THE UNIVERSITY OF CHICAGO

TOPICS AND METHODS IN LABOR ECONOMICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
HIEU TRUNG NGUYEN

CHICAGO, ILLINOIS
AUGUST 2020

To Him, the source of all truth and love

To my loving parents, Phuong Vu and Thanh Nguyen

To my dear friend and mentor, Philip River

Table of Contents

List of Figures

List of Tables

Acknowledgments

Graduate school has been the most challenging and rewarding experience in my life so far. Therefore, I want to thank first and foremost the Kenneth C. Griffin Department of Economics at the University of Chicago for providing such a rigorous education.

Completing the dissertation was a difficult yet gratifying process. I am proud of the opportunity to make a contribution by way of my original ideas and hard work. However, I could not have done them all on my own. I am indebted to my thesis committee, Steven D. Levitt, Yana Gallen, and Alessandra Voena for their patience, support, and guidance in finishing my dissertation. I want to especially thank Yana Gallen for her continual support and encouragement even when I was discouraged and hitting roadblocks in my research. I thank James B. McDonald and Keith Vorkink for their kindness and guidance in my undergraduate work, thus allowing me to matriculate at a competitive PhD program in Economics. I also thank Azeem Shaikh, Alexander Torgovitsky and Derek Neal for their invaluable suggestions and guidances on my earlier dissertation drafts.

I thank my parents, Thanh Nguyen and Phuong Vu for their love, care and continual encouragement throughout my life. Their unwavering support and belief has nurtured me into who I am today. I especially thank my good friend and mentor, Philip River, for his patience, understanding and guidance throughout my graduate school experience. His wiseness, sharp insights and example of inner strength inspired me greatly. I thank him for challenging me to come to know myself and to grow fully into who I can be.

Finally, I thank Him for continually enriching my life with meanings and purposes.

Chapter 1

Fast Education and Fast Emigration?

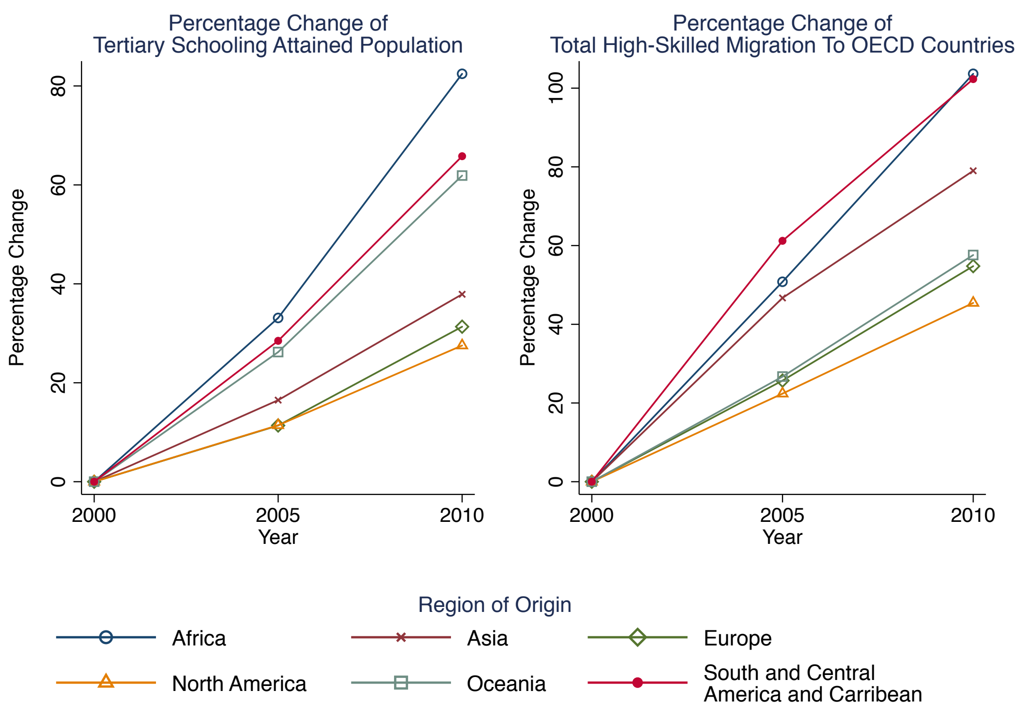
Abstract

High-skilled migration and higher-education attainment rose globally in the period between 2000 and 2010. These trends raise the question of whether increases in higher-education attainment led to increases in high-skilled emigration in source countries during this period? To the contrary, I found a strong negative correlation between percentage changes of high-skilled emigration rates and percentage changes of high-skilled proportions in non-OECD countries. Using the IAB Brain Drain, Barro-Lee and World Development Indicators Datasets, I estimated the elasticity of high-skilled emigration rates with respect to high-skilled percentages in non-OECD countries with ordinary least square, least absolute shrinkage and selection operator (LASSO) and post-double-selection LASSO. The estimates ranged between -0.78 and -0.98.

1.1 Introduction

High-skilled migration[1] and higher-education attainment have been increasing all over the world in recent decades. The plots below show these increases in different regions of the world, with the largest in Africa and South and Central America and the Caribbean between 2000 and 2010:

Figure 1.1: Percentage changes of tertiary schooling attained population and total high-skilled migration to OECD countries by region between 2000 and 2010



Note: Percentage Changes Are Relative to Year 2000
Source: IAB Brain Drain Dataset and Barro-Lee Education Attainment Dataset

These trends raise an interesting, unexplored question: Do increases in higher-education attainment lead to increases in the propensity to emigrate among high-skilled individuals in source countries? What is the elasticity of high-skilled emigration rate with respect to high-

1. High-skilled migrants here are defined as individuals with foreign citizenship at birth who have completed formal education beyond high school.

skilled proportion[2]? Given the continuing rises in higher education globally, this parameter is key to assessing the impact of brain drain on countries' economic development (especially developing countries) and to identifying worldwide bearers of training costs or benefits of high-skilled labors.

Using the IAB Brain Drain and the Barro-Lee Educational Attainment Datasets, I calculate the percentage changes of high-skilled emigration rates and percentage changes of high-skilled proportion for 122 non-OECD countries between 2000 and 2010. I found a negative Pearson correlation coefficient of -0.68 between them. I then estimate the elasticity of high-skilled emigration rates with respect to high-skilled proportion using ordinary least square, least absolute shrinkage and selection operator (LASSO) and post-double-selection LASSO. The least square estimator includes GDP per capita, total population, total trade volume, total migration stock, CO2 emission per capita and their respective percentage changes between between 2000 and 2010 as control variables. The LASSO and post-double-selection LASSO estimators select appropriate controls from a set of 534 time series indicators from the World Development Indicators Dataset[3]. The estimates are robust and range between -0.78 and -0.98.

This chapter contributes to the recent international migration and brain drain literature. Docquier and Marfouk (2006) produced improved estimates of emigration stocks and emigration rates by educational attainment and origin countries for years 1990 and 2000. They showed that more educated groups have higher emigration rates, a phenomenon known as positive selection. Docquier, Lowell, Marfouk (2009) documented a rising feminization of high-skilled migration between 1990 and 2000. The authors also found strong correlation

2. Within the context of this study, I define these two terms as follow. High-skilled emigration rate is defined as the number of high-skilled migrants (to OECD destination countries) divided by the total number of high-skilled migrants and high-skilled non-migrants in an origin country. High-skilled proportion is defined as the total high-skilled migrants and non-migrants divided by the total population in an origin country. See section 1.2.2 for more details

3. The World Development Indicator Dataset has a total of 1600 time-serie indicators. All the indicators with more than 3% missing observations were dropped from the choosing set to ensure enough observations for LASSO and post-double-selection LASSO estimators

between the migration stocks of high-skilled men and high-skilled women across origin countries. Docquier and Rapoport (2012) surveyed the literature on the magnitude, intensity and determinants of the brain drain in recent decades. The authors use a stylized growth model to analyze how brain drain potentially affects source countries' economies through various channels. John and McKenzie (2011) surveyed recent empirical evidence on key questions of brain drain.

This chapter is also related to a strand of literature on the worldwide boom of higher education attainment. Snyder, Dillow, and Hoffman (2008) documented recent trends in males and females' college enrollment, and their respective completion of bachelor's degrees. Becker, Hubbard, Murphy (2010) found that differences in total costs of college for women and men primarily explained the overtaking of men by women in higher education.

Finally, this chapter is related to a growing literature on LASSO-related methodologies. LASSO was invented by Tibshirani (1996) and has been commonly used in building models for prediction. It performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. From a large set of potential controls, the LASSO selects a set of variables that are correlated with covariates that appear in the true model. While the LASSO is powerful for selecting predictors from a large set of potential regressors, it does not allow researchers to make inference about the coefficient estimates. Belloni et al. (2012, 2014, 2016) developed the post-double-selection LASSO that allows practitioners to make inference with standard errors, confidence intervals, p-values of the coefficient estimates. Ahrens et al (2018, 2019) provided the implementation of these estimator via the package lassopack and pdslasso in Stata.

The rest of the chapter is structured as follows. Section 2 describes the datasets and the empirical methodologies. Section 3 discusses the results. Section 4 concludes the chapter.

1.2 Data and Empirical Methodology

1.2.1 Data

For my empirical analyses, I use the IAB Brain Drain Dataset on international migration, the Barro-Lee Dataset on global educational attainment and the World Banks's World Development Indicators Database on countries' development indicators.

The IAB Brain Drain Dataset contains data on the total number of foreign-born individuals aged 25 years and older, living in each of the 20 OECD destination countries, by year, gender, country of origin and educational level [4]. The 20 OECD countries include Australia, Austria, Canada, Chile, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and United States.

The Barro-Lee Dataset collects education attainment information from censuses/surveys across countries as compiled by the UNESCO agency, the European Statistical Offices and other national statistic agencies. The authors of the dataset computed the distribution of education attainment in the adult population over age 15 and over age 25 by gender at seven levels of schooling: no formal education, incomplete primary, complete primary, lower secondary, upper secondary, incomplete tertiary, and complete tertiary education.[5] The dataset include these education statistics for 146 countries.

The World Development Indicators Dataset is a compilation of relevant, high-quality, and internationally comparable statistics about global development and poverty. The database contains 1,600 time series indicators for 217 economies and more than 40 country groups. The indicators cover topics of agriculture and food security, climate change, economic growth, education, energy and extractives, environment and natural resources, financial sector de-

4. Educational levels are distinguished in low, medium and high skilled.

5. See http://www.barrolee.com/ for more details

velopment, gender health, nutrition and population, macroeconomic vulnerability and debt, poverty, private sector development, public sector management, social development, social protection and labor, trade, urban development.

1.2.2 Methodology

I first derive the expressions for the high-skilled emigration rates and the high-skilled proportions. I denote the stock of total natives population originally from country i in year t as N_t^i (equal to the sum of residents and migrants). The stock of high-skilled natives is denoted as $N_{t,h}^i$. The stock of high-skilled natives can be disaggregated into the stock of high-skilled migrants ($M_{t,h}^i$) and high-skilled non-migrants (residents, $R_{t,h}^i$) as $N_{t,h}^i = M_{t,h}^i + R_{t,h}^i$.

