfied with the approaches in [Dube et al., 2010] and [Allegretto et al., 2011], [Neumark et al., 2014b] proposed a synthetic control-style matching estimator as an alternative way to address the concerns about unobserved confounders influencing the aggregate panel data. Their synthetic control approach produces large negatively elasticity estimates. However, [Allegretto et al., 2013] have shown that a synthetic control approach that more closely follows the approach outlined in [Abadie et al., 2010], with cleaner identification of treatment vs controls groups and longer pre- and post-treatment windows, produces small elasticity estimates that are not statistically different from zero. The two camps have continued to debate the Census division-by-period fixed effects,
state-specific time trends, border discontinuity approach, and synthetic controls in follow-up work [Neumark et al., 2014a, Allegretto et al., 2013, Allegretto et al., 2015]. [Allegretto et al., 2015] also use the double-selection post-LASSO approach advanced by [Belloni et al., 2014] as a way of letting the data determine which, if any, additional controls should be included beyond the traditional two-way fixed effects. This procedure selects Census division-by-period fixed effects from one Census division and 29 state-specific linear time trends (no higher order trends) and produces small elasticity estimates that are not statistically different from zero.

The contribution of this paper is to bring a different econometric approach to the data. The factor model estimators advanced by [Pesaran, 2006] and [Bai, 2009] are well-suited to address the issues described above: the methods in these papers allow for consistent estimation of regression parameters under the presence of multiple unobserved common factors influencing the data. The unobserved common factors are allowed to be correlated with the independent variables as well, which makes the estimators very useful for studies with aggregate panel data, as both the dependent and independent variables are likely to be influenced by unobservable confounders in such studies. The intuition for the estimators in [Pesaran, 2006] and [Bai, 2009] is to use either cross section averages of all the variables or principal components to proxy for the unobserved common factors in the data, so that they can be controlled for directly. Essentially, the factor model estimators apply a very flexible structure to the error term of a given specification which embeds many other structures, including, for example, state-specific time trends and Census division-by-period fixed effects. By using the factor model estimators, rather than ordinary least squares (OLS), to estimate the traditional two-way fixed effects specification, it is possible to control for unobservable heterogeneity without having to make specific assumptions about the form of the unobserved heterogeneity. Importantly, the factor model estimators have been shown to perform well even when there are no unobserved common factors in the error term. This means that using the factor model estimators will produce estimates similar to OLS if the specification is correct. As described below in Section
2.3, the factor model estimators both address the concerns associated with each camp in the recent minimum wage-employment debate and have specific advantages over each of the other approaches used in the literature.

Minimum wage-employment elasticity estimates based on the factor model estimators are significantly different than estimates based on OLS. Using the same datasets from the papers discussed above, OLS estimates based on the two-way fixed effects specification replicate the -0.2 to -0.1 elasticity estimates from the literature. The factor model estimators produce elasticity estimates that are much smaller than OLS and not statistically different from zero: restaurant elasticity estimates are in the range of -0.01 to -0.05 while teenage elasticity estimates are in the range of -0.03 to -0.07. These small elasticity estimates are robust to different data sources, different assumptions about the number of common factors in the data, different time and cross section samples of the data, and different specifications that model time trends explicitly. Analysis of the factor structure estimated from the data suggests the common factors are capturing time trends, among other things, and also suggests time-varying regional heterogeneity in the effect of the common factors, which could roughly be approximated by a Census division-by-period fixed effect. However, the factor structure also captures factors that appear to be unrelated to time trends and shows some neighboring or same-Census division counties that appear to experience very different effects from common factors. Thus, while analysis of the factor structure does lend support for preferring specifications with state-specific time trends and/or Census division-by-period fixed effects as in [Dube et al., 2010] and [Allegretto et al., 2011] over ones without them, it is also supports the broad points in [Neumark et al., 2014b] and [Neumark et al., 2014a] that time trends may not always be appropriate and that proximate places do not always make ideal control groups. Simulations at the end of the paper show two key findings: First, the pattern of results discussed above, in which OLS produces relatively large negative elasticity estimates and the factor model estimators produce estimates close to zero, cannot be produced from the data if the two-way fixed effects specification is correct. Second, the OLS estimate
of the minimum wage-employment elasticity produced from the two-way fixed effects specification is negatively biased by the common factors present in the data.

The remainder of the paper is organized as follows: Section 2 describes the factor model setup of which the [Pesaran, 2006] and [Bai, 2009] estimators make use, describes the estimators themselves, and discusses the relative merits of the factor model approach over other approaches from the literature. Section 3 describes how the variables are constructed and provides summary statistics for the data. Section 4 discusses the results. Section 5 discusses the simulations. Section 6 concludes.

1.2 Empirical approach

1.2.1 Multi-factor error structure

The factor model setup is based on a model in which the error term is characterized by a multi-factor error structure. Specifically, the traditional error term in a regression equation is decomposed into time-specific "common factors" that can affect all cross section units, heterogeneous "factor loadings" that represent how a common factor affects a particular cross section unit, and an idiosyncratic error term. In the analysis below, the factor structure will be applied to the traditional two-way fixed effects (one fixed effect for each time period and one for each state or county) specification, which is the most saturated specification on which the literature has been able to agree. Factor model estimators from [Pesaran, 2006] and [Bai, 2009] will then be used for estimation.

The traditional specification for estimating the effect of minimum wages on employment, originating from [Neumark and Wascher, 1992], is given by

\[
\ln(E_{it}) = \beta \ln(MW_{it}) + \Gamma X_{it} + \alpha_i + \delta_t + \varepsilon_{it}.
\]  

(1.1)

Depending on the dataset used in the analysis below, \(E_{it}\) is either a count of the number of restaurant/teenage employees in county \(i\) and period \(t\) or the fraction of
teenagers employed in state $i$ and period $t$. $MW_{it}$ is the higher of the federal and state minimum wage in state or county $i$ and period $t$. Employment and the minimum wage are measured in logs so that $\beta$ represents the minimum wage-employment elasticity. The term $X_{it}$ is a vector of control variables defined in Section 3 that are intended to proxy for supply and demand forces on employment. Unit and period fixed effects are represented by $\alpha_i$ and $\delta_t$, respectively. Several studies have estimated this two-way fixed effects specification with OLS and no additional controls for unobserved heterogeneity and found large negative effects of minimum wages on employment [Neumark and Wascher, 1992, Neumark and Wascher, 2007, Sabia, 2009]. The identification assumption for this specification is that minimum wage variation is uncorrelated with the error term $\epsilon_{it}$, conditional on the two-way fixed effects and other controls.

The difference with the factor model approach is that it allows for the presence of cross section dependence remaining in the error term $\epsilon_{it}$. Cross section dependence is the tendency of outcomes, or residuals in this case, to be correlated across areas. This dependence could be spatial, caused by similarity in geographic characteristics. However, unlike spatial econometric methods, factor models also allow this dependence to depart from geographic proximity, which could occur if two areas experience the same industry-specific shock because of industry specialization, even if they are not neighbors. Cross section dependence is problematic for inference [Andrews, 2005], but will also cause bias if unobserved common factors are correlated with the regressors. The factor model approach facilitates the control of cross section dependence with time-specific common factors that can have heterogeneous effects over areas:

$$\epsilon_{it} = \lambda_t' f_t + u_{it}$$

The log-log specification imposes a linear relationship on the size of the minimum wage hike and the size of the employment change: doubling the size of the minimum wage hike will double the employment effect. In reality, the relationship may not be linear. Firms may be more able to absorb small minimum wage hikes without changing employment than large minimum wage hikes. The reason why this specification is used is that it has been the common approach in the recent literature and because there is relatively little variation in the size of minimum wage hikes during the time frame of analysis in this paper: most minimum wage hikes from 1990-2013 were in the range of 5-15%.

$^2$More formally, $E(\epsilon_{it}|ln(MW_{it}), X_{it}, \alpha_i, \delta_t) = 0$. 

$^3$More formally, $E(\epsilon_{it}|ln(MW_{it}), X_{it}, \alpha_i, \delta_t) = 0$. 

The log-log specification imposes a linear relationship on the size of the minimum wage hike and the size of the employment change: doubling the size of the minimum wage hike will double the employment effect. In reality, the relationship may not be linear. Firms may be more able to absorb small minimum wage hikes without changing employment than large minimum wage hikes. The reason why this specification is used is that it has been the common approach in the recent literature and because there is relatively little variation in the size of minimum wage hikes during the time frame of analysis in this paper: most minimum wage hikes from 1990-2013 were in the range of 5-15%.
where \( f_t \) is an \((r \times 1)\) vector of unobserved common factors and \( \lambda_i \) is an \((r \times 1)\) vector of factor loadings that capture unit-specific effects of the common shocks. These common factor may be thought of as omitted variables. As discussed below, the factor model estimators from [Pesaran, 2006] and [Bai, 2009] allow these unobserved common factors to be correlated with the regressors. Therefore, the identification assumption is no longer that minimum wage variation is uncorrelated with \( \varepsilon_{it} \), but that it is uncorrelated with \( u_{it} \). That is, minimum wage variation is uncorrelated with the error term, conditional on the two-way fixed effects, other controls, and the common factors and factor loadings\(^4\).

It is possible to think of examples of omitted variables that could cause bias in the OLS estimate of \( \beta \) from equation (1) in either direction, but could also be captured by the factor structure in equation (2). [Neumark et al., 2014b] argued that minimum wages may be more likely to be raised when labor markets are tight, citing [Baskaya and Rubinstein, 2011]. This would suggest that minimum wage hikes are associated with positive employment shocks, which would cause positive endogeneity bias in the OLS estimate of \( \beta \) from equation (1). These types of general macroeconomic shocks could be captured by the factor model structure: a common factor could represent a macroeconomic trend and the factor loadings represent how the macroeconomic trend affects each county/state, which would vary depending on characteristics of the area such as demographics and industrial composition. Alternatively, technological change could produce negative endogeneity bias. [Smith, 2011] studied teenage employment rates from 1980 to 2009 and showed that job polarization due to technological change pushes middle-skill adults into low-skill jobs traditionally held by teenagers, thus lowering teenage employment. [Allegretto et al., 2013] showed that between 1990 and 2007, high minimum wage states experienced greater job polarization, on average, than low minimum wage states. Combining these results suggests that high minimum wage states have experienced greater job polarization, which puts downward pressure on teenage employment. This would cause negative endogeneity bias in the OLS

\(^4\)More formally, \( \text{E}(u_{it}|\ln(MW_{it}), X_{it}, \alpha_i, \delta_t, \lambda_i' f_t) = 0 \).
estimate of $\beta$ from equation (1), but could be captured by the factor structure: a common factor could track skill-biased technological change at the country level, which the factor loadings capture the heterogeneous effects of technological change across areas, which would vary according to characteristics such as education level and industrial composition.

It is worth noting that the factor structure shown above can be rewritten to incorporate lagged common factors. This is appealing in the context of the minimum wage-employment data, given that it is reasonable to assume that both employment and minimum wages may be slow in responding to economic conditions due to social norms against laying off workers and the delay between when minimum wage hikes are approved and actually implemented. In this sense, it is intuitive to think that employment and minimum wages may have a lagged response to common factors. The factor model can be rewritten to incorporate lagged responses by rewriting a dynamic factor model as a static factor model, with the error term in equation (1) now taking the form $\varepsilon_{it} = \Lambda_iF_t + u_{it}$, where $F_t = (f_t', f_{t-1}', ..., f_{t-s}')'$ is an $(r(s + 1) \times 1)$ vector of common factors, $\Lambda_i = (\lambda_{i0}', \lambda_{i1}', ..., \lambda_{is}')'$ is an $(r(s + 1) \times 1)$ vector of factor loadings, and $s$ represents the number of lagged factors.

1.2.2 Factor model estimators

There are two commonly used approaches for estimating regression equations with a multi-factor error structure, each of which will be used in the analysis below. The first method is the common correlated effects approach from [Pesaran, 2006]. This method does not attempt to estimate the common factors and factor loadings directly. Rather, Pesaran shows that, in a large $N$ and large $T$ setting, you can proxy for the factors with cross sectional averages of the dependent and independent variables. This estimator has the added benefit that it can be computed by ordinary least squares applied to regressions where the observed explanatory variables are augmented with cross sectional averages of the dependent and independent variables. Pesaran proposes
two versions of this method: the common correlated effects mean group (CCEMG) estimator and the common correlated effects pooled (CCEP) estimator. The CCEMG estimator allows minimum wages to have heterogeneous effects over areas by estimating a separate regression coefficient for each cross section unit. The individual slope coefficients can then be averaged to obtain a single mean group estimate. The CCEP estimator is a generalized version of the standard fixed effects estimator that estimates a single pooled regression coefficient but still allows the common factors to have heterogeneous effects over areas. Standard errors are calculated using equations (58) and (69) in Pesaran (2006) for the CCEMG and CCEP estimators, respectively. The variance of the CCEMG estimator is estimated non-parametrically as the variance of the individual slope coefficients. The variance of the CCEP estimator takes on the common sandwich estimator form and is also based on the variance of the individual slope coefficients. More details are provided in the appendix. Confidence intervals and significance reported in the results section are based on bootstrapped t-statistics using the wild cluster bootstrap-t procedure from [Cameron et al., 2008], clustered at the state level.

The second method is the interactive fixed effects (IFE) approach from [Bai, 2009], which does involve directly estimating the common factors and factor loadings. This is done by jointly estimating the regression coefficients and the factor structure in an iterative process. The IFE approach is based on the fact that, given the common factors and factor loadings, the regression coefficients can be estimated using OLS after subtracting the factor structure from the data, and given the regression coefficients, the factors and factor loadings can be estimated by performing principal component analysis on the regression residuals. However, the regression coefficients and factor structure are both unknown in practice. Therefore, Bai proposes an iter-

Because the IFE estimator actually estimates the factor structure, the number of common factors, \( r \), must be pre-specified. One approach is to use the information criteria from [Bai and Ng, 2002], which estimates the number of strong factors in the data. Alternatively, IFE results could be provided for different numbers of common factors. [Moon and Weidner, 2015] showed that the IFE estimator still performs well if the number of common factors is over-estimated, but can suffer if too few factors are included.
tive procedure in which, given an initial guess of either the regression coefficients or
the common factors and factor loadings, one iterates between estimating one, given
the other, until the percent change in the sum of squared residuals falls below a
pre-specified threshold. A threshold of $10^{-9}$ is used in this paper. Bias-correction
for serial correlation, cross-sectional correlation, and heteroskedasticity is performed
using equations (23) and (24) in Bai (2009). Standard errors are calculated using
Theorem 4 in Bai (2009). The variance of the IFE estimator takes on the standard
sandwich estimator form, based on the square of the residuals. More details are
provided in the appendix. Confidence intervals and significance reported in the re-
sults section are based on bootstrapped t-statistics using the *wild cluster bootstrap-t*
procedure from [Cameron et al., 2008], clustered at the state level.

Neither estimator is considered to dominate the other, which is why both are used
in the analysis. The most significant difference between the two is the assumption
about the number of common factors. The CCE estimators assume that the number
of common factors is less than or equal to the number of dependent plus independent
variables, whose cross-sectional averages proxy for the common factors. The IFE
estimator assumes that the number of common factors is known to the researcher.
[Pesaran, 2006] and other papers have shown that the CCE estimator continues to
perform well even when the number of common factors exceeds the number of cross
sectional averages, although this flexibility comes at a cost: the factor loadings must
take on a random coefficients form. The IFE assumption that the number of common
factors is known may seem more restrictive, but in practice IFE results can just
be reported for many potential number of common factors. Additionally, [Moon
and Weidner, 2015] have shown that the IFE estimator still performs well when

---

$[\text{Bai, 2009}]$ suggests initiating the iteration two different ways: (1) with the OLS estimates of
the regression coefficients, ignoring the factor structure, and (2) with the principal components
estimates of the factor structure from the raw data, ignoring the independent variables. Then,
one keeps whichever set of results has the lowest final sum of squared residuals. Analysis was also
performed using the individual slope coefficients from the CCE approach as an initial guess of the
regression coefficients. The results were nearly identical between the starting methods, but the
results generally converged faster when using the CCE estimates as the starting values. Therefore,
analysis below is based on the CCE starting method.
the number of common factors is over-estimated. Therefore, the most significant assumption of the IFE estimator may be that the factors that are most relevant for bias and inference are among the strongest factors, so that they are picked out even when a small number of factors is assumed.

[Westerlund and Urbain, 2015] also address the relative merits of the two approaches. The authors are actually comparing the relative merits of principal components versus cross sectional averages as a means of controlling for cross section dependence in regression errors rather than comparing IFE and CCE exactly; the estimators they consider are slightly different from those considered in [Pesaran, 2006] and [Bai, 2009], and they therefore caution against extrapolating too widely the conclusions to the use IFE and CCE. Nonetheless, they find that the principal components approach is more efficient when $\beta = 0$, while the cross sectional averages approach is expected to perform best when $\beta \neq 0$. They also find that each estimator performs better when the cross section dimension is larger than the time dimension. For examples of other applied work using these estimators, see [Bailey et al., 2016], [Kim and Oka, 2014], [Cavalcanti et al., 2014], [Eberhardt et al., 2013], [Hagedorn et al., 2013], and [Bertoli and Moraga, 2013].

1.2.3 Factor model approach versus other approaches in literature

As described in the introduction, three different methods have been used in the recent literature that have produced conflicting results and/or much debate about their appropriateness: 1) adding Census division-by-period fixed effects and/or state-specific time trends to the two-way fixed effects specification 2) using a border discontinuity approach to identify the effect based on policy changes between two neighboring counties and 3) using synthetic controls. The factor model approach discussed above has specific advantages over each of these approaches and can address the broad concerns from both sides of the debate.
The advantage of the factor model approach over the inclusion of Census division-by-period fixed effects and/or state-specific time trends is that the factor model approach embeds these controls as special cases of the factor structure: they can be rewritten as the inner product of a vector of time-specific common shocks, \( f_t = (t, \delta_t)' \), and unit-specific factor loadings, \( \lambda_i = (\alpha_i, \zeta_{CD})' \), where \( t \) is a time trend, \( \delta_t \) is a time period dummy variable, \( \alpha_i \) is a state dummy variable, and \( \zeta_{CD} \) is a Census division dummy variable. When multiplied together, this produce state-specific time trends and Census division-by-period fixed effects\(^7\). So the factor model approach embeds these controls, but the important difference is that the factor approach lets the data determine the form of the common shock, \( f_t \), and the nature of the spatial correlation in \( \lambda_i \), whereas modeling state-specific trends and Census division-by-period fixed effects explicitly imposes a fixed form for the unobserved heterogeneity ex ante. This is significant because, while incorrectly imposing Census division-by-period fixed effects and/or state-specific time trends could discard valid identifying information or cause bias, the factor model estimators have been shown to perform well in terms of bias and efficiency even when there are no common factors in the data. This means that the factor model estimators should produce elasticity estimates very similar to OLS even if the traditional two-way fixed effects specification is correctly specified. Therefore, the factor model approach allows for a much more flexible form of unobserved heterogeneity than either only using two-way fixed effects or modeling Census division-by-period fixed effects and state-specific time trends explicitly.

The border discontinuity approach moves away from the panel specification debate and instead tries to embed the case study approach into the panel data setting by comparing employment in cross-state neighboring counties before and after one state implements a minimum wage hike. In [Dube et al., 2010], this is done by limiting the

\(^7\)The two-way fixed effects themselves could also be rewritten as a special case of the factor structure, where \( f_t = (1, \delta_t, t, \delta_t)' \) and \( \lambda_i = (\alpha_i, 1, \alpha_i, \zeta_{CD})' \). However, the analysis below will model them separately for two reasons: First, the goal of this paper is to start with the most saturated specification that the literature can agree on and let the factor structure handle the rest. Second, it is more efficient to model a variable explicitly if it is part of true DGP. While Census division-by-period fixed effects and state-specific time trends are debatable, there is almost certainly unobserved differences across areas and time that will influence employment and minimum wages.
sample to cross-state contiguous counties and then estimating the traditional two-way fixed effects specification with an additional fixed effect for each pair of cross-state contiguous counties, which produces small and insignificant elasticity estimates. Two separate, but related, critiques have been raised with respect to this approach. First, as discussed in the introduction, [Neumark et al., 2014b, Neumark et al., 2014a] have argued that this approach throws out too much valid identifying information, with most of their argument based on the weight that a synthetic control approach places on local areas. This critique is related to the primary drawback of the border discontinuity approach, which is that it relies on the non-testable assumption that confounding factors are removed when comparing contiguous areas. While it seems likely that contiguous areas would be more similar in terms of their unobservable characteristics than two random counties, there may be many cases in which they are not. [Slichter, 2015] addresses this issue of imperfect neighbor matches by extending the control group from only neighbors to include neighbors-of-neighbors, neighbors-of-neighbors-of-neighbors, and so on. However, this approach is still based on the imperfect assumption that similarity in observable characteristics implies similarity in unobservable characteristics. The advantage of the factor model approach is that, rather than assuming that unobserved confounders are removed by comparing local areas, the factor model approach obtains estimates of the unobserved confounders and controls for their presence directly.

The second critique of the border discontinuity approach is that it is not necessarily obvious whether this approach improves or worsens policy endogeneity. In work that uses the border discontinuity approach to study employment flows, [Dube et al., 2014] conjecture that policy endogeneity is likely removed by the border discontinuity approach because state-level minimum wage policy is likely to respond to state-wide shocks, which may not be the same shocks that occur at the border. Alternatively, [Neumark et al., 2014a] conjecture that this approach worsens policy endogeneity because more of the policy variation comes from state-level hikes, which are more likely to be endogenous to state-level employment than federal minimum wage hikes.
The factor model approach provides a way to improve policy endogeneity that may be better than the border discontinuity approach. The conjecture in [Allegretto et al., 2015] that the state-level shocks that state minimum wages respond to are not the same shocks that occur at the border is both untested and unlikely to always be true. Therefore, the factor model approach of obtaining estimates of the unobserved confounders directly may be a more robust approach. Furthermore, the factor model approach is not subject to the critique from [Neumark et al., 2014a] that it may worsen policy endogeneity by changing the identifying variation, because it uses the same variation as the two-way fixed effects approach.

The factor model approach also has advantages over the synthetic control approach. The synthetic control approach discussed in [Abadie et al., 2010] is also based on a factor model structure. They show that, under fairly standard conditions, the ability to match covariates and a long period of pre-treatment outcomes using synthetic controls implies that the factor structure is also matched. In this sense, the linear factor model estimators described above and synthetic controls may be seen as complimentary ways to estimate specifications that include time-varying heterogeneity. However, recent work by [Gobillon and Magnac, 2015] suggests that the IFE estimator from [Bai, 2009] may be preferred to synthetic controls for the purpose of regional policy evaluation: from a theoretical perspective, if the true model is a linear factor model as in equations (1)-(2), then the synthetic control estimator from [Abadie et al., 2010] is equivalent to interactive fixed effects only when the matching variables observable covariates and unobservable factor loadings for the treated areas belong to the support of observable covariates and unobservable factor loadings for the untreated areas. Monte carlo simulations and an empirical application also seem to support interactive fixed effects over synthetic controls.

Furthermore, there are drawbacks associated with the synthetic control approach that are specific to the minimum wage setting. The synthetic control approach in [Abadie et al., 2010] was designed for the setting in which there is a single area

\[\text{Dube and Zipperer, 2015}\] find evidence that this support requirement may not be satisfied for the minimum wage application (p. 26).
receiving a one-time policy treatment. Then, a policy effect is estimated by first constructing a synthetic control that is a weighted combination of untreated areas, or “donor units,” which best matches the treated area in the pre-treatment period and then simply taking the difference between the post-treatment outcome for the treated area and the post-treatment outcome for the synthetic control. Two important requirements for this approach are long pre- and post-treatment windows and many untreated areas. However, these two requirements are difficult to satisfy in the minimum wage setting. Federal minimum wage hikes are difficult to analyze because nearly every state is treated, leaving very few untreated areas with which to construct a synthetic control. State minimum wage hikes are also difficult to analyze because they occur so frequently across states and within states over time that it is difficult to construct long pre- and post-treatment windows during which many other states are untreated and no additional treatment occurs within the treated state. This difficulty has led to two different synthetic control approaches in the minimum wage literature. [Neumark et al., 2014b] construct a synthetic control for each instance of a minimum wage increase and then pool all of the real and synthetic data together to perform OLS with two-way fixed effects and an additional fixed effect for each pair of real and synthetic data. They use 4-quarter windows and, in some cases, include other areas that are being treated in the control group. [Allegretto et al., 2015] and [Dube and Zipperer, 2015] apply an approach more closely assigned with [Abadie et al., 2010] by analyzing individual minimum wage hikes separately, using longer windows, and including only untreated areas in the control group. However, this leaves them with only 29 usable instances of a minimum wage hike from 1979-2013 out of 215 total instances. It is apparent that, in addition to the fact that synthetic controls have produced conflicting results in the literature, synthetic controls are not very applicable to the minimum wage setting to begin with due to the common, recurring, and continuous treatment. The linear factor model estimators described above, however, are entirely amenable to this setting.
The newest approach in the debate is the double-selection post-LASSO approach used in [Allegretto et al., 2015], originally developed in [Belloni et al., 2014]. There are similarities between this approach the factor model approach. Both LASSO regressions and factor models are commonly used for data reduction purposes. The double-selection post-LASSO approach and the linear factor model estimators from [Pesaran, 2006] and [Bai, 2009] use these data reduction abilities to apply a parsimonious structure to a traditional regression equation in settings that could contain many potential variables. The double-selection post-LASSO procedure first uses LASSO regressions to select covariates that predict either the dependent variable or the treatment variable of interest. Then, these covariates are included in a simple OLS regression. Similar to how the linear factor model estimators can embed state-specific time trends and Census division-by-period fixed effects, thus allowing the data to determine if they are appropriate, the double-selection post-LASSO approach applied in [Allegretto et al., 2015] first uses the LASSO-driven procedure to determine which, if any, state-specific time trends and Census division-by-period fixed effects should be included in addition to the two-way fixed effects.

However, the double-section post-LASSO approach can only select from covariates that are specified by the researcher in the LASSO regressions, thus requiring that the researcher is aware of and can account for all relevant covariates and forms of unobserved heterogeneity. Endogeneity bias remains a threat in the form of unobservable or omitted variables not included in the LASSO regression and unobserved heterogeneity structures that OLS cannot estimate, such as interactive fixed effects. The factor model approach remains a more robust and flexible approach, as the researcher does not have to specify variables or fixed effects for the factor model estimators to choose from and the factor model approach allows for unit fixed effects to vary over time. Therefore, while both the double-selection post-LASSO and factor model approaches allow the data determine if Census division-by-period fixed effects and state-specific time trends are appropriate, the factor model approach can also
account for variables that may be omitted from the LASSO regressions in [Allegretto et al., 2015] and allow unit fixed effects to vary over time.

In addition to having advantages over each of the other approaches used in the recent literature, the factor model approach can satisfy the primary concerns from each side of the debate. The primary concern in [Dube et al., 2010] and [Allegretto et al., 2011] was that the traditional two-way fixed effects specification was not sufficient to address unobserved confounders in the data, so they proposed new methods to generate credible estimates of employment effects. [Neumark et al., 2014b, Neumark et al., 2014a] have also proposed methods to generate more credible estimates than those provided by the two-way fixed effects specification, but they caution that the methods in [Dube et al., 2010] and [Allegretto et al., 2011] may discard too much valid identifying information or, in the case of the border discontinuity approach, discard too much data altogether. They have also argued that these methods may worsen policy endogeneity by changing the identifying variation from both state and federal minimum wage variation to only state minimum wage variation, which is more likely to be endogenously determined. The factor model estimators provide a middle ground: They are designed to facilitate the control of unobserved confounders for the purpose of generating unbiased estimates of regression coefficients, and they do so without discarding data, without changing the identifying variation to within Census divisions or across state borders, and they use both state and federal minimum wage variation.

1.2.4 Limitations of the factor model approach

The primary drawback of the factor model approach is that the estimators discussed above can only control for unobserved heterogeneity that fits into the interactive fixed effects form in equation (2). Most relevantly, this means that there cannot be variation in the factor loadings over time. This could occur when some characteristics for a particular area change, such that the area experiences a different effect
over time from a given common factor. For example, using the technological change example from above, if a state’s industrial composition or education level changed over time such that it had more high skill jobs to attract high skill workers or had more high skill workers to attract high skill jobs, then that state’s true factor loading for the technological change factor could change signs. As the model is written in equation (2), the factor model estimators cannot address this type of identification threat.

However, there is some evidence to suggest that the factors model estimators may still perform well in this setting. [Bates et al., 2013] show that the principal components estimation of common factors still performs well even when there is time variation in the factor loadings, including i.i.d. variation, random walk variation, and structural breaks that affect all or some units. While this is a test of the estimation of common factors rather than the estimation of regression coefficients in specifications that include estimated common factors, it does suggest that the interactive fixed effects estimator, which is based on principal components estimation of common factors, may still perform well. Nonetheless, one way to deal with the possibility of time variation in the factor loadings is to estimate the model for separately for different period of time. If the estimates of the minimum wage-employment elasticity are similar across time periods, then the results are likely not being caused by inaccurate estimation of factor loadings due to time instability. This strategy will be used below as a robustness check.

1.3 Data

1.3.1 Data sources

Following the recent minimum wage-employment literature, two different low-skill groups are analyzed in this study: restaurant workers and teenagers. Restaurant workers and teenagers are the two most commonly studied populations in the minimum wage-employment literature [Baleman and Wolfson, 2014] and they have been
the focus of the recent debate in the literature. Based on CPS Outgoing Rotation Group calculations in [Allegretto et al., 2015], during the period 1979-2014, 40.2% of working teenagers earned within 10% of the minimum wage and teenagers accounted for 22.7% of all workers earning within 10% of the minimum wage (down from 32.2% in 1979). During the time period 2000-2014, 28.3% of restaurant workers earned within 10% of the minimum wage and restaurant workers accounted for 28.6% of all workers earning within 10% of the minimum wage.

Analysis of restaurant employment is based on data from the Quarterly Census of Employment of Wages (QCEW) as in [Dube et al., 2010], updated to include more recent years (1990-2010)\(^9\). The QCEW has also been used in the follow-up studies [Neumark et al., 2014b, Neumark et al., 2014a, Allegretto et al., 2013, Allegretto et al., 2015]. The QCEW provides quarterly county-level payroll data by industry based on ES-202 filings that establishments submit for the purpose of calculating payroll taxes related to unemployment insurance. The county-quarter restaurant employment dependent variable is constructed from both Full Service Restaurants (NAICS 7221) and Limited Service Restaurants (NAICS 7222) and measures the total number of full service and limited service restaurant employees. The control variables are the county-quarter total private sector employment and the county population. The employment variables are constructed from the QCEW and the county population comes from the county-level Census Bureau population data which is produced annually. Data is available for the entire time frame of analysis for 1,371 counties\(^{10}\). Robustness checks that analyze the effect of minimum wage hikes on restaurant worker earnings are based on average weekly wages, which are constructed by dividing the county-quarter total restaurant payroll by the county-quarter number of restaurant employees. A minimum wage variable is merged to the dataset, which is always the higher of the

\(^9\)The time period of analysis stops in 2010 because there were classification changes to the four-digit NAICS industry codes, which [Dube et al., 2010] use to identify restaurants, beginning in 2011. Results that include more recent years are very similar to the results shown in this paper.

\(^{10}\)For consistency with [Dube et al., 2010], results are based on a balanced panel.
state and federal minimum wage. Summary statistics for the dataset of analysis on
restaurant workers are shown in Table 3.1.

Analysis of teenage employment is based on data from the CPS Outgoing Rotation
Groups (CPS ORG) as in [Allegretto et al., 2011], updated to include more recent
years (1990-2013). CPS ORG data has been used in each of the follow-up studies
on teenage employment [Neumark et al., 2014b, Neumark et al., 2014a, Allegretto
et al., 2013, Allegretto et al., 2015]. State-quarter observations are constructed by
aggregating the CPS ORG individual-level data up to the state-quarter level. The
state-quarter teenage employment dependent variable is the fraction of teenagers,
ages 16-19, that are employed. The control variables are the state-quarter relative
size of the teenage population and state-quarter unemployment rate, also constructed
from the CPS-ORG. Robustness checks that analyze the effect of minimum wage hikes
on teenage earnings are based on average hourly earnings, which are based only on
those who were working and paid between $1 and $100 per hour in 2009 dollars and
are constructed by aggregating the CPS ORG individual hourly earnings data to the
state-quarter level. A minimum wage variable is merged to the dataset, which is
always the higher of the state and federal minimum wage. Summary statistics for the
dataset of analysis on teenagers are shown in Table 3.1.

