

The MIT Undergraduate Journal of Economics

Volume XXIII

2023-2024

The Celtic Tiger & The Resource Rich: How Corporate Tax Policy and Oil Wealth Defined the Economic Growth of Ireland and Egypt
Diana Degnan

Right-to-Work Laws and Innovative Output
Paul Irvine

From Guac to Glock: Exploring the Consequences of Avocado Production on Organized Crime in Mexico
Brian Liu

Assessing the Impact of Police Officer Line-of-Duty Deaths and Force Size on Civilian Arrests
Srinidhi Narayanan

Dropping the SAT Subject Tests
Quentin Smith

An Empirical Analysis of the 3-Points-for-a-Win Rule in the English Premier League
Jeremy Smithline

**The MIT Undergraduate Journal of
Economics Volume XXIII**

2023-2024

Mailing Address:

The MIT Undergraduate Journal of Economics
Massachusetts Institute of Technology, Building E52-301
Cambridge, MA 02139

Foreword

“Money... must always be scarce with those who have neither wherewithal to buy it, nor credit to borrow it.”

- *Adam Smith*

As MIT undergraduate economics students progress through their coursework, they are continuously introduced to new economic topics, constantly learning the ideas and models of established economists, and relentlessly being challenged to think differently about the observable phenomena around them. It is this enthusiasm for learning that led undergraduates at MIT to proceed in their own research—to experience the excitement of asking a question and striving to answer it. We hope that this year’s papers highlight the vigor with which our undergraduate students pursue economic research and the rigor with which they present their ideas.

The publication of this Journal is made possible by the support of many people. We especially thank Professor David Autor for selecting the articles for this year’s publication.

These relevant student papers demonstrate the enduring importance of rigorous economic research in the days ahead.

The MIT Undergraduate Journal of Economics Volume XXIII

2023-2024

Contents

The Celtic Tiger & The Resource Rich: How Corporate Tax Policy and Oil Wealth Defined the Economic Growth of Ireland and Egypt

Diana Degnan

Right-to-Work Laws and Innovative Output

Paul Irvine

From Guac to Glock: Exploring the Consequences of Avocado Production on Organized Crime in Mexico

Brian Liu

Assessing the Impact of Police Officer Line-of-Duty Deaths and Force Size on Civilian Arrests

Srinidhi Narayanan

Dropping the SAT Subject Tests

Quentin Smith

An Empirical Analysis of the 3-Points-for-a-Win Rule in the English Premier League

Jeremy Smithline

The Celtic Tiger & The Resource Rich

*How Corporate Tax Policy and Oil Wealth
Defined the Economic Growth of Ireland and
Egypt*

14.05 Intermediate Macroeconomics

Diana Degnan

1. Introduction

On paper, Ireland and Egypt look like vastly different countries. Ireland is a member of the European Union with a parliamentary democracy and one of the most highly developed and fastest growing economies in all of Western Europe. Egypt, on the other hand, is a democratic republic in North Africa with a developing economy that derives most of its revenue from agriculture and oil production. However, despite their social, economic, and political differences, Egypt and Ireland have a shared colonial history. Both Ireland and Egypt were colonies of the British Empire for a large part of their history; Ireland gained independence from Britain in late 1921, and Egypt declared independence in early 1922 (Beatty, 2018). Since declaring independence, both Ireland and Egypt have faced internal political and economic turmoil. Northern Ireland experienced The Troubles, a period of sectarian violence between Protestant unionists and Roman Catholic nationalists, from the late 1960s to 1998, and Egypt experienced a bloody coup that led to President Gamal Abdel Nasser's regime from 1952 to 1970 (Wallenfeldt, 2013; John, 2013). How did Egypt and Ireland diverge from their shared colonial histories to the two very different structures and levels of economic development they have today?

In this paper, I argue that Egypt and Ireland's different economic and political structures led the two countries along vastly different growth paths. While capital formation explains the overwhelming majority of Egypt's GDP growth since the 1960s, TFP growth is the primary factor explaining Ireland's overall growth. For both countries, globalization and increasing contact with foreign nations – through foreign direct investment for Ireland and trade for Egypt – were the underlying fundamental causes of each country's growth. Ireland's corporate tax structure, as well as its open and privatized economy, incentivized increased FDI inflows, which eventually led to GDP growth through TFP. In Egypt, on the other hand, its natural resource wealth and the remnants of its centrally-controlled economy resulted in a volatile period of high capital accumulation and GDP growth when oil prices were high, followed by economic stagnation.