The high-skilled emigration rate for the source country i is then defined as:

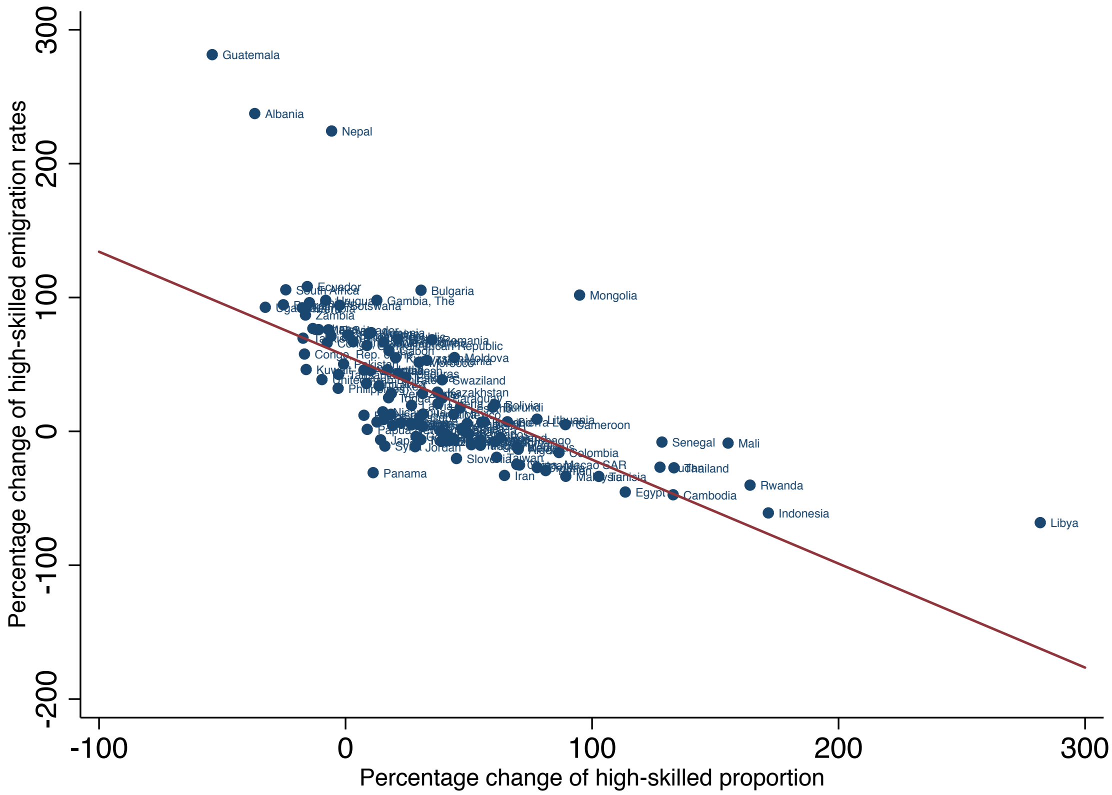
$$m_{t,h}^i = \frac{M_{t,h}^i}{N_{t,h}^i}$$

The high-skilled proportion for the source country i is defined as:

$$p_{t,h}^i = \frac{N_{t,h}^i}{N_t^i}$$

Figure 1.1 (below) shows the scatter plot of percentage change of high-skilled emigration rates versus percentage change of high-skilled proportion for 122 non-OECD origin countries between 2000 and 2010. The plot shows a clear negative relationship (with a Pearson correlation coefficient of -0.68) between the percentage changes of high-skilled emigration rates and percentage changes of high-skilled proportion between 2000 and 2010. Source countries that saw increases in tertiary education attainment among their native population also tend to see decreases in their high-skilled emigration rates and vice versa. This suggests a negative elasticity of high-skilled emigration rates with respect to high-skilled proportion for this period.

Figure 1.2: Scatter plot of percentage change of high-skilled emigration rates versus percentage change of high-skilled proportion between 2000 and 2010



Pearson Correlation coefficient $\rho = -0.68$
Source: IAB Brain Drain Dataset and Barro-Lee Education Attainment Dataset

However, this negative correlation could be biased due to omitted variables that affect both changes in high-skilled emigration rates and changes in higher education attainment. For example, countries with fast growing economies are able to provide ample attractive employment opportunities locally and hence, more likely to retain their high-skilled labors. In contrast, countries with slow growing economies might see their high-skilled labor force shrunk even further because there are not enough attractive opportunities for them at home compared to abroad.

To control for omitted variables bias, I estimate the following model with ordinary least square, LASSO and post-double-selection LASSO:

$$\Delta m_h^i = \beta_0 + \beta_1 \Delta p_h^i + X_i \Gamma + \epsilon_i \tag{1.1}$$

where Δm_h^i is the percentage change of high-skilled emigration rates and Δp_h^i is the percentage change of high-skilled proportion for country i. β_1 is the elasticity of high-skilled emigration rates with respect to high-skilled proportion. X_i is the set of control variables.

In the least square analysis, the control variables set X_i includes GDP per capita in year 2000, percentage changes of GDP per capita, total trade volume (as percentage of GDP) in year 2000, percentage change of total trade volume, total population in year 2000, percentage change of total population, total migration stock from country i to OECD countries in year 2000, CO2 emission in metric ton per capita in year 2000, percentage change of CO2 emission. All percentage-change variables are changes between 2000 and 2010. These control variables are important determinants of high-skilled migration as suggested by the international migration literature. Other important variables include geographic distances, languages and cultures proximity between countries. However, these variables are time-invariant and hence, dropped out of the model since I am looking at differences over time.

In the LASSO and post-double-selection LASSO estimators, the controls set X_i includes the full set of indicators from World Development Indicators World Bank Database, covering topics from agriculture and food security to urban development (more details in the data section). Similar to the OLS specification, I include all the indicators' values for year 2000 and their percentages changes between 2000 and 2010. I dropped all indicators and their percentage changes with more than 3% of missing observations[6] to ensure that there are enough observations for the LASSO and post-double-selection LASSO to select the covariates. The final set of high-dimensional controls include 534 different covariates.

Finally, the empirical model above potentially suffers from reverse causation of Δm_h^i

6. For robustness checks, I also experimented with 1% and 5% levels.

on Δp_h^i. For example, increases in high-skilled emigration rates might induce increases in higher education attainment as individuals invest to prepare for the prospect of migration in the future, and vice versa. In this case, reverse causation will generate positive correlation between Δm_h^i on Δp_h^i. Ordinary least square and post-double-selection LASSO estimates will be biased upward (toward 0). When reverse causation is a big problem, these estimates will provide upper bounds for the true elasticity measure.

The elasticity estimates are shown in the next section. Both standardized and raw coefficients are reported.

1.3 Results

Table 1.1: Estimates of $\hat{\beta}_1$ from OLS with and without controls, LASSO and post-double-selection LASSO

Estimator	OLS	OLS	OLS	LASSO	PDS LASSO
$\hat{\beta}_1$	-0.774***	-0.794***	-0.952***	-0.879	-0.975***
	(-10.13)	(-10.10)	(-9.56)	-	(-6.22)
Standardized $\hat{\beta}_1$	-0.679	-0.698	-0.702	-	-
Controls	N	Y	N	Y	Y
Observations	122	113	96	96	96

t and z statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.1 shows estimates of $\hat{\beta}_1$ for ordinary least square with no control, ordinary least square with controls, LASSO, post-double-selection LASSO, and ordinary least square as baseline comparison for LASSO and PDS LASSO (with no control and same number of non-missing observations). Full results of ordinary least square with various set of controls and post-double-selection LASSOs are included in the appendix.

The estimates are robust and similar in magnitude across different estimators. In the first two columns, the difference between least square with controls and without controls is mainly due to the difference in the number of observations (see appendix). Omitted variable bias is small because $\hat{\beta}_1$ is not significantly changed when including a full set of controls (see appendix). Furthermore, only GDP per capita in 2000 has a statistically significant coefficient.

The last three columns show estimates for LASSO, post-double-selection LASSO, and baseline least square for comparison (with no control). The LASSO estimator selected 93 control variables. However, the LASSO estimates can neither be interpreted as elasticity nor be made inference. Post-double-selection LASSO selected one control variable in this case - the percentage change of methane emission between 2000 and 2010. $\hat{\beta}_1$ is not significantly changed when controlling for this variable (-0.975 vs -0.952). The difference in the t-statistic (for OLS, -9.56) and the z-statistic (for PDS LASSO, -6.22) is mainly due to the added variance of variable-selection errors in the post-double-selection LASSO.

As a robustness check, I also ran post-double-selection LASSO including all development indicators with less than 5% of missing values (as opposed to 3% level for the estimates above, see the appendix). In this case, post-double-selection LASSO selected one control variable - percentage change of merchandise imports from low and middle-income economies in Europe & Central Asia. $\hat{\beta}_1$'s estimate is not significantly changed when controlling for this variable.

Altogether, these results suggest that omitted variable bias is not a significant problem. On average, one percentage increase in high-skilled proportion lead to around 0.78 to 0.98 percentage decrease in high-skilled emigration rates in non-OECD countries between 2000 and 2010. The standardized coefficient of $\hat{\beta}_1$ for ordinary least square with no control (i.e. the Pearson correlation coefficient) and ordinary least square with controls are also similar. They further suggest that this effect is important economically.

1.4 Conclusion

Between 2000 and 2010, non-OECD source countries with increases in high-skilled proportion were also likely to experience decreases in high-skilled emigration rates and vice versa. I estimated the elasticity of high-skilled emigration rates with respect to high-skilled proportion using ordinary least square without controls, ordinary least square with controls, LASSOs and post-double-selection LASSOs. The estimates range between -0.78 and -0.98. Standardized coefficients range between -0.679 and -0.702, suggesting that the effects are important economically. Future work may delve deeper in explaining this negative elasticity and its implications.

References

[1] Ahrens, A., Hansen, C.B., Schaffer, M.E. 2018. *"LASSOPACK: Stata module for lasso, square-root lasso, elastic net, ridge, adaptive lasso estimation and cross-validation."* http://ideas.repec.org/c/boc/bocode/s458458.html

[2] Ahrens, A., Hansen, C.B., Schaffer, M.E. 2018. *"pdslasso and ivlasso: Programs for post-selection and post-regularization OLS or IV estimation and inference."* http://ideas.repec.org/c/boc/bocode/s458459.html

[3] Ahrens, A., Hansen, C.B., Schaffer, M.E. 2019. *"lassopack: Model selection and prediction with regularized regression in Stata."* arxiv:1901.05397.

[4] Becker, Gary S., William H. J. Hubbard, and Kevin M. Murphy. 2010. *"The Market for College Graduates and the Worldwide Boom in Higher Education of Women."* American Economic Review, 100 (2): 229-33.

[5] *"Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain."* Econometrica 80(6): 2369-2429. http://dx.doi.org/10.3982/ECTA9626.

[6] Belloni, A., V. Chernozhukov, and C. Hansen. 2014. *"Inference on treatment effects after selection among high-dimensional controls."* Review of Economic Studies 81: 608-650. https://doi.org/10.1093/restud/rdt044.

[7] Belloni, A., V. Chernozhukov, C. Hansen, and D. Kozbur. 2016. *"Inference in High Di-*

mensional Panel Models with an Application to Gun Control." Journal of Business and Economic Statistics 34(4): 590-605. https://doi.org/10.1080/07350015.2015.1102733.

[8] Docquier, Frédéric, Marfouk, Abdeslam, 2006. *"International migration by educationalattainment, 1990-2000."* In: Ozden, Caglar, Schiff, Maurice (Eds.), International Migration, Remittances, and the Brain Drain. The World Bank and Palgrave McMillan, Washington, DC, pp. 151-200.

[9] Docquier, Frédéric , Lowell, B. Lindsay and Marfouk, Abdeslam. 2009. *"A Gendered Assessment of Highly Skilled Emigration."* Population and Development Review, 35: 297-321. doi:10.1111/j.1728-4457.2009.00277.x

[10] Docquier, Frédéric, and Hillel Rapoport. 2012. *"Globalization, Brain Drain, and Development."*. Journal of Economic Literature, 50 (3): 681-730.

[11] Dumont, Jean-Christophe and Martin, John P. and Spielvogel, Gilles. 2007. *"Women on the Move: The Neglected Gender Dimension of the Brain Drain. "* IZA Discussion Paper No. 2920.

[12] Snyder, T.D., Dillow, S.A., and Hoffman, C.M. (2009). *"Digest of Education Statistics 2008 (NCES 2009-020)."* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

[13] *"Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso."* Journal of the Royal Statistical Society. Series B (Methodological) 58(1): 267-288. http://www.jstor.org/stable/2346178.