As robustness checks, two other datasets are used to analyze teenage employment:
CPS basic monthly and Quarterly Workforce Indicators (QWI). These two datasets
have been used in more recent minimum wage-employment studies [Allegretto et al.,
2013, Allegretto et al., 2015]. The CPS basic monthly files do not have data on earn-
ings, which is necessary for the crucial robustness check on the effect of minimum wage
hikes on earnings, but it has a much larger monthly sample of employment outcomes.
The QWI is a newer dataset which provides county-level teenage employment counts.
The data is a public-use aggregation of the matched employer-employee Longitudinal
Employer Household Dynamics (LEHD) database, which is provided via a partner-
ship between the Census Bureau and the Labor Market Information (LMI) offices.
Both of these datasets are taken directly from the data associated with [Allegretto et al., 2015]\textsuperscript{11}. Summary statistics for these datasets are also shown in Table 3.1.

1.3.2 Cross section dependence and data size

As discussed in the introduction and Section 2.1, the presence of unobserved common factors can cause outcomes or residuals to be correlated across areas, known as cross section dependence, which can be problematic for inference and estimation. The factor model approach is commonly used to model the presence of strong, as opposed to weak, cross section dependence. Weak cross section dependence can be thought of as arising from the fact that geographically proximate places will have similar characteristics, due to integrated geography and labor markets, which will cause correlation in outcomes between neighboring areas. Strong cross section dependence, on the other hand, is typically thought of as arising from unobserved forces ("factors") that influence outcomes in heterogeneous ways between areas. Spatial econometric methods, which deal with cross section dependence by assuming some specific geographic structure for the correlation between areas ex ante, are an alternative way to address issues related to cross section dependence. However, factor models have two important advantages over spatial econometric methods. The first is that factor models are intended to capture the presence of strong cross section dependence, whereas spatial econometric methods typically require that the cross section dependence is only weak. The second is that the factor model approach assumes no geographic relationship for the correlation between areas before estimation, thus allowing outcomes and residuals to be correlated in ways that depart from geographic proximity.

To validate the use of factor model methods, Table 3.1 shows the results of a test for strong cross section dependence in the data using the cross section dependence (CD) test from [Pesaran, 2015]. This test is based on the average of pair-wise correlations in the data, with greater correlations indicating greater cross section de-

\textsuperscript{11}The data can be downloaded here: https://arindube.com/working-papers/
dependence and thus producing a larger test statistic. The null hypothesis of the test, which is distributed standard normal, is that there is only weak cross section dependence, while the alternative is that the cross section dependence is strong. The null hypothesis of weak cross section dependence is rejected at the one-percent level for each of the datasets for restaurant and teenage employment. This suggests the presence of common factors influencing employment across areas and validates the factor model approach. The fact that there is less strong cross section dependence in the state-level CPS datasets is not surprising, given that the unit of analysis occurs at a more aggregated level.

Because CPS data does not allow for reliable estimates at the county level, the teenage dataset has a relatively small cross section dimension of 51, containing all 50 states and Washington D.C. As discussed in Section 2.2, one of the results from [Westerlund and Urbain, 2015] is that both the IFE and CCE estimators perform better when \( N > T \), which is not the case for the CPS datasets. It is therefore possible that the factor model estimators may not be able to capture the common factors as reliably for the CPS datasets as they can for the QCEW and QWI and therefore may not be able to fully remove any cross section dependence and bias that is caused by common factors. Nonetheless, the IFE and CCE estimators still provide significant improvements over traditional OLS methods when common factors exist in the data and still perform well without the presence of common factors in the data even when \( N < T \), as shown in [Pesaran, 2006] and [Bai, 2009] and in the simulations in Section 5.

1.4 Results

1.4.1 Minimum wage-employment elasticity

All OLS results are based on the traditional two-way fixed effects specification in equation (1). Factor model results are based on the two-way fixed effects specification in equation (1) and the multi-factor error structure in equation (2). Results
are reported separately for each of the datasets. The results tables first show the OLS estimate of the two-way fixed effects specification, and then show results using each of the factor model estimators described in Section 2. Confidence intervals and significance reported in the tables are based on bootstrapped t-statistics using the *wild cluster bootstrap-t* procedure from [Cameron et al., 2008], clustered at the state level.

The effect of minimum wage hikes on restaurant employment is shown in Table 1.3. Column (1) shows that the OLS estimate of the traditional two-way fixed effects specification is in line with other estimates from the literature, with a large elasticity of -0.138 that is statistically significant. However, the factor model estimators in columns (2)-(4) produce very different elasticity estimates. The CCEMG, CCEP, and IFE estimators produce elasticity estimates of -0.013, -0.013, and -0.042, respectively. None of the factor model estimates are statistically different from zero. In addition to producing smaller elasticity estimates, the factor model estimators produce more precise estimates: the confidence intervals are much tighter for the factor model estimators than OLS. This is consistent with the presence of common factors that the OLS estimator is not controlling for, which would cause bias and inefficiency in the OLS estimator. Interestingly, the factor model estimates for the other covariates are much more similar to OLS in terms of both the point estimate and the confidence interval length; it appears that the unobserved common factors are correlated with the minimum wage more so than total private sector employment or population.

Table 1.3 also shows residual diagnostics that test for the presence of strong cross section dependence in the residuals. There is still strong cross section dependence in the OLS residuals, which have a test statistic of 26.98, suggesting the presence of unobserved common factors remaining in the OLS residuals. The factor model estimators do a better job of controlling for common factors and removing cross section dependence from the data, with test statistics of 8.46, 17.19, and -0.30 for the CCEMG, CCEP, and IFE estimators, respectively.
The effect of minimum wage hikes on teenage employment is shown in Table 1.4. Column (1) shows that the OLS estimate of the minimum wage-employment elasticity is once again in line with other estimates from the literature, with a large elasticity of -0.178 that is statistically significant. Just as with the restaurant employment dataset, the factor model estimators in columns (2)-(4) produce very different elasticity estimates than OLS: the CCEMG, CCEP, and IFE estimators produce elasticity estimates of -0.040, -0.065, and -0.036, respectively, none of which are statistically different from zero. Regarding the other covariates, the factor model estimators produce slightly smaller estimates than OLS for the effect of the unemployment rate and produce estimates of the opposite sign for the effect of the teenage population share.

Although the factor model estimators produce significantly smaller elasticity estimates than OLS for teenage employment, they do not produce tighter confidence intervals or further remove cross section dependence from the residuals. This may be due to the relatively small size of the cross section dimension of the teenage employment dataset, an issue discussed in Section 3.2: it is possible that the relatively small cross section dimension may lead to imprecise estimates of the common factors, which could leave the residuals contaminated with cross section dependence while still removing bias from the parameter estimates. This seems especially plausible for the $\sqrt{NT}$-consistent IFE estimator, whose estimates of the common factors via principal components are only $\sqrt{N}$-consistent. Nonetheless, the main concern in this paper is removing bias from the minimum wage-employment elasticity, rather than cross section dependence from the residuals. The factor model estimators do still appear to be capturing the presence of common factors, because the factor model estimates are very different from the OLS estimates. Section 5 will confirm that the factor model estimators would produce elasticity estimates similar to OLS if there were no common factors in the data, even for the teenage dataset.
1.4.2 Alternative teenage employment data sources

As discussed in Section 3.1, CPS basic monthly files and the QWI have been used to analyze teenage employment in more recent studies. The advantage of the CPS basic monthly files is that they provide a much larger sample of monthly employment outcomes than the CPS ORG. The advantage of the QWI data is that it provides county-level teenage employment counts, which could address the cross section dimension issue associated with the CPS data for teenage employment.

Table 1.5 shows the results for these two datasets. Panel A is based on CPS basic monthly files. Column (1) shows that the OLS estimate is once again consistent with the literature, with a negative and statistically significant elasticity. Similar to the CPS ORG results, the factor model estimators in columns (2)-(4) produce much smaller elasticity estimates, but do not produce tighter confidence intervals or remove more of the cross section dependence from the residuals.

Panel B is based on the QWI data. In this case, even the OLS estimator produces a small elasticity that is not statistically different from zero. [Allegretto et al., 2015] find significant negative effects when using this dataset and only two-way fixed effects, but they analyze only contiguous counties for the border discontinuity approach, rather than the entire sample. The factor model estimators also produce elasticity estimates that are small and close to zero. The CCE estimates are statistically significant, but they are still smaller in magnitude than the traditional OLS estimates, which are in the range of -0.1 to -0.2. Unlike teenage results based on state-level CPS data, the factor model estimators now produce much tighter confidence intervals than OLS and remove more of the cross section dependence from the residuals than OLS. This confirms that the lack of improvement in confidence interval length and cross section dependence associated with the factor model results for the CPS data were likely due to the relatively small cross section dimension.
1.4.3 Number of common factors for IFE estimation

One important feature of the IFE procedure is the selection of the number of common factors. As discussed in Section 2.2, one way to determine the number of common factors is to use the information criteria from [Bai and Ng, 2002], which estimates the number of strong common factors in the data. However, the information criteria proved to be uninformative, as it picked the minimum number of factors allowed in some cases and the maximum allowed in other cases. This has occurred in other empirical papers using this method [Kim and Oka, 2014, Bailey et al., 2016]. Therefore, in the analysis above, the IFE results were based on the minimum number of factors needed to explain approximately 90% of the variation in the residuals. This meant 4 common factors for restaurant employment, 8 common factors for CPS ORG and CPS basic monthly teenage employment, and 6 common factors for QWI teenage employment. This can be seen in Table 1.6, which shows the relative importance of each common factor, $R^2_p$. The relative importance of each common factor is calculated as the fraction of the total variance in the residuals explained by factors 1 to $p$, given as the sum of the first $p$ largest eigenvalues of the second moment matrix of the OLS residuals divided by the sum of all eigenvalues.

As a robustness check, Table 1.7 shows IFE results based on 2-8 common factors. For restaurant employment, the IFE estimate of the minimum wage-employment elasticity is invariant to the number of factors, with elasticity estimates that remain small and not statistically significant for different numbers of factors. The teenage minimum wage-employment elasticity is not entirely invariant to the number of common factors for either CPS dataset: while the elasticity estimates are not statistically significant for any number of common factors, the IFE estimates are somewhat large when only 2 or 3 common factors are included. Once 5 common factors are included, the IFE estimates are much more similar to the CCE estimates reported in Table 1.4 and Table 1.5 and they remain relatively small for up to 8 common factors. The fact that it takes a larger number of common factors for IFE elasticity estimates based
on CPS data to become small is consistent with the results in Table 1.6 that it takes a larger number of common factors to explain the variation in the CPS residuals. Figure 1.2 plots the IFE elasticity estimate for 1-10 common factors for each dataset. The fact that the IFE estimates are generally more stable for higher numbers of common factors is consistent with the result in [Moon and Weidner, 2015] that the IFE estimator still performs well when the number of common factors is over-estimated, but that the IFE estimator can be biased when the number of common factors is under-estimated.

1.4.4 Accounting for the difference between OLS and factor model estimators

The previous results showed that the factor model estimators consistently produce elasticity estimates that are smaller than the traditional two-way fixed effects OLS estimates. The goal of this section is to attempt to shed some light on what the factor model estimators are capturing in the error term of equation (1) that OLS cannot account for by analyzing the estimated factor structure from the IFE estimator. No direct economic interpretation can be given to the common factors, as they are defined purely in a statistical sense: the factors are the eigenvectors that correspond to the largest eigenvalues of the second moment matrix of the regression residuals. Additionally, while the product $\lambda_i f_i$ is identifiable, the factors and factor loadings themselves are identifiable only up to a sign change. Nonetheless, the factor structure does have some interpretable patterns that are relevant to the recent debate about the appropriateness of state-specific time trends and Census division-by-period fixed effects.

Figure 1.4 and Figure 1.5 show the first common factor from the IFE estimator for restaurant employment and teenage employment from the QWI, respectively. The common factor for restaurant employment looks similar to a linear time trend with seasonality. The common factor for teenage employment from the QWI looks similar
to a quadratic time trend with seasonality. This produces a structure very similar to county-specific time trends when each factor is multiplied by its factor loadings. This lends some support to the inclusion of time trends in the two-way fixed effects specification, but only for restaurant employment and teenage employment from the QWI. Additionally, these common factors that resemble unit-specific time trends represent only a portion of what remains in the error term of the two-way fixed effects specification; it takes several more factors to explain a large fraction of the variance in the OLS residuals for these datasets and the QWI IFE estimates do not become significantly different from OLS until additional factors are included, as seen in Table 1.6 and Table 1.7. The other factors are not shown because they do not have such interpretable or relevant patterns.

Figures 1.6-1.9 plot the combined effect of the unobserved common factors for a given time period. This is constructed as the inner product of the \((1 \times r)\) vector of factor loadings for each cross section unit, \(\lambda_i\), and the \((r \times 1)\) vector of common factors for a given time period, \(f_t\). Each of these figures shows time-varying regional clustering in the effect of the unobserved common factors. In Figure 1.6, for example, much of the West Coast, Midwest, and Northeast experience positive effects on employment from the unobserved common factors in 1990q1. However, in 2010q1, the Southwest, parts of the South and Southeast, and parts of the Northeast experience positive employment effects from the unobserved common factors. Similar patterns of time-varying regional clustering appear in Figures 1.7-1.9.

Interestingly, this time-varying regional clustering could roughly be approximated by a Census division-by-period fixed effect. For example, in Figure 1.6 for 1990q1, the darker regions align fairly well with the Pacific, East North Central, Middle Atlantic, and New England Census divisions. However, the figures also illustrate the flexibility of the factor approach over a Census division-by-period fixed effect or the border discontinuity approach: while regional clustering does occur, there are also many cases in which same-division states or counties on opposite sides of a state border experience very different effects from the common factors. For example, in Figure 1.6...
for 1990q1, Ohio clearly stands out on the map from the states around it; it appears to experiencing a very different shock than other same-division states and counties on the opposite side of the border from Ohio appear to be poor control groups for counties in Ohio for a border discontinuity approach. Another interesting feature of the figures is that in Figure 1.6 for 2010q1, the areas experiencing positive effects from the common factors are primarily clustered around the major cities in the US. The is a remarkable result given that the factor structure does not know where counties are relative to each other on a map when it estimates their factor loadings.

1.4.5 Throwing out the baby with the bathwater?

The factor model results presented above are more similar to results from the literature that include Census division-by-period fixed effects and state-specific time trends than results based on only two-way fixed effects. Additionally, analysis of the factor structure estimated from the IFE estimator revealed some similarities to unit-specific time trends and Census division-by-period fixed effects. One of the main critiques of Census division-by-period fixed effects and state-specific time trends has been that they ”throw out the baby with the bathwater” [Neumark et al., 2014b]. That is, they potentially discard too much valid identifying variation in pursuit of ideal counterfactuals. This critique is mostly driven by the fact that synthetic controls sometimes place very little weight on same-division states, suggesting same-division states may not provide a better counterfactual than a randomly selected state.

The same critique could potentially be applied to the factor model approach: the factors from the IFE procedure and the cross section averages which proxy for factors in the CCE estimator may explain a lot of the identifying variation themselves, leaving a small amount variation in the data with which to estimate the minimum wage-employment elasticity. In order to address this critique, Table 1.8 shows the fraction of the variation in both the dependent employment variables and the minimum wage variable that is explained by each of Census division-by-period fixed effects and state-
specific time trends, the common factors from the IFE estimator, and the cross section averages from the CCE estimator. This is computed by regressing each variable on one of the three sets of controls for unobserved heterogeneity and reporting the r-squared.

Census division-by-period fixed effects and state-specific time trends explain a lot of the variation in the data. For the restaurant employment dataset, they explain 17.4% of the variation in log(employment) and 94.4% of the variation in log(minimum wage). For the CPS ORG teenage employment dataset, they explain 62.9% of the variation in log(employment/population) and 93.5% of the variation in log(minimum wage). Results are very similar for the other data sources for teenage employment.

The common factors from the IFE estimator and the cross section averages from the CCE estimator consistently explain a smaller fraction of the variance in the data across datasets and for both the employment and minimum wage variables. The difference between the fraction explain by Census division-by-period fixed effects and state-specific time trends compared to the fraction explained by factor model controls is largest for the employment variables, but there is also a significant reduction for the minimum wage variable. It is not obvious how much variation being removed by controls for unobserved heterogeneity would be too much, but the factor model approach does at least do a better job of not removing too much variation than Census division-by-period fixed effects and state-specific time trends.

1.4.6 Robustness checks

Earnings elasticity

One potential explanation for finding no significant effect of minimum wage hikes on employment is that minimum wage hikes are not binding wage floors. The fact that a significant portion of teenage and restaurant workers earn within 10% of the minimum wage, as mentioned in Section 3, along with the fact that the average size of a minimum wage hike since 1990 is approximately 10%, suggests that minimum wage hikes likely are binding for a significant number of low-wage workers. To test this
further, Table 1.9 shows the effect of minimum wage hikes of worker earnings. These regressions are based on the same specifications as the employment results, except with earnings as the dependent variable and with log(total private sector average weekly wages) instead of log(total private sector employment) as a control variable for restaurant earnings regressions.

OLS results in column (1) shows positive and statistically significant minimum wage-earnings elasticity estimates for restaurant workers and teenage workers. The elasticity estimates are in the range of 0.1 to 0.2, but they could be driven by unobserved common factors that were influencing the employment results. However, the factor model estimates in columns (2)-(4) are also positive and statistically significant. The factor model estimates are very similar in magnitude to the OLS estimates, except for the QWI teenage employment dataset, with which the factor model estimators produce even larger elasticity estimates than OLS.\(^{12}\)

These results suggest that minimum wage hikes are binding for restaurant workers and teenager workers and that they receive a 1% to 3% increase in earnings from a 10% increase in the minimum wage. The earnings results also confirm that the common factors and cross section averages do not explain so much variation in the data that there is too little left to be able to estimate large and statistically significant effects from minimum wages.

**State-specific time trends**

Because one of the debates in the minimum wage-employment literature has been about the lack of robustness to the presence and order of state-specific time trends, Table 1.10 and Table 1.11 replicate these results and also show how sensitive the factor model estimates are to time trends by adding state-specific time trends to equation (1). In this setup, the factor model estimators are applied to a specification which already models state-specific time trends in addition to unit and period fixed effects.

\(^{12}\)Robustness of the IFE results to the number of common factors is shown in Figure 1.3; the results are almost entirely invariant to the number of common factors.
As discussed previously, the factor model setup can capture deterministic trends as a special case of the factor structure without suffering from any potential bias resulting from specifying state-specific trends when they are not appropriate. Therefore, from a theoretical standpoint, the inclusion of time trends is not necessary, although it would me more efficient to include them if they are part of the true underlying data-generating process; the results from Table 1.4 remain the primary results, given that they will capture time trends only if they show up in the data.

Table 1.10 shows that the OLS estimate of the restaurant minimum wage-employment elasticity does vary across specifications. The large negative estimate from the two-way fixed effects specification disappears when a linear time trend is included, but a smaller negative estimate that is statistically different than zero returns for higher order state-specific trends. The factor model estimators, on the other hand, are much more robust to the presence and order of state-specific time trends: only the CCEP estimate with a 3rd order polynomial state-specific time trend shows any considerable change in magnitude or significance and all of the factor model estimates remain smaller than the OLS estimate from the traditional two-way fixed effects specification in Table 1.3.

Table 1.11 shows the results for teenage employment. The OLS estimate of the teenage minimum wage-employment elasticity is even more sensitive to the presence and order of state-specific time trends: the large negative estimate from the two-way fixed effects specification becomes much smaller when linear or quadratic state-specific time trends are included, but large negative estimates that are almost as large as the two-way fixed effects estimate return for higher order trends. The factor model estimates also show some variance across specifications, but not as much as the OLS estimates. The factor model estimators remain smaller than OLS for each specification and are always smaller than the traditional two-way fixed effects OLS estimate in Table 1.4. The variance across specifications for the factor model estimates of the teenage minimum wage-employment elasticity is likely another result of the relatively small cross section dimension of the CPS data.
Sub-samples of the data

Section 2.4 discussed the primary limitation of the factor model approach, which is the assumption of time-invariant factor loadings. While there is some reason to believe that at least the IFE estimator may still perform well in this setting, this nonetheless posses the most obvious threat to identification of regression parameters, as incorrect modeling of the error term could affect the estimation of regression parameters. One robustness check is to estimate the model for sub-samples of the data: if the time variation in the factor loadings was of the structural break variety and thus time-invariant within sub-samples of the time dimension of the data, sub-sample analysis could help address this issue. The trade off, however, is that sample size concerns become a bigger issue.

Sub-sample results are shown in Tables 1.12 and 1.13 for restaurant and teenage employment, respectively. Three sub-samples are considered for each dataset: (1) the first half of the time dimension of the data, (2) the second and third quarters of the time dimension of the data, and (3) the last half of the time dimension of the data. Analysis is performed for restaurant employment and CPS ORG teenage employment\textsuperscript{13}. For restaurant employment, OLS shows a large negative effect in the 1995q2-2005q3 sub-sample. The factor model estimates do not show any large negative effects. The CCEP estimate is negative and statistically significant in the 1990q1-2000q2 sub-sample, but the magnitude is -0.060, which is much smaller than the traditional negative estimates found in the literature. For teenage employment, OLS shows large negative effects in earlier sub-samples. The factor model estimators find similar results, although the magnitude of the negative effect in the first half of the sample is smaller. Overall, the factor model estimates remain both smaller than OLS and below the traditional -0.1 to -0.3 range.

\textsuperscript{13}The dimension of the sub-samples are $N = 1371$, $T = 42$ for restaurant employment and $N = 51$, $T = 48$ for teenage employment. Splitting the QWI sample is not feasible given the already short time frame. CPS basic monthly results are very similar to the CPS ORG results.
Falsification test

Table 1.14 shows the results of a falsification test based on the manufacturing industry. Only 2.8% of manufacturing workers earn within 10% of the minimum wage [Dube et al., 2010]. The manufacturing industry therefore should not experience significant employment or earnings effects from minimum wages hikes. Reassuringly, the factor model estimators find no statistically significant effect on employment. The CCEP estimator does find a positive and statistically significant effect on earnings. This could just be a statistical artifact, although if there was going to be any effect on manufacturing workers, it would likely be a small increase in earnings, given evidence that there can be spillovers to workers making more than the minimum wage [Lopresti and Mumford, 2015].

Pre-existing trends

Support for the Census division-by-period fixed effects, state-specific trends, and the border discontinuity approach in [Dube et al., 2010] and [Allegretto et al., 2011] comes from evidence that these controls remove negative pre-existing trends in employment for states that raise their minimum wage relative to states that do not. The way that the authors show this is to include leads of the minimum wage variable, in which case the traditional two-way fixed effects approach produces large negative leading effects that are removed when the additional controls are included. The authors argue that finding negative effects prior to the policy change reflects spurious pre-trends due to minimum wage changes tending to occur at times and places with unusually low employment growth. The leading effect results have been another topic of debate in the follow up work.

Figure 1.10 and Figure 1.11 show elasticity estimates for specifications that include leads in the minimum wage variable. Results are based on the same model specified in equation (1), except with 1-, 2-, and 3-year leads in the minimum wage variable included. The elasticity reported at each point in time in the figures is the cumulative
elasticity, which is the sum of the contemporaneous elasticity estimates for each year until that point in time.

Figure 1.10 shows the results for restaurant employment. Similar to the results reported in [Dube et al., 2010], OLS produces a negative trend in the leading effect of the minimum wage: the elasticity one year before the minimum wage increase occurs is -0.15, which is similar to the contemporaneous effect shown in Table 1.3. This negative trend is removed by the factor model estimators, which produce flat leading effects that are closer to zero. Figure 1.11 shows the results for teenage employment.

There is again a negative trend in the leading effect for OLS, with an elasticity of nearly -0.2 three years before the minimum wage increase and -0.3 one year before the increase. The factor model estimators do not fully remove the negative leading effects for teenage employment, but they do decrease their magnitude.

**Contiguous counties**

For comparison with the border discontinuity approach in [Dube et al., 2010], Table 1.15 shows the OLS and factor model estimates based on the sample of contiguous counties from [Dube et al., 2010]. Column (1) shows an OLS estimate of -0.106, which is not statistically significant but is at the low end of the traditional -0.1 to -0.3 range of estimates. This is consistent with the results from [Dube et al., 2010]; it is not until they add county-pair-by-period fixed effects for each pair of contiguous counties that they find very small elasticity estimates. However, each of the factor model estimators produces small elasticity estimates in the range of -0.01 to -0.03 based on only the two-way fixed effects specification. The factor model estimators also produce much tighter confidence intervals and remove more cross section dependence from the residuals than OLS.
1.5 Simulation

This section assesses the relative ability of OLS and the factor model estimators to estimate the minimum wage-employment elasticity under different assumptions about the unobserved heterogeneity in the data. Specifically, the performance of OLS, CCEMG, CCEP, and IFE are compared with and without the presence of common factors in the data-generating process. The goals of these simulations are (1) to confirm the precision of the factor model estimates of the minimum wage-employment elasticity when no common factors exist in the data and (2) to confirm the direction of the bias in the OLS estimate of the minimum wage-employment elasticity caused by the common factors. The simulations use the same data from the results section in the DGP, but impose different assumptions on the unobserved heterogeneity to simulate new employment observations.

The first simulation analyzes the performance of the OLS, CCEP, CCEMG, and IFE estimators with only state/county and period fixed effects representing the unobserved heterogeneity in the DGP. This DGP uses the OLS results as the true value of the coefficients in the DGP, with independent and identically distributed (IID) normal errors. For the restaurant employment DGP, these coefficients come from the OLS estimates in Table 1.3 and the error variance is computed using the residuals from this specification. For the teenage employment DGP, the true value of the coefficients and the error variance come from the OLS estimates in Table 1.4. The simulation is performed for 1,000 repetitions for each dataset.

The results from this simulation are shown in Table 1.16. Columns (1) and (4) report the median of the estimates for each of the estimators for restaurant and teenage employment, respectively, and columns (2)-(3) and (5)-(6) report the 95% range of the estimates. All four estimators perform well without the presence of

\[ y_{it}^* = \hat{\beta} \ln(MW_{it}) + X_{it} \hat{\Gamma} + \hat{\alpha}_i + \hat{\delta}_t + v_{it}. \]

The independent variables are the same variables from the main results section, the parameters are from the OLS results reported in Tables 3 and 4, and \( v_{it} \) is an idiosyncratic error term whose variance is determined by the variance of the OLS residuals. State/county and period fixed effects are recovered from the data and included in the DGP.
factors in the DGP; the median estimate of the minimum wage-employment elasticity is near the true value for each estimator. The factor model estimators also perform well in terms of the 95% range of the estimates, with only slightly wider ranges than OLS. There are two key takeaways from this simulation. The first key takeaway is that the factor model estimators perform well without the presence of factors in the DGP, even with the small cross section dimension of the teenage dataset. The second key takeaway is that the pattern of results in this simulation does not match the pattern of results in Section 4.1: the factor model results were very different from OLS in Section 4.1, but they are very similar here. Overall, these results show that the factor model estimators would produce minimum wage-employment elasticity estimates similar to OLS if the two-way fixed effects specification was correct.

The second simulation analyzes the performance of the OLS, CCEP, CCEMG, and IFE estimators with state/county and period fixed effects and common factors representing the unobserved heterogeneity in the DGP. This DGP uses the coefficients, common factors, and factor loadings from the IFE estimation, with independent and identically distributed (IID) normal errors. For the restaurant employment DGP, the true value of the coefficients and factor structure comes from the IFE estimates in Table 1.3 and the error variance is computed using the residuals from this specification. For the teenage employment DGP, the true value of the coefficients, factor structure, and the error variance come from the IFE estimates in Table 1.4. The simulation is performed for 1,000 repetitions for each dataset.

The results from this simulation are shown in Table 1.17. For restaurant employment, the CCEP, CCEMG, and IFE estimators all perform well. The OLS estimator, however, shows consistent and severe negative bias across repetitions: the true value of the coefficient for the minimum wage-employment elasticity is not even in the 95% range of the OLS estimates. For teenage employment, the OLS estimator once again

---

15The DGP is $y_{it} = \beta \ln(MW_{it}) + X_{it}\hat{\Gamma} + \hat{\alpha}_i + \hat{\delta}_t + \hat{\lambda}_i^\prime \hat{f}_t + v_{it}$. The independent variables are the same variables from the main results section, the parameters are from the IFE estimation reported in Tables 3 and 4, and $v_{it}$ is an idiosyncratic error term whose variance is determined by the variance of the IFE residuals. State/county and period fixed effects are recovered from the data and included in the DGP.
shows significant negative bias, with the 95% confidence range not containing the true value of the minimum wage-employment elasticity coefficient. The CCEP and CCEMG estimators also show some negative bias for the teenage dataset, although not as much as OLS. This is consistent with the discussion in Section 3.2 that the factor model estimators may not be able to remove all of the bias caused by common factors in the teenage employment data due to the relatively small cross section dimension of the data. This simulation produces two key takeaways. The first key takeaway is that the OLS estimate of the minimum wage-employment elasticity is negatively biased when the common factors for restaurant and teenage employment are included in the DGP. The second key takeaway is that the pattern of results from this simulation matches the pattern of results seen in Table 1.3 and Table 1.4: the OLS estimates of the minimum wage-employment elasticity are much larger in magnitude than the factor model estimates both in Table 1.3 and Table 1.4 and in simulations with common factors included in the DGP.

In summary, the simulations show that the CCEP, CCEMG, and IFE estimators would produce minimum wage-employment elasticity estimates similar to OLS if state/county and period fixed effects fully represented the unobserved heterogeneity in the underlying data generating process. When common factors are included in the DGP, the OLS estimate of the minimum wage-employment elasticity is negatively biased, while the factor model estimators perform much better. These results suggest that the presence of common factors in the true underlying DGP can cause the different estimates of the minimum wage-employment elasticity seen across approaches in Table 1.3 and Table 1.4.

1.6 Discussion and conclusion

The recent minimum wage-employment debate in the literature has focused on how to generate credible estimates of employment effects when using aggregated panel data. Doing so is very difficult, as both outcomes and minimum wage policy are likely
to be correlated across areas due to unobservable confounders. Many approaches have been proposed in order to control for or remove these confounders, including saturating the two-way fixed effects specification with additional controls, using a border discontinuity approach to identify the effect based on policy discontinuity at state borders, and using synthetic controls. Each approach has its drawbacks. More importantly, the various approaches have produced inconsistent results, and it is not obvious which approach is best or which set of results to believe.

The factor model estimators from [Pesaran, 2006] and [Bai, 2009] are very well-suited to address the issues in the minimum wage-employment literature: they are intended to control for unobservable factors in panel data in order to generate unbiased estimates of regression parameters for observed covariates. These estimators have specific advantages over the other approaches that have recently been used in the literature. They also satisfy the main concerns from each side of the recent debate, which is that they facilitate the control of unobserved confounders without discarding a significant amount of identifying variation, changing the identifying variation, or discarding data altogether.

The factor model estimators produce minimum wage-employment elasticity estimates in the range of -0.01 to -0.05 for restaurant employment and -0.03 to -0.07 for teenage employment. These results are generally robust to a number of robustness checks, including alternative sources for teenage employment data and different assumptions about the number of common factors in the data. Furthermore, the pattern of relatively large negative OLS elasticity estimates and small factor model estimates cannot be replicated in simulations which include only two-way fixed effects as the true unobserved heterogeneity, suggesting that the traditional two-way fixed effects specification is not correct. The simulations also confirm that common factors from the minimum wage-employment data cause negative bias in the OLS estimate of the minimum wage-employment elasticity.

Overall, the factor model results suggest that there has been little to no effect of minimum wage hikes on restaurant or teenage employment over the last three
decades. However, the size of the minimum wage hike is important. Most minimum wage hikes in the U.S. during this time have been in the range of 5% to 15%. The results in this study, or any other study based on data from past U.S. minimum wage hikes, may not be informative about the effects of larger hikes.
Fig. 1.1.: The top figure shows the average minimum wage in each state from 1980-2015. The bottom figure shows the total number of minimum wage hikes in each state from 1980-2015.
Fig. 1.2.: Each figure shows the interactive fixed effects (IFE) estimate of the minimum wage-employment elasticity for different numbers of common factors. Zero common factors is equivalent to OLS. Corresponding confidence intervals are shown in Table 1.7.