1. Growth Accounting:

To conduct the growth accounting analysis, I used the following function and equation:

$$\text{Aggregate Production Function: } Y(t) = K(t)^\alpha (A(t) * H(t) * L(t))^{1-\alpha}$$

The equation above represents the Cobb-Douglas production function with labor-augmenting technological progress and human capital. This equation applies very generally, describing the output of an economy both at and away from the steady-state and regardless of whether labor grows or consumers save at a constant rate.

$$\text{Growth Accounting Equation: } \frac{Y_{t+1} - Y_t}{Y_t} = \alpha \frac{K_{t+1} - K_t}{K_t} + (1 - \alpha) \frac{L_{t+1} - L_t}{L_t} + (1 - \alpha) \frac{H_{t+1} - H_t}{H_t} + x(t)$$

The above growth accounting equation represents the discrete-time approximation of the above Cobb-Douglas production function.

As my datasource, I used annual data for Egypt and Ireland over the time frame of 1960-2019, sourced from the Penn World Table (PWT) online database. I used real GDP at constant national prices for Y, population for L, capital stock at constant national prices for K, and the index of human capital per person for H. I chose to use population for L rather than employment because employment data was spotty for Egypt across certain time periods. Making the assumption that the data is in perfect competition, I used the equation 1-labor share to compute α for each year, where labor share was the PWT variable for the share of labor compensation in GDP at current national prices. I then used the average of all α_t s across the entire time period 1960-2019 as α for the growth accounting procedure. Year-over-year growth was calculated for Y, K, L, and H using the following general growth formula: $\frac{X_{t+1} - X_t}{X_t}$. To calculate each factor's contribution to overall GDP growth, capital stock growth was multiplied by the average α , and L and H were multiplied by a factor of $(1-\alpha)$, as is consistent with the Growth Accounting Equation detailed above. To calculate the contribution of the residual, or Total Factor Productivity (TFP), I subtracted capital's contribution, population's contribution, and human capital's contribution from GDP growth. The remaining value is TFP's contribution to GDP growth.

As a baseline point of comparison, I plotted Ireland’s and Egypt’s log GDP per capita over the entire time period, 1960-2019. As can be seen in Figure 1 below, Ireland’s log GDP per capita (green) is much higher than that of Egypt (red) and grew at a significantly higher rate between the mid 1990s and early 2000s. These high-level trends in log GDP per capita between Ireland and Egypt help contextualize the results of the growth accounting analysis, which are displayed in the figures below for each country. Figure 2 and Figure 3 visualize each factor’s average contribution to GDP growth across each decade.