Appendix

Table 2: Ordinary Least Square Regression Results

	$\Delta m^i_{t,h}$	$\Delta m^i_{t,h}$	$\Delta m^i_{t,h}$	$\Delta m^i_{t,h}$	$\Delta m^i_{t,h}$
$\Delta p^i_{t,h}$	-0.774***	-0.788***	-0.802***	-0.797***	-0.794***
	(-10.13)	(-10.13)	(-9.95)	(-10.74)	(-10.10)
$\Delta GDPpercapita$			0.139		0.0258
			(1.05)		(0.18)
$\Delta tradevolume$			0.0643		0.00522
			(0.63)		(0.05)
$\Delta totalpopulation$			-0.00790		-0.220
			(-0.03)		(-0.55)
$\Delta CO2emissionpercapita$			0.0414		-0.0374
			(0.35)		(-0.32)
GDP per capita in 2000				-0.0000158**	-0.0000166**
				(-2.86)	(-2.83)
Trade volume in 2000				-0.00167*	-0.00163
				(-1.99)	(-1.80)
Total population in 2000				4.65e-11	3.53e-11
				(0.10)	(0.07)
Total migration stock in 2000				-8.17e-08	-7.73e-08
				(-1.28)	(-1.16)
CO2 emission pc in 2000				0.0122	0.0126
				(1.45)	(1.44)
Population growth in 2000				-0.0349	-0.0143
				(-1.22)	(-0.34)
Constant	0.567***	0.598***	0.536***	0.874***	0.899***
	(12.50)	(12.66)	(5.40)	(8.43)	(6.24)
Observations	122	113	113	114	113

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Post-double-selection LASSO including all development indicators with less than 3% of missing observations, OLS with and without the selected control variable

	OLS	OLS	PDS LASSO
	$\Delta m^i_{t,h}$	$\Delta m^i_{t,h}$	$\Delta m^i_{t,h}$
$\Delta p^i_{t,h}$	-0.952***	-0.975***	-0.975***
	(-9.56)	(-9.32)	(-6.22)
Δmethane_emission		-0.00194	-0.00194
		(-0.73)	(-1.14)
Constant	0.662***	0.668***	0.668***
	(12.37)	(12.32)	(9.07)
Observations	96	96	96

t and z statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Post-double-selection LASSO including all development indicators with less than 5% of missing observations, OLS with and without the selected control variable

	OLS	OLS	PDS LASSO
	$\Delta m^i_{t,h}$	$\Delta m^i_{t,h}$	$\Delta m^i_{t,h}$
$\Delta p^i_{t,h}$	-0.964***	-1.041***	-1.041***
	(-9.03)	(-9.42)	(-6.46)
Δmerchandise_imports		0.000573*	0.000573***
		(2.09)	(4.66)
Constant	0.667***	0.681***	0.681***
	(11.13)	(11.57)	(8.19)
Observations	70	70	70

t and z statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 2

The effect of high school exit exams on educational attainment and employment: new evidence from a reform in the state of Georgia

Abstract

In 2015, the state of Georgia eliminated exit exams from high school graduation requirements in an education reform. I study this reform's impact on 4-year high school graduation rate, post-secondary institutions enrollment and employment rate among new high school graduates. Using the state of Florida as control, I estimate weighted difference-in-difference analyses at the school level for graduation rates and at the district level for post-secondary institutions enrollment and employment rates. The reform increased graduation rate by 4% for all students, and by 1.4% for White, 5.98% for Black and 5.59% for Hispanic subgroups. The reform decreased post-secondary institutions enrollment rate by 1.8% but had no statis-

tically significant impact on employment rate among new high school graduates who don't pursue post-secondary education. In the pre-periods, the trends are parallel for graduation rates, post-secondary institutions enrollment rates and employment rates between Georgia and Florida. I conclude that although exit exams elimination increased 4-year graduation rates (especially among Black and Hispanic students), the additional diploma holders are less likely to enroll in post-secondary institutions and no more likely to be employed when choosing not to pursue postsecondary education.

2.1 Introduction

In the United States (U.S.), a majority of high school students have been familiar with high school exit exams (HSSEs) in recent decades. For instance, in 2012 twenty six states in the U.S. required their students to pass some form of exit exams for high school graduation. These twenty six states together enrolled about 75% of all public high school students and 84% of all students of color in the U.S.(Center on Education Policy [CEP]).

Exit exams were first adopted in the 1970s in the form of minimum competency exams (MCE). They were advocated to address high unemployment and high inflation rates in the U.S. by way of shoring up educational skills and workforce quality. By 1982, thirty-nine states had passed MCE legislations, and nineteen of those states included MCE in their high school graduation requirement. In the 1990s, states then adopted more rigorous standard-based exit exams to ensure that students mastered higher level of skills. However, since 2012, the number of states requiring HSEE has declined rapidly. In fact, only eleven states have exit exams in place for the high school class of 2020. This new policies wave reflects current widespread belief that exit exams bring few positives but many negatives to students, especially the most disadvantaged groups.

As part of this new policies wave, in March 2015, the state of Georgia eliminated exit exams from high school graduation requirements in an education reform. In this chapter, I study this reform's impact on 4-year adjusted cohort graduation rates (ACGR), post-secondary institutions enrollment and employment rate of new high school graduates. Using the state of Florida as control, I estimated the impact on 4-year ACGR[1] from weighted difference-in-difference regressions at the schools level over the period from 2012 to 2018. I found that the reform increased the 4-year ACGR rate by 4% for all students, and by 1.4% for White, 5.98% for Black and 5.59% for Hispanic subgroups. I then estimated the impact on post-secondary institutions enrollment and employment rate from weighted difference-in-

1. See section 2.2.3 for more details on how this rate is calculated.

difference regressions at the district level[2] over the period from 2011 to 2016. I found that the reform decreased post-secondary institution enrollment by 1.8 % but had no statistically significant impact on employment rate among new high school graduates who don't pursue post-secondary education. These results suggest that although exit exams elimination increased 4-year ACGR (especially among Black and Hispanic students), the additional high school diploma holders are less likely to enroll in post-secondary institutions and are no more likely to be employed if choosing not to pursue post-secondary education.

This chapter contributes to the empirical literature on the impact of exit exams. While much previous researches have studied exit exams adoption, this chapter is among the firsts to study exit exams elimination. Furthermore, in order to increase transparency and accountability in high school education across different states, the U.S. Department of Education has required states to submit standardized and quality reports of 4-year ACGR and post high-school graduation outcomes since the school year 2010-2011[3]. This chapter is among the firsts to study these new reports. Finally, no other work has studied the 2015 exit exams elimination reform in the state of Georgia.

A number of papers have investigated the effect of HSEE on students' achievement scores, drop out, completion rates and other outcome variables. Jacob (2001) found graduation tests to have no significant impact on 12th-grade math or reading achievement, after controlling for prior student achievement and various student-level, school-level, and state-level variables. The author also found that graduation tests increased the probability of dropping out among the lowest-ability students but had no appreciable effect for the average students. Grodsky et al (2009) used nationally representative data to study the effects of state HSEE on students achievement in mathematics and reading between 1971 and 2004. They found that state HSEE did not have any significant impact on students achievement in either reading or mathematics either at the mean or at the 10th, 20th, 80th or 90th percentiles of the

2. The post high-school outcomes data are only available at the district level. See section 2.3.1 for more details.

3. See section 2.2.3 for more details

achievement distribution.

Dongshu Ou (2009) used a regression discontinuity design and data from the state of New Jersey to study the effect of failing the HSEE on early drop-out probability. They found HSEE to lead to higher dropout probability among minority and low-income students. Reardon et al (2010) used a regression discontinuity design and student-level data from four large California public school districts to estimate the effect of failing high school exit exam on subsequent student achievement, course taking, persistence and graduation. They found no evidence of significant effect of failing the exam on high school course-taking, achievement, persistence, or graduation for students with test scores near the exit exam passing score. Papay et al (2010) used a regression discontinuity design to evaluate the impact of HSEE on the probability of graduation. The authors found that low-income urban students who just fail the mathematics examination have an 8 percentage point lower graduation rate. They also found that failing the mathematics test does not affect the likelihood of on-time grade promotion, but increased the drop out rate in the year following the test by 4 percentage points. Ahn (2014) used regression discontinuity to analyze the impact of passing a mathematics exit exam (Algebra I) on students' decision of choosing between a more rigorous college-preparatory math curriculum and an easier career-track math curriculum. The author found a 5 percentage point gap in the probability of selecting the rigorous curriculum between 9th grade students who just passed and those who just failed the exam.

Dee and Jacob (2007) analyzed the impact of HSEE using Census and Common Core of Data. They found that exit exams significantly reduced the probability of completing high school, particularly for black students. Papay et al (2014) studied the impact of state-mandated HSEE in Massachusetts on the probability of college enrollment. They found that barely failing an exit exam, for students on the margin of passing, reduces the probability of college attendance several years after the test. Holme et al (2010) reviewed 46 unique studies on the impact of exit exams policies on student achievement, graduation, postsecondary outcomes, and school responses. They concluded that exit exams policies have produced few of the expected benefits and have been associated with costs for the most disadvantaged students.

The rest of the chapter is structured as follows. Section 2 describes the public education system in the states of Georgia and Florida, and details of the reform. Section 3 describes the data and empirical strategy. Section 4 show the results and discussion. Section 5 concludes the chapter.

2.2 Institutional Background

2.2.1 Public Education System in the state of Georgia and Florida

Both Georgia and Florida public education system (prekindergarten through grade 12) operates within districts governed by locally elected school boards and superintendents. In 2013, public schools in Georgia enrolled 1,703,332 students in 2,387 schools in 218 school districts, and public schools in Florida enrolled 2,692,162 students in 4,269 schools and 76 school districts. Out of these totals, 478,840 students enrolled in 450 high schools in Georgia, and 854,453 students enrolled in 1,118 high schools in Florida. Additionally, 109,365 teachers were staffed in public schools in Georgia, compared to 176,537 teachers staffed in public schools in Florida.

Georgia also had roughly one administrator for every 280 students, while Florida had one administrator for every 327 students. The national average was one administrator for every 295 students in 2013. On average, Georgia spent $9,099 per pupil (ranked 37th highest in the nation), and Florida spent $8,433 per pupil (ranked 42nd highest in the nation). The similarity in public education systems and cultures, as well as proximity in geography, between these two states make Florida a suitable control state for my empirical analyses.

2.2.2 Georgia Graduation Tests Waiver Reform

On March 30, 2015, Georgia Governor Nathan Deal signed House Bill 91 into law. This law provided that students shall no longer be required to earn a passing score on any graduation tests to earn a high school diploma in the state of Georgia. This law covers all subjects, forms and versions of the Georgia High School Graduation Tests (English Language Arts, Mathematics, Science, and Social Studies), Georgia High School Writing Test, and Basic Skills Tests (Reading, Mathematics, and Writing). This law also contained additional details and requirements regarding students who were no longer enrolled and whose sole reason for not receiving a high school diploma was due to not passing any part of the graduation tests. These students may petition the local board of education for a diploma. This law was effective immediately for the graduating class of 2015 and beyond.

Hence, all students who entered grade nine for the first time after July 1, 1991, and through June 30, 2011, must pass the Georgia High School Graduation Tests (GHSGT) to earn a high school diploma. Students who entered grade nine for the first time after July 1, 2011 are not required to pass the GHSGT to earn a high school diploma.

2.2.3 Adjusted Cohort Graduation Rate

In an effort to increase transparency and accountability in high school education across different states, the U.S. Department of Education [section 1111(h) of ESEA] amended a regulation in October 2008. The amendment requires all states and local educational agencies (LEAs) receiving Title I funds (money for schools with a certain percentage of low-income students) to begin calculating and reporting the Adjusted Cohort Graduation Rate (ACGR) starting with 2010-2011 school year's data.

To calculate the ACGR, states identify the "cohort" of first-time 9th graders in a particular school year, and adjust this number by adding any students who transfer into the cohort after 9th grade and subtracting any students who transfer out, emigrate to another country, or pass away. The ACGR is the percentage of the students in this cohort who graduate within four years. States calculate the ACGR for individual schools and districts and for the state as a whole using detailed data that track each student over time. Having the same data collection requirements across state improves confidence and accountability for the aggregate high school graduation rates over time.

In addition, the regulation required states to set goals to improve graduation rates, and school districts to intervene in high schools where students from low-income families, students of color, and other traditionally underserved students had consistently low graduation rates.

2.3 Empirical Strategy

2.3.1 Data

I collect publicly available reports from Georgia and Florida Department of Education websites on 4-year ACGR at the school level, post-secondary institutions enrollment and employment rate[4] of new high school graduates at the district level. The ACGR reports are available for years from 2012 to 2018, and the high school graduates outcomes reports are available for years from 2011 to 2016.

Schools with missing 4-year ACGR reports were dropped from the sample, as similar to districts with missing high school graduates outcomes reports. The ACGR reports at the school level are also available for White, Black and Hispanic subgroups, but the high school graduates outcomes reports at the district level are not.