Fig. 1.3.: Each figure shows the interactive fixed effects (IFE) estimate of the minimum wage-employment elasticity for different numbers of common factors. Zero common factors is equivalent to OLS.
Fig. 1.4.: The figure shows the first common factor for restaurant employment. Common factors are estimated jointly with the regression coefficients in the interactive fixed effects (IFE) procedure.
Fig. 1.5.: The figure shows the first common factor for teenage employment (QWI). Common factors are estimated jointly with the regression coefficients in the interactive fixed effects (IFE) procedure.
Fig. 1.6.: Each figure plots the combined effect of the common factors for the specified time period on each county’s log employment ($\hat{\lambda}_i \hat{f}_t$, 5 factors). Common factors and factor loadings are estimated jointly with the regression coefficients in the interactive fixed effects (IFE) procedure. Counties without data in the figure had missing observations in the raw data and were therefore not included in the balanced sample used for analysis.
Fig. 1.7.: Figures plots $\lambda_i f_t$, 8 factors. See Figure 1.6 for details.
Fig. 1.8.: Figure plots $\hat{\lambda}_i \hat{f}_t$, 8 factors. See Figure 1.6 for details.
Fig. 1.9.: Figure plots $\hat{\lambda}_i \hat{f}_t$, 6 factors. See Figure 1.6 for details.
Fig. 1.10.: This figure plots lead effects of minimum wage hikes. Results are based on equation (1) with 1-, 2-, and 3-year leads in the minimum wage variable included. The elasticity reported at each point in time is the cumulative elasticity, which is the sum of the contemporaneous elasticity estimates for each year until that point in time.
Fig. 1.11.: See Figure 1.10 for details.
Table 1.1.: Partial review of recent panel studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Approach</th>
<th>Elasticity</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neumark and Wascher (1992)</td>
<td>Teenagers</td>
<td>Traditional</td>
<td>-0.140**</td>
<td>Negative</td>
</tr>
<tr>
<td>Neumark and Wascher (2007)</td>
<td>Teenagers</td>
<td>Traditional</td>
<td>-0.136*</td>
<td>Negative</td>
</tr>
<tr>
<td>DLR (2010)</td>
<td>Restaurants</td>
<td>Census division-by-period fixed effects</td>
<td>-0.023</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CDxP FE &amp; state-specific linear trends</td>
<td>0.054</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contiguous county pairs</td>
<td>0.016</td>
<td>None</td>
</tr>
<tr>
<td>ADR (2011)</td>
<td>Teenagers</td>
<td>Census division-by-period fixed effects</td>
<td>-0.036</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>State-specific linear trends</td>
<td>-0.034</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CDxP FE &amp; state-specific linear trends</td>
<td>0.047</td>
<td>None</td>
</tr>
<tr>
<td>NSW (2014a, 2014b)</td>
<td>Restaurants</td>
<td>Synthetic controls</td>
<td>-0.063***</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>Teenagers</td>
<td>State-specific 5th-order trend</td>
<td>-0.185**</td>
<td>Negative</td>
</tr>
<tr>
<td>ADRZ (2013, 2015)</td>
<td>Teenagers</td>
<td>Synthetic controls</td>
<td>-0.036</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Double-selection post-LASSO</td>
<td>-0.012</td>
<td>None</td>
</tr>
</tbody>
</table>

The elasticity result is taken directly from the results reported in each study. For Neumark and Wascher (2007), this elasticity is constructed using the employment-population ratio in Table 1 and the employment coefficient in Table 2, specification 1. The “traditional” approach refers to using two-way fixed effects for time and location, with no additional controls for regional heterogeneity or selection of states experiencing minimum wage hikes. DLR=[Dube et al., 2010], ADR=[Allegretto et al., 2011], NSW=Neumark, Salas, and Wascher, ADRZ=Allegretto, Dube, Reich, and Zipperer. The synthetic control approach in NSW pools all synthetic and real data together and then estimates the two-way fixed effects specification with a fixed effect for each set of synthetic and real observations. Significance levels are as follows: *10 percent, **5 percent, ***1 percent.
Table 1.2: Data sources

<table>
<thead>
<tr>
<th>Panel A: Restaurant employment (QCEW)</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>CD test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restaurant employment</td>
<td>4,786</td>
<td>11,168</td>
<td>4992.8***</td>
</tr>
<tr>
<td>Restaurant average weekly wages</td>
<td>$170.77</td>
<td>$43.72</td>
<td>6939.0***</td>
</tr>
<tr>
<td>Total private sector employment</td>
<td>68,289</td>
<td>174,797</td>
<td></td>
</tr>
<tr>
<td>Total private sector average wages</td>
<td>$481.37</td>
<td>$136.80</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>181,719</td>
<td>423,564</td>
<td></td>
</tr>
<tr>
<td>Minimum wage</td>
<td>$5.26</td>
<td>$1.07</td>
<td></td>
</tr>
<tr>
<td>T (1990-2010)</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1,371</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Teenage employment (CPS ORG)</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>CD test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction of teenagers employed</td>
<td>0.41</td>
<td>0.12</td>
<td>205.2***</td>
</tr>
<tr>
<td>Average hourly wage for teens</td>
<td>$8.26</td>
<td>$0.86</td>
<td>88.82***</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>5.68</td>
<td>2.14</td>
<td></td>
</tr>
<tr>
<td>Relative size of teenage population</td>
<td>0.09</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Minimum wage</td>
<td>$5.58</td>
<td>$1.26</td>
<td></td>
</tr>
<tr>
<td>T (1990-2013)</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Teenage employment (CPS basic monthly)</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>CD test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction of teenagers employed</td>
<td>0.41</td>
<td>0.12</td>
<td>253.9**</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>5.58</td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td>Relative size of teenage population</td>
<td>0.07</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Minimum wage</td>
<td>$5.58</td>
<td>$1.26</td>
<td></td>
</tr>
<tr>
<td>T (1990-2013)</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D: Teenage employment (QWI)</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>CD test statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teenage employment</td>
<td>1,299</td>
<td>4,029</td>
<td>6856.8***</td>
</tr>
<tr>
<td>Teenage average weekly wages</td>
<td>$474.61</td>
<td>$868.95</td>
<td>6485.5***</td>
</tr>
<tr>
<td>Total private sector employment</td>
<td>41,046</td>
<td>153,885</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>357,940</td>
<td>1,237,330</td>
<td></td>
</tr>
<tr>
<td>Teenage population</td>
<td>13,233</td>
<td>44,986</td>
<td></td>
</tr>
<tr>
<td>Minimum wage</td>
<td>$5.98</td>
<td>$0.98</td>
<td></td>
</tr>
<tr>
<td>T (2000-2011)</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2,189</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the restaurant employment analysis, county-quarter employment and wage data come from the Quarterly Census of Employment and Wages (QCEW) and population data comes from the annual Census Bureau estimates. For teenage employment analysis using CPS ORG and CPS basic monthly files, all variables are constructed by aggregating the individual-level CPS files to the state-quarter level. For teenage employment analysis using the Quarterly Workforce Indicators (QWI) dataset, county-quarter employment and wage data come from the QWI and population data comes from the annual Census Bureau estimates. A minimum wage variable that is always the higher of the state and federal minimum wage is added to each dataset. Tests for cross section dependence are based on the test in [Pesaran, 2015].
### Table 1.3: Minimum wage-employment elasticity - restaurant employment

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>CCEMG</td>
<td>CCEP</td>
<td>IFE</td>
</tr>
<tr>
<td>Dependent variable: log(employment)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(Minimum wage)</td>
<td>-0.138*</td>
<td>-0.013</td>
<td>-0.013</td>
<td>-0.042</td>
</tr>
<tr>
<td></td>
<td>[-0.297,0.019]</td>
<td>[-0.042,0.026]</td>
<td>[-0.046,0.028]</td>
<td>[-0.085,0.015]</td>
</tr>
<tr>
<td>log(Total private sector emp.)</td>
<td>0.512***</td>
<td>0.704***</td>
<td>0.585***</td>
<td>0.519***</td>
</tr>
<tr>
<td></td>
<td>[0.430,0.595]</td>
<td>[0.667,0.742]</td>
<td>[0.515,0.653]</td>
<td>[0.424,0.601]</td>
</tr>
<tr>
<td>log(Population)</td>
<td>0.587***</td>
<td>0.373***</td>
<td>0.412***</td>
<td>0.296***</td>
</tr>
<tr>
<td></td>
<td>[0.432,0.742]</td>
<td>[0.184,0.566]</td>
<td>[0.285,0.547]</td>
<td>[0.138,0.436]</td>
</tr>
<tr>
<td>Period fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>County fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CD test statistic</td>
<td>26.98***</td>
<td>8.46***</td>
<td>17.19***</td>
<td>-0.30</td>
</tr>
<tr>
<td>$T_xN$</td>
<td>115,164</td>
<td>115,164</td>
<td>115,164</td>
<td>115,164</td>
</tr>
</tbody>
</table>

Each column uses a different estimator applied to the traditional two-way fixed effects specification shown in equation (1). IFE results are based on 4 common factors. OLS standard errors are clustered at the state level. Standard errors for CCEMG, CCEP, and IFE are calculated according to [Bai, 2009] and [Pesaran, 2006] and are described in the appendix. The confidence intervals and significance reported for CCEMG, CCEP, and IFE are based on bootstrapped t-statistics following the *wild cluster bootstrap-t* procedure in [Cameron et al., 2008], clustered at the state level. Significance levels are as follows: *10 percent, **5 percent, ***1 percent. Residual diagnostics for strong cross section dependence are based on the test in [Pesaran, 2015].
Table 1.4.: Minimum wage-employment elasticity - teenage employment

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>CCEMG</td>
<td>CCEP</td>
<td>IFE</td>
</tr>
<tr>
<td>Dependent variable: log(employment/population)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(Minimum wage)</td>
<td>-0.178**</td>
<td>-0.040</td>
<td>-0.065</td>
<td>-0.036</td>
</tr>
<tr>
<td></td>
<td>[-0.323,-0.033]</td>
<td>[-0.214,0.135]</td>
<td>[-0.191,0.061]</td>
<td>[-0.157,0.097]</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>-3.608***</td>
<td>-2.660***</td>
<td>-2.805***</td>
<td>-1.787***</td>
</tr>
<tr>
<td>Teen population share</td>
<td>-0.154</td>
<td>0.482</td>
<td>0.274</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td>[-0.709,0.401]</td>
<td>[-0.104,1.068]</td>
<td>[-0.223,0.771]</td>
<td>[-0.233,0.773]</td>
</tr>
<tr>
<td>Period fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CD test statistic</td>
<td>-5.96***</td>
<td>-6.01***</td>
<td>-6.03***</td>
<td>-6.69***</td>
</tr>
<tr>
<td>$T_{xN}$</td>
<td>4,896</td>
<td>4,896</td>
<td>4,896</td>
<td>4,896</td>
</tr>
</tbody>
</table>

IFE results are based on 8 common factors. See Table 1.3 for additional details.
Table 1.5.: Minimum wage-employment elasticity - teenage employment, other data sources

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>CCEMG</td>
<td>CCEP</td>
<td>IFE</td>
</tr>
<tr>
<td>Panel A: CPS basic monthly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log (Minimum wage)</td>
<td>-0.116*</td>
<td>0.001</td>
<td>-0.066</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>[-0.245,0.014]</td>
<td>[-0.182,0.184]</td>
<td>[-0.179,0.047]</td>
<td>[-0.017,0.159]</td>
</tr>
<tr>
<td>CD test statistic</td>
<td>-6.06***</td>
<td>-5.19***</td>
<td>-6.02***</td>
<td>-6.78***</td>
</tr>
<tr>
<td>TxN</td>
<td>4,896</td>
<td>4,896</td>
<td>4,896</td>
<td>4,896</td>
</tr>
<tr>
<td>Panel B: Quarterly Workforce Indicators</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log (Minimum wage)</td>
<td>-0.019</td>
<td>-0.089**</td>
<td>-0.039**</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td>[-0.277,0.239]</td>
<td>[-0.164,-0.014]</td>
<td>[-0.077,-0.001]</td>
<td>[-0.110,0.042]</td>
</tr>
<tr>
<td>CD test statistic</td>
<td>95.85***</td>
<td>51.81***</td>
<td>76.63***</td>
<td>20.93***</td>
</tr>
<tr>
<td>TxN</td>
<td>105,072</td>
<td>105,072</td>
<td>105,072</td>
<td>105,072</td>
</tr>
<tr>
<td>Period fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>County/State fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

IFE results are based on 8 common factors for the CPS basic monthly data and 6 common factors for the QWI data. Panel A also controls for state-quarter unemployment rate, state-quarter relative size of the teenage population, and state-quarter mean demographics for sex, age, race, Hispanic heritage, and marital status. Panel B controls for county-quarter total population, teenage population, and total private sector employment. See Table 1.3 for additional details.
Table 1.6.: Summary statistics for common factors

<table>
<thead>
<tr>
<th>Factor #p</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Restaurant employment</td>
<td>AR1((\hat{f}_{pt}))</td>
<td>0.990</td>
<td>0.975</td>
<td>0.527</td>
<td>0.447</td>
<td>0.988</td>
<td>0.969</td>
<td>0.954</td>
<td>0.884</td>
<td>0.882</td>
</tr>
<tr>
<td></td>
<td>(R^2_p)</td>
<td>0.538</td>
<td>0.707</td>
<td>0.785</td>
<td>0.857</td>
<td>0.899</td>
<td>0.928</td>
<td>0.951</td>
<td>0.971</td>
<td>0.987</td>
</tr>
<tr>
<td>Panel B: Teen employment (CPS ORG)</td>
<td>AR1((\hat{f}_{pt}))</td>
<td>0.496</td>
<td>0.295</td>
<td>-0.014</td>
<td>-0.009</td>
<td>0.021</td>
<td>0.031</td>
<td>-0.015</td>
<td>0.008</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>(R^2_p)</td>
<td>0.195</td>
<td>0.374</td>
<td>0.470</td>
<td>0.561</td>
<td>0.648</td>
<td>0.730</td>
<td>0.806</td>
<td>0.879</td>
<td>0.944</td>
</tr>
<tr>
<td>Panel C: Teen employment (CPS basic monthly)</td>
<td>AR1((\hat{f}_{pt}))</td>
<td>0.825</td>
<td>0.086</td>
<td>0.440</td>
<td>0.281</td>
<td>0.430</td>
<td>0.363</td>
<td>0.382</td>
<td>0.263</td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td>(R^2_p)</td>
<td>0.274</td>
<td>0.477</td>
<td>0.563</td>
<td>0.640</td>
<td>0.713</td>
<td>0.775</td>
<td>0.836</td>
<td>0.899</td>
<td>0.950</td>
</tr>
<tr>
<td>Panel D: Teen employment (QWI)</td>
<td>AR1((\hat{f}_{pt}))</td>
<td>0.827</td>
<td>0.032</td>
<td>0.917</td>
<td>0.941</td>
<td>0.901</td>
<td>0.876</td>
<td>-0.059</td>
<td>0.844</td>
<td>0.622</td>
</tr>
<tr>
<td></td>
<td>(R^2_p)</td>
<td>0.377</td>
<td>0.586</td>
<td>0.734</td>
<td>0.809</td>
<td>0.861</td>
<td>0.899</td>
<td>0.932</td>
<td>0.960</td>
<td>0.982</td>
</tr>
</tbody>
</table>

The first 10 common factors for each data source come from the interactive fixed effects (IFE) results with 10 pre-specified common factors. \(R^2_p\) shows the relative importance of each factor, calculated as the fraction of the total variance of the residuals explained by factors 1 to \(p\). This is given as the sum of the first \(p\) largest eigenvalues of the sample second moment matrix of the OLS residuals divided by the sum of all eigenvalues. \(\text{AR1}(\hat{f}_{pt})\) is the first order autocorrelation coefficient for the given factor.
Table 1.7.: IFE estimates for different numbers of common factors

<table>
<thead>
<tr>
<th>Number of common factors</th>
<th>Panel A: Restaurant employment</th>
<th>Panel B: Teen employment (CPS ORG)</th>
<th>Panel C: Teen employment (CPS basic monthly)</th>
<th>Panel D: Teen employment (QWI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>log(Minimum wage)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>-0.016</td>
<td>-0.035</td>
<td>-0.042</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td>[-0.056,0.029]</td>
<td>[-0.072,0.019]</td>
<td>[-0.085,0.015]</td>
<td>[-0.068,0.028]</td>
</tr>
<tr>
<td></td>
<td>CD test statistic</td>
<td>29.02***</td>
<td>-0.30</td>
<td>-0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.40</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>log(Minimum wage)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.104</td>
<td>-0.093</td>
<td>-0.069</td>
<td>-0.064</td>
</tr>
<tr>
<td></td>
<td>[-0.233,0.038]</td>
<td>[-0.224,0.045]</td>
<td>[-0.196,0.076]</td>
<td>[-0.179,0.072]</td>
</tr>
<tr>
<td></td>
<td>CD test statistic</td>
<td>-6.45***</td>
<td>-6.48***</td>
<td>-6.52***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-6.64***</td>
<td>-6.66***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-6.72***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>log(Minimum wage)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.072</td>
<td>-0.069</td>
<td>-0.078</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>[-0.143,0.039]</td>
<td>[-0.139,0.035]</td>
<td>[-0.140,0.029]</td>
<td>[-0.052,0.118]</td>
</tr>
<tr>
<td></td>
<td>CD test statistic</td>
<td>-6.29***</td>
<td>-6.71***</td>
<td>-6.74***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-6.77***</td>
<td>-6.77***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-6.78***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>log(Minimum wage)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.098**</td>
<td>-0.071*</td>
<td>-0.009</td>
<td>-0.031</td>
</tr>
<tr>
<td></td>
<td>[-0.170,-0.023]</td>
<td>[-0.146,0.002]</td>
<td>[-0.076,0.073]</td>
<td>[-0.108,0.044]</td>
</tr>
<tr>
<td></td>
<td>CD test statistic</td>
<td>20.17***</td>
<td>17.00***</td>
<td>21.42***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>18.75***</td>
<td>20.93***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13.25***</td>
<td>12.49***</td>
</tr>
</tbody>
</table>

Each column is a separate interactive fixed effects (IFE) estimate of the minimum wage-employment elasticity based on the traditional two-way fixed effects specification shown in equation (1). Each columns assumes a different number of pre-specified common factors for the IFE procedure. See Table 1.3 for additional details about standard errors, inference, and CD test statistics.
Table 1.8.: Proportion of variance ($R^2$) explained by controls for unobserved heterogeneity

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDxP fixed effects and state-specific time trends</td>
<td>Common factors estimated from IFE procedure</td>
<td>cross section averages from CCE procedure</td>
</tr>
<tr>
<td>Panel A: Restaurant employment</td>
<td>log(Employment)</td>
<td>0.174</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>log(Minimum wage)</td>
<td>0.944</td>
<td>0.837</td>
</tr>
<tr>
<td>Panel B: Teenage employment (CPS ORG)</td>
<td>log(Employment/population)</td>
<td>0.629</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>log(Minimum wage)</td>
<td>0.935</td>
<td>0.643</td>
</tr>
<tr>
<td>Panel C: Teenage employment (CPS ORG)</td>
<td>log(Employment/population)</td>
<td>0.693</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>log(Minimum wage)</td>
<td>0.935</td>
<td>0.715</td>
</tr>
<tr>
<td>Panel D: Teenage employment (QWI)</td>
<td>log(Employment)</td>
<td>0.189</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>log(Minimum wage)</td>
<td>0.885</td>
<td>0.753</td>
</tr>
</tbody>
</table>

Each column shows the proportion of the variance in the data for the given dependent variable that is explained by the controls for unobserved heterogeneity. This is computed by regressing the dependent variable on Census division-by-period fixed effects and state-specific time trends for column (1), regressing the dependent variable on the common factors estimated from the interactive fixed effects (IFE) procedure for column (2), and regressing the dependent variable on the cross section averages used in the common correlated effects (CCE) procedure for column (3).
Table 1.9.: Minimum wage-earnings elasticity estimates

<table>
<thead>
<tr>
<th></th>
<th>(1) OLS</th>
<th>(2) CCEMG</th>
<th>(3) CCEP</th>
<th>(4) IFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Restaurant average weekly wages</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(Minimum wage)</td>
<td>0.209***</td>
<td>0.231***</td>
<td>0.222***</td>
<td>0.145***</td>
</tr>
<tr>
<td></td>
<td>[0.160,0.257]</td>
<td>[0.175,0.264]</td>
<td>[0.199,0.251]</td>
<td>[0.091,0.210]</td>
</tr>
<tr>
<td></td>
<td>5.12***</td>
<td>6.00***</td>
<td>2.88***</td>
<td>0.21</td>
</tr>
<tr>
<td>Panel B: Teenage hourly earnings (CPS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(Minimum wage)</td>
<td>0.104***</td>
<td>0.097**</td>
<td>0.110***</td>
<td>0.158***</td>
</tr>
<tr>
<td></td>
<td>[0.041,0.167]</td>
<td>[0.010,0.188]</td>
<td>[0.034,0.184]</td>
<td>[0.083,0.232]</td>
</tr>
<tr>
<td></td>
<td>-6.18***</td>
<td>-6.10***</td>
<td>-6.20***</td>
<td>-6.77***</td>
</tr>
<tr>
<td>Panel C: Teenage average weekly wages (QWI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(Minimum wage)</td>
<td>0.193***</td>
<td>0.307***</td>
<td>0.294***</td>
<td>0.308***</td>
</tr>
<tr>
<td></td>
<td>[0.143,0.243]</td>
<td>[0.235,0.383]</td>
<td>[0.244,0.331]</td>
<td>[0.253,0.354]</td>
</tr>
<tr>
<td></td>
<td>108.05***</td>
<td>101.32***</td>
<td>74.31***</td>
<td>18.18***</td>
</tr>
<tr>
<td>Period fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Each column uses a different estimator applied to the traditional two-way fixed effects specification to estimate the minimum wage-earnings elasticity. The dependent variable is county-quarter average weekly wages for restaurant workers in Panel A, individual hourly earnings for teenage workers in the CPS aggregated to the state-quarter level in Panel B, and county-quarter average weekly wages for teenagers in the QWI in Panel C. Independent variables are the same as those for employment regressions, except log(Total private sector employment) is replaced with log(Total private sector average weekly wages) for restaurant earnings regressions.
<table>
<thead>
<tr>
<th>Polynomial order for state-specific trend</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>-0.138*</td>
<td>-0.041**</td>
<td>-0.024</td>
<td>-0.052***</td>
<td>-0.040**</td>
<td>-0.031*</td>
</tr>
<tr>
<td>[-0.297,0.019]</td>
<td>[-0.075,-0.008]</td>
<td>[-0.058,0.009]</td>
<td>[-0.085,-0.020]</td>
<td>[-0.071,-0.009]</td>
<td>[-0.067,0.003]</td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>-0.013</td>
<td>-0.003</td>
<td>-0.012</td>
<td>-0.013</td>
<td>0.001</td>
<td>-0.007</td>
</tr>
<tr>
<td>[-0.042,0.026]</td>
<td>[-0.039,0.022]</td>
<td>[-0.044,0.019]</td>
<td>[-0.045,0.018]</td>
<td>[-0.032,0.029]</td>
<td>[-0.035,0.022]</td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>-0.013</td>
<td>-0.003</td>
<td>-0.018</td>
<td>-0.041***</td>
<td>-0.013</td>
<td>-0.007</td>
</tr>
<tr>
<td>[-0.046,0.028]</td>
<td>[-0.036,0.025]</td>
<td>[-0.049,0.014]</td>
<td>[-0.073,-0.021]</td>
<td>[-0.038,0.022]</td>
<td>[-0.034,0.024]</td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>-0.042</td>
<td>-0.048</td>
<td>-0.041</td>
<td>-0.044</td>
<td>-0.033</td>
<td>-0.025</td>
</tr>
<tr>
<td>[-0.085,0.015]</td>
<td>[-0.087,0.014]</td>
<td>[-0.084,0.015]</td>
<td>[-0.085,0.013]</td>
<td>[-0.080,0.019]</td>
<td>[-0.078,0.023]</td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>-0.042</td>
<td>-0.048</td>
<td>-0.041</td>
<td>-0.044</td>
<td>-0.033</td>
<td>-0.025</td>
</tr>
<tr>
<td>CCEMG</td>
<td>-0.013</td>
<td>-0.003</td>
<td>-0.012</td>
<td>-0.013</td>
<td>0.001</td>
<td>-0.007</td>
</tr>
<tr>
<td>[-0.042,0.026]</td>
<td>[-0.039,0.022]</td>
<td>[-0.044,0.019]</td>
<td>[-0.045,0.018]</td>
<td>[-0.032,0.029]</td>
<td>[-0.035,0.022]</td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>-0.042</td>
<td>-0.048</td>
<td>-0.041</td>
<td>-0.044</td>
<td>-0.033</td>
<td>-0.025</td>
</tr>
<tr>
<td>CCEP</td>
<td>-0.013</td>
<td>-0.003</td>
<td>-0.012</td>
<td>-0.013</td>
<td>0.001</td>
<td>-0.007</td>
</tr>
<tr>
<td>[-0.046,0.028]</td>
<td>[-0.036,0.025]</td>
<td>[-0.049,0.014]</td>
<td>[-0.073,-0.021]</td>
<td>[-0.038,0.022]</td>
<td>[-0.034,0.024]</td>
<td></td>
</tr>
<tr>
<td>IFE</td>
<td>-0.042</td>
<td>-0.048</td>
<td>-0.041</td>
<td>-0.044</td>
<td>-0.033</td>
<td>-0.025</td>
</tr>
<tr>
<td>[-0.085,0.015]</td>
<td>[-0.087,0.014]</td>
<td>[-0.084,0.015]</td>
<td>[-0.085,0.013]</td>
<td>[-0.080,0.019]</td>
<td>[-0.078,0.023]</td>
<td></td>
</tr>
</tbody>
</table>

Each column shows the OLS, common correlated effects (CCE), and interactive fixed effects (IFE) estimates for different specifications. Column (1) is the traditional two-way fixed effects result without state-specific time trends shown in Table 1.3. Columns (2)-(6) add flexible state-specific time trends, beginning with linear state-specific time trends and extending to 5th-order polynomial state-specific time trends. IFE results are based on 4 common factors. See Table 1.3 for additional details about standard errors and inference.
Table 1.11.: State-specific trend robustness checks - teenage employment

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None</td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
<td>4th</td>
<td>5th</td>
</tr>
<tr>
<td>OLS</td>
<td>-0.178**</td>
<td>-0.074</td>
<td>-0.045</td>
<td>-0.142**</td>
<td>-0.108</td>
<td>-0.125*</td>
</tr>
<tr>
<td></td>
<td>[-0.323,-0.033]</td>
<td>[-0.194,0.045]</td>
<td>[-0.167,0.077]</td>
<td>[-0.303,-0.019]</td>
<td>[-0.275,0.059]</td>
<td>[0.270,-0.020]</td>
</tr>
<tr>
<td>CCEMG</td>
<td>-0.040</td>
<td>-0.025</td>
<td>0.029</td>
<td>-0.129</td>
<td>-0.088</td>
<td>-0.106</td>
</tr>
<tr>
<td></td>
<td>[-0.214,0.135]</td>
<td>[-0.198,0.147]</td>
<td>[-0.119,0.193]</td>
<td>[-0.321,0.075]</td>
<td>[-0.293,0.129]</td>
<td>[0.310,0.089]</td>
</tr>
<tr>
<td>CCEP</td>
<td>-0.065</td>
<td>-0.043</td>
<td>0.009</td>
<td>-0.090</td>
<td>-0.018</td>
<td>-0.065</td>
</tr>
<tr>
<td></td>
<td>[-0.191,0.061]</td>
<td>[-0.173,0.088]</td>
<td>[-0.106,0.125]</td>
<td>[-0.260,0.080]</td>
<td>[-0.192,0.155]</td>
<td>[-0.221,0.091]</td>
</tr>
<tr>
<td>IFE</td>
<td>-0.036</td>
<td>0.041</td>
<td>0.028</td>
<td>-0.079</td>
<td>-0.064</td>
<td>-0.094</td>
</tr>
<tr>
<td></td>
<td>[-0.157,0.097]</td>
<td>[-0.053,0.137]</td>
<td>[-0.074,0.128]</td>
<td>[-0.220,0.054]</td>
<td>[-0.174,0.062]</td>
<td>[-0.208,0.098]</td>
</tr>
</tbody>
</table>

IFE results are based on 8 common factors. See Table 1.10 for additional details.
Table 1.12: Sub-sample robustness checks - restaurant employment

<table>
<thead>
<tr>
<th></th>
<th>(1) 1990q1-2000q2</th>
<th>(2) 1995q2-2005q3</th>
<th>(3) 2000q3-2010q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>-0.018</td>
<td>-0.137**</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>[-0.177,0.141]</td>
<td>[-0.264,-0.010]</td>
<td>[-0.095,0.019]</td>
</tr>
<tr>
<td>CCEMG</td>
<td>0.120</td>
<td>0.088</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>[-0.335,0.095]</td>
<td>[-0.219,0.042]</td>
<td>[0.013,-0.049]</td>
</tr>
<tr>
<td>CCEP</td>
<td>-0.060**</td>
<td>-0.003</td>
<td>-0.020</td>
</tr>
<tr>
<td></td>
<td>[-0.004,-0.118]</td>
<td>[0.039,-0.043]</td>
<td>[0.015,-0.059]</td>
</tr>
<tr>
<td>IFE</td>
<td>-0.052</td>
<td>-0.015</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>[0.015,-0.013]</td>
<td>[0.056,-0.055]</td>
<td>[0.063,-0.053]</td>
</tr>
</tbody>
</table>

Each column shows the OLS, common correlated effects (CCE), and interactive effects (IFE) estimates for sub-samples of the data. The dimensions of the data in the sub-samples are $N = 1371$ and $T = 42$. IFE results are based on 4 common factors. See Table 1.3 for additional details about standard errors and inference.
Table 1.13.: Sub-sample robustness checks - teenage employment

<table>
<thead>
<tr>
<th></th>
<th>(1) 1990q1-2001q4</th>
<th>(2) 1996q1-2007q4</th>
<th>(3) 2002q1-2013q4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td>-0.205**</td>
<td>-0.102</td>
<td>-0.069</td>
</tr>
<tr>
<td></td>
<td>[-0.409,-0.001]</td>
<td>[-0.249,0.045]</td>
<td>[-0.261,0.131]</td>
</tr>
<tr>
<td><strong>CCEMG</strong></td>
<td>-0.020</td>
<td>-0.113</td>
<td>-0.045</td>
</tr>
<tr>
<td></td>
<td>[-0.519,0.485]</td>
<td>[-0.364,0.145]</td>
<td>[-0.349,0.258]</td>
</tr>
<tr>
<td><strong>CCEP</strong></td>
<td>-0.156</td>
<td>-0.107</td>
<td>-0.066</td>
</tr>
<tr>
<td></td>
<td>[-0.376,0.073]</td>
<td>[-0.310,0.112]</td>
<td>[-0.282,0.159]</td>
</tr>
<tr>
<td><strong>IFE</strong></td>
<td>-0.142</td>
<td>-0.102</td>
<td>-0.069</td>
</tr>
<tr>
<td></td>
<td>[-0.416,0.186]</td>
<td>[-0.303,0.109]</td>
<td>[-0.268,0.143]</td>
</tr>
</tbody>
</table>

Each column shows the OLS, common correlated effects (CCE), and interactive effects (IFE) estimates for sub-samples of the data. The dimensions of the data in the sub-samples are $N = 51$ and $T = 48$. IFE results are based on 8 common factors. See Table 1.3 for additional details about standard errors and inference.
### Table 1.14: Falsification test: employment and earnings minimum wage elasticity estimates for the manufacturing industry

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>(1) OLS</th>
<th>(2) CCEMG</th>
<th>(3) CCEP</th>
<th>(4) IFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment</td>
<td>-0.013</td>
<td>0.015</td>
<td>0.026</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>[-0.176,0.149]</td>
<td>[-0.028,0.051]</td>
<td>[-0.021,0.063]</td>
<td>[-0.038,0.041]</td>
</tr>
<tr>
<td>Earnings</td>
<td>-0.096</td>
<td>0.086</td>
<td>0.085**</td>
<td>-0.045</td>
</tr>
<tr>
<td></td>
<td>[-0.314,0.121]</td>
<td>[-0.078,0.235]</td>
<td>[0.013,0.161]</td>
<td>[-0.139,0.063]</td>
</tr>
</tbody>
</table>