Figure 1: Ireland and Egypt’s Log GDP Per Capita, 1961-2019

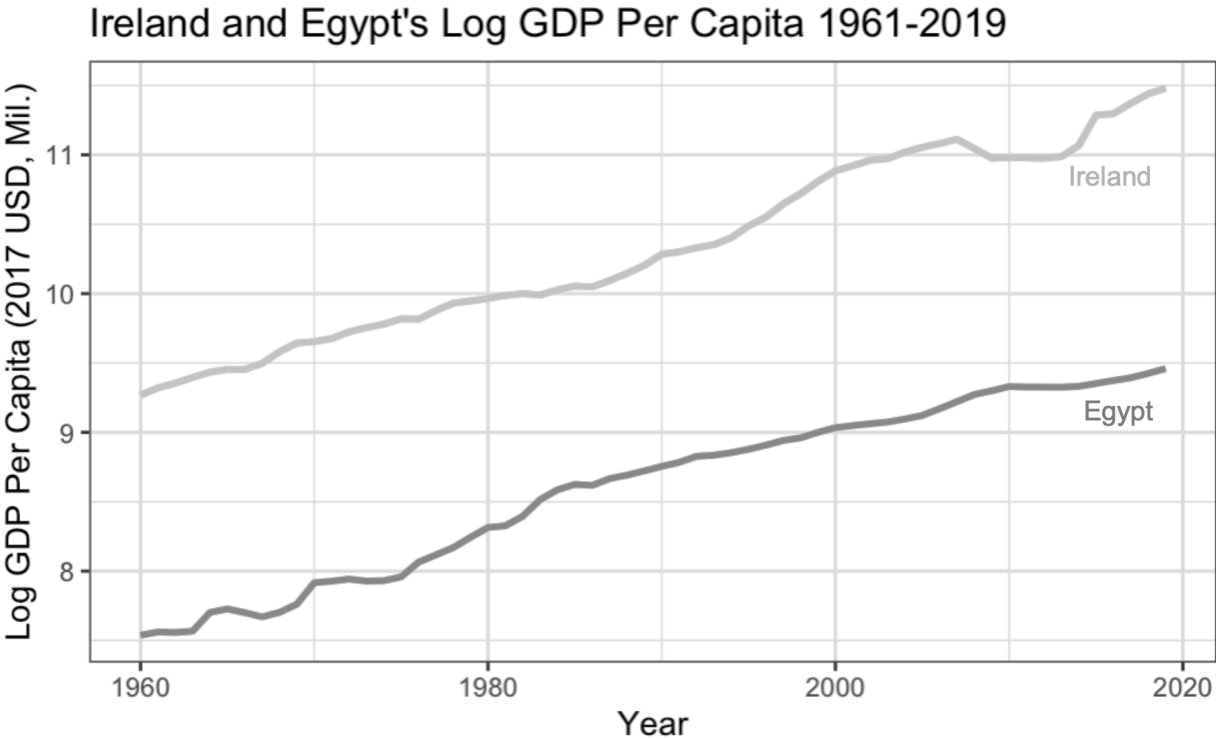


Figure 2: Growth Accounting Results, Ireland

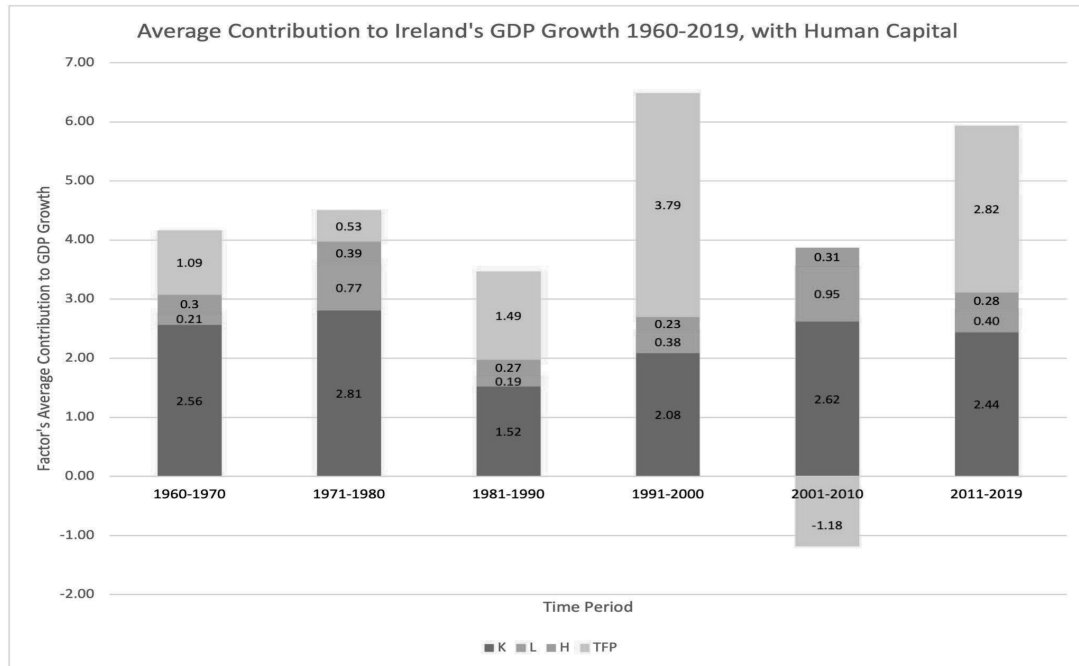


Figure 3: Growth Accounting Results, Egypt

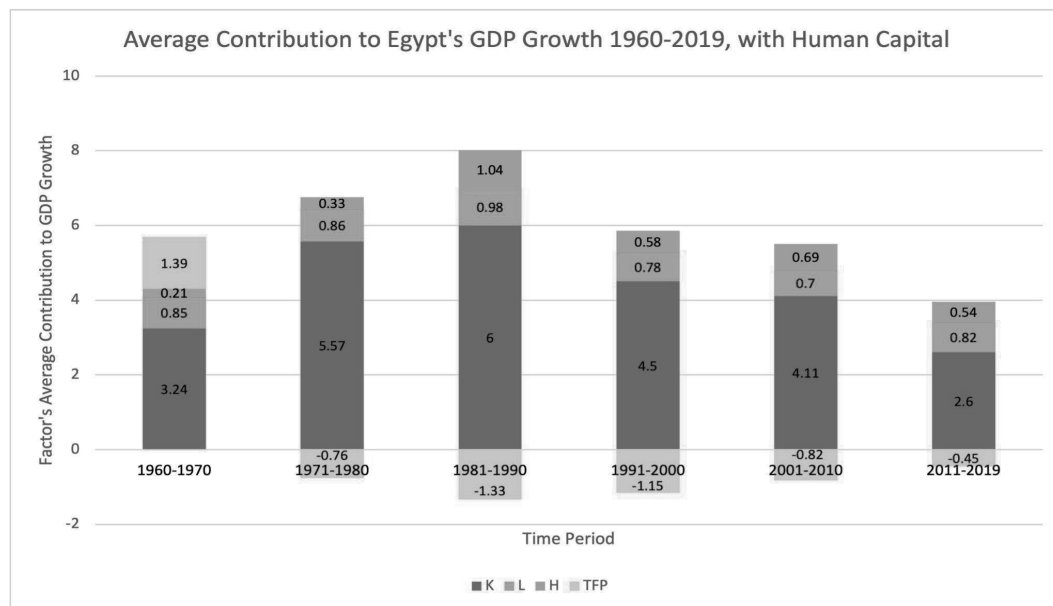


Figure 1 and Figure 2 reveal the results of the growth accounting analysis for Ireland. Figure 1 shows that Ireland has sustained a relatively high level of GDP growth since 1960, with a dip from 2001-2010 likely due to the Irish Economic Recession of 2008-2009, during which a real estate bubble prompted a banking crisis that resulted in a large economic downturn (IMF, 2018). From Figure 2, it is

clear that while Ireland's capital stock growth (blue) has played a consistent and significant role in its GDP growth, TFP (yellow) is often the factor most responsible for the country's growth. Population (orange) and human capital growth (gray) both contribute relatively little to Ireland's overall GDP growth.

Figure 1 and Figure 3 reveal the results of the growth accounting analysis for Egypt. Figure 1 shows that Egypt has experienced a slowly increasing level of GDP per capita growth since the 1960s. From Figure 3, it is clear that capital stock growth (blue) explains nearly all of Egypt's GDP growth, especially from 1971-1990. Capital accumulation's contribution to Egypt's GDP growth experienced a large jump up from the 1960s to the 1970s, and again from the 1970s to 1980s. However, since then, capital accumulation's contribution to GDP growth has been decreasing. Although its contribution is small, population growth (orange) contributes a steady amount to GDP growth, with population growth's average contribution increasing this past decade. Human capital growth's contribution (gray) to Egypt's GDP growth is also small and exhibits substantial variation. The TFP residual (yellow) is consistently negative or near zero across the entire time frame, thus, TFP explains a much smaller portion of GDP growth in Egypt than it does in Ireland.