4. Of students not continuing post-secondary education.

2.3.2 Empirical Strategy

Using Florida schools and districts as controls, I estimate the impact of HSEE elimination using difference-in-difference estimators:

$$y_{it} = \beta_0 + \beta_1 Georgia_i * Post_t + \sum_{i=1}^{N} \lambda_i + \sum_{t=t_0}^{T} \alpha_t + \epsilon_{it}$$

where y_{it} is 4-year ACGR for school i in year t, post-secondary institutions enrollment rate[5] and employment rate[6] for district i in year t. $Post_t$ is a dummy variable indicating whether the graduating class is in 2015 or after. $Georgia_i$ is a dummy variable indicating whether school i or district i is in the state of Georgia. α_t's are year fixed effects and λ_i's are school or district fixed effects.

To test for the parallel trend assumption, I also report results of the following model specification:

$$y_{it} = \beta_0 + \beta_1 Georgia_i + \sum_{t=t_0}^{T} \alpha_t + \sum_{t=t_0+1}^{T} \gamma_t Georgia_i * \alpha_t + \epsilon_{it}$$

where $Georgia_i$, α_t and y_{it} are defined similarly as above. If the parallel trend assumption holds, we should see $\gamma_t = 0$ for $t < 2015$.

I also report results for the following model specification, which show the differences between treatment and control group for each year through parameters γ_t's:

$$y_{it} = \beta_0 + \sum_{t=t_0+1}^{T} \gamma_t Georgia_i * \alpha_t + \sum_{i=1}^{N} \lambda_i + \sum_{t=t_0}^{T} \alpha_t + \epsilon_{it}$$

where $Georgia_i$, α_t, λ_i, y_{it} and γ_t are defined as previously.

Finally, all above specifications are weighted by schools' graduating cohort size for 4-year

5. Of new high school graduates one year following high school graduation

6. Of new high school graduates who don't continue post-secondary education

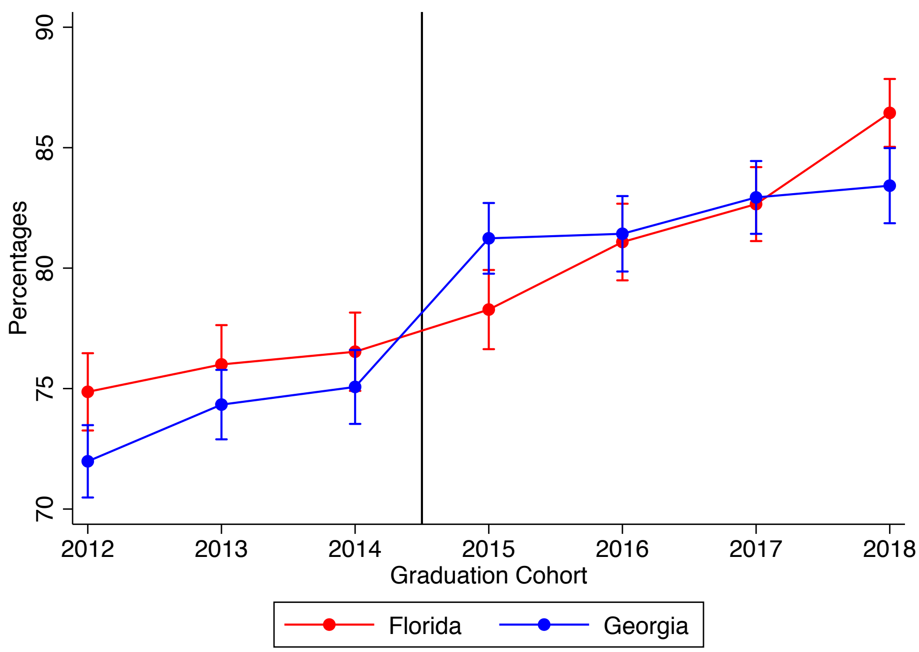
ACGR, and by districts' graduating cohort size for post-secondary institutions enrollment and employment rate. I show the regression results in the next section.

2.4 Results and Discussions

2.4.1 Effect of the reform on 4-year Adjusted Cohort Graduation Rates

Figure 2.1 plot the weighted mean and 95% confidence interval of 4-year ACGR for all students at the school level between Florida and Georgia:

Figure 2.1: 4-year ACGR for all students between Florida and Georgia



Note: The mean percentages are weighted by schools' graduating cohort size
Source: Data from Florida and Georgia Department of Education

The plot shows a clear parallel trend of 4-year ACGR between Georgia and Florida before the policy change in 2015 (denoted by a vertical line). Florida had higher 4-year ACGR compared to Georgia before 2015, but Georgia overtook Florida in 2015. Georgia's

4-year ACGR subsequently declined and fell below Florida's in 2018. The plot suggests that exit exams eliminations temporarily increase 4-year ACGR rate at the school level. Table 2.1 below show regression results for the difference-in-difference estimator:

Table 2.1: Effect of the reform on 4-year Adjusted Cohort Graduation Rates for all students

Dependent variable: 4-year ACGR	(1)	(2)	(3)
Post=1 × Georgia=1	3.974***		
	(15.88)		
Georgia=1		-2.884*	-319.7
		(-2.41)	(.)
Georgia=1 × classyear=2013		1.216	0.926
		(0.73)	(1.78)
Georgia=1 × classyear=2014		1.422	1.830***
		(0.81)	(3.69)
Georgia=1 × classyear=2015		5.841**	6.683***
		(3.14)	(14.26)
Georgia=1 × classyear=2016		3.227	5.139***
		(1.75)	(10.89)
Georgia=1 × classyear=2017		3.160	4.950***
		(1.69)	(9.96)
Georgia=1 × classyear=2018		-0.135	2.866***
		(-0.07)	(5.49)
Constant	73.97***	74.86***	196.3
	(397.83)	(88.75)	(.)
Years FE	Y	Y	Y
School FE	Y	N	Y
Observations	8475	8475	8475

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All regressions are weighted by schools' graduating cohort size

Source: Data from Florida and Georgia Department of Education

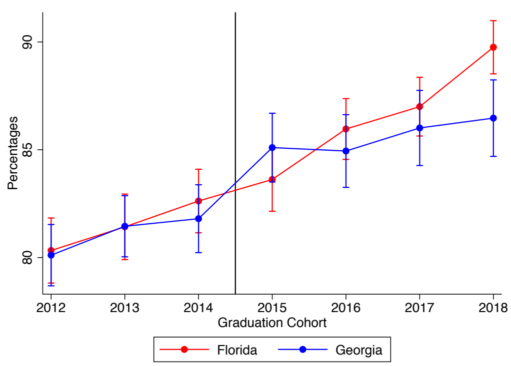
The first column of the table shows result for the first model specification. The reform is estimated to increase 4-year ACGR by about 4 % among all students in the period from 2012 to 2018. The effect is statistically significant at the 0.1 % level. The second column of the table shows that the parallel trend assumption is satisfied. The coefficients of the interaction terms between Georgia and years dummies for 2013 and 2014 are not statistically different from zero. The interaction term between Georgia and year-2015 dummies is large and statistically significant, confirming the effect of the reform. The last column of the table show results for the third specification and further confirm the results of the first two specifications. These results suggest that exit exams are difficult obstacles for a group of student in getting their high school diplomas.

2.4.2 Effect of the reform on 4-year Adjusted Cohort Graduation Rates by racial subgroups

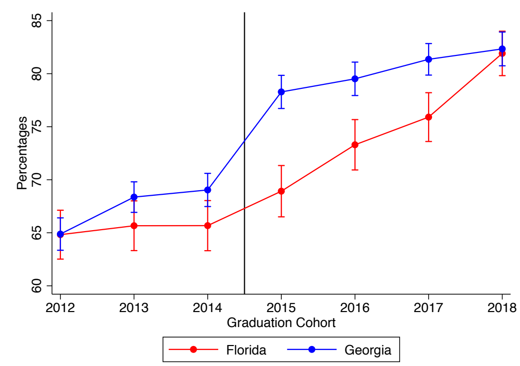
Figure 2.2 plots the weighted mean and 95% confidence interval of the 4-year ACGR for White, Black and Hispanic subgroups at the school level between Florida and Georgia.

Figure 2.2: 4-year ACGR By Racial Subgroups Between Georgia and Florida

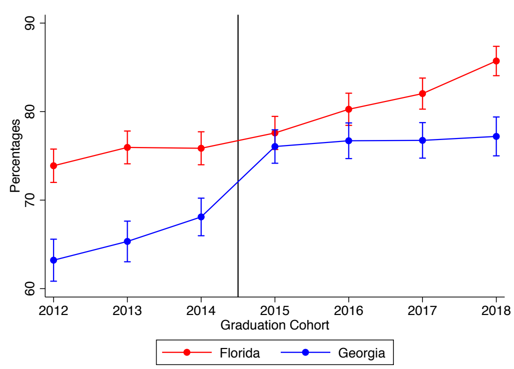
(a) White



(b) Black



(c) Hispanic



Note: The mean percentages are weighted by schools' graduating cohort sizes
Source: Data from Florida and Georgia Department of Education

The plots show clear parallel trends of 4-year ACGR for White and Hispanic subgroups between Georgia and Florida, but less so for Black subgroup. However, the figures clearly show that exit exams elimination increased 4-year ACGR for all three subgroups in 2015 in Georgia compared to Florida. Table 2.2 below show regression results for the difference-in-difference estimators:

Table 2.2: Effect of the reform on 4-year Adjusted Cohort Graduation Rates for White, Black, and Hispanic Subgroups

Dependent variable: 4-year ACGR	White	White	Black	Black	Hispanic	Hispanic
Georgia=1	-0.216		0.0560		-10.67***	
	(-0.18)		(0.03)		(-5.41)	
Post=1 × Georgia=1		1.386***		5.980***		5.595***
		(5.51)		(14.76)		(10.12)
Georgia=1 × classyear=2013	0.244		2.644		0.0478	
	(0.14)		(1.05)		(0.02)	
Georgia=1 × classyear=2014	-0.604		3.309		2.913	
	(-0.33)		(1.26)		(1.06)	
Georgia=1 × classyear=2015	1.694		9.303**		9.138***	
	(0.89)		(3.23)		(3.46)	
Georgia=1 × classyear=2016	-0.806		6.163*		7.114**	
	(-0.41)		(2.20)		(2.72)	
Georgia=1 × classyear=2017	-0.772		5.392*		5.387*	
	(-0.36)		(2.03)		(2.04)	
Georgia=1 × classyear=2018	-3.072		0.368		2.152	
	(-1.42)		(0.14)		(0.79)	
Constant	80.33***	80.16***	64.82***	65.32***	73.89***	72.26***
	(111.83)	(447.18)	(43.37)	(198.71)	(69.74)	(228.19)
Years FE	Y	Y	Y	Y	Y	Y
School FE	N	Y	N	Y	N	Y
Observations	6632	6632	6435	6435	5259	5259

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All regressions are weighted by schools' graduating cohorts sizes

Source: Data from Florida and Georgia Department of Education

The second, fourth and sixth columns show results of the first specification for the three subgroups. Exit exams elimination is estimated to increase 4-year ACGR by about 1.39% for White, 5.98% for Black and 5.59% for Hispanic in the period from 2012 to 2018. All three estimates are statistically significant at the 0.1% level. The first, third and fifth columns show that parallel trend assumption is satisfied for all three subgroups. Although the coefficients of the interaction terms between Georgia and years dummies for 2013 and 2014 are not statistically different from zero for all three subgroups, the evidence is stronger for Hispanic and White than Black. The interaction terms between Georgia and year-2015 dummies are large and statistically significant for Black and Hispanic, confirming the large effect of the reform for these two subgroups in 2015.

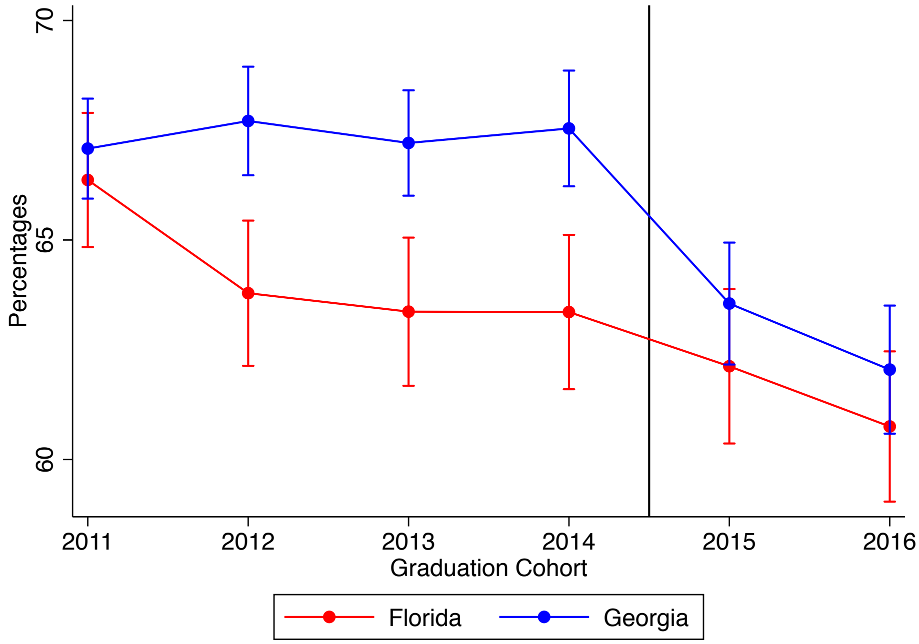
Overall, the results suggest that exit exams elimination had a larger effect on Black and Hispanic than White. In other words, high school exit exams appear to challenge a larger group of Black and Hispanic students compared to White. This finding agrees with previous researches concluding that the most disadvantaged students bear the largest cost of exit exams adoption.