Period fixed effects: Yes, Yes, Yes, Yes
State fixed effects: Yes, Yes, Yes, Yes

Each column uses a different estimator applied to the traditional two-way fixed effects specification shown in equation (1) for the given dependent variable. The employment dependent variable is county-quarter manufacturing employment. The earnings dependent variable is county-quarter average weekly wages for manufacturing workers. Data on county-quarter employment and earnings for manufacturing workers comes from the QCEW. Employment regressions control for total private sector employment and county population. Earnings regressions control for total private sector average weekly wages and county population. IFE results are based on 4 common factors. See Table 1.3 for additional details about standard errors and inference.
Table 1.15.: Minimum wage-employment elasticity - restaurant employment, contiguous counties sample

<table>
<thead>
<tr>
<th></th>
<th>(1) OLS</th>
<th>(2) CCEMG</th>
<th>(3) CCEP</th>
<th>(4) IFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Minimum wage)</td>
<td>-0.106</td>
<td>-0.025</td>
<td>-0.011</td>
<td>-0.034</td>
</tr>
<tr>
<td></td>
<td>[-0.236,0.024]</td>
<td>[-0.088,0.038]</td>
<td>[-0.056,0.034]</td>
<td>[-0.090,0.024]</td>
</tr>
<tr>
<td>log(Total private sector emp.)</td>
<td>0.288***</td>
<td>0.600***</td>
<td>0.468***</td>
<td>0.425***</td>
</tr>
<tr>
<td></td>
<td>[0.198,0.379]</td>
<td>[0.533,0.667]</td>
<td>[0.370,0.566]</td>
<td>[0.292,0.548]</td>
</tr>
<tr>
<td>log(Population)</td>
<td>0.778***</td>
<td>0.224</td>
<td>0.476***</td>
<td>0.181</td>
</tr>
<tr>
<td></td>
<td>[0.612,0.944]</td>
<td>[-0.084,0.568]</td>
<td>[0.239,0.713]</td>
<td>[-0.117,0.442]</td>
</tr>
<tr>
<td>Period fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>County fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CD test statistic</td>
<td>7.15***</td>
<td>1.15</td>
<td>2.22**</td>
<td>-1.25</td>
</tr>
<tr>
<td>$T_xN$</td>
<td>42,000</td>
<td>42,000</td>
<td>42,000</td>
<td>42,000</td>
</tr>
</tbody>
</table>

Results correspond to the results in Table 1.3, except based on the contiguous county sample from [Dube et al., 2010].
Table 1.16.: Minimum wage-employment elasticity simulation results - no factors in DGP

<table>
<thead>
<tr>
<th></th>
<th>(1) Rest. Employment</th>
<th>(2)</th>
<th>(3)</th>
<th>(4) Teen. Employment</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median 2.5% 97.5%</td>
<td>Median 2.5%</td>
<td>97.5%</td>
<td>Median 2.5% 97.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True value</td>
<td>-0.138</td>
<td></td>
<td></td>
<td>-0.178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.139 -0.152 -0.128</td>
<td>-0.177 -0.223 -0.134</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCEMG</td>
<td>-0.139 -0.167 -0.111</td>
<td>-0.179 -0.248 -0.107</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCEP</td>
<td>-0.138 -0.159 -0.120</td>
<td>-0.177 -0.234 -0.124</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFE</td>
<td>-0.138 -0.153 -0.124</td>
<td>-0.177 -0.225 -0.131</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table reports simulation results for the case without common factors in the data generating process. The DGP is \( y_{it} = \beta \ln(MW_{it}) + X_{it} \Gamma + \alpha_i + \delta_t + v_{it} \), where the independent variables are the same variables used in the results section, the parameters are from the OLS results for the traditional two-way period and location fixed effects specification reported in Table 1.3 and Table 1.4, and \( v_{it} \) is an idiosyncratic error term whose variance is equal to the variance of the OLS residuals. The number of repetitions is 1,000.
Table 1.17.: Minimum wage-employment elasticity simulation results - factors in DGP

<table>
<thead>
<tr>
<th></th>
<th>Restaurant Employment</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>2.5%</td>
<td>97.5%</td>
</tr>
<tr>
<td><strong>True value</strong></td>
<td>-0.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>-0.175</td>
<td>-0.203</td>
<td>-0.135</td>
</tr>
<tr>
<td>CCEMG</td>
<td>-0.045</td>
<td>-0.092</td>
<td>0.007</td>
</tr>
<tr>
<td>CCEP</td>
<td>-0.041</td>
<td>-0.087</td>
<td>-0.002</td>
</tr>
<tr>
<td>IFE</td>
<td>-0.043</td>
<td>-0.072</td>
<td>-0.011</td>
</tr>
</tbody>
</table>

This table reports simulation results for the case with common factors in the data generating process. The DGP is $y_{it} = \hat{\beta}ln(MW_{it}) + X_{it}\hat{\Gamma} + \hat{\alpha}_i + \hat{\delta}_t + \hat{\chi}_i f_t + v_{it}$, where the independent variables are the same variables used in the results section, the parameters are from the IFE results for the traditional two-way period and location fixed effects specification reported in Table 1.3 and Table 1.4, and $v_{it}$ is an idiosyncratic error term whose variance is equal to the variance of the IFE residuals. The number of repetitions is 1,000.
CHAPTER 2. THE LASTING IMPACT OF HIGH SCHOOL VALUE-ADDED

2.1 Introduction

Many states and school districts have recently begun using value-added models to evaluate teachers and schools\(^1\). These models are intended to estimate the impact that teachers and schools have on student test scores. This typically involves estimating the gains or losses in terms of normalized test scores that students experience while under the instruction of a particular teacher or while in enrolled in a particular school. Some states and school districts have tied raises, tenure, and school funding to value-added scores. Some also release these value-added scores to the public\(^2\).

The use of value-added models to evaluate teachers and schools has caused two debates. The first is whether value-added models provide unbiased estimates of the impact that teachers and schools have on student test scores. The second is a public policy debate of whether these value-added models should actually be a part of the evaluation process for teachers and schools. Much work has been done on the first debate. A number of papers have raised concerns that value-added models may be biased by an inability to capture the sorting of students into classrooms and schools based on unobservable characteristics that are correlated with test scores [Baker et al., 2010, Rothstein, 2010, Paufler and Amrein-Beardsley, 2014]. However, there is lots of evidence and a growing consensus that modern value-added models do a good job of capturing these unobservable characteristics and providing unbiased value-added estimates [Kane and Staiger, 2008, Kane et al., 2013, Deming, 2014, Chetty et al., 014a, Koedel et al., 2015].

---

\(^1\)See [Hershberg, 2015] for an overview of which states rely on some form of value-added models for teacher and school evaluation and how this assessment is implemented.

\(^2\)In Tennessee, for example: http://www.tn.gov/education/topic/report-card.
The second debate has not approached a consensus and is still a very divisive topic. One critique against the use of value-added models to evaluate schools and teachers is that value-added estimates may be too unstable to be used for high-stakes personnel decisions [Baker et al., 2010, Newton et al., 2010]. Another important question related to the public policy debate is whether value-added estimates capture lasting impacts that teachers and schools are having on students. Are teachers and schools that have high value-added estimates actually improving the cognitive ability of their students, or are they simply better at teaching to the test? The former would be a valuable argument for public policy use of value-added estimates. Alternatively, if higher value-added teachers and schools are simply better at teaching to the test and have no lasting impact on student outcomes, then their appropriateness as a way of evaluating teachers and schools should become even more heavily scrutinized.

I address this second debate by first using K-12 test score data for students in Florida and North Carolina to estimate value-added scores for high schools in these two states. I use value-added models that have been shown to provide unbiased estimates of value-added scores in prior work. I confirm the performance of these models by including tests which suggest that the value-added models appear to sufficiently capture the sorting of students into high schools based on unobservable characteristics that are correlated with test scores. I then link these value-added scores for Florida and North Carolina high schools to a second dataset on college performance. This dataset includes detailed college transcript data as well as some background data on the student, including their high school. This allows me to determine the impact of attending a higher value-added high school on college performance.

The only other paper that analyzes the lasting impact of value-added on adult outcomes is [Chetty et al., 014b]. They study the lasting impact of value-added measured at the teacher level in grades 4-8 and find statistically significant positive effects of having a higher value-added teacher on college attendance, college quality, and early adulthood earnings. This study is the most similar to mine because it analyzes the lasting impact of value-added on adult outcomes, but there are also
several differences. First, their study analyzed the lasting impact of value-added during grades 4-8, whereas I am studying the lasting impact of value-added that occurs during high school. Studies have found that the earliest years of a student’s schooling are the most influential [Cunha and Heckman, 2010]. This suggests that value-added that occurs during high school may have less impact on a student’s adult outcomes. The results from [Chetty et al., 014b] therefore do not answer the question of how value-added that occurs during high school affects adult outcomes.

A second difference is that their study analyzed the lasting impact of being in a high value-added teacher’s classroom, whereas this study analyzes the lasting impact of attending a high value-added school. This difference is an important distinction between their study and mine and, more generally, an important addition to the value-added literature. Several states and school districts evaluate schools using value-added methods, just as they do for teachers, and attach school-level funding to value-added scores. Additionally, teachers without test score data, such as preschool and special education teachers, in some cases have their school’s value-added score substituted into their evaluation. Yet, the value-added literature focuses primarily on teacher value-added. Finally, my paper uses adult outcome variables that were not studied in [Chetty et al., 014b]: college GPA and graduation.

I find that higher value-added high schools have positive and statistically significant impacts on the future college GPA of their students. The effect is largest in earlier semesters. A one standard deviation increase in high school value-added increases first-semester GPA by 0.022 points. The impact becomes smaller and eventually insignificant for later semesters, which is consistent with the value-added model framework. Ultimately, a one standard deviation increase in high school value-added increases final GPA by 0.017 points, which is approximately a one letter grade increase over the course of a college career. The impact on graduation is not statistically significant. While these results are somewhat small in magnitude, one of the identifying assumptions is that higher value-added high schools do not cause selection into or out of the universities in the sample. In reality, higher value-added high schools do likely
push lower-ability students into the sample of universities and push higher-ability students out of the sample and into better universities. In fact, as mentioned above, [Chetty et al., 014b] find evidence that higher value-added teachers do in fact increase both college attendance and university quality. This would cause negative bias in the results in this paper. Therefore, these estimates likely provide a lower-bound on the impact of attending a higher value-added high school on college performance. I also find that these results are generally consistent across race and gender, although the impact is somewhat larger for White students. Additionally, quantile regressions suggest that the lasting impact on GPA is larger for students on the lower-end of the ability distribution.

The remainder of the paper is organized as follows: Section 2 discusses the conceptual framework that modern value-added models are based on. Section 3 describes how the value-added scores are estimated and how I estimate the impact of attending a higher value-added high school on college performance. Section 4 describes the datasets used. Section 5 tests some of the assumptions of the models. Section 6 presents and discusses the results. Section 7 concludes.

2.2 Model background

Value-added modeling typically involves linear regression analysis which predicts student test score performance based on a combination on prior achievement and demographics. The details for the specification and estimation of value-added scores and their impact on college outcomes will be discussed in Section 3. First, I discuss a statistical framework for student achievement adopted from [Todd and Wolpin, 2003]. Under some conditions discussed below, the linear regression value-added models can be derived from this framework.
Student achievement in a given year is a function of school inputs and other factors:

\[ A_i[t] = A[S_i(t), F_i(t), \alpha_0, \varepsilon_i]. \] (2.1)

\( A_i[t] \) is a measure of achievement for student \( i \) at time \( t \), which is a function of the history of family inputs \((F_i(t))\), school inputs \((S_i(t))\), ability \((\alpha_0)\), and an idiosyncratic error term \((\varepsilon_i)\), where \( S_i/F_i(t) \) represents a detailed history from each year until time \( t \). The intuition behind the common value-added models is that lagged achievement measures can be a sufficient statistic for detailed input histories and ability. In this setting, the achievement production function can be rewritten as a function of contemporaneous inputs and a baseline test score:

\[ A_{ijt} = A[s_{ij(t)}, f_{it}, A_{i,t-1}[S_i(t-1), F_i(t-1), \alpha_0], \varepsilon_{ijt}]. \] (2.2)

which models current academic achievement as a function of current school inputs \((s_{ij(t)})\), which is a measure of school quality for school \( j \) that student \( i \) attends in year \( t \), current family inputs \((f_{it})\), lagged achievement \((A_{i,t-1})\), and an idiosyncratic error term\(^3\).

Applied value-added work traditionally assumes that the arguments in (2) are additively separable, which leads to the equation:

\[ A_{ijt} = s_{ij(t)} + \alpha f_{it} + \gamma A_{i,t-1} + \varepsilon_{ijt}. \] (2.3)

Difficult to measure prior input histories and unobservable individual ability are not in (3) because they are, in principle, captured in the lagged achievement score. Applied value-added models typically make one additional assumption, which is that current

---

\(^3\)Here, school quality is indexed by \((i, t)\) to indicate that it is the school that student \( i \) attended in year \( t \). Below, when I estimate the impact of high school quality on college performance, I use the average value-added score for each school during the entire time frame; I do not use within-school variation in value-added scores. Thus, in these regressions, I index school quality only by \((i)\) to indicate which high school the college student attended.
family inputs such as parental effort do not respond to contemporaneous school inputs, in which case the model of student achievement is further simplified to:

$$A_{ijt} = s_{j(i,t)} + \gamma A_{i,t-1} + \varepsilon_{ijt}$$  

which is the basis for applied value-added work. There are a variety of estimation methods that are often used to recover the school inputs from this model. These methods will be discussed in the next section.

Several conditions are required in order to link the value-added model in equation (4) to the underlying structural cumulative achievement model in (1). Among them are grade invariance in the education production function, geometric decay in the impact of prior inputs, and geometric decay in the impact of unobservable individual ability that is equal to the rate of geometric decay for prior inputs. A detailed discussion of these assumptions can be found in [Todd and Wolpin, 2003] and [Sass et al., 2014]. [Sass et al., 2014] show that these conditions are usually not met. However, as they point out, the failure of typical value-added models to link to a structural interpretation says little about the informational value of the value-added measures. Ultimately, the extent to which value-added measures capture useful information about school inputs is an empirical question. Another assumption of this model that is standard in applied value-added work is that school quality is time-invariant, conditional on the roster and experience of the teachers and administrators in the school. This rules out the possibility that school value-added fluctuates from year-to-year for reasons unrelated to school faculty composition effects or that it depends on characteristics of the students.

Before moving to the college achievement function, it should be mentioned that much of the applied work on value-added is concerned with value-added from teachers, rather than entire schools. Thus, in equation (4), ‘school inputs’ becomes a measure of teachers’ impacts on test scores. However, the focus in this paper is on measuring school-level value-added and the ‘school inputs’ will be a measure of the school’s impact on test scores. The use of school-level value-added is a necessity, given that
the college outcome data only allows me to link students to high schools, rather than teachers. But it is not a concern; the use of school value-added, rather than teacher value-added, is likely most appropriate at the high school level. Within any school, there will be variation in teacher quality. For earlier grades, where students often only have instruction from a single teacher, the entire school’s value-added score would be a very noisy measure of the value-added that each individual student receives. In high school, however, students take courses with many different instructors; over the course of their high school career, students will likely receive instruction from a large fraction of the school’s teachers. This makes the use of teacher value-added less appropriate, as a student will often take only one course from a specific teacher in their entire high school career. When only school effects are included without teacher effects, the school effect is likely made up of an accumulation of teacher effects, but they also likely capture the effects of the administration, leadership, and polices associated with the school itself. In this sense, the interpretation for a school value-added score can be seen as the accumulated value-added that the student receives from all of the various teachers, administrators, and policies associated with the school.

The model for college achievement is analogous to the model for academic achievement shown in equation (1). College achievement is a function of school inputs, family inputs, ability, and an idiosyncratic error term:

$$A_{it}^C = A^C[S_i(t), F_i(t), \alpha_0^C, \nu_{it}^C].$$

(2.5)

The significant difference between this equation and that in (1) is that there is a new ability term, $\alpha_0^C$, that is specific to college outcomes. The two models are allowed to have different abilities because of the difference in outcome measures between the high school value-added achievement models and the college achievement outcomes I will be using. As is the tradition in the high school value-added literature, I will be estimating the high school value-added model by using data on standardized test scores as outcome variables. The college outcome variables will be GPA and graduation. Some of the unobserved ability that impacts standardized test scores, such as
cognitive ability, will also impact college GPA and graduation; both involve cognitive skills. Therefore, $\alpha_{i0}$ and $\alpha_{i0}^C$ may be correlated. However, GPA and graduation likely require other types of ability as well, such as non-cognitive skills like motivation, adaptability, self-restraint, and persistence, which are less important in determining standardized test scores.

I rely on the same assumptions discussed above for the high school value-added model to rewrite (5) as a linear function of high school quality and a proxy that is a sufficient statistic for detailed school and family histories and ability:

$$A_{ijt}^C = \beta s_{j(i)} + \lambda A_{i,t-1}^C + \nu_{ijt}^C. \quad (2.6)$$

Details on the proxy and the interpretation of $\beta$ will be discussed in section 3.2.

2.3 Methodology

2.3.1 Estimating the impact of high schools on test scores

I first focus on estimating the impact of high schools on test scores. The specification used to estimate the impact of high schools corresponds to equation (4) from the previous section. Relying on the assumptions described in the previous section that family inputs are not endogenously determined and that the lagged test score is a sufficient statistic for input histories and unobserved ability, I can estimate the impact of high schools on test scores. To do this, I begin by estimating the following equation:

$$A_{ijt} = \gamma f(A_{i,t-1}) + \phi X_{ijt} + \epsilon_{ijt} \quad (2.7)$$

$$\epsilon_{ijt} = s_{j(i,t)} + u_{it}, \quad (2.8)$$
where \( f(A_{i,t-1}) \) is a control function of lagged test scores that includes a cubic in both 8th grade reading and math scores and \( X_{it} \) is a vector of controls for the student’s gender, ethnicity, free/reduced lunch status, disability status, and the grade in which the student took the high school exam\(^4\). The error term is decomposed into a measure of school quality \( (s_j(i,t)) \) and an idiosyncratic error term \( (u_{it}) \). This equation is directly comparable to the one presented in equation (4), but with student characteristics included to help capture the unobserved student ability, \( \alpha_{it0} \), from the structural model. Together with the lagged test scores, these controls should capture most the individual student’s ability and history of school and family inputs. The idiosyncratic error term captures heterogeneity that is orthogonal to lagged test scores and individual characteristics.

The identification assumption of the model is that students are not sorted into schools based on unobservables determinants of test scores. If this assumption fails, then the estimate of school quality will capture both school quality and the correlation between school quality and unobserved determinants of test scores. Note, of course, that this assumption allows for sorting of students into schools. Sorting into high schools based on characteristics such as ability and family characteristics definitely happens. The identification assumption just requires that the lagged test scores and observable characteristics as sufficient to capture the unobserved characteristics on which the sorting occurs. Section 5.1 will discuss a test of this assumption.

Many different methods have been used to estimate value-added models like the one presented in (7)-(8) [Koedel et al., 2015]. One methodology selection decision is the use of gain-score versus lagged-score models. The model presented in (7)-(8) is a lagged-score model, while a gain-score model would be one where the the lagged-score was moved to the left-hand side of the equation such that the dependent variable becomes \( (A_{it} - A_{i,t-1}) \). Thus, the gains-score model imposes the restriction that the parameter on the lagged-score in equation (7) is equal to one. A benefit of the gains-

\(^4\)I control for the grade in which the high school exam occurred because not all high school students take high school end-of-course exams in the same grade and the timing of the high school exam could impact test scores [Clotfelter et al., 2012].
score model is that it avoids the complication of having an independent variable that is measured with error. However, the current consensus is that lagged-score models outperform the gain-score specification [Koedel et al., 2015]. Another methodological issue is the use of one-step versus two-step models. The model presented above, which includes the school effects in the residual, is a two-step model, while a model with fixed effects for each school would be a one-step model. There is not a preferred specification between these two options in the literature. Reported correlations between the two have been shown to be as high as 0.98 [Chetty et al., 014a].

I use a two-step lagged-score value-added approach that averages school value-added scores across years with a shrinkage step. This is the most commonly used value-added approach and has been shown to produce unbiased estimates of value-added scores [Chetty et al., 014a, Deming, 2014]5. The mechanics of estimating the value-added scores for schools is as follows:

Step 1: Estimate equation (7), recover the residuals, $\hat{\epsilon}_{ijt}$, and calculate the average residual for each school, $\bar{\epsilon}_{jt}$. This step is performed separately for each subject and year of high school test scores, such that each school has an average residual for each subject and each year in the data, $\bar{\epsilon}_{jkt}$, where $k$ is the subject.

Step 2: Shrink the school-subject-year effect estimates. It is common in applied value-added work to apply a Bayesian shrinkage factor, which shrinks the estimates of teacher/school value-added towards zero based on how noisy the value-added score is for a given teacher/school [Kane and Staiger, 2008, Chetty et al., 014a]. Following this tradition, I apply a shrinkage estimator, but I apply a more general version that also incorporates information across subjects for a given school. As discussed in [Lefgren and Sims, 2012], in a setting where one has value-added scores across multiple subjects for a given school, the best guess of a school’s ability to increase Algebra 1 scores should incorporate information regarding the school’s ability to increase

---

5The consensus of averaging value-added scores across years is based on the finding from many studies that value-added estimates are less biased when they are averaged over many years [McCaffrey et al., 2009, McCaffrey et al., 2010, Koedel and Betts, 2011, Deming, 2014]; much of the within-school/teacher variation over time is noise [Kane and Staiger, 2002]. Average value-added scores therefore leverage a larger degree of true variation in school quality.
Biology scores. This shrinkage can be performed by ordinary least squares (OLS) regressions of a school’s value-added score for one subject-year on it’s value-added scores for all other subject-year value-added scores:

$$\tilde{\epsilon}_{jkt} = \pi_0 + \pi_1 \bar{\epsilon}_{j,1,-t} + \pi_2 \bar{\epsilon}_{j,2,-t} + \ldots + \pi_K \bar{\epsilon}_{j,K,-t} + \mu_{jkt}$$

The coefficients in the regression represent the weights from the theoretical shrinkage function from equation (7) in [Lefgren and Sims, 2012]. The intuition for this procedure is as follows. If the value-added scores for a particular school are imprecisely measured and noisy, then there will be less correlation over time and subjects. This will cause estimates of $\pi_p^k$ that are close to zero and will therefore shrink the estimated value-added score for that school toward zero. Including separate regressors for each subject and for each distance from the current year’s value-added score allows for a flexible shrinkage function in which some subjects and some years can have larger weight than others. The predicted value for each school-subject-year serves as the post-shrinkage estimate of the school-subject-year value-added scores.

**Step 3:** Calculate the average school effects. This is done by averaging the post-shrinkage residuals across subjects and years for each school. Letting $\tilde{\epsilon}_{jkt}$ be the post-shrinkage predicted school-subject-year estimate for school $j$ in subject $k$ and year $t$, the average school value-added is:

$$\hat{s}_j = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{K} \sum_{k=1}^{K} \tilde{\epsilon}_{jkt}.$$ 

The test scores for each subject are normalized by state and year to mean 0 and standard deviation 1 before being used in the value-added models. Therefore, the value-added estimates, $\hat{s}_j$, are scaled in units of student test score standard deviations. For example, if a high school had an average value-added estimate of 0.20, that would mean that the high school increases it’s students test scores by 0.20 standard deviations, on average.
2.3.2 Estimating the impact of high schools on college performance

The primary purpose of this paper is to better understand whether value-added scores that are commonly used to evaluate schools and teachers are capturing a lasting impact that schools and teachers have on students. In order to answer this question, I include the value-added score for each student’s high school as an independent variable in regressions for college performance:

\[ A_{ijt}^C = \beta \hat{s}_{j(i)} + \lambda A_{i,t-1} + \phi_1 X_{it} + \phi_2 \bar{X}_{j(it)} + v_{ijt}, \quad (2.9) \]

where \( A_{ijt}^C \) is the college outcome variable (either GPA or graduation), \( \hat{s}_{j(i)} \) is the value-added score for the high school that student \( i \) attended, \( A_{i,t-1} \) is the lagged score that proxies for the history of school and family inputs and unobserved ability, \( X_{it} \) is a vector of controls for the student’s race, gender and age, and \( \bar{X}_{j(it)} \) is a vector of controls for the student’s high school’s mean fraction male, fraction Black, fraction Asian, fraction other race, fraction on free-or-reduced lunch, and mean incoming test scores. This equation is directly comparable to equation (6), but with controls for student characteristics and mean characteristics for the student’s high school added to help capture unobserved student ability from the college achievement model presented in section 2. The analysis in section 4 will also include major and university fixed effects in order to control for differences in difficulty across majors and universities. The coefficient \( \beta \) in this regression captures the correlation between high school value-added and college performance.

I use high school GPA as the lagged score that is meant to proxy for historical school and family inputs and unobserved ability. The unobserved ability that influences high school GPA is likely correlated with the unobserved ability that influences college GPA. The ideal lagged variable would be a measure of GPA before entering high school. However, I cannot link the students in the college achievement dataset to the K-12 dataset. For each student in the college outcome dataset, the only pre-college information I have are the high school that the student attended, ACT and
SAT scores, and high school GPA and rank. The inclusion of high school GPA would be problematic if high schools that raise their students’ test scores also systematically raise their students’ GPA. Then, the impact of value-added in equation (9) could get absorbed into the high school GPA variable. However, this is not a significant concern because, while standardized test scores are intended to measure a student’s cognitive skills, high school GPA may be largely determined by non-cognitive skills like motivation, adaptability, self-restraint, and persistence. If I had GPA data in addition to the test score data for K-12 students, then it would be possible to test whether high schools that are good at raising test scores also raise GPAs, but I do not have this data. However, [Jackson, 2012] has shown that the ability of a teacher to raise test scores shows little correlation with their ability to raise non-cognitive skills. Section 4.3 will provide some evidence that high school GPA is not impacted by a school’s ability to raise test scores.

There are two important assumptions for identification of equation (9). The first is the usual assumption that there is no correlation between student unobservables and the school value-added scores. This allows for there to be sorting of students into schools based on ability, but high schools with better students cannot have higher value-added scores, conditional on student and high school characteristics concluded in the regressions. Section 5.1 will show evidence that value-added scores are in fact uncorrelated with unobservable determinants of test scores. The second important assumption is that there is no sorting into or out of the universities in the college achievement dataset based on value-added scores. If high value-added schools changed the ability distribution of students in the sample by pushing students that would typically attend the universities in the sample to better universities and pushing worse ability students into the sample of universities, then that would cause negative bias in the estimated impact of attending a higher value-added high school on college performance. This assumption will be tested in section 5.2.

An alternative approach to the one discussed above would be to estimate an equation analogous to (7)-(8), where I estimate the impact of high schools on college
performance directly. This would produce a college performance value-added score for each high school, rather than estimating the lasting impact of attending a higher test score value-added school. I choose not to take this approach for two reasons. First, estimating the lasting impact of attending a high school that is good at raising test scores more closely relates to the public policy question of whether high school value-added scores are capturing anything that has a lasting impact on students. The second reason is that estimating the impact of high schools on college performance directly requires a stronger assumption; it requires that there is no correlation between unobservable determinants of college performance and the high school that the student attended, conditional on high school GPA and other covariates, while estimating the impact of high school value-added on college performance only requires that there is no correlation between unobservables and the high school value-added scores.

2.4 Data

I study the lasting impact of attending a higher value-added high school by linking two K-12 student-level datasets with a dataset on college performance at several universities in the United States. The datasets cannot be linked at the student-level; they can only be linked at the high school-level. Therefore, I use the two K-12 student-level datasets to estimate value-added scores for all high schools in the datasets. Then, I match these value-added scores to the dataset on college performance by matching the high school that each college student attended to the estimated value-added score for that high school.

2.4.1 K-12 test score data

The two K-12 student-level datasets that I use to measure high school value-added scores come from the North Carolina Education Research Data Center at Duke University and the PK20 Education Data Warehouse at the Florida Department of
Education. North Carolina students in grades 3-8 take annual math and reading standardized tests. High school students have end-of-course exams in Algebra 1, Algebra 2, English 1, Biology, Chemistry, and History. The dataset also contains information on ethnicity, gender, date of birth, receipt of special education services, limited English proficiency, and receipt of free or reduced lunch. The dataset contains information on 3.1 million students over grades 3-12 from 1997-2012. Florida students in grades 8 and 10 take end-of-course exams in Math, Reading, Writing, and Science. The dataset also contains information on ethnicity, gender, date of birth, receipt of special education services, limited English proficiency, and receipt of free or reduced lunch. This dataset contains information on students from 1997-2010.

As described in the previous section, I use this data to estimate high school value-added scores for each high school. The value-added models are estimated separately for each state. The student demographic covariates included in the value-added model are the ones listed in the previous paragraph. I also include a control variable for the grade in which the student took the high school exam. All test scores are normalized by subject, grade, and year before estimation. After estimating the value-added scores, I match these value-added scores, as well as other high school-level variables, to the dataset on college performance by matching on high school.

2.4.2 College performance data

The dataset on college performance comes from the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD). MIDFIELD contains information on 1,248,363 college students at 11 different universities in the United States from 1987-2005. Of the 11 universities in the dataset, 6 are from either Florida or North Carolina. This dataset includes detailed student-level information including start date, final date, graduation, GPA by semester, final GPA, majors, transfers, ACT scores, SAT scores, high school GPA, high school class rank, high

\footnote{Chemistry and History in North Carolina and Science in Florida are not included in the value-added estimates because these exams are only in the data for a small fraction of the years.}
school ID, ethnicity, gender, and age. I use the high school IDs to match value-added scores to the high schools in the data. I also match other high school-level variables such as percent of students on free or reduced lunch, gender and ethnic characteristics, and mean incoming 8th grade normalized test scores. The percent of students receiving free or reduced lunches and the racial composition of the school serve as measures of the school’s social and economic setting. The mean incoming test scores help capture sorting of students into schools based on unobservable characteristics that are correlated with test scores.

2.4.3 Dataset of analysis

Summary statistics for the linked dataset of analysis are shown in table 3.1. Of the 1,248,363 students in the MIDFIELD dataset, 103,250 are from high schools in Florida or North Carolina that could be matched to a value-added score. Table 3.1 shows summary statistics for the primary outcome variables, college GPA and graduation, as well as gender, ethnicity, high school GPA, SAT and ACT scores, and the value-added score and mean characteristics of the student’s high school. The summary statistics are shown for all students in the data and separately according to whether the student is from a high school that has a value-added score above the median value or below the median value.

The average final cumulative GPA for students in the linked dataset of analysis is 2.34 and 42 percent graduated. Students in the linked sample were 53% female, 27% Black, 6% Hispanic, and 4% Asian. The average starting age was 18.508 years.

7Despite the use of the word engineering in the title, MIDFIELD collects data on students from all majors.
8The sample size for the graduated variable is smaller than the sample size for college GPA because the sample for graduation is limited to students that enrolled at least 5 years before MIDFIELD stopped tracking students in 2006. This gives all students in the regressions for graduation time to graduate. A similar issue arises when looking at final cumulative GPA. Therefore, in GPA regressions later in the paper, I restrict the sample based on how many years the student is in the data.
9In the analysis below, I limit the sample to traditional college students than enroll as freshmen in the fall. This removes transfer students and early enrollees. Excluding these students does not significantly change the results.
The mean high school value-added score is 0.003 for all observations with a standard deviation of 0.071\(^{10}\). The majority of students in the sample come from one of six different universities in the MIDFIELD dataset that are located in either Florida or North Carolina: the University of Florida, North Carolina State University, Florida State University, the University of North Carolina at Charlotte, North Carolina A&T University, and Florida A&M University. Other universities in the sample include Georgia Tech, Virginia Tech, the University of Colorado, Purdue University, and Clemson University. In the analysis below, I pool students from North Carolina and Florida high school students together. Results are similar if I run the analysis separately for students from each state or control for which state the student is from.