2.4.3 Effect of the reform on post-secondary institution enrollment rates

Figure 2.3 plots the weighted mean and 95% confidence interval of post-secondary institutions enrollment rates (one year following high school graduation) at the district level between Georgia and Florida. The plot shows a clear parallel trend of post-secondary institutions enrollment rates between Florida and Georgia in years 2012, 2013 and 2014 despite deviating from year 2011. The reform also appears to decrease Georgia enrollment rate significantly compared to Florida in 2015, but not in 2016.

Table 2.3 show regression results for the difference-in-difference estimators.The first column shows results of the first model specification. Exit exams elimination is estimated to decrease post-secondary institutions enrollment rates by about 1.8% in the time period from 2011 to 2016. This effect is statistically significant at the 0.1% level. The second column shows that the parallel trend assumption is satisfied. Although the coefficients of the interaction terms between Georgia and 2012, 2013, 2014 years dummies are statistically different from zero, they are very similar in magnitude. They reflect deviation in post-secondary insti-

Figure 2.3: Post-secondary Institution Enrollment Rates Between Florida and Georgia



Note: The mean percentages are weighted by districts' graduating cohort sizes
Source: Data from Florida and Georgia Department of Education

tutions enrollment rates from the base year 2011 (as can be seen from figure 2.3). I conclude that the parallel trend assumption is satisfied for the three years prior to the policy change (2012, 2013 and 2014). The last column show results of the third specification (similar to the second specification but also include school fixed effects - see section 2.3.3 for more details) and further confirm the results of the first two columns.

These results suggest that the additional (marginal) diploma holders are less likely to be enrolled in postsecondary institution, one year following graduation, compared to existing diploma holders. This conclusion follows from the findings in previous sections that exit exams elimination increased 4-year ACGR by 4%.

Table 2.3: Effect of the reform on Post-secondary Institutions Enrollment Rates

Dependent variable: Post-secondary Enrollment Rate	(1)	(2)	(3)
Post=1 × Georgia=1	-1.784***		
	(-4.58)		
Georgia=1		0.714	0
		(0.37)	(.)
Georgia=1 × classyear=2012		3.208	2.995***
		(1.12)	(4.24)
Georgia=1 × classyear=2013		3.128	2.950***
		(1.06)	(4.61)
Georgia=1 × classyear=2014		3.467	3.347***
		(1.13)	(5.01)
Georgia=1 × classyear=2015		0.714	0.755
		(0.23)	(1.10)
Georgia=1 × classyear=2016		0.582	0.374
		(0.19)	(0.51)
Constant	66.65***	66.37***	66.66***
	(208.51)	(45.76)	(235.13)
Years FE	Y	Y	Y
School FE	Y	N	Y
Observations	1528	1528	1528

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

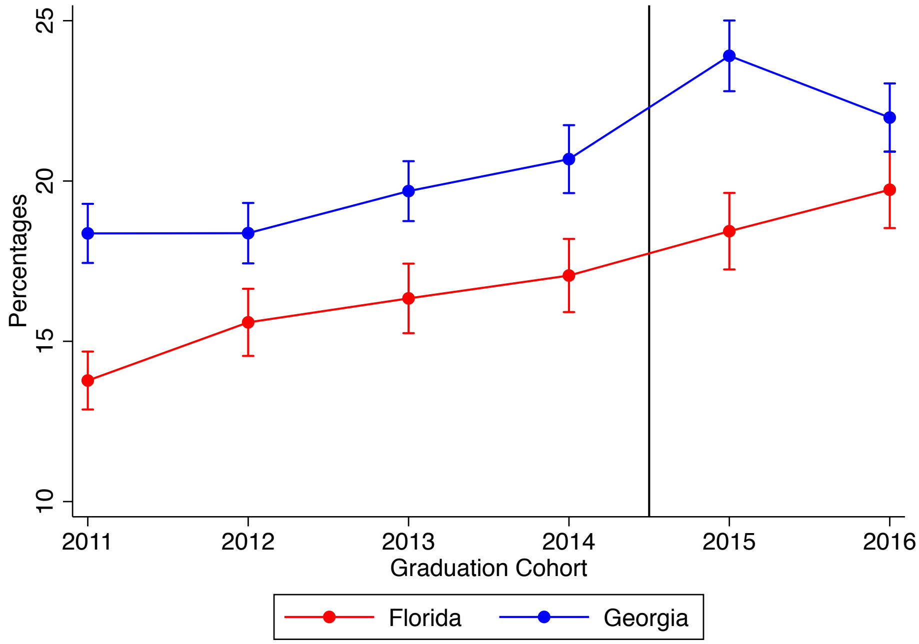
Note: All regressions are weighted by districts graduating cohorts' sizes.

Source: Data from Florida and Georgia Department of Education.

2.4.4 Effect of the reform on employment rates of new high school graduates

Figure 2.4 plots the weighted mean and 95% confidence interval for the employment rate (of new high school graduates who do not pursue postsecondary education, one year following high school graduation) at the district level between Georgia and Florida:

Figure 2.4: Employment Rate of New High School Graduates Between Florida and Georgia



Note: The mean percentages are weighted by district graduating cohorts' size
Source: Data from Florida and Georgia Department of Education

The plot shows a clear parallel trend of employment rate between Florida and Georgia in years from 2011 to 2014, before the policy change. The reform appears to increased the employment rate in Georgia compared to Florida in 2015, but then declined in 2016.

Table 2.4 show regression results for the difference-in-difference estimators. The first column shows result of the first specification. Exit exams elimination is estimated to increase employment rate (of new high school graduates who don't continue postsecondary education) by 0.24% in the period from 2011 to 2016. The effect is small and not statistically significant. The second column shows that the parallel trend assumption is satisfied. The coefficients of the interaction terms between Georgia and 2012, 2013 and 2014 years dummies are negative but not statistically different from zero. The third column confirms findings from the first two columns.

These results suggest that the additional (marginal) diploma holders are no more likely to be employed one year following high school graduation, compared to existing diploma holders

Table 2.4: Effect of the reform on employment rate of new high school graduates

Dependent variable: Employment Rate	(1)	(2)	(3)
Post=1 × Georgia=1	0.235		
	(0.72)		
Georgia=1		4.590**	0
		(3.27)	(.)
Georgia=1 × classyear=2012		-1.807	-1.646**
		(-0.88)	(-3.03)
Georgia=1 × classyear=2013		-1.243	-1.131*
		(-0.60)	(-2.08)
Georgia=1 × classyear=2014		-0.959	-0.869
		(-0.43)	(-1.66)
Georgia=1 × classyear=2015		0.881	0.874
		(0.39)	(1.60)
Georgia=1 × classyear=2016		-2.339	-2.189***
		(-1.05)	(-3.58)
Constant	15.73***	13.78***	15.75***
	(69.67)	(16.37)	(69.20)
Years FE	Y	Y	Y
School FE	Y	N	Y
Observations	1518	1518	1518

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All regressions are weighted by districts graduating cohorts' sizes.

Source: Data from Florida and Georgia Department of Education.

who don't continue postsecondary education. This conclusion follows from the findings in previous sections that exit exams elimination increased 4-year ACGR by 4%.

2.5 Conclusion

In March 2015, the state of Georgia eliminated exit exams from high school graduation requirements in an education reform. In this chapter, I study this reform's impact on 4-year adjusted cohort graduation rates (ACGR), post-secondary institutions enrollment and employment rate of new high school graduates. Using the state of Florida as control, I estimated the impact on 4-year ACGRfrom weighted difference-in-difference regressions at the schools level over the period from 2012 to 2018. I found that the reform increased the 4-year ACGR rate by 4% for all students, and by 1.4% for White, 5.98% for Black and 5.59% for Hispanic subgroups. I then estimated the impact on post-secondary institutions enrollment and employment rate from weighted difference-in-difference regressions at the district level over the period from 2011 to 2016. I found that the reform decreased post-secondary institution enrollment by 1.8 % but had no statistically significant impact on employment rate among new high school graduates who don't pursue post-secondary education. These results suggest that although exit exams elimination increased 4-year ACGR (especially among Black and Hispanic students), the additional high school diploma holders are less likely to enroll in post-secondary institutions and are no more likely to be employed if choosing not to pursue post-secondary education.

References

[1] Ahn, Tom, 2014. *"A regression discontinuity analysis of graduation standards and their impact on students? academic trajectories,"* Economics of Education Review, Elsevier, vol. 38(C), pages 64-75.

[2] Dee, T., & Jacob, B.A. (2007). *"Do High School Exit Exams Influence Educational Attainment or Labor Market Performance?"* In A. Gamoran (Ed), Standards-Based Reform and Children in Poverty: Lessons for "No Child Left Behind". Brookings Institution Press.

[3] Dongshu Ou, 2009. *"To Leave or Not to Leave? A Regression Discontinuity Analysis of the Impact of Failing High School Exit Exam."* CEE Discussion Papers 0107, Centre for the Economics of Education, LSE.

[4] Grodsky, E., Warren, J. R., & Kalogrides, D. (2009). *"State High School Exit Examinations and NAEP Long-Term Trends in Reading and Mathematics, 1971-2004."* Educational Policy, 23(4), 589-614.

[5] Holme, J. J., Richards, M. P., Jimerson, J. B., & Cohen, R. W. (2010). *"Assessing the Effects of High School Exit Examinations."* Review of Educational Research, 80(4), 476-526.

[6] Jacob, B. A. (2001). *"Getting Tough? The Impact of High School Graduation Exams."* Educational Evaluation and Policy Analysis, 23(2), 99-121.

[7] Reardon, S. F., Arshan, N., Atteberry, A., & Kurlaender, M. (2010). *"Effects of Failing a High School Exit Exam on Course Taking, Achievement, Persistence, and Graduation."* Educational Evaluation and Policy Analysis, 32(4), 498-520.

[8] Papay, J. P., Murnane, R. J., & Willett, J. B. (2010). *"The Consequences of High School Exit Examinations for Low-Performing Urban Students: Evidence From Massachusetts."* Educational Evaluation and Policy Analysis, 32(1), 5-23.

[9] John P. Papay, Richard J. Murnane & John B. Willett (2014) *"High-School Exit Examinations and the Schooling Decisions of Teenagers: Evidence From Regression-Discontinuity Approaches."* Journal of Research on Educational Effectiveness, 7:1, 1-27.

Chapter 3

Semi-parametric Estimation of Tobit Type-3 Sample Selection Model

Abstract

I propose a new semi-parametric estimator for the Tobit Type-3 sample selection model. I first generalize the mode regression estimator to its bivariate counterpart. I then apply this bivariate mode regression estimator to the Tobit Type-3 model under two separate sets of assumptions. First set will be computationally simpler but requires more restrictive assumptions, while second set will be more computationally taxing yet requires less restrictive assumptions. I then show that the estimators are identified and strongly consistent under both sets of assumptions.

3.1 Introduction

Sample selection models are frequently used in empirical studies of labor supply, migration, trade and health insurance. The two most commonly used models are the Type-2 and Type-3 Tobit models (Amemiya, 1985: Chapter 10). The difference between the two is that the selection variable is a binary variable (e.g., participant or non-participant) in the Type-2 and a censored or truncated variable in the Type-3. A classical example of the Type-2 Tobit model is Gronau's (1973) female labor supply model. The labor supply model in Heckman (1974) is a variation of Type-3 Tobit model, and also Udry (1994). A more recent example is Mulligan and Rubinstein (2008).

The increment of information available in the Tobit Type-3 model (compared to Type-2) allows for both identifying the model parameters under weaker restrictions and obtaining more efficient estimators. Hence, it is worthwhile to study estimation techniques of the Tobit Type-3 model, independent of the Tobit-Type 2 model. The Tobit Type-3 model is formally represented as a system of two equations:

$$Y_1^* = X_1\beta_1 + \epsilon_1$$
$$Y_2^* = X_2\beta_2 + \epsilon_2$$

where the first equation is the selection equation, and the second equation is the outcome equation. The dependent variables (Y_1^*, Y_2^*) are observed only if the selection variable Y_1^* is positive.

The Tobit Type-3 model is traditionally estimated by maximum likelihood methods, most commonly the Heckman's two-step method (Heckman, 1979). However, consistency of such likelihood-based parametric estimators relies heavily upon correct specification of the error terms' distribution function, in the presence of limited dependent variables. Lee (1982) adopted a rich parametric family to mitigate the misspecification problem to a certain extent.

Other semi-parametric estimation procedures using various moment-based approaches have been proposed (see Lee (1994), Chen (1997), Honore et al (1997), Li and Wooldridge (2002), and Zhou and Pan (2015)).

However, all of the semi-parametric estimators above require the strong assumption of statistical independence between the error terms and the covariates, which is unlikely to hold in practice. In this chapter, I propose an alternative semi-parametric estimator that does not require the full statistical independence assumption. More specifically, I first extend the mode regression estimator (Lee (1989) and Lee (1993)) to its bivariate counterpart. I then apply this bivariate mode regression estimator to the Tobit Type-3 model. I develop the estimators under two separate sets of assumption. First set is computationally simpler but requires more restrictive assumptions, while second set is more computationally taxing yet requires less restrictive assumptions. I then show that the estimators are identified and strongly consistent under both sets of assumptions.

The remainder of the chapter is structured as follows. Section 2 extends the mode regression estimator to its bivariate counterpart, introduces the model's assumptions and the estimators. Section 3 then shows proofs of identification and section 4 shows proofs of strong consistency for the parameters. Section 5 briefly discuss asymptotic distribution and choice of bandwidth.

3.2 The model and the estimator

3.2.1 Bivariate extension of the mode regression estimator

I first extend the mode regression estimator (Lee (1989)) to its two-dimensional counterpart. I propose the following loss function for a bivariate random variable $\boldsymbol{y}$ and predictor $\boldsymbol{p}$:

$$L(\boldsymbol{y}, \boldsymbol{p}) = 1[|y_1 - p_1| \leq w].1[|y_2 - p_2| \leq w]$$

where $\boldsymbol{y} = (y_1, y_2)$ and $\boldsymbol{p} = (p_1, p_2)$.

This loss function is uniquely maximized at $\boldsymbol{p} = mode(\boldsymbol{y})$ if $f(y_1, y_2)$ is locally bi-radial symmetric (up to a square of length w) around its unique global mode.

Next, consider a system of two-equations model:

$$y_1 = x_1\beta_1 + \epsilon_1$$

$$y_2 = x_2\beta_2 + \epsilon_2$$

For ease of notations, I denote $\boldsymbol{y} = (y_1, y_2)$, $\boldsymbol{x} = (x_1, x_2)$ and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2)$. If $mode(\boldsymbol{\epsilon}) = \mathbf{0}$, then $mode(\boldsymbol{y}|\boldsymbol{x}) = (x_1\beta_1, x_2\beta_2)$. Hence, if $f(\boldsymbol{\epsilon})$ is locally bi-radial symmetric around its unique global mode, then similar as above, $(x_1\beta_1, x_2\beta_2)$ uniquely maximizes the following expected loss function:

$$E_{\boldsymbol{x}}E_{\boldsymbol{y}|\boldsymbol{x}}1[|y_1 - x_1\beta_1| \leq w].1[|y_2 - x_2\beta_2| \leq w]$$

The finite-sample counterpart for this objective function is

$$(1/T)\sum_t 1[|y_{1t} - x_{1t}\beta_1| \leq w].1[|y_{2t} - x_{2t}\beta_2| \leq w]$$

I can also generalize the bivariate mode regression estimator above (with the square kernel) by using other kernel weight functions $K()$

$$(1/T)\sum_t K(y_{1t} - x_{1t}\beta_1, y_{2t} - x_{2t}\beta_2)$$

For example, the population objective function for the bivariate mode regression estimator would be

$$Q(\beta_1, \beta_2) = E[[w^2 - (y_1 - x_1\beta_1)^2][w^2 - (y_2 - x_2\beta_2)^2]1[|y_1 - x_1\beta_1| \leq w].1[|y_2 - x_2\beta_2| \leq w]]$$

Its finite-sample counterpart would be

$$Q_T(\beta_1, \beta_2) =$$
$$(1/T)\sum_t [w^2 - (y_{1t} - x_{1t}\beta_1)^2][w^2 - (y_{2t} - x_{2t}\beta_2)^2]1[|y_{1t} - x_{1t}\beta_1| \le w].1[|y_{2t} - x_{2t}\beta_2| \le w]$$

3.2.2 The model

I then apply the bivariate mode regression estimator above to the Tobit type-3 sample selection model. First, I describe the model's assumptions:

Assumption 1. System of two linear models and truncated random sample

$$y_{1t}^* = x_{1t}\beta_1 + \epsilon_{1t}$$
$$y_{2t}^* = x_{2t}\beta_2 + \epsilon_{2t}$$

where (y_{1t}^*, y_{2t}^*) are observed if $y_{1t}^* > 0$, $x_{1t} = (1, x_{1t}^1, x_{1t}^2, \ldots, x_{1t}^{k_1})$ and $x_{2t} = (1, x_{2t}^1, x_{2t}^2, \ldots, x_{2t}^{k_2})$. The first equation is the selection equation and the second equation is the outcome equation. For ease of notations, I denote $\boldsymbol{y} = (y_1, y_2)$, $\boldsymbol{x} = (x_1, x_2)$ and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2)$

Assumption 2. Monotonicity and unimodality of $\boldsymbol{\epsilon}|\boldsymbol{x}$

I consider 2 separate cases below. Case 1 will be computationally easier but requires more restrictive assumptions. Case 2 is less restrictive but will be more computationally taxing. I develop two separate estimators for each case.

Case 1: $\boldsymbol{\epsilon}|\boldsymbol{x}$ has a joint density which is unimodal at (0,0). Furthermore, for all (x_1, x_2), the conditional joint density is bi-radial symmetric for a square of length w (for a suitably chosen w) at the unique global mode (0,0), and is increasing up to the (0,0) and decreasing after the (0,0) along any line that pass through (0,0).

Additionally, $\epsilon_1|x_1$ also has a density which is unimodal at 0. Furthermore, for all x_1, the conditional density of $\epsilon_1|x_1$ is symmetric at 0 (around the interval [-w,w]), and is also increasing up to the mode (0) and decreasing after the mode.

Case 2: $\boldsymbol{\epsilon}|\boldsymbol{x}$ has a joint density which is unimodal at (0,0). Furthermore, for all (x_1, x_2), the conditional joint density is bi-radial symmetric for a square of length w (for a suitably chosen w) at the unique global mode (0,0), and is increasing up to the (0,0) and decreasing after the (0,0) along any line that pass through (0,0).

Assumption 3. Conditional mode restriction

Under both case 1 and case 2, $mode(y_1^*, y_2^*|x_1, x_2) = (x_1\beta_1, x_2\beta_2)$ since $mode(\epsilon_1, \epsilon_2|x_1, x_2) = (0, 0)$ and the linear models in assumption 1.

Additionally, under case 1, $mode(y_1^*|x_1) = x_1\beta 1$ since $mode(\epsilon_1|x_1) = (0, 0)$.

Assumption 4. Wide support of $\boldsymbol{\epsilon}|\boldsymbol{x}$ distribution

The support of $\boldsymbol{\epsilon}|\boldsymbol{x}$ is wider than $[-w, w] \times [-w, w]$ for all (x_1, x_2) and for a w suitably chosen.

Assumption 5. Full rank conditions

$E(x_1 x_1^T 1[x_1\beta_1 \geq w])$ and $E(x_2 x_2^T 1[x_1\beta_1 \geq w])$ are positive definitive matrices.

Assumption 6. Compact parameter space

The parameter space B of (β_1, β_2) is a compact space.

3.2.3 The estimators

Case 1. Under this case, $mode(\epsilon_1|x_1) = 0$, and $mode(\epsilon_1, \epsilon_2|x_1, x_2) = (0, 0)$. As a result, $mode(y_1^*, y_2^*|x_1, x_2) = (x_1\beta_1, x_2\beta_2)$ and $mode(y_1^*|x_1) = x_1\beta 1$. I propose a two-step estimator

Step 1. I estimate β_1 using the univariate mode regression estimator proposed by Lee (1989)

$$\beta_1 = \arg\max_{b_1} E1[|y_1 - max(x_1 b_1, w)| \leq w]$$

Its finite-sample estimate $\hat{\beta}_1$ is

$$\hat{\beta}_1 = \arg\max_{b_1}(1/T)\sum_t 1[|y_{1t} - max(x_{1t} b_1, w)| \leq w]$$

Step 2. I estimate β_2 by maximizing the following population objective function (using the identified β_1 from step 1)

$$\beta_2 = \arg\max_{b_2} E1[x_1\beta_1 \geq w]1[|y_1 - x_1\beta_1| \leq w].1[|y_2 - x_2 b_2| \leq w]$$

Its finite-sample estimate $\hat{\beta}_2$ is:

$$\hat{\beta}_2 = \arg\max_{b_2}(1/T)\sum_t 1[x_{1t}\hat{\beta}_1 \geq w]1[|y_{1t} - x_{1t}\hat{\beta}_1| \leq w].1[|y_{2t} - x_{2t}b_2| \leq w]$$

Case 2. Under this case, $mode(\epsilon_1, \epsilon_2|x_1, x_2) = (0,0)$ and so $mode(y_1^*, y_2^*|x_1, x_2) = (x_1\beta_1, x_2\beta_2)$. I propose a two-step estimator

Step 1. For each pair of (x_1, x_2), find $p(x_1, x_2)$ such that:

$$p(x_1, x_2) = \arg\max_p E1[|y_1 - w| \leq w|x_1]1[|y_2 - p| \leq w|x_2]$$

Its finite-sample estimate $\hat{p}$ is:

$$\hat{p}(x_1, x_2) = \arg\max_p(1/N)\sum_{\substack{x_{1t}=x_1,\\ x_{2t}=x_2}} 1[|y_{1t} - w| \leq w]1[|y_{2t} - p| \leq w]$$

where N is the number of observations in the finite-sample such that $x_{1t} = x_1$ and $x_{2t} = x_2$.