One other variable worth discussing is the high school GPA variable. As discussed in section 3.2, I use high school GPA as a proxy for unobserved ability and detailed input histories in the main analysis below. This is potentially problematic if high schools that are good at raising test scores also tend to have students with higher GPAs. However, looking in table 3.1, we see that students from higher value-added high schools have almost exactly the same average high school GPA as students from lower value-added high schools; the average GPA is 3.55 for higher value-added high schools and 3.54 for lower value-added high schools. This is very different than what we see for ACT and SAT scores, which is strictly a measure of cognitive skills; students from high value-added high schools have higher ACT and SAT scores. This supports the use of high school GPA as a proxy for student ability in the analysis below because high school GPA appears to be a measure of non-cognitive skills, rather than cognitive skills, so that it is not an outcome variable of attending a higher value-added high school.

\(^{10}\)These means and standard deviations are based on the full dataset of analysis, in which most high schools are linked to many students and thus contribute to the mean and standard deviation multiple times. The actual mean and standard deviation of the value-added scores for high schools linked to students in the MIDFIELD data are -0.02 and 0.077.
2.5 Testing the assumptions for the analysis

2.5.1 Correlation between high school value-added and unobservables

As discussed in section 3, the identification assumption of the value-added model is that high school value-added scores are uncorrelated with unobservable determinants of test scores. I test this assumption by estimating the correlation between the high school value-added scores and a variable that is typically left out of value-added models: prior test scores.

The value-added models control for once-lagged test scores, $A_{i,t-1}$, so they cannot be used to test for selection. But twice-lagged scores, $A_{i,t-2}$, can be used. As explained in section 3, the value-added model allows for there to be sorting into schools based on unobservable determinants of test scores, such as family inputs and ability, but the value-added model requires that lagged test scores and other observable characteristics sufficiently capture the unobservable characteristics on which sorting occurs. If this holds, then high school value-added scores should be uncorrelated with 7th grade test scores. Similarly, adding 7th grade test scores to the value-added model should not significantly change the value-added scores.

To test this assumption, I regress 7th grade test scores in reading and math on the average value-added score of the high school that the 7th grader would eventually attend. I include student-level controls for gender, race, free or reduced lunch, and a cubic in 8th grade math and reading scores and include high school controls for mean demographics, mean free or reduced lunch receipt, and a cubic in mean 8th grade math and reading scores. Year fixed effects are also included\(^{11}\). If 7th grade test scores appear to be impacted by the value-added of their future high school, then this would be strong evidence that the value-added model is not sufficiently capturing the unobservable characteristics on which students sort into high schools. Indeed, the tendency for teacher/school value-added scores to significantly impact lagged test scores has been one of the main critiques of value-added models [Rothstein, 2010].

\(^{11}\)This test is similar to the one used in [Chetty et al., 014a].
Tables 2.2 and 2.3 show the results of this test for the application of the value-added model to the North Carolina and Florida test score data, respectively. For each table, the first several columns show the impact of attending a higher value-added high school on each of the high school-level test scores used to estimate the value-added scores. The tables then show the impact of attending a higher value-added high school on 7th grade reading and math scores. Finally, the last columns show the impact of value-added on high school test scores with the 7th grade reading and math scores added as additional controls.

In both table 2.2 and table 2.3, attending a higher value-added high school has a positive and statistically significant impact on high school exam scores. As described in section 3, the value-added scores are scaled in units of student test score standard deviations. Thus, the interpretation for the coefficient of 1.47 for the effect of value-added on Algebra 1 scores in table 2.2 is that high schools that increase student test scores by an average of 1 standard deviation increase their students’ Algebra 1 scores, specifically, by 1.47 standard deviations.

Table 2.2 and table 2.3 both show that attending a higher value-added high school has no significant impact on 7th grade reading and math scores. For North Carolina, the coefficients are very close to zero and not statistically significant. For Florida, the coefficients on value-added are statistically significant, but they are still very small and they are actually negative. This suggests that higher value-added high schools might have slightly lower ability students. Thus suggests that there may be negative bias when estimating the effect of attending a higher value-added high school on college outcomes. Nonetheless, the results support the assumption that once-lagged test scores and other covariates are good proxies for unobservable characteristics that are correlated with test scores and the sorting of students into high schools.

The last columns of table 2.2 and table 2.3 show how the estimated effect of attending a higher value-added high school on high school test scores changes when 7th grade test scores are added to the regression as independent variables. The results show that, despite the fact that 7th grade test scores are strong predictors of high
school test scores, the coefficient for the high school’s value-added score is essentially unchanged. This again suggests that the once-lagged test scores and other covariates sufficiently capture unobservable student characteristics that are correlated with test scores and the sorting of students into high schools.

2.5.2 Selection into or out of universities in the sample

An important assumption required in order to identify the impact of attending a higher value-added high school on college performance is that there is no change in the underlying distribution of student ability in the dataset of analysis as a result of attending a higher value-added high school. I only observe students that select into the schools in the dataset. Therefore, if higher value-added high schools have positive transcript effects for students, then there could be negative selection bias caused by higher value-added high schools pushing low ability students into the schools in the sample and pushing high ability students out of the schools in the sample and into better schools. This is a difficult assumption to test, but I am able to provide some indirect evidence on the assumption by testing for selection across universities within the sample (rather than selection into or out of the sample) and by testing the correlation between the number of students in the sample from each school and the school’s value-added score.

Results for these selection tests are shown in table 2.4. Column (2) tests the effect of value-added on selection into or out of Historically Black Colleges and Universities (HBCUs) for Black students in the dataset. Florida A&M and North Carolina A&T are HBCUs. I look at the effect first by pooling all observations and then by separating the effect by high school type, according to their mean incoming 8th grade test scores. This is done in an attempt to separate the effect of value-added on attending an HBCU by student ability, as students from high schools with higher mean incoming 8th grade test scores should, on average, be of higher ability. The numbers in the table represent marginal effects from a probit model with student controls for age, gender, and race,
mean high school demographics and incoming test scores, and year fixed effects. The results show some evidence of an impact of value-added on the sorting of students into or out of HBCUs. Overall, students from higher value-added schools are less likely to attend HBCUs, although the effect is only marginally significant. This effect appears to be primarily occurring for the sample of schools with lower incoming test scores.

Column (3) looks at the effect of value-added on selection across universities in the sample from North Carolina. One could argue that there is a somewhat natural ranking of the North Carolina schools in the dataset in terms of quality: (1) North Carolina State University (2) UNC Charlotte and (3) North Carolina A&T. I therefore estimate an ordered logit of this quality ranking on value-added scores, including student controls for age, gender, and race, mean high school demographics and incoming test scores, and year fixed effects. There is again some evidence of selection occurring here, as the coefficient for value-added is positive and statistically significant, suggesting that students from higher value-added high schools attend better schools in the sample. This effect is primarily occurring within the sample of high schools with higher incoming test scores.

Column (4) looks at the effect of a school’s value-added score on the number of students in the sample from that high school. This specification attempts to actually test selection into and out of the sample of universities in the dataset by exploiting the idea that students from high schools with low mean incoming 8th grade test scores will, on average, be of relatively low ability, while students from high schools with high mean incoming 8th grade test scores will, on average, be of relatively high ability. If there is selection bias into and out of the sample, it would be caused by higher value-added schools pushing low ability students into the sample of schools in the dataset and pushing high ability students out of the sample of schools. One therefore might expect to find positive effects of a high school’s value-added on the number of students from that high school in the dataset for high schools with low mean incoming 8th grade test scores and negative effects of value-added on the number of students
in the sample from that high school for high schools with high mean incoming 8th grade scores. To test this, I regress the number of students from each high school on its value-added score with controls for age, gender, and race, mean high school demographics and incoming test scores, and year fixed effects. The results of this test do not show any evidence of selection effects. A one standard deviation increase in value-added changes the number of students in the sample from each school in a given year by no more than 2.5 students. The sign of the effect changes across the distribution of incoming test score percentiles with no pattern and none of the estimates are statistically significant. The lack of any significant patterns could just mean that the selection of high ability students out of the sample for each school offsets the selection of low ability students into the sample.

Overall, the results in table 2.4 show some evidence of selection, although the evidence is for selection across different universities within the sample rather than selection into or out of the sample. To the extent that selection exists, driving low ability students into the sample and high ability students out of the sample, it will cause negative bias in the estimate of the effect of attending a higher value-added high school on college performance.

2.6 Results

This section shows the results for the analysis of the lasting impact of attending a higher value-added high school on college GPA and graduation. All of the analysis is restricted to students that originally enrolled as freshmen, thus excluding transfer students\textsuperscript{12}. The sample of students for the graduation regressions is restricted to students that enrolled at least 5 years before MIDFIELD stopped tracking students in 2006. The lasting impact of attending a higher value-added high school on GPA is shown separately for each of the first 6 semesters and then for the final cumulative

\textsuperscript{12}This is because I look at the lasting impact of attending a higher value-added high school by semester. Students that transfer into the schools in the sample do not have GPA information for previous semesters at other schools.
GPA. The sample for GPA regressions is restricted to students that were enrolled for 6-12 semesters and do not have any missing GPA information in the first 6 semesters. The purpose of requiring at least 6 semesters is so that there is a consistent sample of students across regressions for GPAs from different semesters. Students enrolled for more than 12 semesters are dropped in order to limit the sample to students with a traditional college graduation time-line\textsuperscript{13}.

I first present results for all students. I then separate the results for GPA according to whether the student graduated. For those that graduated, the sample is again restricted to those that were enrolled for 6-12 semesters. For those that did not graduate, the sample is not restricted by the total number of semesters attended. Therefore, the number of observations in regressions by semester decreases for later semesters for the non-graduation sample. Regressions that condition on graduation status restrict the sample to students enrolled at least 5 years before MIDFIELD stopped tracking students in 2006.

The coefficient for the value-added variable shown in the results tables represents the impact of increasing the value-added score for the student’s high school by 1 unit. A more intuitive interpretation is the impact of increasing the value-added score for the student’s high school by one standard deviation. This can be found by multiplying the coefficients reported in the tables by the standard deviation of the high school value-added scores, which is 0.077, as discussed in section 4. The high school value-added scores are scaled in units of student test score standard deviations. Thus, increasing a high school’s value-added score by one is associated with increasing student test scores by a full standard deviation, while increasing a high school’s value-added score by 0.077 is associated with increasing student test scores by 0.077 standard deviations.

All of the analysis presented below includes student controls for age, gender, ethnicity, and high school GPA. High school controls matched from the K-12 datasets include the fraction of the enrollment that is male, Black, Asian, Hispanic, and on free

\textsuperscript{13}Only 4\% of students in the sample have a total semester count larger than 12.
or reduced lunch as well as the mean 8th grade math and reading scores for incoming students. University, major, and year fixed effects are also included. Standard errors are clustered by high school.

### 2.6.1 GPA and Graduation

Column (1) of table 2.5 shows the lasting impact of attending a higher value-added high school on the probability of graduating from college. The coefficient is 0.009, meaning that a one standard deviation increase in high school value-added scores increases the probability of graduating from college by 0.0007, or less than one one-thousandth of a percentage point. The effect is not statistically significant.

The results for GPA tell a different story. Columns (2)-(7) of table 2.5 show the impact of attending a higher value-added high school on GPA in semesters 1-6, respectively. Here, we see that attending a higher value-added high school does have a lasting impact, although the impact becomes smaller as the time since high school increases. The coefficient for value-added in semester 1 is 0.29, statistically significant at the one-percent level. This corresponds to an increase in GPA of 0.022 points from a one standard deviation increase in high school value-added.

The effect remains stable throughout the first three semesters. The coefficient in semester 2 is again 0.29, statistically significant at the one-percent level. In semester 3, the coefficient actually increases slightly to 0.32, again statistically significant at the one-percent level. However, the impact begins to decrease after semester 3. The coefficient is 0.23 and significant at the five-percent level in semester 4. The coefficients for semesters 5 and 6 are 0.11 and 0.04, repetitively, neither of which are statistically significant. Column (8) shows the impact of attending a higher value-added high school on the student’s final cumulative GPA. The coefficient is 0.22, statistically significant at the one-percent level. This represents an increase in final GPA by 0.017 points from a one standard deviation increase in high school value-added.
Tables 2.6 and 2.7 show the impact of attending a higher value-added high school on college GPA separately for students that did and did not graduate from college. The purpose of separating the effect according to whether the student graduated is to begin to see if there is a different impact for different types of students. However, the results for both students that graduated and those that did not are very similar to the results that pool all students together in table 2.5; there is a positive and statistically significant impact of attending a higher value-added high school on GPA. The effect is largest in semesters 1-4 and then begins to fade out and is no longer statistically significant. The impact on final cumulative GPA is positive and statistically significant at the one-percent level for both students that did and did not graduate, with coefficients of 0.31 and 0.41, respectively.

Overall, the results in table 2.5, table 2.6, and table 2.7 show that there is a positive and statistically significant impact of attending a higher value-added high school on college GPA. The effect is largest in earlier semesters. This result is consistent with an implied, although not discussed, feature of the value-added models discussed earlier, which is that the impact of school inputs can vary with the temporal distance between the time the inputs were applied and the time of the outcome measure [Todd and Wolpin, 2003]. This result is also consistent with studies which find that the lasting impact of teacher value-added on achievement in later K-12 grades fades out with the temporal distance since the student was exposed to the teacher [Carrell and West, 2010, Brian A. Jacob and Sims, 2010, Rothstein, 2010, Chetty et al., 014a].

Although the impact on GPA is positive and statistically significant, the impact is small. The impact of increasing high school value-added by one standard deviation is associated with an increase in first semester GPA by 0.022 points and final GPA by 0.017 points. This is approximately equivalent to a one letter grade increase in exactly one course during a student’s entire time in college. When put in that context, it is

\footnote{Conditioning on graduation could cause negative bias in the estimated impact of attending a higher value-added high school on college GPA if graduation is impacted by value-added. This would push low ability students from higher value-added high schools into the graduated sample. However, given the previous results for graduation, this is likely a very small concern.}
not surprising that there is no significant impact of attending a higher value-added high school on the probability of graduating from college.

2.6.2 Gender and racial heterogeneity

Table 2.8 shows the impact of attending a higher value-added high school on the probability of graduation and GPA separately for males, females, White students, and Black students. Each row-column entry is a separate regression and shows only the coefficient and standard error for the value-added variable. Each regression includes the same control variables as in the previous analysis.

These results are very similar to the results shown in table 2.5, which showed the results for all students. There is no significant impact of attending a higher value-added high school on the probability of finishing college for any gender or race, but there is a positive impact of attending a higher value-added high school on college GPA. The impact on GPA is largest in early semesters for each race and gender and then becomes smaller and statistically insignificant in later semesters.

There is not much difference in the results for males and females. The magnitude of the impact of value-added is very similar for the two groups and very similar to the overall results shown in table 2.5. A one standard deviation increase in high school value-added increases final cumulative GPA by 0.017 and 0.015 points for males and females, respectively, each statistically significant at the five-percent level. Most of this impact occurs during the first four semesters.

There are significant differences for White and Black students. The results for White students are larger in magnitude than for Black students and than the overall results in table 2.5. A one standard deviation increase in high school value-added increases final cumulative GPA by 0.022 points for White students. Most of the impact comes in the first five semesters. For Black students, the results also show a positive impact on final GPA that mostly occurs in the first few semesters, but the impacts are small; the impact of a one standard deviation increase in high school
value-added on final GPA is 0.010 and not statistically significant. It appears that the reason for the smaller and insignificant impact of final cumulative GPA for Black students is due to a quicker fade out of the impact of high school value-added for Black students. Black and White students experience similar impacts of a one standard deviation increase in high school value-added in semester 1: an increase of 0.022 points for Black students and 0.025 for White students. But the impact fades out much more quickly after semester 1 for Black students than for White students; the impact of a one standard deviation increase in high school value-added in semester 2 and semester 3 is 0.013 and 0.007 points, respectively, for Black students and 0.026 and 0.033 for White students.

2.6.3 Quantile regressions

I use quantile regressions to determine if value-added has different impacts for students of different ability levels. The first set of results are shown in table 2.9. This table shows the results of quantile regressions for the effect of attending a higher value-added high school on college GPA by semester. It appears that the largest impact of attending a higher value-added high school is for lower ability students; for each of the first three semesters, the coefficient on the value-added variable is generally larger for lower quantiles, although it is statistically significant for every quantile. This is a fairly intuitive result, as one could imagine that low ability students would have more to gain from attending a higher value-added high school than high ability students who are already great students.

However, this pattern ceases beginning in semester 4. The coefficient on the value-added variable for the 10th quantile becomes smaller than the other low quantiles beginning in the 4th semester. The coefficient for the 20th quantile does the same beginning in semester 5. Similarly, the quantile regression results for final GPA show that value-added appears to have larger impacts for lower ability students except for the very lowest quantiles. One possible explanation is that the very lowest
quantile includes students that eventually drop out of college. These students are likely different not only in terms of their GPA quantile, but also in terms of other unobservable characteristics that make them more likely to struggle through college and eventually drop out.

Table 2.10 shows quantile regression results for students that graduated. Here, there is a more consistent pattern that attending a higher value-added high school has larger impacts for lower-ability students, particularly when looking at final GPA in column (7). However, this pattern does not hold for every individual semester. For the first three semesters, the impact is similar across quantiles, with slightly larger coefficients for both the low and high quantiles. After the first three semesters, the impact fades out quickly for the higher ability students, but remains large and often statistically significant for lower quantiles.

Overall, these results provide some evidence that the lasting impact of attending a higher value-added high school is larger for low ability students. However, it should be remembered that the low ability students enrolled in state universities in Florida and North Carolina on average likely come from near the top of the distribution of student ability in their own high school.

2.7 Conclusion

This paper addresses the question of whether attending a higher value-added high school has lasting impacts on students later in life. This is a policy-relevant question, as many states and school districts have begun using value-added scores to evaluate teachers and schools. The use of these value-added models has sparked debate about whether they provide unbiased estimates of the impact that teachers have on test scores and debate about whether they are appropriate metrics for evaluating teachers even if they are unbiased. While the growing consensus is that modern value-added models can provide unbiased estimates of teacher impacts, their value as a public policy tool is still heavily debated.
One important piece for understanding whether value-added models are an appropriate way to evaluate teachers and schools is to test whether having better value-added teachers or schools is correlated with better outcomes in adulthood. If they are correlated, this suggests that the test score gains associated with higher value-added teachers and schools represent gains in cognitive ability that have a lasting impact on the student. Alternatively, if they are not correlated, higher value-added teachers and schools may simply be those that are better at teaching to the test, rather than having any impact on the ability of the student.

I address this public policy question by estimating value-added scores for high schools in Florida and North Carolina and then linking these value-added scores to a second dataset that includes detailed college transcript data, including high school. I find that attending a higher value-added high school has lasting impacts on college GPA, which primarily occur during early semesters. A one standard deviation increase in high school value-added increases first semester GPA by 0.022 points and final GPA by 0.017 points. This is approximately equal to one letter grade increase in one course over a college career. These results are generally consistent across race and gender, although the impact is slightly larger for White students.

While the results are small in magnitude, this is not entirely surprising given what we know about early educational inputs mattering more than later inputs [Cuncha and Heckman, 2010]. Additionally, much of the impact of attending a higher value-added high school may result from boosting college attendance and getting students into better universities, in addition to improving their performance in college. [Chetty et al., 014b] find evidence of exactly that, which would cause negative bias in my estimates of the impact of attending a higher value-added high school on college performance by changing the underlying distribution of student ability in my sample of universities. Therefore, the estimates that I find may be interpreted as a lower-bound on the impact of attending a higher value-added high school on college performance. Finding positive and statistically significant effects of attending a higher value-added high school despite this possible negative selection bias provides more strong support
for the idea that high value-added teachers and schools can have significant lasting impacts on their students.
Table 2.1.: Summary statistics for the dataset of analysis

<table>
<thead>
<tr>
<th></th>
<th>(1) Less than median VA</th>
<th>(2) Greater than median VA</th>
<th>(3) All Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Female</td>
<td>0.54</td>
<td>0.499</td>
<td>0.53</td>
</tr>
<tr>
<td>Black</td>
<td>0.29</td>
<td>0.452</td>
<td>0.25</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.05</td>
<td>0.225</td>
<td>0.06</td>
</tr>
<tr>
<td>Asian</td>
<td>0.04</td>
<td>0.191</td>
<td>0.04</td>
</tr>
<tr>
<td>other race</td>
<td>0.02</td>
<td>0.136</td>
<td>0.02</td>
</tr>
<tr>
<td>start age</td>
<td>18.53</td>
<td>2.877</td>
<td>18.49</td>
</tr>
<tr>
<td>HS GPA</td>
<td>3.54</td>
<td>0.638</td>
<td>3.55</td>
</tr>
<tr>
<td>SAT</td>
<td>1091.18</td>
<td>184.814</td>
<td>1123.36</td>
</tr>
<tr>
<td>ACT</td>
<td>23.63</td>
<td>4.215</td>
<td>24.20</td>
</tr>
<tr>
<td>final GPA</td>
<td>2.32</td>
<td>0.966</td>
<td>2.36</td>
</tr>
<tr>
<td>graduated</td>
<td>0.42</td>
<td>0.493</td>
<td>0.43</td>
</tr>
<tr>
<td>VA</td>
<td>-0.05</td>
<td>0.041</td>
<td>0.05</td>
</tr>
<tr>
<td>HS-% male</td>
<td>0.47</td>
<td>0.035</td>
<td>0.47</td>
</tr>
<tr>
<td>HS-% Black</td>
<td>0.30</td>
<td>0.243</td>
<td>0.27</td>
</tr>
<tr>
<td>HS-% Asian</td>
<td>0.02</td>
<td>0.023</td>
<td>0.03</td>
</tr>
<tr>
<td>HS-% Hispanic</td>
<td>0.13</td>
<td>0.145</td>
<td>0.13</td>
</tr>
<tr>
<td>HS-% free/reduced lunch</td>
<td>0.27</td>
<td>0.119</td>
<td>0.22</td>
</tr>
<tr>
<td>HS-mean 8th grade reading scores</td>
<td>0.04</td>
<td>0.299</td>
<td>0.17</td>
</tr>
<tr>
<td>HS-mean 8th grade math scores</td>
<td>0.03</td>
<td>0.301</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Summary statistics are shown for all students in the data and according to whether the student is from a high school with a value-added score above or below the median value-added score. Value-added is scaled in units of student test score standard deviations. Demographics, high school GPA, SAT and ACT scores, college GPA, and graduation status come from the MIDFIELD dataset. High school value-added scores and high school characteristics for the high school that the student attended are matched from the data on test scores from Florida and North Carolina. The method for estimating value-added scores are described in section 3.1.
Table 2.2.: Test for correlation between value-added scores and unobservables - North Carolina

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Algebra 1</td>
<td>Algebra 2</td>
<td>English 1</td>
<td>Biology</td>
<td>7th Reading</td>
<td>7th Math</td>
<td>Algebra 1</td>
<td>Algebra 2</td>
<td>English 1</td>
<td>Biology</td>
</tr>
<tr>
<td>VA</td>
<td>1.47***</td>
<td>2.21***</td>
<td>0.24***</td>
<td>0.84***</td>
<td>0.03</td>
<td>-0.02</td>
<td>1.46***</td>
<td>2.21***</td>
<td>0.24***</td>
<td>0.84***</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.078)</td>
<td>(0.045)</td>
<td>(0.078)</td>
<td>(0.021)</td>
<td>(0.026)</td>
<td>(0.063)</td>
<td>(0.085)</td>
<td>(0.044)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>7th reading</td>
<td>-0.03***</td>
<td>0.01</td>
<td>0.29***</td>
<td>0.22***</td>
<td>0.0044</td>
<td>0.0054</td>
<td>0.0040</td>
<td>0.0042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th math</td>
<td>0.15***</td>
<td>0.27***</td>
<td>0.07***</td>
<td>0.13***</td>
<td>0.0054</td>
<td>0.0072</td>
<td>0.0045</td>
<td>0.0048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1845488</td>
<td>877846</td>
<td>2130103</td>
<td>1579277</td>
<td>3504036</td>
<td>3507403</td>
<td>1607523</td>
<td>762399</td>
<td>1856951</td>
<td>1349098</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.443</td>
<td>0.394</td>
<td>0.577</td>
<td>0.499</td>
<td>0.692</td>
<td>0.727</td>
<td>0.446</td>
<td>0.406</td>
<td>0.609</td>
<td>0.520</td>
</tr>
</tbody>
</table>

Each column reports the coefficients from an OLS regression, with standard errors clustered by high school. The regressions are estimated using the K-12 test score data from North Carolina. Each column is a regression of the test score shown in the column heading on the average value-added score for the student’s high school, along with several covariates. Value-added is scaled in units of student test score standard deviations. Columns (1)-(4) show the effect of value-added on the different high school test scores used to estimate school value-added scores. Columns (5)-(6) show the effect of high school value-added on 7th grade test scores. Columns (7)-(10) show the effect of high school value-added on the high school test scores with additional controls for 7th grade test scores. The covariates include student controls for gender, race, free or reduced lunch, and a cubic in 8th grade math and reading scores and include high school controls for mean demographics, mean free or reduced lunch receipt, and a cubic in mean 8th grade math and reading scores. Year fixed effects are also included.
Table 2.3.: Test for correlation between value-added scores and unobservables - Florida

<table>
<thead>
<tr>
<th></th>
<th>(1) Math</th>
<th>(2) Reading</th>
<th>(3) Writing</th>
<th>(4) 7th reading</th>
<th>(5) 7th math</th>
<th>(6) Math</th>
<th>(7) Reading</th>
<th>(8) Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA</td>
<td>0.70***</td>
<td>0.78***</td>
<td>0.86***</td>
<td>-0.12***</td>
<td>-0.06*</td>
<td>0.68***</td>
<td>0.79***</td>
<td>0.66***</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.055)</td>
<td>(0.070)</td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.060)</td>
<td>(0.060)</td>
<td>(0.061)</td>
</tr>
<tr>
<td>7th reading</td>
<td>0.03***</td>
<td>0.28***</td>
<td>0.08***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0021)</td>
<td>(0.0011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th math</td>
<td>0.26***</td>
<td>0.07***</td>
<td>0.06***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>3141458</td>
<td>3539915</td>
<td>2279756</td>
<td>6604671</td>
<td>6603112</td>
<td>2241308</td>
<td>2595526</td>
<td>1751434</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.678</td>
<td>0.629</td>
<td>0.140</td>
<td>0.669</td>
<td>0.721</td>
<td>0.707</td>
<td>0.663</td>
<td>0.113</td>
</tr>
</tbody>
</table>

Each column reports the coefficients from an OLS regression, with standard errors clustered by high school. The regressions are estimated using the K-12 test score data from Florida. See table 2.2 for more details.
Table 2.4.: Selection across universities within sample and selection out of sample

<table>
<thead>
<tr>
<th></th>
<th>Mean Students Per School</th>
<th>Attends HBCU</th>
<th>NC University Quality</th>
<th>Students in Sample from each HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>incoming test score percentile</td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>All Schools</td>
<td>28.54</td>
<td>-0.22*</td>
<td>2.43***</td>
<td>-0.03</td>
</tr>
<tr>
<td>≤5th Percentile</td>
<td>23.72</td>
<td>-1.52**</td>
<td>-36.39***</td>
<td>1.51</td>
</tr>
<tr>
<td>5th-10th</td>
<td>20.16</td>
<td>-0.21</td>
<td>0.94</td>
<td>-2.29</td>
</tr>
<tr>
<td>10th-25th</td>
<td>22.32</td>
<td>-0.45**</td>
<td>1.58</td>
<td>-2.15</td>
</tr>
<tr>
<td>75th-90th</td>
<td>37.89</td>
<td>-0.02</td>
<td>2.74**</td>
<td>1.77</td>
</tr>
<tr>
<td>90th-95th</td>
<td>48.83</td>
<td>0.69*</td>
<td>5.78***</td>
<td>2.39</td>
</tr>
<tr>
<td>≥95th</td>
<td>50.73</td>
<td>1.29***</td>
<td>9.96***</td>
<td>-0.12</td>
</tr>
</tbody>
</table>

Column (1) shows the average number of students from each school in a given year for all schools and separately according to the school’s average 8th grade scores for incoming students. Column (2) shows the probit model results for the effect of high school value-added on attending one of the two HBCUs in the dataset. Column (3) shows the ordered logit results for university quality in North Carolina. The three schools in the sample from North Carolina were given the following quality ranking: (1) North Carolina State University (2) UNC Charlotte (3) North Carolina A&T. Column (4) shows the regression results for the effect of high school value-added on the number of students in the sample from each high school. All specifications include student controls for age, gender, and race, mean high school demographics and incoming test scores, and year fixed effects. Column (2) includes only Black students. Results in column (4) represent the effect of increasing high school value-added by one standard deviation. Standard errors are clustered by high school.
Table 2.5.: The impact of attending a higher value-added high school on graduation and GPA

<table>
<thead>
<tr>
<th></th>
<th>(1) Graduated</th>
<th>(2) Semester 1</th>
<th>(3) Semester 2</th>
<th>(4) Semester 3</th>
<th>(5) Semester 4</th>
<th>(6) Semester 5</th>
<th>(7) Semester 6</th>
<th>(8) Final GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA</td>
<td>0.009</td>
<td>0.29***</td>
<td>0.29***</td>
<td>0.32***</td>
<td>0.23**</td>
<td>0.11</td>
<td>0.04</td>
<td>0.22***</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.110)</td>
<td>(0.092)</td>
<td>(0.120)</td>
<td>(0.101)</td>
<td>(0.110)</td>
<td>(0.110)</td>
<td>(0.073)</td>
</tr>
<tr>
<td>age</td>
<td>0.02***</td>
<td>-0.002</td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.03***</td>
<td>0.03***</td>
<td>0.03***</td>
<td>0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.007)</td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.008)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>female</td>
<td>0.06***</td>
<td>0.10***</td>
<td>0.07***</td>
<td>0.06***</td>
<td>0.08***</td>
<td>0.10***</td>
<td>0.09***</td>
<td>0.11***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.06***</td>
<td>-0.16***</td>
<td>-0.14***</td>
<td>-0.21***</td>
<td>-0.17***</td>
<td>-0.14***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.020)</td>
<td>(0.019)</td>
<td>(0.022)</td>
<td>(0.021)</td>
<td>(0.022)</td>
<td>(0.022)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.08***</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.07***</td>
<td>-0.12***</td>
<td>-0.06***</td>
<td>-0.02</td>
<td>-0.05***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.025)</td>
<td>(0.027)</td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.032)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Asian</td>
<td>-0.01</td>
<td>-0.10***</td>
<td>-0.07***</td>
<td>-0.05</td>
<td>-0.06**</td>
<td>-0.13***</td>
<td>-0.09***</td>
<td>-0.10***</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.027)</td>
<td>(0.024)</td>
<td>(0.031)</td>
<td>(0.028)</td>
<td>(0.031)</td>
<td>(0.035)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>other race</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.035)</td>
<td>(0.034)</td>
<td>(0.038)</td>
<td>(0.044)</td>
<td>(0.045)</td>
<td>(0.042)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>HS GPA</td>
<td>0.17***</td>
<td>0.35***</td>
<td>0.63***</td>
<td>0.59***</td>
<td>0.58***</td>
<td>0.55***</td>
<td>0.52***</td>
<td>0.48***</td>
</tr>
<tr>
<td></td>
<td>(0.0084)</td>
<td>(0.019)</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.012)</td>
</tr>
</tbody>
</table>

Each column is a separate regression of value-added scores on different outcome variables, as indicated in the column headings. In addition to the control variables shown in the table, each column also controls for high school characteristics including the mean fraction of the enrollment that is male, Black, Asian, Hispanic, and on free or reduced lunch as well as the mean incoming 8th grade math and reading scores. Each column also includes university, major, and year fixed effects. All results are based on students that originally enrolled as freshman, thus excluding transfer students. The sample for graduation results is further limited to students that enrolled at least five year before MIDFIELD stopped tracking students in 2006. The sample for GPA regressions is limited to students who were enrolled for 6-12 semesters and did not have missing GPA observations for any semesters. Graduation results are based on a probit model and show marginal effects. Value-added is scaled in units of student test score standard deviations. Standard errors are clustered by high school.
Table 2.6.: The impact of attending a higher value-added high school on graduation and GPA: students that graduated

<table>
<thead>
<tr>
<th></th>
<th>(1) Semester 1</th>
<th>(2) Semester 2</th>
<th>(3) Semester 3</th>
<th>(4) Semester 4</th>
<th>(5) Semester 5</th>
<th>(6) Semester 6</th>
<th>(7) Final GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA</td>
<td>0.44***</td>
<td>0.40***</td>
<td>0.33**</td>
<td>0.42***</td>
<td>0.21</td>
<td>0.16</td>
<td>0.31***</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.130)</td>
<td>(0.140)</td>
<td>(0.140)</td>
<td>(0.160)</td>
<td>(0.140)</td>
<td>(0.074)</td>
</tr>
<tr>
<td>age</td>
<td>0.02*</td>
<td>0.04**</td>
<td>-0.01</td>
<td>0.012</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.016)</td>
<td>(0.022)</td>
<td>(0.022)</td>
<td>(0.017)</td>
<td>(0.031)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>female</td>
<td>0.10***</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.03</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.05***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.025)</td>
<td>(0.031)</td>
<td>(0.026)</td>
<td>(0.028)</td>
<td>(0.030)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Black</td>
<td>0.07*</td>
<td>0.01</td>
<td>-0.08</td>
<td>-0.10**</td>
<td>-0.11**</td>
<td>-0.11**</td>
<td>-0.08***</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.045)</td>
<td>(0.048)</td>
<td>(0.048)</td>
<td>(0.047)</td>
<td>(0.044)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.19**</td>
<td>0.06</td>
<td>-0.25*</td>
<td>-0.09</td>
<td>-0.17</td>
<td>-0.21*</td>
<td>-0.11**</td>
</tr>
<tr>
<td></td>
<td>(0.095)</td>
<td>(0.110)</td>
<td>(0.130)</td>
<td>(0.097)</td>
<td>(0.110)</td>
<td>(0.110)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>Asian</td>
<td>0.10</td>
<td>-0.08</td>
<td>0.06</td>
<td>-0.01</td>
<td>0.03</td>
<td>-0.09</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.064)</td>
<td>(0.062)</td>
<td>(0.065)</td>
<td>(0.067)</td>
<td>(0.081)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>other race</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.11</td>
<td>0.08</td>
<td>0.05</td>
<td>0.12*</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.056)</td>
<td>(0.074)</td>
<td>(0.072)</td>
<td>(0.075)</td>
<td>(0.064)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>HS GPA</td>
<td>0.061</td>
<td>0.38***</td>
<td>0.44***</td>
<td>0.40***</td>
<td>0.39***</td>
<td>0.42***</td>
<td>0.25***</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.031)</td>
<td>(0.039)</td>
<td>(0.031)</td>
<td>(0.034)</td>
<td>(0.033)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Observations</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.200</td>
<td>0.129</td>
<td>0.131</td>
<td>0.116</td>
<td>0.094</td>
<td>0.097</td>
<td>0.189</td>
</tr>
</tbody>
</table>

This table limits the analysis to students that graduated. See the footnotes in Table 2.5 for further details.
<table>
<thead>
<tr>
<th></th>
<th>(1) Semester 1</th>
<th>(2) Semester 2</th>
<th>(3) Semester 3</th>
<th>(4) Semester 4</th>
<th>(5) Semester 5</th>
<th>(6) Semester 6</th>
<th>(7) Final GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA</td>
<td>0.31*</td>
<td>0.54***</td>
<td>0.40*</td>
<td>0.53**</td>
<td>0.04</td>
<td>0.19</td>
<td>0.41**</td>
</tr>
<tr>
<td>age</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03**</td>
<td>0.03</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.02*</td>
</tr>
<tr>
<td>female</td>
<td>0.10***</td>
<td>0.10***</td>
<td>0.06**</td>
<td>0.05*</td>
<td>0.10***</td>
<td>0.04</td>
<td>0.10***</td>
</tr>
<tr>
<td>Black</td>
<td>0.01</td>
<td>0.04</td>
<td>-0.06</td>
<td>-0.09**</td>
<td>-0.24***</td>
<td>-0.13***</td>
<td>-0.08***</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.01</td>
<td>-0.07</td>
<td>-0.07</td>
<td>-0.06</td>
<td>-0.8</td>
<td>-0.04</td>
<td>-0.04</td>
</tr>
<tr>
<td>Asian</td>
<td>-0.06</td>
<td>-0.08</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.19**</td>
<td>-0.09</td>
<td>-0.06</td>
</tr>
<tr>
<td>other race</td>
<td>-0.10</td>
<td>0.04</td>
<td>0.04</td>
<td>0.01</td>
<td>-0.06</td>
<td>0.02</td>
<td>-0.05</td>
</tr>
<tr>
<td>HS GPA</td>
<td>0.47***</td>
<td>0.74***</td>
<td>0.64***</td>
<td>0.57***</td>
<td>0.52***</td>
<td>0.44***</td>
<td>0.54***</td>
</tr>
</tbody>
</table>

Observations: 8785 8281 7139 6526 5910 5465 8813
R-squared: 0.397 0.289 0.235 0.195 0.191 0.143 0.422

This table limits the analysis to students that did not graduate. See the footnotes in table 2.5 for further details.
Table 2.8.: The impact of attending a higher value-added high school on graduation and GPA: gender and racial heterogeneity

<table>
<thead>
<tr>
<th></th>
<th>(1) Graduated</th>
<th>(2) Semester 1</th>
<th>(3) Semester 2</th>
<th>(4) Semester 3</th>
<th>(5) Semester 4</th>
<th>(6) Semester 5</th>
<th>(7) Semester 6</th>
<th>(8) Final GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-0.05</td>
<td>0.25*</td>
<td>0.32***</td>
<td>0.36**</td>
<td>0.22</td>
<td>0.12</td>
<td>0.14</td>
<td>0.22**</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.140)</td>
<td>(0.110)</td>
<td>(0.150)</td>
<td>(0.140)</td>
<td>(0.141)</td>
<td>(0.152)</td>
<td>(0.086)</td>
</tr>
<tr>
<td>Female</td>
<td>0.06</td>
<td>0.31**</td>
<td>0.25**</td>
<td>0.27*</td>
<td>0.23*</td>
<td>0.09</td>
<td>-0.07</td>
<td>0.20**</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.131)</td>
<td>(0.134)</td>
<td>(0.162)</td>
<td>(0.141)</td>
<td>(0.130)</td>
<td>(0.130)</td>
<td>(0.088)</td>
</tr>
<tr>
<td>White</td>
<td>0.03</td>
<td>0.33***</td>
<td>0.34***</td>
<td>0.43***</td>
<td>0.22*</td>
<td>0.22*</td>
<td>0.07</td>
<td>0.29***</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.122)</td>
<td>(0.113)</td>
<td>(0.132)</td>
<td>(0.132)</td>
<td>(0.124)</td>
<td>(0.120)</td>
<td>(0.084)</td>
</tr>
<tr>
<td>Black</td>
<td>0.09</td>
<td>0.29</td>
<td>0.17</td>
<td>0.09</td>
<td>0.26</td>
<td>-0.05</td>
<td>0.06</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.190)</td>
<td>(0.150)</td>
<td>(0.211)</td>
<td>(0.172)</td>
<td>(0.191)</td>
<td>(0.230)</td>
<td>(0.120)</td>
</tr>
</tbody>
</table>

This table shows the impact of attending a higher value-added high school on GPA and graduation separately by gender and race. Each row-column entry is a separate regression and shows only the coefficient and standard error (in parentheses) for the value-added variable. Each regression includes the same control variables as the previous analysis. See the footnotes in table 2.5 for further details.
Table 2.9.: Quantile regression impact of attending a higher value-added high school on GPA by semester

<table>
<thead>
<tr>
<th></th>
<th>(1) Semester 1</th>
<th>(2) Semester 2</th>
<th>(3) Semester 3</th>
<th>(4) Semester 4</th>
<th>(5) Semester 5</th>
<th>(6) Semester 6</th>
<th>(7) Final GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>q10</td>
<td>0.49*** (0.13)</td>
<td>0.48*** (0.15)</td>
<td>0.72*** (0.20)</td>
<td>0.07 (0.16)</td>
<td>-0.01 (0.25)</td>
<td>-0.14 (0.22)</td>
<td>0.15 (0.096)</td>
</tr>
<tr>
<td>q20</td>
<td>0.35*** (0.078)</td>
<td>0.32*** (0.12)</td>
<td>0.57*** (0.12)</td>
<td>0.34*** (0.089)</td>
<td>0.13 (0.17)</td>
<td>0.11 (0.12)</td>
<td>0.26*** (0.082)</td>
</tr>
<tr>
<td>q30</td>
<td>0.39*** (0.097)</td>
<td>0.33*** (0.071)</td>
<td>0.43*** (0.077)</td>
<td>0.33*** (0.085)</td>
<td>0.29** (0.15)</td>
<td>0.16 (0.13)</td>
<td>0.26*** (0.071)</td>
</tr>
<tr>
<td>q40</td>
<td>0.24*** (0.084)</td>
<td>0.26*** (0.095)</td>
<td>0.27*** (0.098)</td>
<td>0.33*** (0.076)</td>
<td>0.21 (0.13)</td>
<td>0.13 (0.12)</td>
<td>0.29*** (0.064)</td>
</tr>
<tr>
<td>q50</td>
<td>0.23*** (0.078)</td>
<td>0.19** (0.089)</td>
<td>0.19** (0.083)</td>
<td>0.35*** (0.11)</td>
<td>0.20 (0.13)</td>
<td>0.07 (0.10)</td>
<td>0.26*** (0.056)</td>
</tr>
<tr>
<td>q60</td>
<td>0.18** (0.083)</td>
<td>0.27*** (0.071)</td>
<td>0.25*** (0.088)</td>
<td>0.27** (0.12)</td>
<td>0.22* (0.12)</td>
<td>0.07 (0.11)</td>
<td>0.28*** (0.056)</td>
</tr>
<tr>
<td>q70</td>
<td>0.22** (0.089)</td>
<td>0.24*** (0.083)</td>
<td>0.18* (0.095)</td>
<td>0.19* (0.11)</td>
<td>0.20* (0.11)</td>
<td>0.03 (0.12)</td>
<td>0.22*** (0.045)</td>
</tr>
<tr>
<td>q80</td>
<td>0.25*** (0.081)</td>
<td>0.26*** (0.078)</td>
<td>0.089 (0.11)</td>
<td>0.15 (0.11)</td>
<td>0.06 (0.092)</td>
<td>-0.01 (0.088)</td>
<td>0.16*** (0.059)</td>
</tr>
<tr>
<td>q90</td>
<td>0.21*** (0.082)</td>
<td>0.36*** (0.13)</td>
<td>0.19** (0.083)</td>
<td>0.16 (0.073)</td>
<td>-0.04 (0.057)</td>
<td>0.02 (0.057)</td>
<td>0.09 (0.065)</td>
</tr>
</tbody>
</table>

Observations: 23608

This table shows quantile regressions for the impact of attending a higher value-added high school on GPA. See table 2.5 for further details about covariates and standard errors.
Table 2.10.: Quantile regression impact of attending a higher value-added high school on GPA by semester, students that graduated

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>q10</td>
<td>0.52**</td>
<td>0.61***</td>
<td>0.28</td>
<td>0.73**</td>
<td>0.24</td>
<td>0.18</td>
<td>0.56***</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.14)</td>
<td>(0.26)</td>
<td>(0.33)</td>
<td>(0.44)</td>
<td>(0.18)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>q20</td>
<td>0.60***</td>
<td>0.37**</td>
<td>0.40**</td>
<td>0.78***</td>
<td>0.61**</td>
<td>0.43**</td>
<td>0.44***</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.18)</td>
<td>(0.20)</td>
<td>(0.24)</td>
<td>(0.25)</td>
<td>(0.20)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>q30</td>
<td>0.58***</td>
<td>0.44**</td>
<td>0.18</td>
<td>0.57***</td>
<td>0.37*</td>
<td>0.60***</td>
<td>0.34***</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.17)</td>
<td>(0.22)</td>
<td>(0.21)</td>
<td>(0.22)</td>
<td>(0.20)</td>
<td>(0.092)</td>
</tr>
<tr>
<td>q40</td>
<td>0.44**</td>
<td>0.48***</td>
<td>0.13</td>
<td>0.57***</td>
<td>0.39***</td>
<td>0.35*</td>
<td>0.27**</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.15)</td>
<td>(0.22)</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.19)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>q50</td>
<td>0.32**</td>
<td>0.30**</td>
<td>0.17</td>
<td>0.47***</td>
<td>0.39**</td>
<td>0.24</td>
<td>0.25*</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.13)</td>
<td>(0.20)</td>
<td>(0.15)</td>
<td>(0.18)</td>
<td>(0.19)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>q60</td>
<td>0.37***</td>
<td>0.35**</td>
<td>0.30</td>
<td>0.32*</td>
<td>0.46**</td>
<td>0.22</td>
<td>0.22**</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.17)</td>
<td>(0.23)</td>
<td>(0.16)</td>
<td>(0.20)</td>
<td>(0.19)</td>
<td>(0.092)</td>
</tr>
<tr>
<td>q70</td>
<td>0.37**</td>
<td>0.28</td>
<td>0.54**</td>
<td>0.25</td>
<td>0.21</td>
<td>0.07</td>
<td>0.25**</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.26)</td>
<td>(0.18)</td>
<td>(0.18)</td>
<td>(0.14)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>q80</td>
<td>0.42**</td>
<td>0.37*</td>
<td>0.43*</td>
<td>0.28</td>
<td>0.16</td>
<td>0.25*</td>
<td>0.28**</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.22)</td>
<td>(0.24)</td>
<td>(0.21)</td>
<td>(0.12)</td>
<td>(0.15)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>q90</td>
<td>0.49**</td>
<td>0.51**</td>
<td>0.45**</td>
<td>0.26</td>
<td>0.07</td>
<td>0.07</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.24)</td>
<td>(0.22)</td>
<td>(0.19)</td>
<td>(0.090)</td>
<td>(0.19)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Observations</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
<td>3656</td>
</tr>
</tbody>
</table>

This table shows quantile regressions for the impact of attending a higher value-added high school on GPA for students that graduated. See table 2.5 for further details about covariates and standard errors.
CHAPTER 3. THE IMPACT OF STUTTERING ON ADULT LABOR MARKET OUTCOMES

3.1 Introduction

The economics literature has long been interested in quantifying the impact of various disabilities and chronic health conditions on adult labor market outcomes. This includes Attention Deficit Hyperactive Disorder [Fletcher, 2014], depression [Ettnner et al., 1997, Fletcher, 2013], migraine headaches [Rees and Sabia, 2015], psychiatric disorders [Chatterji et al., 2011], cancer [Heinesen and Kolodziejczyk, 2013], general health [Smith, 2009], obesity [Norton and Han, 2008], physical disabilities that onset during adulthood [Haveman and Wolfe, 1990, Bound and Waidmann, 2002, Charles, 2003, Meyer and Mok, 2013], and chronic disabilities [Baldwin and Johnson, 1994, Baldwin and Johnson, 1995, Kidd et al., 2000]. Our study adds to this literature by using data from the National Longitudinal Study of Adolescent to Adult Health to quantify the impact of stuttering on adult labor market outcomes. This dataset includes detailed background and labor market information for respondents, as well as information on whether the respondent is a person who stutters.

Presently, it is estimated that 70 million people stutter worldwide and 3 million in the United States. Stuttering is a complex and heterogeneous disorder that extends beyond the speech disfluencies produced by both children and adults diagnosed with the disorder. Research has shown that chronic, persistent stuttering leads to the development of negative thoughts, attitudes, and emotions about speaking and one’s sense of efficacy as a communicator [Guitar, 2014]. Considering the ubiquitous nature of communication, it is likely that communication differences and difficulties would affect labor market outcomes for people who stutter (PWS) through various channels. However, no study has adequately quantified the impact of stuttering on adult labor
market outcomes or attempted to account for how various channels influence outcome differences between PWS and people who do not stutter (PWNS).

Although no study has adequately quantified the impact of stuttering on labor market outcomes, there is an abundance of evidence that PWS face difficulties in the labor market [Rice and Kroll, 1994, Rice and Kroll, 1997, Crichton-Smith, 2002, Hay-how et al., 2002, Klein and Hood, 2004, Palasik et al., 2012, Bricker-Katz et al., 2013]. These studies include self-reports of reduced earnings as a result of stuttering, reports of PWNS believing negative stereotypes about PWS, and reports of reduced occupational opportunities for PWS. These hardships could arise from many sources, including differences in background characteristics, differences in education levels, self-selection out of specific occupations, differences in personality characteristics such as self-esteem or self-efficacy, differences in job performance, and employer discrimination in the form of barriers to entry into specific occupations or decreased promotion possibilities within occupations\(^1\).

After controlling for detailed personal and family background characteristics, we find that stuttering has significant impacts on labor market outcomes and that stuttering has heterogeneous effects for males and females; for males, stuttering is associated with a significant decrease in the probability of being employed, decrease in the probability of being in the labor force, and increase in the probability of receiving public assistance such as welfare, but no impact on the probability of being underemployed conditional on employment. For females, the opposite is true; stuttering has no effect on the probability of being employed or in the labor force, but females who stutter are significantly more likely to be underemployed. Both males and females who stutter experience a reduction in hourly earnings of approximately 14%. Blinder-Oaxaca decomposition shows that much of the difference in hourly earnings can be explained by differences in education, occupation, and self-stigma, which is a variable that we construct to capture differences in self-esteem and self-efficacy that may develop as a

---

\(^1\)The United Kingdom has advocacy groups that support PWS in the workforce, such as the British Stammering Association and the Defence Stammering Network, but no such organizations currently exist in the United States.
person stutters throughout his or her lifetime. However, there remains an unexplained difference, which is often attributed to discrimination in decomposition studies.

Estimating the causal impact of stuttering may be difficult if stuttering is correlated with unobserved confounders at the individual or family level. Previous research on stuttering in the economics literature has shown that while stuttering appears to be associated with a decrease in human capital acquisition, these effects disappear when approaches intended to address difficult-to-measure heterogeneity are added [Rees and Sabia, 2014]2. We use similar approaches, which include adding controls for ADHD and other co-occurring learning disabilities or using propensity score matching. Unlike in [Rees and Sabia, 2014], we find that the effect of stuttering on labor market outcomes is robust to these approaches. The heterogeneous effect of stuttering by gender, which was not found in [Rees and Sabia, 2014], may also be interpreted as evidence that the effect of stuttering on labor market outcomes is causal; if the results were spurious and related to unobserved individual or family confounders, then the spurious relationship may be expected to occur for all labor market outcome variables or at least in the same outcome variables for males and females. Furthermore, the result that stuttering impacts more labor market outcome variables for males than females is consistent with evidence that females view stuttering as less of a handicap in the labor market than males [Klein and Hood, 2004].

The remainder of the paper is organized as follows: Section 2 provides background information on stuttering and the existing knowledge about employment experiences of PWS. Section 3 describes the data. Section 4 discusses the methodology used in the analysis. Section 5 presents the results. Section 6 concludes.

2[Rees and Sabia, 2014] use ordinary least squares and find that stuttering is associated with a significant decrease in high school GPA, high school graduation, and college attendance. However, they find that the effects become small and insignificant when any of the following three approaches are added: (1) adding controls for ADHD and learning disabilities, (2) using propensity score matching, or (3) adding family fixed effects. These three approaches all produce similar results that are much smaller than initial OLS results.
3.2 Background

3.2.1 What is stuttering and who stutters?

Stuttering is a neurodevelopmental disorder involving many different brain systems active for speech - including language, motor, and emotional networks [National Stuttering Asso Meeting, 2015]. Given that stuttering is an intricate human behavior, the cause of stuttering is both dynamic and multifactorial in nature [Smith, 1999]. Several factors have been implicated as potential contributors to stuttering etiology, including genetics and heritability, sex, and co-occurrence of additional speech and language difficulties in childhood. PWS are more likely than typical speakers to have a relative who stutters [Ambrose et al., 1993]. Additionally, stuttering across the lifespan is more common in males than females [Ambrose et al., 1997].

All speakers produce disfluencies, or breaks in the smooth transitions within and between words, during speech production. Typical speech disfluencies produced by both people who do and do not stutter include interjections (e.g. He wants the umbrella), phrase repetitions (I want I want the car) and revisions (Where did I put Oh, there it is). On the other hand, so-called ”stuttering-like” disfluencies (SLDs) are seen with greater frequency in PWS. Stuttering-like disfluencies include sound and syllable repetitions (e.g. b-b-boy), sound prolongations (e.g. hhhhhhhhello), and blocks, which are characterized by cessation in voicing and articulator movement (e.g. no), [Guitar, 2014]. An important feature of stuttering is that it is highly variable within and between PWS. In general, PWS produce fluent speech more often than disfluent speech.

The lifetime incidence of stuttering, or the number of people who report experiencing stuttering at some point in their life, is estimated to be between 5% and 9% [Mnsson, 2000, Felsenfeld et al., 1997, Dworzynski et al., 2007]. The prevalence, or the number of people who report experiencing stuttering at a given time, typically hovers around just less than 1% [Craig et al., 2002], but has been reported to be as high as 2.12% [Gillespie and Cooper, 1973]. The incidence of stuttering is relatively
high because stuttering emerges and is fairly common in the preschool years, with one study suggesting that 17.7% of preschoolers experience stuttering [Mnsson, 2000]. However, approximately three out of four preschoolers who experience stuttering exhibit so-called unassisted or "spontaneous" recovery. Unassisted recovery refers to a natural cessation of stuttering-like behaviors without therapeutic intervention, and typically occurs within one to two-years post-onset. The preschoolers who do not recover (approximately one out of four) will persist in stuttering across the lifespan [Yairi and Ambrose, 1992].

As a person stutters throughout his or her lifetime, the disorder and it’s behavioral manifestations evolves into a more complex problem that may include cognitive (negative thoughts) and affective (negative emotions) components. Holistically, stuttering frequently affects social interactions and can negatively impact how people view themselves as communicators and general members of society. For some PWS, these can interfere more with communicating than the behavior itself.

3.2.2 Stuttering, employment, and stigma

Employment experiences of PWS

Several studies have explored occupational experiences of PWS [Rice and Kroll, 1994, Rice and Kroll, 1997, Crichton-Smith, 2002, Hayhow et al., 2002, Klein and Hood, 2004, Palasik et al., 2012, McAllister et al., 2012, Bricker-Katz et al., 2013]. In a survey study with 282 PWS, over half of participants agreed that employers had misjudged their skillsets because of their stuttering. Sixteen percent of participants indicated that a potential employer directly told them that they were not hired for a job because of stuttering [Rice and Kroll, 1994]. In a similar survey, over 40% of the 232 PWS polled agreed that their job choice and earnings had been affected by stuttering. Specifically, 50% indicated that they sought out employment that required little speaking, 41% agreed that they would have a different job if they did not stutter, and 38% believed they were earning less money than they would if they
did not stutter. In terms of getting hired and promoted, 70% of participants believed that their chances were decreased, and 20% had declined a job or promotion because of their stuttering [Klein and Hood, 2004]. [Palasik et al., 2012]s survey findings also demonstrate that at least some PWS experience work related difficulties and 43% of the 184 PWS polled reported experiencing occupational discrimination.

These studies yield mixed findings as to whether specific groups of people are more adversely affected in the workplace by their stuttering. The majority of studies indicate that males, non-Caucasians, and people without prior speech-therapy experience are more likely to believe that stuttering negatively impacted their employability and job performance [Klein and Hood, 2004]. However, one study did not find differences between these groups and their counterparts [Palasik et al., 2012].

To date, only one group of authors has attempted to conduct a quantitative comparison of labor market outcomes between PWS and PWNS. [McAllister et al., 2012] used a longitudinal British cohort dataset that is similar to ours in an attempt to study the impact of stuttering on labor market outcomes. However, they recode the earnings variable into a dichotomous variable indicating whether the respondent earns more or less than the average level of earnings in the dataset, which provides an incomplete analysis of the impact of stuttering on earnings. Additionally, they have no measure of employment status, labor force participation, or underemployment, which is where much of the impact of stuttering on labor market outcomes may occur. Given the discrepancy between reports of negative occupational experiences from PWS and the null findings in their large-scale study, it is clear that more research is needed to capture the nuanced difficulties reported by PWS.

**Public-stigma**

One of the best constructs for framing the factors that contribute to differences in labor market outcomes between PWS and PWNS is stigma and how it impacts PWS. Current models of stigma have broken the phenomenon down into two forms,
public-stigma and self-stigma [Corrigan et al., 2009]. Stigma refers to a personal mark or characteristic that devalues or discredits an individual in society [Goffman, 1963, Jones et al., 1984]. In the workplace, PWS not only have to manage their stuttering, but they also have to manage the stigma associated with stuttering.

Stereotypes, prejudice, and discrimination comprise public-stigma [Corrigan et al., 2011, Boyle and Blood, 2015]. Stereotypes about PWS have been studied extensively in the speech-language pathology literature [Kalinowski et al., 1996]. Research has shown that teachers, college students, speech-language pathologists, and general members of the public hold negative stereotypes about PWS [Cooper and Cooper, 1996, Dorsey and Guenther, 2000, Craig et al., 2003, Boyle et al., 2009], including that they are afraid, nervous, anxious, passive, avoidant, guarded, tense, introverted, shy, embarrassed, frustrated, insecure, timid, self-derogatory, reticent, or sensitive [Kalinowski et al., 1996, MacKinnon et al., 2007].

Employer-held stereotypes of PWS have also been specifically explored [Hurst and Cooper, 1983]. In this study, 644 employers were surveyed about their attitudes toward employees who stutter. Eighty-four percent of employers indicated that they believed stuttering decreases a persons employability at least somewhat, with 10% of total respondents indicating that employability is decreased a great deal. Additionally, 29% of employers indicated that stuttering interferes with job performance, and 40% reported that stuttering interferes with promotion opportunities. When asked who should be hired if two equally qualified applicants applied for employment, one being a PWS and one being a PWNS, 61.9% indicated that the PWNS should be hired [Hurst and Cooper, 1983].

One way in which stereotypes and prejudices can lead to discrimination is through role entrapment of the stigmatized group. Role entrapment occurs when members of the majority group (e.g. PWNS) define roles that are appropriate for members of the minority group (e.g. PWS) [Smart, 2001, Gabel et al., 2004]. Occupational role entrapment results in limited career choices for PWS. For example, 43.6% of employers agreed that PWS should seek employment where little speaking is required [Hurst
and Cooper, 1983]. In another study, 385 college students were asked to indicate appropriate careers for PWS. Of the 43 total career choices, 20 choices (including attorney, judge, speech-language pathologist, and minister) were rated as significantly less appropriate for PWS than PWNS [Gabel et al., 2004].

**Self-stigma**

Self-stigma is another form of stigma. Self-stigma occurs when a person who possesses a marked characteristic or trait internalizes and embeds an associated public-stigma into his or her self-concept, and often results in reduced self-esteem and self-efficacy [Smart, 2001, Corrigan and Watson, 2002]. The phenomenon of self-stigma is not inherent to the individual, but develops through sociocultural interaction [Barreto and Ellemers, 2010, Boyle and Blood, 2015]. Additionally, self-stigma is not a phenomenon specific to PWS. Drawing from the broader disability literature, many people with a variety of disabilities have received and accepted society’s message that they are inferior [Smart, 2001]. This phenomenon is more likely to occur in people with uncommon disabilities, as they may have fewer opportunities to disprove the stereotypes and may be isolated from positive models of others with the same disability [Smart, 2001].

In the qualitative research literature, some PWS report awareness of stereotypes related to stuttering. PWS have said that they sometimes believe others view them as mentally defective, stupid, introverted, not very bright, weird, incompetent, a failure, socially crippled, inferior, not normal, or having something major wrong with [them] [Corcoran and Stewart, 1998, Klompas and Ross, 2004, Plexico et al., 2009, Boyle, 2013, Bricker-Katz et al., 2013]. Some people applied these stereotypes to themselves and reported feeling hopeless or suffering from reduced self-worth, self-esteem or self-identity [Klompas and Ross, 2004, Plexico et al., 2009].

[Boyle, 2013] developed the Self-Stigma of Stuttering Scale (4S), a psychometrically sound assessment tool that was normed from responses from 291 PWS. In this
sample, 86% of participants demonstrated high levels of awareness of stigma associated with stuttering, 19% agreed that stigma associated with stuttering may apply to other PWS, and 39% indicated that the stigma applied to themselves. In this sample, stigma at all levels (awareness, agreement, and application) was associated with increased levels of anxiety and depression, as well as decreased feelings of hope, empowerment, quality of life, and social support [Boyle, 2015]. No research has ever been conducted that attempts to capture how self-stigma may affect occupational outcomes in PWS.

### 3.3 Data

Data for the study comes from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States during the 1994-95 school year, which was conducted by the Carolina Population Center at the University of North Carolina at Chapel Hill. The Add Health cohort has been followed into young adulthood with four in-home interviews, the most recent of which occurred in 2008, when the respondents were aged 24-32. Add Health combines longitudinal survey data on respondents social, economic, psychological, and physical well-being with contextual data on the family, neighborhood, community, school, friendships, peer groups, and romantic relationships. Add Health data has been used to study the effect of a variety of worker characteristics on labor market outcomes, such as obesity [Norton and Han, 2008], attractiveness [Fletcher, 2009], adolescent depression [Fletcher, 2013], ADHD [Fletcher, 2014], and migraine headaches [Rees and Sabia, 2015].

---

3Eighty high schools from the US were selected using sampling and stratification methods to ensure that the selected schools were representative of US schools with respect to region of country, urbanicity, size, type, and ethnicity. Selected schools that did not enroll 7th graders helped to identify feeder schools that contribute students to the high school. Fifty-two such schools were added, bringing the total number of schools in the study to 132, covering 80 communities in the US. See [http://www.cpc.unc.edu/projects/addhealth/design](http://www.cpc.unc.edu/projects/addhealth/design) for additional information on the Add Health sampling process.
During Wave I of the in-home interviews, a core sample of 12,105 students were selected at random from the 132 schools. Supplemental samples of students were selected with a special emphasis on collecting information from ethnic minorities, students with physical disabilities involving the use of their limbs, and siblings. Additionally, the survey was given to every student enrolled at 16 of the 132 schools. The total number of respondents was 20,745. Three waves of follow-up in-home interviews have been conducted since the initial wave. Wave II was conducted one year after the first wave, during the 1995-96 school year. Wave III occurred in 2001-02, when respondents were between 18 and 26 years old. The most recent wave, Wave IV, occurred during 2007-08 when respondents were 24-32 and includes detailed labor market information.

3.3.1 The measure of stuttering

Wave III and Wave IV ask respondents, "Do you have a problem with stuttering or stammering?" Approximately 4% of the sample (208 out of the 5,114 respondents in the public-use data) answered yes to this question. As discussed in Section 2, the consensus is that around 1% of the adult population stutters, although some studies have reported higher estimates. Most likely, there are people in the sample who answered yes to the question about stuttering that would not be clinically diagnosed as a PWS. This could occur because the respondent stuttered as a child and has since spontaneously recovered, but still identified as a PWS in the survey due to misunderstanding the longitudinal nature of stuttering. It could also occur because some people confuse "normal" disfluencies (e.g., interjections, phrase repetitions, and revisions that occasionally occur during normal speech) with stuttering.

We made some attempts to separately identify people who actually stutter from those who used to stutter and those who confuse normal disfluencies with stuttering. Wave IV asks respondents if they would describe their stuttering as mild, moderate,
or severe. Keeping only those who reported their stuttering as moderate or severe greatly reduces the estimated prevalence of stuttering. Only including this group in the analysis, or showing how the effect of stuttering differs according to stuttering severity, was considered. However, this was unfeasible given that only 28 people in the public-use dataset reported moderate or severe stuttering. Additionally, self-assessment of stuttering severity may be correlated with factors such as self-esteem, which would also influence labor market outcomes and thus introduce endogeneity into the measure of stuttering. We also attempted to address this issue by limiting the sample of PWS to those who reported that they stutter in both Wave III and Wave IV. Given that both the onset of and spontaneous recovery from stuttering typically occur during childhood, a respondent who reports that they stutter in one wave but not the other is more likely to be a person who is confusing normal disfluencies with stuttering. This slightly decreases the estimated prevalence, but the results using this sample are nearly identical to those reported later in the paper and are therefore not shown.

Including PWNS in the sample of PWS would result in attenuation bias in the estimated effect of stuttering on labor market outcomes toward zero. Given that 4% is much higher than the general consensus on the prevalence of stuttering into adulthood, a large fraction of our sample of PWS may be PWNS. If this is the case, then the results presented later in the paper may underestimate the impact of stuttering on labor market outcomes.