Step 2. I estimate (β_1, β_2) by maximizing the following population objective function

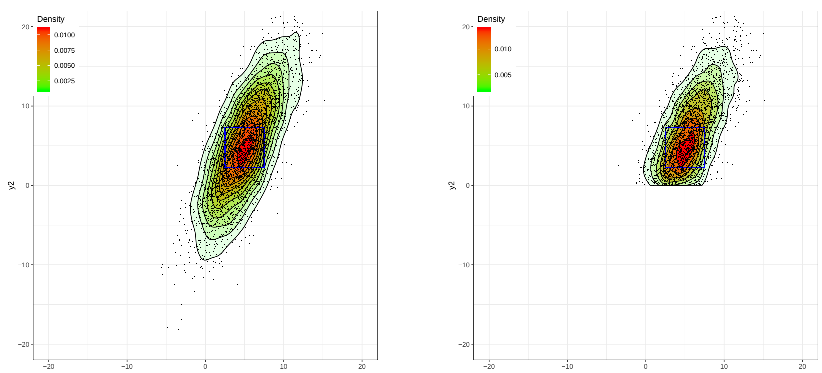
$$\begin{aligned} Q(\beta_1, \beta_2) =& E1[x_1\beta_1 \geq w]1[|y_1 - x_1\beta_1| \leq w]1[|y_2 - x_2\beta_2| \leq w]+ \\ & 1[x_1\beta_1 < w]\arg\max_p E1[|y_1 - w| \leq w|x_1]1[|y_2 - p| \leq w|x_2] \end{aligned}$$

The corresponding finite-sample counterpart is

$$\begin{aligned}Q_T(\beta_1, \beta_2) =&(1/T)\sum_t 1[x_{1t}\beta_1 \geq w]1[|y_{1t} - x_{1t}\beta_1| \leq w]1[|y_{2t} - x_{2t}\beta_2| \leq w]+\\ &1[x_{1t}\beta_1 < w]\arg\max_{p_t}(1/N_t)\sum_{\substack{x_{1k}=x_{1t},\\ x_{2k}=x_{2t}}} 1[|y_{1k} - w| \leq w]1[|y_{2k} - p_t| \leq w]\end{aligned}$$

3.2.4 Discussion

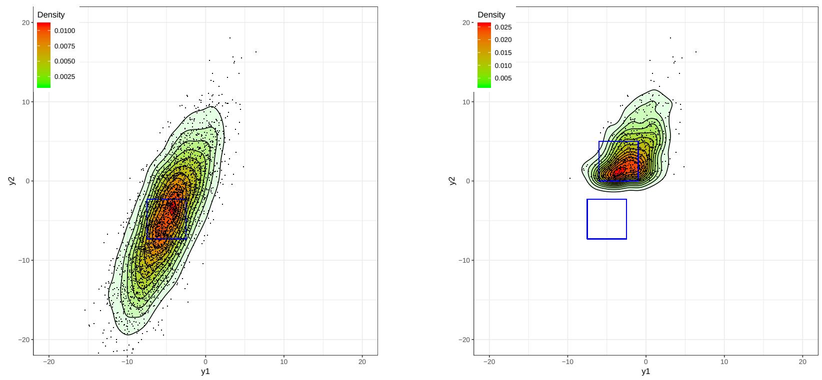
Next, I discuss the intuition of the proposed estimators via an illustrative example. In the example below, $y2$ is the selection (truncated) variable, and the error terms are bivariate normally distributed. The figures depict which point my proposed estimators identify when sample selection (truncation) censors the true mode of the joint distribution of ($y1$, $y2$) versus when truncation does not.



(a) Nontruncated Data (b) Truncated Data

Figure 3.1: Truncation does not censor the true mode of the joint distribution

In figure 3.1, truncation does not censor the true mode of the joint distribution of ($y1$, $y2$). That is, the mode of the uncensored joint distribution is the same as the mode of the censored joint distribution. In this case, the proposed estimators will be able to identify the true mode of the joint distribution (pre-censored distribution).



(a) Nontruncated Data

(b) Truncated Data

Figure 3.2: Truncation censors the true mode of the joint distribution

In figure 3.2, truncation censors the true mode of the joint distribution of $(y1, y2)$. That is, the mode of the uncensored joint distribution is different from the mode of the censored joint distribution. In this case, the proposed estimators will not be able to identify the true mode of the joint distribution (pre-censored distribution). The estimator will identify the mode of the post-censored joint distribution instead.

The example above highlights the main intuition that identification of the slope parameters (β_1, β_2) is possible if there is enough probability density in the region where truncation does not censor the true mode of the joint distribution. In that case, the proposed estimators will identify the slope parameters using data from that region where we can learn useful information about the true mode. I formally prove identification and strong consistency for the proposed estimators in the next sections.

3.3 Identification

The identification of the bivariate mode estimators will be shown by proving that the population objective functions achieves an unique global maximum at (β_1, β_2) for each case.

Theorem 1. *The parameters (β_1, β_2) are identified for case 1 under the model's assumptions.*

Proof. First, identification of β_1 follows from Theorem 1 in Lee (1989), since β_1 uniquely maximizes

$$E1[|y_1 - max(x_1 b, w)| \leq w]$$

under the model's assumptions. Next, I show that $x_2\beta_2$ maximizes

$$E_{\boldsymbol{y}|\boldsymbol{x}} 1[x_1\beta_1 \geq w] 1[|y_1 - x_{1t}\beta_1| \leq w].1[|y_2 - x_2 b| \leq w]$$

for a given (x_1, x_2) and the identified β_1 from above. I rewrite the above expression as

$$\begin{aligned}
&[F_{\boldsymbol{y}|\boldsymbol{x}}(x_1\beta_1 + w, x_2 b + w) - F_{\boldsymbol{y}|\boldsymbol{x}}(x_1\beta_1 + w, x_2 b - w) \\
&- F_{\boldsymbol{y}|\boldsymbol{x}}(x_1\beta_1 - w, x_2 b + w) + F_{\boldsymbol{y}|\boldsymbol{x}}(x_1\beta_1 - w, x_2 b - w)] 1[x_1\beta_1 \geq w]
\end{aligned}$$

For any x_2, and any x_1 such that $x_1\beta_1 \geq w$, $x_2\beta_2$ uniquely maximizes the above expression because $(x_1\beta_1,\ x_2\beta_2)$ is the unique global mode of joint distribution of (y_1, y_2). The monotonicity and unimodality assumption (Assumption 2, case 1 of the model) ensures that the square with the center located at the global mode $(x_1\beta_1,\ x_2\beta_2)$ will capture the most probability mass compared to any other location of (y_1, y_2).

For any x_2, and any x_1 such that $x_1\beta_1 < w$, the expression is equal to 0, and hence any value of $x_2 b$ will maximizes the expression.

Hence, when taking the expectation operator over all values of x_1 and x_2, $x_2\beta_2$ will uniquely maximize

$$E1[x_1\beta_1 \geq w] 1[|y_{1t} - x_{1t}\beta_1| \leq w].1[|y_{2t} - x_{2t} b| \leq w]$$

if $P(x_1\beta_1 \geq w) > 0$.

If $P(x_1\beta_1 \geq w) = 0$, then β_2 is not identified. However, this case is ruled out by Assumption 5. The full rank conditions (Assumption 5) also rules out the case of another $\gamma \neq \beta_2$ such that $P(x_2\beta_2 = x_2\gamma) = 1$. As a result, β_2 uniquely maximize $Q(b)$. Hence,

(β_1, β_2) are identified for case 1 under the model assumptions.

$\square$

Theorem 2. *The parameters (β_1, β_2) are identified for case 2 under the model's assumptions.*

Proof. I first show that $(x_1\beta_1, x_2\beta_2)$ maximizes

$$E_{\boldsymbol{y}|\boldsymbol{x}} 1[x_1b_1 \geq w]1[|y_1 - x_1b_1| \leq w]1[|y_2 - x_2b_2| \leq w] + \\ 1[x_1b_1 < w] \max_p E1[|y_1 - w| \leq w|x_1]1[|y_2 - p| \leq w|x_2]$$

for a given x_1 and x_2. I rewrite the expression above as:

$$\begin{aligned}
&[F_{\boldsymbol{y}|\boldsymbol{x}}(x_1b_1 + w, x_2b_2 + w) - F_{\boldsymbol{y}|\boldsymbol{x}}(x_1b_1 + w, x_2b_2 - w) \\
&- F_{\boldsymbol{y}|\boldsymbol{x}}(x_1b_1 - w, x_2b_2 + w) + F_{\boldsymbol{y}|\boldsymbol{x}}(x_1b_1 - w, x_2b_2 - w)]1[x_1b_1 \geq w] \\
&+ 1[x_1b_1 < w]\{\max_p F_{\boldsymbol{y}|\boldsymbol{x}}(2w, p + w) - F_{\boldsymbol{y}|\boldsymbol{x}}(2w, p - w) \\
&- F_{\boldsymbol{y}|\boldsymbol{x}}(0, p + w) + F_{\boldsymbol{y}|\boldsymbol{x}}(0, p - w)\}
\end{aligned}$$

For any x_2, and any x_1 such that $x_1\beta_1 \geq w$, $(x_1\beta_1, x_2\beta_2)$ is uniquely optimal since the second term is zero and the first term captures the probability mass of the square centered at the unique global mode $(x_1\beta_1, x_2\beta_2)$. This is uniquely optimal because of the monotonicity and unimodal assumption (Assumption 2 case 2). Choosing x_1b_1 such that $x_1b_1 < w$ is not optimal in this case because the first term will be zero while the second term will be strictly less than the value of the first term with $(x_1\beta_1, x_2\beta_2)$.

For any x_2, and any x_1 such that $x_1\beta_1 < w$, choosing any x_1b_1 such that $x_1b_1 < w$ (including $x_1\beta_1$) is at least weakly preferred to choosing any x_1b_1 such that $x_1b_1 \geq w$. This is because the square with the center located at the unique true global mode is censored in this case. Due to the monotonicity Assumption 2, the expression $max_p F_{\boldsymbol{y}|\boldsymbol{x}}(2w, p + w) - F_{\boldsymbol{y}|\boldsymbol{x}}(2w, p - w) - F_{\boldsymbol{y}|\boldsymbol{x}}(0, p + w) + F_{\boldsymbol{y}|\boldsymbol{x}}(0, p - w)$ will locate the (uncensored) square with the maximum probability mass, although the maximum is not necessarily unique.

Hence, when taking the expectation of the expression above over $\boldsymbol{x}$, $(x_1\beta_1, x_2\beta_2)$ will uniquely maximizes

$$E1[x_1b_1 \geq w]1[|y_1-x_1b_1| \leq w]1[|y_2-x_2b_2| \leq w]+1[x_1b_1 < w]\max_p E1[|y_1-w| \leq w]1[|y_2-p| \leq w]$$

if $P(x_1\beta_1 > w) > 0$. If $P(x_1\beta_1 \geq w) = 0$, then (β_1, β_2) are not identified. This case is ruled out by Assumption 5. The full rank conditions further rules out cases of $\alpha \neq \beta_1$ and $\gamma \neq \beta_2$ such that $P(x_1\beta_1 = x_1\alpha) = 1$ and $P(x_2\beta_2 = x_2\gamma) = 1$ when $x_1\beta_1 > w$.
Hence, the parameters (β_1, β_2) are identified for case 2 under the model's assumptions. □

3.4 Strong consistency

Theorem 3. *Strong consistency of the bivariate mode estimator (b_{1T}, b_{2T}) for case 1 under the model's assumptions.*

Proof. Strong consistency of b_{1T} follows from Theorem 2 of Lee (1989) under the model's assumptions. I show strong consistency of b_{2T} in three steps below:

Step 1. Almost sure uniform convergence of $Q_T(b_2)$ to $Q(b_2)$. $Q_T(b_2)$ converges a.s. and uniformly in b_2 to $Q(b_2)$:

$$Q(b_2) = E1[x_1\beta_1 \geq w]1[|y_{1t} - x_{1t}\beta_1| \leq w].1[|y_{2t} - x_{2t}b_2| \leq w]$$

Let I be the class of the indicator functions $1[|y_{2t} - x_{2t}b_2| \leq w]$, indexed by b_2. The graph of a real-valued function $h(x_1, x_2, y_1, y_2)$ is defined as (see Pollard, p.27)

$$G_h \equiv \{(x_1, x_2, y_1, y_2, z) : 0 \leq z \leq h(x_1, x_2, y_1, y_2) \text{ or } h(x_1, x_2, y_1, y_2) \leq z \leq 0\}$$

in $R^{k_1+k_2+3}$, where k_1 is the dimension of x_1 and k_2 is the dimension of x_2. Then I claim that the class of the "graphs" of the functions in I have "polynomial discrimination". A class of sets in $R^{k_1+k_2+3}$ has polynomial discrimination if the class is not too variable to pick up all 2^n subsets of a finite subset of $R^{k_1+k_2+3}$ with n elements (see Pollard, p.17, for the precise definition). For instance, on a real line R^1, the class of the sets composed of the connected closed intervals cannot pick up some subsets of a set with n elements: ordering the elements by 1,2,...,n, the class cannot pick up 2 and 4 without also picking up 3 (any interval including 2 and 4 should also include 3).

Back to the class of the graphs of the functions in I, note that G_h is the intersection of the two half spaces $\{y_2 - x_2 b - w \leq 0\}$ and $\{y_2 - x_2 b + w \geq 0\}$. The class of the half spaces has polynomial discrimination and so does the class formed by intersections of the half spaces. Hence, the class of G_h with $h = 1[|y_{2t} - x_{2t} b| \leq w]$ has polynomial discrimination.