3.3.2 Outcome variables

The labor market outcome variables are log hourly earnings and indicator variables for whether the respondent is employed, in the labor force (employed or unemployed but looking for a job), underemployed, and on public assistance. Each of these outcome variables comes from Wave IV. The earnings data come from the following question: "Now think about your personal earnings. In [previous year] how much
income did you receive from personal earnings before taxes, that is, wages or salaries, including tips, bonuses, and overtime pay, and income from self-employment?" This was converted into an hourly earnings variable using the next question, asked in reference to their current job: "How many hours per week do you usually work at this job?" We then multiplied this number by 52 to get annual hours and then divided annual earnings by annual hours\(^5\).

The outcome variable for whether the respondent is employed is taken from a question that asked respondents if they are currently working for pay at least 10 hours per week. The employed variable is equal to 1 if the respondent answered yes to this question and 0 otherwise. The labor force variable is equal to 1 if the respondent is working for pay at least 10 hours per week or if they reported that they are unemployed but looking for a job and 0 if the respondent neither is working at least 10 hours per week nor is looking for a job. The underemployed variable was created by matching the SOC occupation codes in the Add Health data to the Bureau of Labor Statistics' data on the typical level of education for each occupation. If a respondent had a higher level of education than the typical worker in that occupation, then the respondent was coded as being underemployed. The public assistance variable is equal to 1 if the respondent indicated that they have ever been enrolled in a public assistance program such as welfare and 0 otherwise.

Summary statistics in Table 3.1 and Table 3.2 show the mean and standard deviation for each of these labor market outcomes and are shown separately according to whether the respondent is a PWS.

### 3.3.3 Labor market characteristics, individual characteristics, and family characteristics

Stuttering may impact labor market outcomes through many channels, including through labor market characteristics such as education level, occupation, and per-

\(^5\)Results based on annual earnings are very similar to the results shown later for hourly earnings.
sonality characteristics like self-esteem and self-efficacy. Data on these characteristics will be useful when attempting to account for the cause of differences in labor market outcomes between PWS and PWNS. Table 3.1 shows summary statistics for education levels, average hourly earnings in the respondent’s occupation, and a variable that measures self-stigma, which we construct to account for personality differences between PWS and PWNS.

As described in Section 2, some PWS have been shown to exhibit self-stigma, which occurs when a person who possesses a marked characteristic or trait (in this case, stuttering) internalizes and embeds an associated public-stigma into his or her self-concept, and often results in reduced self-esteem and self-efficacy. In this sense, accounting for self-stigma is similar to other studies that have accounted for self-esteem, which has been shown to have significant impacts on labor market outcomes [Heckman et al., 2006, Fortin, 2008]. Our measure of self-stigma is constructed using answers to a subset of six questions asked in Wave IV of the Add Health data. These questions were chosen because they align closely with common stereotypes about PWS. The exact questions and the stereotype that they align with are shown in the appendix. The response to each of these questions was re-coded into a binary variable (1 or 0), with a 1 indicating that the respondent gave an answer that aligns with the stereotype. This was done for both PWS and PWNS. Each of these six indicator variables were then summed together and re-scaled into a composite variable.

\[\text{We calculate mean occupation earnings by taking the average hourly earnings for each occupation in the Add Health data and then matching this variable to each respondent’s occupation. In the analysis below, occupation differences will be accounted for by using occupation fixed effects.}\]

\[\text{We measure self-stigma in Wave IV because self-stigma occurs over time as a stigmatized group is exposed to and then internalizes the public-stigma into his or her self-concept. Much of this exposure to public stigma may occur when PWS enter the labor force, so we choose to measure self-stigma in Wave IV. However, reverse causality is a potential concern; people may be more likely to answer survey questions in a way that suggests that they are quiet, nervous, lacking assertion, etc. (some of the common stereotypes of PWS, as shown in the appendix) because of struggles in the labor market that are unrelated to any discrimination or stereotyping that they have experienced. Therefore, we also used a measure of self-stigma constructed from Wave I. Results using this measure are similar to those shown in the paper, although self-stigma appears to play a somewhat smaller role in explaining the difference in labor market outcomes between PWS and PWNS when measured using Wave I.}\]
ranging from 0 to 1. PWS, on average, have statistically significant higher levels of this self-stigma index, which suggests that we are capturing differences in self-esteem and self-efficacy that are correlated with stuttering. Additionally, the difference in the response to each of the individual questions between PWS and PWNS was statistically significant at the one-percent level, with PWS responding in ways that aligned with the internalization of public stigmas.

Summary statistics for individual and family characteristics are shown in Table 3.2. These include the respondent’s age, gender, race, ethnicity, attractiveness, height, and weight at Wave IV and information on their parent’s income, education, and marital status at Wave I. We also use the school that the respondent was enrolled in during Wave I to capture difficult-to-measure differences in school and neighborhood quality. Table 3.1 also shows summary statistics for tenure (years at current job) and weekly hours.

3.4 Methodology

3.4.1 Regression analysis and propensity score matching

The first method that we use to quantify the impact of stuttering on labor market outcomes is regression analysis using ordinary least squares (OLS) for hourly earnings regressions and probit estimation for binary outcome variables. We first estimate:

\[ Y_i = \alpha + \beta PWS_i + \varepsilon_i, \]

where \( Y_i \) is one of the outcome variables discussed in the previous section and \( PWS_i \) is an indicator for whether the respondent is a PWS. In this equation, the estimate of \( \beta \) captures the total difference in outcomes between PWS and PWNS without

---

8The Cronbach’s alpha, a measure of reliability that indicates whether the responses to each question comprising the composite index are highly correlated, is 0.43. This is very similar to that of scales used in economics to measure self-esteem and locus of control [Waddell, 2006, Fortin, 2008].
controlling for any characteristics of the respondent, \( \alpha \) is a constant, and \( \varepsilon_i \) is an individual-specific error term.

Next, we add covariates to the equation to control for individual and family background characteristics:

\[
Y_{is} = \alpha + \beta PWS_i + \phi' X_i + z_s + \varepsilon_{is},
\]  

(3.2)

where \( X_i \) is a vector of individual and family background characteristics (age, gender, race, ethnicity, attractiveness, height, weight, parent’s income, parent’s education, and parent’s marital status) and \( z_s \) is a fixed effect for each of the 132 schools in Wave I. The covariates control for differences in observable background characteristics and the school fixed effects capture unobservable differences related to school or neighborhood factors. The estimate of \( \beta \) in this equation captures the difference in outcomes between PWS and PWNS that is due to reasons other than differences in these background characteristics.

If stuttering is uncorrelated with any remaining individual or family heterogeneity left over in the error term \( \varepsilon_{is} \) from equation (2), then the estimate of \( \beta \) will capture the total causal effect of stuttering on the given labor market outcome; after controlling for individual, family, and school/neighborhood background characteristics, any remaining difference in labor market outcomes between PWS and PWNS is likely an effect of stuttering itself. However, it is possible that there are unobservable differences in background characteristics remaining in the error term that are correlated with stuttering and could cause bias in our estimate of the effect of stuttering on labor market outcomes.

One potential confounder not included in equation (2) is ADHD and learning disabilities. Stuttering has been shown to be correlated with both ADHD and some other learning disabilities\(^9\). [Rees and Sabia, 2014] showed that controls for ADHD

\(^9\)Estimates of the prevalence of concomitant ADHD in children who stutter have ranged from 4% to 26% [Riley and Riley, 2000, Conture, 2001, Arndt and Healey, 2001]. This may be similar to or higher than the national prevalence of ADHD in children in the United States, which ranges from 5% to 11% [Association, 2013, Visser et al., 2014]. Phonological and/or language disorders have
and learning disabilities significantly reduce the estimated effect of stuttering on human capital accumulation. Table 3.2 confirms that stuttering is highly correlated with both ADHD and learning disabilities. Therefore, we also estimate the following specification, which adds controls for ADHD and learning disabilities:

$$Y_{is} = \alpha + \beta PW_{Si} + \phi'X_i + \delta ADHD_i + \gamma LD_i + z + \varepsilon_{is},$$

(3.3)

where $ADHD_i$ and $LD_i$ are binary variables indicating ADHD and learning disabilities, respectively. The variable for ADHD comes from a question in Wave IV that asks, "Has a doctor, nurse or other health care provider ever told you that you have or had: attention problems or ADD or ADHD?" The variable for learning disabilities comes from a question in Wave I that asks, "Does (he/she) have a specific learning disability, such as difficulties with attention, dyslexia, or some other reading, spelling, writing, or math disability?"

Another way to address differences between PWS and PWNS in terms of observable characteristics is the use of propensity score matching. This approach will produce more reliable estimates than regression analysis if the two groups lack common support. [Rees and Sabia, 2014] found that using propensity score matching also significantly reduces the estimated effect of stuttering on human capital accumulation. Therefore, we perform nearest neighbor matching, which first estimates the propensity of each respondent to be a PWS based on the covariates described above:

$$Pr(PWS_i = 1) = 1 - \Phi(-\alpha - \phi'X_i - \delta ADHD_i - \gamma LD_i - z).$$

Then, we use nearest neighbor matching with replacement to match each PWS to three PWNS with a similar estimated propensity. The propensity score matching estimates for the effect of stuttering on labor market outcomes are calculated by

also been shown to be more prevalent in children who stutter than in the general population, with estimates of comorbidity with stuttering as high as 42% [Arndt and Healey, 2001]. Comorbidity of stuttering and learning disabilities or literacy disorders is also estimated to be higher than in the general population, with estimated comorbidities of 15.2% and 8.2%, respectively [Blood et al., 2003]
comparing the outcomes of PWS to the outcomes of their match. Specifically, the
estimate is the mean difference in outcomes between PWS and their match. We limit
the sample to matches whose estimated propensity scores were within 0.004. We also
dropped the 2% of PWS whose propensity score was furthest from the propensity
score of their nearest match\(^{10}\).

These methods will estimate the total causal impact of stuttering on labor market outcomes, conditional on individual and family background characteristics and co-
occuring disabilities, assuming that there are no remaining unobserved confounders\(^ {11}\).

3.4.2 Labor market characteristics and Blinder-Oaxaca decomposition

We are also interested in understanding the causes of the remaining difference in labor market outcomes between PWS and PWNS after controlling for individual and family background characteristics and co-occurring disabilities. There are several channels through which stuttering could impact labor market outcomes, including differences in education levels, personality characteristics such as self-esteem and self-
efficacy, occupations, job performance, and employer discrimination. We therefore add several additional covariates to equation (3) in order to control for differences in observable labor market characteristics:

\[
Y_{is} = \alpha + \beta PWS_i + \phi' X_i + \delta ADHD_i + \gamma LD_i + \sigma' LM_i + z_s + \varepsilon_{is}, \tag{3.4}
\]

\(^{10}\)We chose to match with replacement and to match to multiple neighbors ("oversampling") because this approach did the best job of matching covariate means between PWS and PWNS. Results are very similar if we match without replacement, match to a different number of neighbors, or match to a single neighbor. Results are also very similar for different caliper widths and different trimming percentages.

\(^{11}\)[Rees and Sabia, 2014] also use family fixed effects to address potential unobserved differences in family characteristics that may be correlated with stuttering and human capital accumulation. The public-use version of the data does not provide sibling identifiers and therefore does not allow for family fixed effects. However, as discussed in the introduction, [Rees and Sabia, 2014] find that family fixed effects, propensity score matching, and controls for ADHD and learning disabilities all produce very similar results that are significantly different from baseline OLS results. Therefore, the propensity score matching and specifications with controls for ADHD and learning disabilities should be sufficient to indicate whether baseline OLS results are entirely spurious or if there is an effect of stuttering on labor market outcomes.
where $LM_i$ is a vector of controls for education level and self-stigma. For earnings regressions, it also includes occupation fixed effects, job tenure, average weekly hours worked, and limits the sample to full-time workers (at least 35 hours per week).

Results from this equation will show how much of the remaining difference in labor market outcomes between PWS and PWNS after controlling for individual and family characteristics and co-occurring disabilities can be explained by differences in observable labor market characteristics. If a large fraction of the difference can be explained by these labor market characteristics, that does not mean that there is no effect of stuttering on labor market outcomes. Rather, it would indicate that the effect of stuttering on labor market outcomes occurs through these observable labor market characteristics.

The estimate of $\beta$ from equation (4) will show the total difference in labor market outcomes that remains between PWS and PWNS after controlling for all background characteristics and labor market characteristics. But we are not just interested in quantifying the unexplained difference. We are also interested in accounting for the relative importance of each of the background and labor market characteristics in explaining the total difference in earnings\textsuperscript{12}. Understanding the most important factors that account for earnings differences between PWS and PWNS is especially important given that this is a policy relevant question; the relative importance of immutable individual and family background characteristics, labor market characteristics, and unexplained factors informs the relative importance of access to speech therapy, the focus of speech therapy (self-stigma versus fluency), and anti-discrimination policies that support PWS in the workforce. The desire to account for the causes of earnings differences between groups is also relevant in the literature on gender and racial earn-

\textsuperscript{12}As explained in [Gelbach, 2014], accounting for the relative importance of different covariates in explaining a difference in outcomes across groups cannot be done by simply observing how the coefficient of a group dummy variable in a linear regression changes as covariates are added. This issue is known as sequencing sensitivity or path dependence. Rather, decomposition methods are required.
ings differences. This is commonly done by using Blinder-Oaxaca (BO) decomposition [Blinder, 1973, Oaxaca, 1973], which we adopt for the same purpose.\[^{13}\]

Blinder-Oaxaca decomposition is straightforward. Let $X^BO_i$ be a vector that contains some combination of the covariates described above. That is, $X^BO_i = [X'_i, ADHD_i, LD_i, LM'_i, z_s]$, or any subset of those covariates. For simplicity in exposition below, assume that $X^BO_i$ contains $K$ of these covariates. Assuming a linear regression approach as above, we can write hourly earnings as a function of these covariates separately for PWS and PWNS:

$$Y_{gi} = \alpha_g + \sum_{k=1}^{K} X^BO_{ik} \beta_{gk} + \varepsilon_{gi}, \quad g = \text{PWS, PWNS}. \quad (3.5)$$

Then, the overall difference in earnings between PWS and PWNS,

$$\hat{\Delta} = \bar{Y}_{PWNS} - \bar{Y}_{PWS},$$

can be rewritten as:

$$\hat{\Delta} = \sum_{k=1}^{K} (\bar{X}^BO_{PWNS,k} - \bar{X}^BO_{PWS,k}) \hat{\beta}_{PWS,k} \quad (3.6)$$

$$+ \sum_{k=1}^{K} \bar{X}^BO_{PWNS,k} (\hat{\beta}_{PWNS,k} - \hat{\beta}_{PWS,k}) + (\hat{\alpha}_{PWNS} - \hat{\alpha}_{PWS}).$$

where $\hat{\alpha}_g$ and $\hat{\beta}_{gk}$ are the estimated intercept and slope coefficients from the model in (5) for $g = \text{PWS, PWNS}$ and $\bar{X}_{gk}$ is the average value of variable $k$ for group $g$.

The first term on the right-hand side of (6) is usually called the ”explained” effect or ”characteristics” effect in Blinder-Oaxaca decomposition, because it represents the fraction of the difference in outcomes that is due to differences in observable characteristics. The next two terms on the right-hand side of (5) are usually called the ”unexplained” effect. They are also often referred to as the ”coefficients effect” or ”wage effect”, because they represent the fraction of the earnings difference that is

\[^{13}\text{See [O’Neill and O’Neill, 2006] for an example of decomposition methods applied to gender and racial discrimination and [Fortin and Firpo, 2011] for a detailed overview of decomposition methods.} \]
due to differences in the return to characteristics, captured by differences in the coefficients. This unexplained effect can also be interpreted as a treatment effect [Fortin and Firpo, 2011]. In earnings studies, it is commonly assumed to capture employer discrimination, although it also captures the effect of any unobserved differences between groups.

Thus, in this application, we can decompose the total hourly earnings difference between PWS and PWNS into the explained difference due to various background and labor market characteristics and the unexplained difference due to employer discrimination or other unobserved differences between PWS and PWNS. More importantly, we can further breakdown the explained fraction of the earnings difference into the fraction that is explained by each background and labor market characteristic. This provides a detailed accounting for the relative influence of each of the background and labor market characteristics on the total hourly earnings difference between PWS and PWNS.

3.5 Results

The main results are reported in Tables 2-7. Each table reports pooled results for males and females and also reports results separately by gender. Standard errors are clustered by school of residence in Wave I in all tables.

3.5.1 Regression analysis

Table 3.3 shows the baseline regression results. The table shows the impact of stuttering on each of hourly earnings, employment, labor force participation, under-

\footnote{One of the primary drawbacks of Blinder-Oaxaca decomposition - that it inherently follows a partial equilibrium approach - is less of a concern in this application than applications to gender and racial differences. The male (White) wage structure does not provide a proper counterfactual for female (Black) workers in the absence of discrimination, due to general equilibrium wage effects that would occur; changing the female (Black) wage structure would also affect the male (White) wage structure. However, due to the relatively small size of the stuttering population, general equilibrium effects would likely be very small in this setting.}
employment, and receipt of public assistance separately by panel. Pooled results are reported in columns (1)-(2) and results for males and females are reported in columns (3)-(4) and (5)-(6), respectively. The first column for each group shows the total unconditional difference, corresponding to the specification in equation (1). The next column adds controls for background characteristics and corresponds to equation (2). Each panel-by-column result is based on a separate regression. We only report the coefficient on the stuttering variable for the purpose of brevity in the table.

Pooled results in column (1) show that stuttering appears to have a significant impact on hourly earnings, employment status, labor force participation, and receipt of public assistance. The estimated effect of stuttering on each of these outcomes is statistically significant at at least the 5% level. When controls for detailed individual and family background characteristics are added in column (2), the estimated effect of stuttering on labor market outcomes decreases in magnitude by approximately 25% - 40% depending on the outcome variable, but statistically significant differences remain, except for employment status.

However, the pooled results mask significant gender heterogeneity. For males, stuttering is associated with a 19% reduction in hourly earnings, a 12% reduction in the probability of being employed, a 14% reduction in the probability of being in the labor force, and a 12% increase in the probability of having ever received public assistance. Each of these estimates are statistically significant at at least the 5% level. Conditional on being employed, stuttering has no impact, statistically or economically, on the probability of being underemployed.

For females, the results are nearly the opposite; stuttering is associated with a significant effect on hourly earnings and underemployment, but no effect on the probability of being employed, being in the labor force, or having ever received public assistance. The coefficient for the stuttering indicator variable suggests a 16% reduction in hourly earnings and a 10% increase in the probability of being underemployed. Each of these estimates are statistically significant at the 10% level. Stuttering has no impact, statistically or economically, on the other labor market outcomes for females.
3.5.2 Controls for individual heterogeneity

We address concerns about individual-level heterogeneity that may be correlated with stuttering and labor market outcomes by adding controls for ADHD and learning disabilities in Table 3.4. The first column for each group (all, males, females) is based on equation (2), which controls for individual and family background characteristics. The next column adds controls for ADHD and learning disabilities, corresponding to equation (3). Because of missing observations for ADHD and learning disabilities, the sample size decreases compared to Table 3.3.

The magnitude of the effect of stuttering on labor market outcomes decreases by approximately 10% to 30%, depending on the outcome variable and group, when controls for ADHD and learning disabilities are added. Given that stuttering does have a high co-occurrence with ADHD and other learning disabilities, a decrease in magnitude is expected. However, the estimates reported in Table 3.3 do not appear to be entirely caused by other co-occurring disabilities.

For males, stuttering is associated with a statistically significant impact on the same outcome variables as before: hourly earnings, employment, labor force participation, and receipt of public assistance. The effect is also still economically significant; stuttering is associated with a 14% reduction in hourly earnings, a 12% decrease in the probability of being employed, a 14% increase in the probability of being in the labor force, and a 13% increase in the probability of having ever received public assistance.

For females, the results are also very similar to Table 3.3; stuttering is associated with an economically large reduction in hourly earnings of 14% and an increase in the probability of being underemployed of 8%. However, the estimate for the effect of stuttering on these two outcomes is no longer statistically significant. Nonetheless, stuttering does appear to impact these two outcomes for females, even if the estimates are not measured precisely enough to be statistically significant.
3.5.3 Propensity score matching

Table B.2 and B.3 in the appendix shows the means of the covariates before and after matching. In the unmatched sample, PWS are significantly more likely to be male, Black, and Native American. PWS are also less attractive, taller, weigh more, and are more likely to have ADHD or other learning disabilities. In the matched sample, no statistically significant differences between PWS and PWNS remain.

The propensity score matching results are shown in Table 3.5. The first column for each group (all, males, females) shows the mean difference in each labor market outcome variable for the unmatched sample. The next column shows the mean difference for the matched sample. Results based on the unmatched sample are very similar to the results reported in previous tables. Propensity score matching results based on the matched sample are very similar to results reported in Table 3.4.

For males, stuttering is associated with a significant reduction in hourly earnings, a decrease in the probability of being employed, a decrease in the probability of being in the labor force, and an increase in the probability of having ever received public assistance. For females, stuttering is again associated with an economically large reduction in hourly earnings of 9.3% and increase in the probability of being underemployed of 12.2%, although neither is statistically significant, as in Table 3.4.

3.5.4 Labor market characteristics and Blinder-Oaxaca decomposition

The previous results suggest that stuttering is associated with a significant impact on several labor market outcomes. This impact could occur through many possible channels, including differences in education levels, personality characteristics such as self-esteem and self-efficacy, occupations, job performance, and employer discrimination. Therefore, Table 3.6 adds controls for several labor market characteristics in order to better understand the role of these channels. These characteristics include education level and self-stigma. Results for log hourly earnings also include occupation fixed effects, controls for tenure and hours, and are based on full-time workers.
The first column for each group (all, males, females) controls for background characteristics, ADHD, and learning disabilities. The next column adds controls for these labor market characteristics, which corresponds to equation (4).

In almost all cases, the magnitude of the estimated effect of stuttering is reduced when controls for labor market characteristics are included. In several cases, statistical significance is either removed or reduced by the inclusion of these controls. This confirms that a portion of the difference in labor market outcomes between PWS and PWNS can be explained by differences in these labor market characteristics. However, the differences are not entirely removed.

For males, there is still a statistically significant effect of stuttering on the probability of being employed, the probability of being in the labor force, and the probability of having ever received public assistance. Much of the difference in hourly earnings associated with stuttering for males is removed when labor market characteristics are included. For females, there are no statistically significant estimates, but stuttering does seem to still have an economically significant effect on hourly earnings and underemployment; the estimates for these outcomes are large, just imprecisely measured. Unlike for males, much of the effect of stuttering on hourly earnings remains for females after adding controls for labor market characteristics.

Table 3.7 shows Blinder-Oaxaca decomposition of the effect of stuttering on log hourly earnings. The results are based on the same specifications from Table 3.6; the first column for each group (all, males, females) controls for background characteristics, ADHD, and learning disabilities and the next column adds controls for labor market characteristics. Visual summary of the Blinder-Oaxaca results is shown in Figure 3.1.

Panel I of Table 3.7 shows the total unconditional difference in hourly earnings between PWS and PWNS. This corresponds to the results in Table 3.3 without background characteristics, although the estimates are slightly different due to missing observations for ADHD, learning disability, or labor market characteristics. The next two rows show the portion of the total difference that is due to differences in ob-
servable characteristics and the difference that is due to differences in coefficients, respectively. The difference due to coefficients is equivalent to the estimates for the impact of stuttering on hourly earnings in Table 3.6.

The purpose of the Blinder-Oaxaca results is to account for the relative importance of various background and labor market characteristics in explaining the hourly earnings difference between PWS and PWNS. This is shown in Panel II, which shows the contribution of each covariate to the explained difference. For brevity in the table we group each of the background controls, ADHD, and learning disability together. The table then shows the fraction of the difference that is explained by differences in occupation, education level, self-stigma, tenure, and hours.

For males, column (4) in Panel I shows that 15.2 of the 21.3 percentage point reduction in hourly earnings can be explained by observable differences in background characteristics, ADHD and learning disabilities, and labor market characteristics. Panel II shows that much of that explained difference comes from differences in occupation; there is an 11.1 percentage point reduction in hourly earnings associated with stuttering for males that is attributable to differences in occupation. Education level and self-stigma also play a significant role; differences in education contribute a 3.9 percentage point reduction and differences in self-stigma contribute a 3.7 percentage point reduction. Each of these three covariates explain a statistically significant portion of the difference in earnings between males who stutter and males who do not stutter. The other covariates contribute negatively, meaning that conditional on equivalent occupations, education, and self-stigma, males who stutter have, on average, better background characteristics, more job tenure, and work more hours.

For females, column (6) in Panel I shows that only 8.2 percentage points of the 21.4 percentage point reduction in hourly earnings can be explained by differences in background characteristics, ADHD and learning disabilities, and labor market characteristics. Panel II shows that this explained portion is almost entire due to differences in education, which contributes a 7.4 percentage point reduction in hourly earnings.
earnings. None of the other covariates contribute significantly to the hourly earnings difference.

Table 3.8 shows how the Blinder-Oaxaca decomposition results change when self-stigma is removed from the specification. Differences in self-esteem and self-efficacy captured in the self-stigma variable could impact labor market outcomes in two ways: (1) by causing PWS to avoid education levels or occupations that require lots of speaking or (2) negatively impacting job performance and interactions with coworkers. For example, PWS may be reticent to speak up when they have knowledge to contribute to a discussion, less willing to seek assistance when needed, or less willing to accept promotions that involve increased speaking demands [Klein and Hood, 2004]. Seeing how the decomposition results change when self-stigma is removed should shed light on what our measure of self-stigma is capturing and how it impacts labor market outcomes; if our measure of self-stigma is impacting hourly earnings through differences in education levels and occupations, then the contribution of education and occupation to the hourly earnings difference should become larger when self-stigma is excluded. Alternatively, if self-stigma is mostly capturing difficult-to-measure impacts of stuttering on job performance and interactions, then the occupation and years of school components should remain mostly unchanged and the component previously explained by self-stigma should become unexplained.

Results in Table 3.8 show that the contribution of education and occupation to the hourly earnings difference are almost entirely unchanged when self-stigma is removed; much of what was previous explained by self-stigma becomes unexplained. For males, 2.3 of the 3.7 percentage points attributed to self-stigma in column (3) become unexplained in column (4). Much of the remainder becomes attributed to background characteristics. Self-stigma explained very little for females, but most of what it did explain in column (5) becomes unexplained in column (6). These results suggest that our measure of self-stigma captures difficult-to-measure impacts of stuttering on job performance and interactions that are mostly orthogonal to other observable characteristics, but are partially correlated with background characteristics.
3.6 Conclusion

Despite an abundance of research suggesting that stuttering is associated with significant hardships in the labor market [Rice and Kroll, 1994, Rice and Kroll, 1997, Crichton-Smith, 2002, Hayhow et al., 2002, Klein and Hood, 2004, Palasik et al., 2012, Bricker-Katz et al., 2013], no study has adequately quantified the impact of stuttering on labor market outcomes. We address this issue by using the National Longitudinal Study of Adolescent to Adult Health, which tracked adolescents into adulthood and collected labor market information and detailed background characteristics, including whether the respondent is a person who stutters.

We find that stuttering does have a significant impact on labor market outcomes and that the impact is different for males and females; for males, stuttering is associated with a decrease in the probability of being employed, a decrease in the probability of being in the labor force (employed or unemployed but looking for a job), and an increase in the probability of having ever received public assistance such as welfare. For females, we find no effect of stuttering on employment, labor force participation, or receipt of public assistance, but we do find that stuttering is associated with a significant increase in the probability of being underemployed (over-educated for their current job). Both males and females experience a significant reduction in hourly earnings associated with stuttering. For males, most of this difference can be explained by differences in occupations, education levels, and self-stigma. A much larger fraction remains unexplained for females.

The primary difficulty in quantifying the impact of stuttering on labor market outcomes is unobserved individual or family heterogeneity; if stuttering is correlated with unobserved confounders that also influence labor market outcomes, then results will be biased. [Rees and Sabia, 2014] showed that although stuttering appeared to be associated with a reduction in high school GPA, high school graduation, and college attendance, the effect of stuttering on these outcomes was removed when the authors controlled for co-occurring disabilities or used propensity score matching.
However, we find that the effect of stuttering on labor market outcomes is robust to these approaches. This result can be reconciled with the null findings in [Rees and Sabia, 2014] under the assumption that differences and difficulties in communication are more of a hardship in the labor market than in the classroom; employer decisions about hires and promotions likely provide more opportunities for discrimination than teacher or school decisions about grades and graduation.

To put the magnitude of these results in perspective, it is useful to compare them to those from studies on other health conditions and labor market outcomes. In specifications that control for ADHD and other co-occurring learning disabilities, we find that stuttering decreases hourly earnings by approximately 14% for both males and females; [Fletcher, 2014] found that ADHD reduces earnings by 33%; [Rees and Sabia, 2015] found that migraine headaches reduce earnings by 30% in specifications that use instrumental variables, but smaller effects using OLS; [Kidd et al., 2000] found that a broad definition of disability - including mobility, manual dexterity, physical co-ordination, continence, ability to lift/carry/move everyday objects, speech, hearing or eyesight, memory/ability to concentrate/learn/understand, or the perception of the risk of physical danger - reduces earnings by 14%; [Fletcher, 2013] found that adolescent depression reduces earnings by 15%. Therefore, the effect of stuttering on labor market outcomes may not be as great as some other health conditions, but the results are consistent with a growing literature that health conditions have significant effects on labor market outcomes.