The expression under the expectation also has three more half spaces, $\{x_1\beta_1 \geq w\}$, $\{y_1 - x_1\beta_1 - w \leq 0\}$ and $\{y_1 - x_1\beta_1 + w \geq 0\}$, in addition to $\{y_2 - x_2 b - w \leq 0\}$ and $\{y_2 - x_2 b + w \geq 0\}$. By repeatedly apply the same argument before, the class of the graphs of $Q(b)$ has polynomial discrimination. The fact that the class of the graphs has polynomial discrimination and that the indicator functions have an uniform upper bound 1 imply that the covering number for the class is bounded (see Pollard, p.27, lemma 25). The covering number is an index of how many 'distinct' members are in I (see Pollard, p.25, for the precise definition). The a.e. uniform convergence of $Q_T(b_2)$ to $Q(b_2)$ then comes from Pollard (p.25, theorem 24) using the boundedness of the covering number.

Step 2. Continuity of $Q(b_2)$. $Q(b_2)$ can be rewritten as

$$Q(b_2) = \int\int 1[x_1\beta_1 \geq w][F_{\boldsymbol{y}|\boldsymbol{x}}(x_1\beta_1 + w, x_2 b_2 + w) - F_{\boldsymbol{y}|\boldsymbol{x}}(x_1\beta_1 + w, x_2 b_2 - w) \\ - F_{\boldsymbol{y}|\boldsymbol{x}}(x_1\beta_1 - w, x_2 b_2 + w) + F_{\boldsymbol{y}|\boldsymbol{x}}(x_1\beta_1 - w, x_2 b_2 - w)]dG_{x_1,x_2}$$

The integrand is a bounded and continuous function of b_2, since $F_{\boldsymbol{y}|\boldsymbol{x}}$ is bounded and continuous in b_2 and $(x_2 b_2 + w)$ and $(x_2 b_2 - w)$ are both continuous in b_2. Then, due to the bounded convergence theorem, $Q(b_2)$ is continuous in b_2.

Step 3. Maximum of $Q(b_2)$. Due to the assumption that the parameter space B is compact (Assumption 6) and the continuity of $Q(b_2)$, $Q(b_2)$ attain its maximum on B. Hence there exists at least one b_2 which maximizes $Q(b_2)$.

By the three steps above and Theorem 1, b_{1T} and b_{2T} are strongly consistent for β_1 and β_2 for case 1 under the model's assumptions.

□

Theorem 4. *Strong consistency of the bivariate mode estimator (b_{1T}, b_{2T}) for case 2 under the model's assumptions.*

Proof. I show consistency of (b_{1T}, b_{2T}) in three standard steps as before.

Step 1. Almost sure uniform convergence of $Q_T(b_1, b_2)$ to $Q(b_1, b_2)$. $Q_T(b_1, b_2)$ converges a.s. and uniformly in (b_1, b_2) to $Q(b_1, b_2)$

$$\begin{aligned}Q(b_1, b_2) =&E1[x_1b_1 \geq w]1[|y_1 - x_1b_1| \leq w]1[|y_2 - x_2b_2| \leq w]\\ &+ 1[x_1b_1 < w] \max_p E1[|y_1 - w| \leq w|x_1]1[|y_2 - p| \leq w|x_2]\end{aligned}$$

The almost sure uniform convergence of the first term $1[x_1b_1 \geq w]1[|y_1-x_1b_1| \leq w]1[|y_2-x_2b_2| \leq w]$ follows from the same arguments of half-spaces, uniform bound and boundedness of the covering number in theorem 3. Consider the second term,

$$1[x_1b_1 < w] \max_p E1[|y_1 - w| \leq w|x_1]1[|y_2 - p| \leq w|x_2]$$

$\frac{1}{T}\sum_t 1[x_{1t}b_1 < w]$ converge almost surely and uniformly in b_1 to $E1[x_1b_1 < w]$ by the same argument of half-spaces, uniform bound and boundedness of the covering number in theorem 3. Consider the sub-expression,

$$\max_p E1[|y_1 - w| \leq w|x_1]1[|y_2 - p| \leq w|x_2]$$

$\frac{1}{N}\sum_{\substack{x_{1k}=x_1\\x_{2k}=x_2}} 1[|y_{1k}-w| \leq w|x_1]1[|y_{2k}-p| \leq w|x_2]$ converges almost surely and uniformly in p to $E1[|y_1-w| \leq w|x_1]1[|y_2-p| \leq w|x_2]$ because they are product of half spaces and uniformly bounded above by 1. As a result, $\max_{p_N} \frac{1}{N}\sum_{\substack{x_{1k}=x_1\\x_{2k}=x_2}} 1[|y_{1k}-w| \leq w|x_1]1[|y_{2k}-p| \leq w|x_2]$ converges almost surely to $\max_p E1[|y_1-w| \leq w|x_1]1[|y_2-p| \leq w|, x_2]$. Furthermore, since the expression does not depend on b_1, it converges almost surely and uniformly in b_1.

Since both terms converge almost surely and uniformly in b_1, and are uniformly bounded by 1, its product

$$\frac{1}{T}\sum_t 1[x_{1t}\beta_1 < w] \max_p (1/N_t) \sum_{\substack{x_{1k}=x_{1t},\\x_{2k}=x_{2t}}} 1[|y_{1k}-w| \leq w|x_{1t}]1[|y_{2k}-p| \leq w|x_{2t}]$$

converges almost surely and uniformly in b_1 to

$$E1[x_1 b_1 < w] \max_p E1[|y_1-w| \leq w|x_1, x_2]1[|y_2-p| \leq w|x_1, x_2]$$

by the continuous mapping theorem. Adding two terms together, $Q_T(b_1, b_2)$ converges a.s. and uniformly in (b_1, b_2) to $Q(b_1, b_2)$.

Step 2. Continuity of $Q(b_1, b_2)$. $Q(b_1, b_2)$ can be rewritten as

$$
\begin{aligned}
Q(b_1, b_2)) = \int\int \{&1[x_1 b_1 \geq w][F_{\boldsymbol{y}|\boldsymbol{x}}(x_1 b_1 + w, x_2 b_2 + w) - F_{\boldsymbol{y}|\boldsymbol{x}}(x_1 b_1 + w, x_2 b_2 - w) \\
&- F_{\boldsymbol{y}|\boldsymbol{x}}(x_1 b_1 - w, x_2 b_2 + w) + F_{\boldsymbol{y}|\boldsymbol{x}}(x_1 b_1 - w, x_2 b_2 - w)] \\
&+ 1[x_1 b_1 < w]\{\max_p F_{\boldsymbol{y}|\boldsymbol{x}}(2w, p + w) - F_{\boldsymbol{y}|\boldsymbol{x}}(2w, p - w) \\
&- F_{\boldsymbol{y}|\boldsymbol{x}}(0, p + w) + F_{\boldsymbol{y}|\boldsymbol{x}}(0, p - w)\} dG_{x_1, x_2} \\
= \iint_{x_1 b_1 \geq w} &\{F_{\boldsymbol{y}|\boldsymbol{x}}(x_1 b_1 + w, x_2 b_2 + w) - F_{\boldsymbol{y}|\boldsymbol{x}}(x_1 b_1 + w, x_2 b_2 - w) \\
&- F_{\boldsymbol{y}|\boldsymbol{x}}(x_1 b_1 - w, x_2 b_2 + w) + F_{\boldsymbol{y}|\boldsymbol{x}}(x_1 b_1 - w, x_2 b_2 - w)\} dG_{x_1, x_2} \\
&+ \iint_{x_1 b_1 < w} \{\max_p F_{\boldsymbol{y}|\boldsymbol{x}}(2w, p + w) - F_{\boldsymbol{y}|\boldsymbol{x}}(2w, p - w) \\
&- F_{\boldsymbol{y}|\boldsymbol{x}}(0, p + w) + F_{\boldsymbol{y}|\boldsymbol{x}}(0, p - w)\} dG_{x_1, x_2}
\end{aligned}
$$

The integrand under the first double integral is a bounded and continuous function of b_1 and b_2 since $F_{\boldsymbol{y}|\boldsymbol{x}}$ is bounded and continuous in (b_1, b_2), and $(x_1 b_1 + w)$, $(x_1 b_1 - w)$, $(x_2 b_2 + w)$ and $(x_2 b_2 - w)$ are all continuous functions of b_1 and b_2. Since the region of integration $\{x_1 b_1 \geq w\}$ is a half space and is also a continuous function of b_1, the first double integral is a continuous function of b_1 and b_2 by the bounded convergence theorem.

The integrand under the second double integral is a bounded function and does not depend on b_1 and b_2, and so is a continuous function of (b_1, b_2) trivially. Since the region of integration $\{x_1 b_1 < w\}$ is a half space and a continuous function of b_1, the second double integral is a continuous function of b_1 and b_2 by the bounded convergence theorem.

Being the sum of the two integrals above, $Q(b_1, b_2)$ is continuous in (b_1, b_2).

Step 3. Maximum of $Q(b_1, b_2)$. Due to the assumption that the parameter space B (Assumption 6) is compact and the continuity of $Q(b_1, b_2)$, $Q(b_1, b_2)$ attain its maximum on B. Hence there exists at least one (b_1, b_2) which maximizes $Q(b)$.

By the three steps above and Theorem 2, b_{1T} and b_{2T} are strongly consistent for β_1 and

β_2 for case 2 under the model's assumptions.

□

3.5 Asymptotic Distribution and Choice of bandwidth w

Kim and Pollard (1987) studied 'nonconventional' estimation methods including high-dimensional mode estimators with indicator functions, Rousseeuw's Least Median of Square, and the maximum score estimation of Manski (1975, 1985). Their major finding is that these estimators are $N^{\frac{1}{3}}$-consistent and their population objective functions can be represented by a Brownian motion process indexed by z plus a deterministic term quadratic in z. Although there exists no closed form solution for the bivariate mode estimators proposed in this chapter, Kim and Pollard give reasons to believe that the finite-sample estimators b_{1T} and b_{2T} converges at the rate of $N^{-\frac{1}{3}}$ to β_1 and β_2 for cases 1 and 2.

The choice of w is related to the bias and efficiency trade off of the bivariate mode estimators. Since there is no closed form solution for the estimators, it is not possible to derive an optimal w. However, one can always proceed by trial and error. If the score (the number of data covered by the modal square; that is, the number of data with $1[|y_1 - x_1\beta_1| \leq w].1[|y_2 - x_2\beta_2| \leq w] = 1$) is too small, there will be many estimates with the same score. If w is too large, many estimates can easily score the maximum $= T$. In both cases, the estimations results wont be good. Hence, the sensible bounds can be found in practice by trying many w's on the dataset.

The bivariate mode estimators proposed in this chapter could be generalized further by substituting other kernel weight functions $K(.)$ instead of the current square kernel indicator function. These kernel generalizations can potentially improve the computation in practice as well as theoretical convergence rates. These generalizations are out of scope for the current chapter but is a worthwhile future research endeavor.

References

[1] Amemiya, T., 1985. *"Advanced Econometrics."* Harvard University Press, Cambridge.

[2] Chen, S., 1997. *"Semiparametric estimation of the Type-3 Tobit model."* J. Econometrics 80, 1-34.

[3] Chernoff, H., 1964. *"Estimation of the mode"* Institute of Statistical Mathematics Annals 16. 31-41.

[4] Heckman, J. (1979). *"Sample Selection Bias as a Specification Error."* Econometrica, 47(1), 153-161. doi:10.2307/1912352

[5] Honore, B.E., Kyriazidou, E., Udry, C., 1997. *"Estimation of Type 3 Tobit models using symmetric trimming and pairwise comparisons."* J. Econometrics 76, 107-128.

[6] Lee, Myoung-jae, 1989. *"Mode regression"* Journal of Econometrics, Elsevier, vol. 42(3), pages 337-349, November.

[7] Lee, Myoung-jae, 1993. *"Quadratic mode regression."* Journal of Econometrics, Elsevier, vol. 57(1-3), pages 1-19.

[8] Lee, L.-F., 1994. *"Semiparametric two-stage estimation of sample selection models subject to Tobit-type selection rules."* J. Econometrics 61, 305-344.

[9] Mulligan, C., Rubinstein, Y. (2008). *"Selection, Investment, and Women's Relative Wages over Time."* The Quarterly Journal of Economics, 123(3), 1061-1110.

[10] Pollard. D., 1984. *"Convergence of stochastic processes"* Springer-Verlag, New York, NY.

[11] Christopher Udry, 1994. *"Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria"* The Review of Economic Studies, Volume 61, Issue 3, July 1994, Pages 495-526

[12] Zhou, Xianbo, Pan, Zhewen, 2015. *"Two-step semiparametric estimation of the Type-3 Tobit model "* Statistics & Probability Letters, Elsevier, vol. 105(C), pages 96-105.