While difficulties in the labor market are not unique to stuttering compared to other disabilities or minority statuses such as gender and race, the experience of PWS is unique in another sense: PWS may have less support than some people with other types of disabilities or minority statuses. Despite the prevalence of stuttering discussed above and the clear impact that stuttering can have on a person’s life, speech therapy for PWS is typically not covered by federal or private insurance companies [Reeves et al., 2006]. Additionally, while there are organizations that fight for disability rights for people with many other types of disabilities, no such organization
exists in the United States for PWS. In our current society, most PWS are neither protected from discrimination by employers nor are they afforded easy access to resources such as speech therapy, which can help PWS manage their stuttering and self-stigma in preparation for job interviews and occupations that require speaking. Given the prevalence of stuttering, the decrease in earnings associated with stuttering may represent significant lost revenue to the government in the form of taxable income. A policy that provided publicly-funded speech therapy for PWS could potentially be a net benefit to the government; increased access to speech therapy could both increase tax revenue and decrease PWS’s receipt of public assistance by helping PWS increase their fluency and reduce self-stigma. These financial benefits may offset the cost of providing publicly-funded speech therapy for PWS.
Table 3.1.: Summary statistics

<table>
<thead>
<tr>
<th>Labor market outcomes</th>
<th>PWNS</th>
<th>PWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employed</td>
<td>0.79</td>
<td>0.71</td>
</tr>
<tr>
<td>Hourly earnings</td>
<td>16.48</td>
<td>12.13</td>
</tr>
<tr>
<td>In labor force</td>
<td>0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>Underemployed</td>
<td>0.50</td>
<td>0.51</td>
</tr>
<tr>
<td>Ever received public assistance</td>
<td>0.24</td>
<td>0.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Labor market characteristics</th>
<th>PWNS</th>
<th>PWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than high school education</td>
<td>0.07</td>
<td>0.16</td>
</tr>
<tr>
<td>High school</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>Some college</td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td>College</td>
<td>0.20</td>
<td>0.14</td>
</tr>
<tr>
<td>More than college</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>Mean occupation earnings</td>
<td>16.34</td>
<td>14.69</td>
</tr>
<tr>
<td>Self-stigma</td>
<td>0.17</td>
<td>0.29</td>
</tr>
<tr>
<td>Tenure</td>
<td>4.26</td>
<td>4.30</td>
</tr>
<tr>
<td>Weekly hours</td>
<td>41.21</td>
<td>40.53</td>
</tr>
</tbody>
</table>

"PWS" = people who stutter. "PWNS" = people who do not stutter. Summary statistics are based on the public-use version of the Add Health data, which consists of 5,114 people in Wave IV. Of those 5,114 respondents, 208 said that they stutter. Each of the labor market outcomes and labor market characteristics are measured in Wave IV of the data, when respondents were 24-32 years old. Individual and family characteristics come from Wave I of the survey when respondents were in grades 7-12. All means and standard deviations are based on unweighted data.
Table 3.2.: Summary statistics (continued)

<table>
<thead>
<tr>
<th></th>
<th>PWNS</th>
<th>PWS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Individual characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>30.00</td>
<td>1.774</td>
</tr>
<tr>
<td>Female</td>
<td>0.54</td>
<td>0.498</td>
</tr>
<tr>
<td>Black</td>
<td>0.23</td>
<td>0.424</td>
</tr>
<tr>
<td>Native American</td>
<td>0.01</td>
<td>0.116</td>
</tr>
<tr>
<td>Asian</td>
<td>0.03</td>
<td>0.172</td>
</tr>
<tr>
<td>Other race</td>
<td>0.06</td>
<td>0.230</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.10</td>
<td>0.304</td>
</tr>
<tr>
<td>Very unattractive</td>
<td>0.03</td>
<td>0.178</td>
</tr>
<tr>
<td>Unattractive</td>
<td>0.04</td>
<td>0.203</td>
</tr>
<tr>
<td>Average attractiveness</td>
<td>0.47</td>
<td>0.499</td>
</tr>
<tr>
<td>Attractive</td>
<td>0.36</td>
<td>0.480</td>
</tr>
<tr>
<td>Very attractive</td>
<td>0.10</td>
<td>0.294</td>
</tr>
<tr>
<td>Height (feet)</td>
<td>5.58</td>
<td>0.327</td>
</tr>
<tr>
<td>Weight (pounds)</td>
<td>186.03</td>
<td>50.819</td>
</tr>
<tr>
<td>ADHD</td>
<td>0.05</td>
<td>0.220</td>
</tr>
<tr>
<td>Learning disability</td>
<td>0.12</td>
<td>0.321</td>
</tr>
<tr>
<td>Family characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family income as adolescent ($10,000s)</td>
<td>4.77</td>
<td>4.841</td>
</tr>
<tr>
<td>Parent-married</td>
<td>0.72</td>
<td>0.424</td>
</tr>
<tr>
<td>Parent-less than high school education</td>
<td>0.18</td>
<td>0.388</td>
</tr>
<tr>
<td>Parent-high school</td>
<td>0.26</td>
<td>0.438</td>
</tr>
<tr>
<td>Parent-some college</td>
<td>0.30</td>
<td>0.456</td>
</tr>
<tr>
<td>Parent-college</td>
<td>0.15</td>
<td>0.359</td>
</tr>
<tr>
<td>Parent-more than college</td>
<td>0.11</td>
<td>0.312</td>
</tr>
</tbody>
</table>

"PWS" = people who stutter. "PWNS" = people who do not stutter. Summary statistics are based on the public-use version of the Add Health data, which consists of 5,114 people in Wave IV. Of those 5,114 respondents, 208 said that they stutter. Each of the labor market outcomes and labor market characteristics are measured in Wave IV of the data, when respondents were 24-32 years old. Individual and family characteristics come from Wave I of the survey when respondents were in grades 7-12. All means and standard deviations are based on unweighted data.
Table 3.3: OLS and probit estimates of the effect of stuttering on labor market outcomes

<table>
<thead>
<tr>
<th></th>
<th>All (1)</th>
<th>Males (2)</th>
<th>All (3)</th>
<th>Males (4)</th>
<th>Females (5)</th>
<th>Females (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel I: Log Hourly Earnings</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.174**</td>
<td>-0.131*</td>
<td>-0.192***</td>
<td>-0.190**</td>
<td>-0.199*</td>
<td>-0.169*</td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td>(0.070)</td>
<td>(0.072)</td>
<td>(0.085)</td>
<td>(0.113)</td>
<td>(0.103)</td>
</tr>
<tr>
<td>Background</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>2346</td>
<td>2346</td>
<td>1167</td>
<td>1167</td>
<td>1179</td>
<td>1179</td>
</tr>
<tr>
<td><strong>Panel II: Employed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.080**</td>
<td>-0.056</td>
<td>-0.125***</td>
<td>-0.127**</td>
<td>-0.017</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.038)</td>
<td>(0.042)</td>
<td>(0.049)</td>
<td>(0.064)</td>
<td>(0.077)</td>
</tr>
<tr>
<td>Background</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>3215</td>
<td>3215</td>
<td>1461</td>
<td>1461</td>
<td>1754</td>
<td>1754</td>
</tr>
<tr>
<td><strong>Panel III: In the Labor Force</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.076**</td>
<td>-0.076**</td>
<td>-0.107***</td>
<td>-0.145***</td>
<td>0.011</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.033)</td>
<td>(0.026)</td>
<td>(0.042)</td>
<td>(0.058)</td>
<td>(0.074)</td>
</tr>
<tr>
<td>PWS</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>3215</td>
<td>3215</td>
<td>1461</td>
<td>1461</td>
<td>1754</td>
<td>1754</td>
</tr>
<tr>
<td><strong>Panel IV: Underemployed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>0.035</td>
<td>0.022</td>
<td>-0.008</td>
<td>-0.001</td>
<td>0.124*</td>
<td>0.105*</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.043)</td>
<td>(0.055)</td>
<td>(0.053)</td>
<td>(0.064)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>Background</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>3765</td>
<td>3765</td>
<td>1748</td>
<td>1748</td>
<td>2017</td>
<td>2017</td>
</tr>
<tr>
<td><strong>Panel V: Public Assistance Receipt</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>0.135***</td>
<td>0.083***</td>
<td>0.143***</td>
<td>0.125***</td>
<td>0.147**</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.032)</td>
<td>(0.040)</td>
<td>(0.042)</td>
<td>(0.061)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>Background</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>3851</td>
<td>3851</td>
<td>1795</td>
<td>1795</td>
<td>2056</td>
<td>2056</td>
</tr>
</tbody>
</table>

Each panel-by-column is a separate regression. Columns (1), (3), and (5) include only an indicator variable for stuttering as the independent variable. Columns (2), (4), and (6) add controls for individual and family background characteristics for gender, age, race, ethnicity, attractiveness, height, weight, parental income, parental education, parental marital status, and fixed effects for school of residence in Wave I. Results for hourly earnings are based on OLS regression. Results for the other outcomes are based on probit estimates and show marginal effects. Standard errors (shown in parentheses) are clustered by school of residence in Wave I. All estimates are based on weighted regression using the Add Health weights. Statistical significance is as follows: * 10% level, ** 5% level, and *** 1% level.
Table 3.4.: Robustness of results to controls for ADHD and learning disabilities

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td><strong>Panel I: Log Hourly Earnings</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.124*</td>
<td>-0.085</td>
<td>-0.190**</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.070)</td>
<td>(0.083)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>2305</td>
<td>2305</td>
<td>1138</td>
</tr>
<tr>
<td><strong>Panel II: Employed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.057</td>
<td>-0.049</td>
<td>-0.130**</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.039)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>3142</td>
<td>3142</td>
<td>1201</td>
</tr>
<tr>
<td><strong>Panel III: In the Labor Force</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.077**</td>
<td>-0.066**</td>
<td>-0.153***</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.033)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>3011</td>
<td>3011</td>
<td>839</td>
</tr>
<tr>
<td><strong>Panel IV: Underemployed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>0.017</td>
<td>0.019</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.043)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>3702</td>
<td>3702</td>
<td>1685</td>
</tr>
<tr>
<td><strong>Panel V: Public Assistance Receipt</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>0.091***</td>
<td>0.080**</td>
<td>0.146***</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.031)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>3732</td>
<td>3732</td>
<td>1556</td>
</tr>
</tbody>
</table>

Each panel-by-column is a separate regression. Columns (1), (3), and (5) include controls for individual and family background characteristics. Columns (2), (4), and (6) add controls for ADHD and other learning disabilities. See the notes from Table 3.3 for additional details.
Table 3.5.: Propensity score matching estimates of the effect of stuttering on labor market outcomes

<table>
<thead>
<tr>
<th></th>
<th>All (1)</th>
<th>PSM (2)</th>
<th>Males (3)</th>
<th>PSM (4)</th>
<th>Females (5)</th>
<th>PSM (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel I: Log Hourly Earnings</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.201***</td>
<td>-0.149*</td>
<td>-0.229***</td>
<td>-0.202**</td>
<td>-0.196*</td>
<td>-0.093</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.083)</td>
<td>(0.086)</td>
<td>(0.102)</td>
<td>(0.112)</td>
<td>(0.156)</td>
</tr>
<tr>
<td>N</td>
<td>2217</td>
<td>300</td>
<td>1083</td>
<td>196</td>
<td>882</td>
<td>104</td>
</tr>
<tr>
<td><strong>Panel II: Employed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.082**</td>
<td>-0.135***</td>
<td>-0.165***</td>
<td>-0.138**</td>
<td>-0.008</td>
<td>-0.046</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.046)</td>
<td>(0.040)</td>
<td>(0.057)</td>
<td>(0.061)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>N</td>
<td>3034</td>
<td>504</td>
<td>1378</td>
<td>324</td>
<td>1502</td>
<td>188</td>
</tr>
<tr>
<td><strong>Panel III: In the Labor Force</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.050*</td>
<td>-0.082**</td>
<td>-0.137***</td>
<td>-0.140***</td>
<td>0.028</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.038)</td>
<td>(0.028)</td>
<td>(0.046)</td>
<td>(0.055)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>N</td>
<td>3034</td>
<td>504</td>
<td>1378</td>
<td>324</td>
<td>1502</td>
<td>188</td>
</tr>
<tr>
<td><strong>Panel IV: Underemployed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>0.023</td>
<td>0.080</td>
<td>-0.036</td>
<td>0.020</td>
<td>0.107*</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.051)</td>
<td>(0.055)</td>
<td>(0.065)</td>
<td>(0.065)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>N</td>
<td>3599</td>
<td>548</td>
<td>1663</td>
<td>336</td>
<td>1826</td>
<td>208</td>
</tr>
<tr>
<td><strong>Panel V: Public Assistance Receipt</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>0.138***</td>
<td>0.086*</td>
<td>0.145***</td>
<td>0.143***</td>
<td>0.167***</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.046)</td>
<td>(0.041)</td>
<td>(0.055)</td>
<td>(0.058)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>N</td>
<td>3696</td>
<td>588</td>
<td>1700</td>
<td>372</td>
<td>1898</td>
<td>216</td>
</tr>
</tbody>
</table>

Results are based on nearest neighbor matching with replacement that matches each person who stutters with three people who do not stutter whose estimated propensity of stuttering was within 0.004. Columns (1), (3), and (5) show the mean difference in outcomes between PWS and PWNS for the unmatched sample. Columns (2), (4), and (6) show the mean difference in outcomes between matched people who stutter and people who do not stutter. The 2% of PWS whose estimated propensity of stuttering were furthest from their match were dropped. Matching variables included individual and family background characteristics. The differences in these characteristics before and after matching are shown in Table B.2.
Table 3.6.: Remaining difference after controlling for labor market characteristics

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td><strong>Panel I: Log Hourly Earnings</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.127*</td>
<td>-0.065</td>
<td>-0.178**</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.066)</td>
<td>(0.083)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Labor market</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>1968</td>
<td>1968</td>
<td>1035</td>
</tr>
<tr>
<td><strong>Panel II: Employed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.050</td>
<td>-0.027</td>
<td>-0.108**</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.041)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Labor market</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>3126</td>
<td>3126</td>
<td>1186</td>
</tr>
<tr>
<td><strong>Panel III: In the Labor Force</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>-0.063*</td>
<td>-0.050</td>
<td>-0.127***</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.034)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Labor market</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>2995</td>
<td>2995</td>
<td>834</td>
</tr>
<tr>
<td><strong>Panel IV: Underemployed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>0.017</td>
<td>0.041</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.046)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Labor market</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>3693</td>
<td>3693</td>
<td>1679</td>
</tr>
<tr>
<td><strong>Panel V: Public Assistance Receipt</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>0.072**</td>
<td>0.052</td>
<td>0.126***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.034)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Background</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ADHD/learn</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Labor market</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>3718</td>
<td>3718</td>
<td>1549</td>
</tr>
</tbody>
</table>

Each panel-by-column is a separate regression. Columns (1), (3), and (5) include controls for individual and family background characteristics and for ADHD and other learning disabilities. Columns (2), (4), and (6) add controls for labor market characteristics. The labor market characteristics include education level and self-stigma for panels II-V. Panel I also includes occupation fixed effects, hours, tenure, and limits the sample to full-time workers. See the notes from Table 3.3 for additional details.
Table 3.7.: Blinder-Oaxaca decomposition of the effect of stuttering on log hourly earnings

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td><strong>Panel I: Explained and Unexplained Difference</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total difference</td>
<td>0.225***</td>
<td>0.225***</td>
<td>0.213***</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.072)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>Difference due to characteristics (explained)</td>
<td>0.097*</td>
<td>0.160**</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.069)</td>
<td>(0.064)</td>
</tr>
<tr>
<td>Difference due to coefficients (unexplained)</td>
<td>0.127**</td>
<td>0.065</td>
<td>0.178***</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.057)</td>
<td>(0.068)</td>
</tr>
<tr>
<td><strong>Panel II: Contribution of Covariates to Explained Difference</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background controls, ADHD, learning disability</td>
<td>0.097*</td>
<td>0.019</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.050)</td>
<td>(0.064)</td>
</tr>
<tr>
<td>Occupation</td>
<td>0.055</td>
<td></td>
<td>0.111**</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td></td>
<td>(0.049)</td>
</tr>
<tr>
<td>Education</td>
<td>0.066***</td>
<td>0.039**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.019)</td>
<td></td>
</tr>
<tr>
<td>Self-stigma</td>
<td>0.027*</td>
<td>0.037**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.018)</td>
<td></td>
</tr>
<tr>
<td>Tenure</td>
<td>-0.007</td>
<td>-0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td>Hours</td>
<td>0.001</td>
<td>-0.007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.010)</td>
<td></td>
</tr>
</tbody>
</table>

Panel I shows the total unconditional difference in hourly earnings, the portion of that difference that is due to differences in observable characteristics, and the portion that is due to differences in coefficients. Panel II shows the contribution of each covariate to the explained difference. Columns (1), (3), and (5) controls for individual and family background characteristics and for ADHD and other learning disabilities. Columns (2), (4), and (6) add controls for labor market characteristics shown in the table.
<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td><strong>Panel I: Explained and Unexplained Difference</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total difference</td>
<td>0.225***</td>
<td>0.225***</td>
<td>0.213***</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.072)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>Difference due to characteristics (explained)</td>
<td>0.160**</td>
<td>0.147**</td>
<td>0.152*</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.072)</td>
<td>(0.090)</td>
</tr>
<tr>
<td>Difference due to coefficients (unexplained)</td>
<td>0.065</td>
<td>0.078</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.058)</td>
<td>(0.077)</td>
</tr>
<tr>
<td><strong>Panel II: Contribution of Covariates to Explained Difference</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background controls, ADHD, learning disability</td>
<td>0.019</td>
<td>0.028</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.050)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>Occupation</td>
<td>0.055</td>
<td>0.057</td>
<td>0.111**</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.038)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>Education</td>
<td>0.066***</td>
<td>0.067***</td>
<td>0.039**</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Self-stigma</td>
<td>0.027*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenure</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Hours</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.008)</td>
<td>(0.010)</td>
</tr>
</tbody>
</table>

Panel I shows the total unconditional difference in hourly earnings, the portion of that difference that is due to differences in observable characteristics, and the portion that is due to differences in coefficients. Panel II shows the contribution of each covariate to the explained difference. Columns (1), (3), and (5) are equivalent to columns (2), (4), and (6) from Table 3.7. Columns (2), (4), and (6) remove self-stigma from the specification.
Fig. 3.1.: The vertical axis shows the percentage point difference in hourly earnings between people who stutter and people who do not stutter. The horizontal axis shows the total difference, the contribution of observable characteristics to the total difference, and the unexplained difference. "Background" includes individual and family background characteristics, ADHD, and other learning disabilities. Results are based on the Blinder-Oaxaca decomposition results in Table 3.7.
LIST OF REFERENCES


APPENDICES
APPENDIX A. THE EFFECT OF MINIMUM WAGES ON EMPLOYMENT: A FACTOR MODEL APPROACH

A.1 Common correlated effects

The individual slope coefficients for the CCE estimator are given by

\[ \hat{b}_i = (X_i'\bar{M}_\omega X_i)^{-1}X_i'\bar{M}_\omega y_i, \]

where \( X_i \) and \( y_i \) are the independent and dependent variables, \( \bar{M}_\omega \) is given by

\[ \bar{M}_\omega = I_T - \bar{H}_\omega (\bar{H}'_\omega \bar{H}_\omega)^{-1} \bar{H}'_\omega, \]

and \( \bar{H}_\omega = (D, \bar{Z}_\omega) \), with \( D \) being a \((T \times 1)\) vector of ones, which produces an individual fixed effect, and \( \bar{Z}_\omega \) being the \((T \times (k+1))\) matrix of cross sectional averages of the dependent and independent variables.

[Pesaran, 2006] then suggests two versions of the CCE estimator. The Common Correlated Effects - Mean Group (CCEMG) is the average of the individual CCE coefficients,

\[ \hat{b}_{CCEMG} = N^{-1} \sum_{i=1}^{N} \hat{b}_i, \]

while the Common Correlated Effects - Pooled (CCEP) is given by

\[ \hat{b}_{CCEP} = \left( \sum_{i=1}^{N} \theta_i X_i'\bar{M}_\omega X_i \right)^{-1} \sum_{i=1}^{N} \theta_i X_i'\bar{M}_\omega y_i. \]

The CCEP assumes that \( \beta_i = \beta \) for all \( i \), although it does allow the slope coefficients of the common effects to differ across \( i \).

The variance of the CCEMG estimator is given by
\[ \hat{\Sigma}_{CCEMG} = (N - 1)^{-1} \sum_{i=1}^{N} (\hat{b}_i - \hat{b}_{CCEMG})(\hat{b}_i - \hat{b}_{CCEMG})'. \]

The variance of the CCEP estimator is given by

\[ \hat{\Sigma}_{CCEP} = N^{-1} \hat{\Psi}^{-1} \hat{R} \hat{\Psi}^{-1}, \]

where \( \hat{\Psi} \) and \( \hat{R} \) are given by

\[ \hat{\Psi} = \sum_{i=1}^{N} N^{-1} \left( \frac{X_i' \bar{M}_\omega X_i}{T} \right) \]

and

\[ \hat{R} = (N - 1)^{-1} \sum_{i=1}^{N} \frac{1/N}{N^{-2}} \left( \frac{X_i' \bar{M}_\omega X_i}{T} \right) (\hat{b}_i - \hat{b}_{CCEMG})(\hat{b}_i - \hat{b}_{CCEMG})' \left( \frac{X_i' \bar{M}_\omega X_i}{T} \right). \]

### A.2 Interactive fixed effects

[Bai, 2009] notes that, given \( F \) and \( \Lambda \), the regression coefficients could be estimated in the usual way after subtracting the factor structure out of the data,

\[ \hat{\beta}(F) = \left( \sum_{i=1}^{N} X_i' X_i \right)^{-1} \sum_{i=1}^{N} X_i' (Y_i - F \lambda_i), \]

and, given \( \beta \), principal component analysis could be used to compute \( F \) and \( \Lambda \) from the pure factor model,

\[ W_i = Y_i - X_i \beta. \]

In practice, both the regression coefficients and the factor structure are unknown. Therefore, [Bai, 2009] proposes an iterative procedure:
Step 1: Ignore the factor structure and estimate the regression coefficients. Given these regression coefficients, estimate the factors and factor loadings. Given these factors and factor loadings, re-estimate the regression coefficients. Iterate until the percent change in the sum of squared residuals in the regression coefficient estimation is less than a specified threshold\(^1\).

Step 2: Ignore the regression coefficients and estimate the factors and factor loadings. Given these factors and factor loadings, estimate the regression coefficients. Given these regression coefficients, re-estimate the factors and factor loadings. Iterate until the percent change in the sum of squared residuals in the factor and factor loading estimation is less than a specified threshold.

Step 3: Keep the estimates from the previous step that had the lowest final sum of squared residuals.

The data is demeaned in both directions before estimation to account for unit and period fixed effects. Bias correction is performed using equations (23) and (24) in [Bai, 2009].

The variance of the IFE estimator is given by

\[ \Sigma_{IFE} = \frac{1}{NT} D_0^{-1} D_Z D_0^{-1}, \]

where \( D_0 = (NT)^{-1} \sum_{i=1}^{N} Z_i'Z_i \), \( D_Z = N^{-1} \sum_{i=1}^{N} \hat{\sigma}_i^2 (T^{-1} \sum_{t=1}^{T} z_i't z_i't) \), with \( \hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_i^2 \), \( Z_i = MFX_i - N^{-1} \sum_{k=1}^{N} (\hat{\gamma}_k'(\hat{L}'\hat{L}/N)^{-1}\hat{\gamma}_k)MFX_k \), and \( \hat{\gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_N)' \).

\(^1\)The CCEMG estimates were also used as an alternative way to initiate step 1. In this case, the initial estimate of the regression coefficients is based on the CCEMG results. Given these estimates, the factor structure is estimated using principal components. The regression coefficients are then re-estimated, given these factors and factor loadings, using the IFE estimate of the regression coefficients described above. Iteration then continues as described in step 1. This approach converges much faster than when the initial estimate of the regression coefficients is based on traditional OLS and finishes with approximately the same final sum of squared residuals. The results are nearly identical.
APPENDIX B. THE IMPACT OF STUTTERING ON ADULT LABOR MARKET OUTCOMES

B.1 Construction of the composite self-stigma index

The composite self-stigma index is constructed using a subset of questions from Wave IV of the Add Health data that correspond to a particular commonly held stereotype about PWS. The stereotypes come from Kalinowski, Stuart, and Armson (1996). The questions and the stereotype that they correspond to are shown in Table B.1.

The first five questions are recorded in a five-option format ("strongly agree", "agree", "neither agree nor disagree", "disagree", "strongly disagree"). These are re-coded into a dichotomous (0 or 1) variable indicating any level agreement that would correspond with the stereotype. For example, for the question "I don’t talk most of the time," which corresponds to the stereotype 'quiet,' the respondent would be given a 1 if they answered "agree" or "strongly agree." For the question "I am relaxed most of the time," which corresponds to the stereotype 'nervous,' the respondent would be given a 1 if they answered "disagree" or "strongly disagree." The sixth question, "You feel you are just as good as other people," is recorded as a four-option format ("never or rarely", "sometimes", "a lot of the time", "most of the time or all of the time"). We re-coded this question into a dichotomous variable with a 1 if the respondent answered "never or rarely." We re-coded the variables into dichotomous variables to avoid potential issues due to anchoring; respondents with the same reactions to a statement may evaluate the strength of their agreement differently depending on where they anchor their scale. However, anchoring is less likely to impact whether the respondent agrees in principle.
Table B.1 shows the fraction of PWS and PWNS who answered each question with a response that corresponds to the stereotype listed. For each question, PWS were much more likely to answer according to the stereotype. The difference in the mean response between PWS and PWNS is statistically significant at the one-percent level for each question.
Table B.1.: Questions used to construct the composite self-stigma index

<table>
<thead>
<tr>
<th>Question</th>
<th>Corresponding Stereotype</th>
<th>PWNS Mean</th>
<th>PWNS SD</th>
<th>PWS Mean</th>
<th>PWS SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don’t talk a lot</td>
<td>Quiet</td>
<td>0.24</td>
<td>0.006</td>
<td>0.38</td>
<td>0.034</td>
</tr>
<tr>
<td>I am relaxed most of the time</td>
<td>Nervous</td>
<td>0.14</td>
<td>0.005</td>
<td>0.21</td>
<td>0.028</td>
</tr>
<tr>
<td>I keep in the background</td>
<td>Lacking Assertion</td>
<td>0.24</td>
<td>0.006</td>
<td>0.42</td>
<td>0.034</td>
</tr>
<tr>
<td>I go out of my way to avoid having to deal with problems in my life</td>
<td>Avoidant</td>
<td>0.19</td>
<td>0.006</td>
<td>0.27</td>
<td>0.031</td>
</tr>
<tr>
<td>There is really no way to solve the problems that I have</td>
<td>Self-derogatory</td>
<td>0.03</td>
<td>0.002</td>
<td>0.09</td>
<td>0.020</td>
</tr>
<tr>
<td>You feel you are just as good as other people</td>
<td>Insecure</td>
<td>0.20</td>
<td>0.006</td>
<td>0.40</td>
<td>0.034</td>
</tr>
</tbody>
</table>

Each question comes from Wave IV of the Add Health dataset. The response to each question is re-coded into a dichotomous (0 or 1) variable, with 1 indicating a response that corresponds with the stereotype. See the appendix text for additional details.
### B.2 Propensity score matching

**Table B.2.: Difference in covariates before and after propensity score matching**

<table>
<thead>
<tr>
<th></th>
<th>Unmatched</th>
<th></th>
<th></th>
<th>Matched</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Person</td>
<td>Person</td>
<td>P-value</td>
<td>Person</td>
<td>Person</td>
<td>P-value</td>
</tr>
<tr>
<td></td>
<td>who</td>
<td>who</td>
<td></td>
<td>who</td>
<td>who</td>
<td></td>
</tr>
<tr>
<td></td>
<td>stutters</td>
<td>does</td>
<td></td>
<td>stutters</td>
<td>does</td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>0.400</td>
<td>0.510</td>
<td>0.046**</td>
<td>0.400</td>
<td>0.436</td>
<td>0.661</td>
</tr>
<tr>
<td>age</td>
<td>30.024</td>
<td>29.927</td>
<td>0.613</td>
<td>29.947</td>
<td>29.996</td>
<td>0.868</td>
</tr>
<tr>
<td>Black</td>
<td>0.388</td>
<td>0.198</td>
<td>0.000***</td>
<td>0.360</td>
<td>0.396</td>
<td>0.656</td>
</tr>
<tr>
<td>Native American</td>
<td>0.035</td>
<td>0.011</td>
<td>0.040**</td>
<td>0.013</td>
<td>0.013</td>
<td>1.000</td>
</tr>
<tr>
<td>Asian</td>
<td>0.024</td>
<td>0.023</td>
<td>0.995</td>
<td>0.027</td>
<td>0.018</td>
<td>0.714</td>
</tr>
<tr>
<td>other race</td>
<td>0.047</td>
<td>0.047</td>
<td>0.993</td>
<td>0.053</td>
<td>0.053</td>
<td>1.000</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.106</td>
<td>0.090</td>
<td>0.619</td>
<td>0.107</td>
<td>0.116</td>
<td>0.864</td>
</tr>
<tr>
<td>unattractive</td>
<td>0.106</td>
<td>0.043</td>
<td>0.006***</td>
<td>0.053</td>
<td>0.076</td>
<td>0.582</td>
</tr>
<tr>
<td>average attractiveness</td>
<td>0.412</td>
<td>0.463</td>
<td>0.356</td>
<td>0.453</td>
<td>0.489</td>
<td>0.665</td>
</tr>
<tr>
<td>attractive</td>
<td>0.412</td>
<td>0.364</td>
<td>0.369</td>
<td>0.413</td>
<td>0.369</td>
<td>0.580</td>
</tr>
<tr>
<td>very attractive</td>
<td>0.035</td>
<td>0.096</td>
<td>0.060*</td>
<td>0.040</td>
<td>0.031</td>
<td>0.771</td>
</tr>
<tr>
<td>height (feet)</td>
<td>5.702</td>
<td>5.611</td>
<td>0.011**</td>
<td>5.687</td>
<td>5.685</td>
<td>0.977</td>
</tr>
<tr>
<td>weight (lbs)</td>
<td>202.400</td>
<td>188.370</td>
<td>0.016**</td>
<td>194.790</td>
<td>213.750</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Results are based on nearest neighbor matching with replacement that matches each person who stutters with three people who do not stutter whose estimated propensity of stuttering was within 0.004. The 2% of PWS whose estimated propensity of stuttering were furthest from their match were dropped.
Table B.3.: Difference in covariates before and after propensity score matching (continued)

<table>
<thead>
<tr>
<th></th>
<th>Unmatched</th>
<th>Matched</th>
<th></th>
<th>Unmatched</th>
<th>Matched</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Person</td>
<td>Person</td>
<td>P-value</td>
<td>Person</td>
<td>Person</td>
<td>P-value</td>
</tr>
<tr>
<td></td>
<td>who</td>
<td>who</td>
<td></td>
<td>who</td>
<td>who</td>
<td></td>
</tr>
<tr>
<td></td>
<td>stutters</td>
<td>does</td>
<td></td>
<td>stutters</td>
<td>does</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>not</td>
<td></td>
<td></td>
<td>not</td>
<td></td>
</tr>
<tr>
<td>learning disability</td>
<td>0.224</td>
<td>0.109</td>
<td>0.001***</td>
<td>0.200</td>
<td>0.169</td>
<td>0.626</td>
</tr>
<tr>
<td>ADHD</td>
<td>0.153</td>
<td>0.050</td>
<td>0.000***</td>
<td>0.107</td>
<td>0.067</td>
<td>0.387</td>
</tr>
<tr>
<td>family income as adolescent</td>
<td>4.739</td>
<td>4.952</td>
<td>0.722</td>
<td>4.955</td>
<td>4.094</td>
<td>0.337</td>
</tr>
<tr>
<td>Parent - HS</td>
<td>0.306</td>
<td>0.267</td>
<td>0.429</td>
<td>0.307</td>
<td>0.311</td>
<td>0.953</td>
</tr>
<tr>
<td>Parent - some college</td>
<td>0.294</td>
<td>0.305</td>
<td>0.825</td>
<td>0.267</td>
<td>0.240</td>
<td>0.710</td>
</tr>
<tr>
<td>Parent - college</td>
<td>0.118</td>
<td>0.164</td>
<td>0.256</td>
<td>0.133</td>
<td>0.142</td>
<td>0.876</td>
</tr>
<tr>
<td>Parent - more than college</td>
<td>0.106</td>
<td>0.119</td>
<td>0.715</td>
<td>0.120</td>
<td>0.102</td>
<td>0.731</td>
</tr>
<tr>
<td>Parent - married</td>
<td>0.659</td>
<td>0.727</td>
<td>0.170</td>
<td>0.693</td>
<td>0.653</td>
<td>0.604</td>
</tr>
</tbody>
</table>

Results are based on nearest neighbor matching with replacement that matches each person who stutters with three people who do not stutter whose estimated propensity of stuttering was within 0.004. The 2% of PWS whose estimated propensity of stuttering were furthest from their match were dropped.
VITA
VITA

Evan S. Totty

Education

Ph.D. Economics, Purdue University
Dissertation: Essays in Labor Economics and Panel Data Analysis
Committee: Mohitosh Kejriwal (co-chair), Kevin Mumford (co-chair), Justin Tobias, John Barron

M.S. Economics, Purdue University
2013

B.S. Economics (minor: Mathematics), Middle Tennessee State University
2011

Research Interests

Labor Economics, Applied Econometrics, Applied Microeconomics

Invited Conference and Seminar Presentations

Institute for Research on Labor and Employment, University of California at Berkeley
2015 Joint Statistical Meetings, Seattle
25th Annual Meeting of the Midwest Econometrics Group, St. Louis
2016 American Economic Association Annual Meeting, San Francisco
Department of Communication Sciences and Disorders, University of Iowa
Fellowships and Awards

Robert W. Johnson Award for Distinguished Research Proposal 2014
Bilsland Dissertation Fellowship 2014-15
Certificate for Distinguished Teaching (4.4/5 or higher average on selected questions from student evaluations) Fall 2013, Summer 2014
Certificate for Outstanding Teaching (4.0-4.39/5) Summer 2013
Certificate for Distinguished Recitation Teaching (4.4/5 or higher) Spring 2012

Research Experience

Research Assistant to Mohitosh Kejriwal Spring 2014, Fall 2015
Research Assistant to Tim Bond Spring 2014
Research Assistant to Tim Cason Fall 2011, Fall 2012

Teaching Experience

Course Instructor
  Macroeconomics (undergrad) Summer 2013, Fall 2013
  Macroeconomics - online (undergrad) Summer 2014
Recitation Instructor
  Principles of Economics (undergrad) Spring 2012
Teaching Assistant
  Intermediate Microeconomic Theory Spring 2013, Spring 2014 (undergrad)
  Game Theory (undergrad) Fall 2012
  Behavioral Economics (undergrad) Fall 2011
  Probability and Statistics (Ph.D.) Fall 2015
Refereeing

*The European Economic Review, Journal of Urban Economics*

References

Mohitosh Kejriwal  
Department of Economics  
Purdue University  
(765) 494-4503  
mkejriwa@purdue.edu

Kevin Mumford  
Department of Economics  
Purdue University  
(765) 494-6773  
mumford@purdue.edu

Justin Tobias  
Department of Economics  
Purdue University  
(765) 494-8570  
jltobias@purdue.edu