Thus, the most noticeable difference in the growth accounting results of the two countries is that while capital stock growth is by far the factor most responsible for GDP growth in Egypt, TFP growth is primarily responsible for Ireland's GDP growth. In the remainder of the paper, I explore the underlying causes of the capital growth in Egypt and the TFP growth in Ireland that largely determine each country's GDP.

2. Ireland's TFP Growth - FDI Inflows

In this section, I identify and explore the underlying fundamental causes of Ireland's TFP growth, particularly the role of foreign direct investment, intellectual property transfers, and Ireland's corporate tax structure.

3.1 Foreign Direct Investment

Foreign direct investment (FDI) is likely a significant underlying cause of Ireland's TFP growth. As established by the growth accounting results, TFP growth is the primary factor explaining Ireland's overall GDP growth. With the exception of 2001-2010, during which the Irish Economic Recession resulted in a sharp decrease in Ireland's per capita GDP, TFP growth has explained an exceptionally high portion of Ireland's GDP growth since the 1980s. Ireland's TFP growth, and thus much of its overall GDP growth, is likely driven by increases in Ireland's foreign direct investment inflows (Cassidy, 2004; IMF, 2016). Ireland's average FDI inflows (Figure 4) and outflows (Figure 5) as a percentage of GDP in each decade are shown below, using data from the World Bank's World Development Indicators database. Due to the unavailability of FDI outflow data prior to 1987, Ireland's average FDI outflows are shown only from 1987-2019. From these two graphs, we see that both FDI inflows and outflows have increased greatly since the turn of the 21st century. Ireland's FDI inflows have risen particularly dramatically, especially in the past decade. Figure 4 shows that FDI inflows between 2011 and 2019 averaged over one-quarter of Ireland's GDP. Thus, it is likely that a substantial portion of Ireland's TFP growth, and resulting GDP growth, is due to the country's dramatic increase in FDI inflows (OECD, 1993).

Figure 4: Ireland's Average FDI Inflows, Percent of GDP

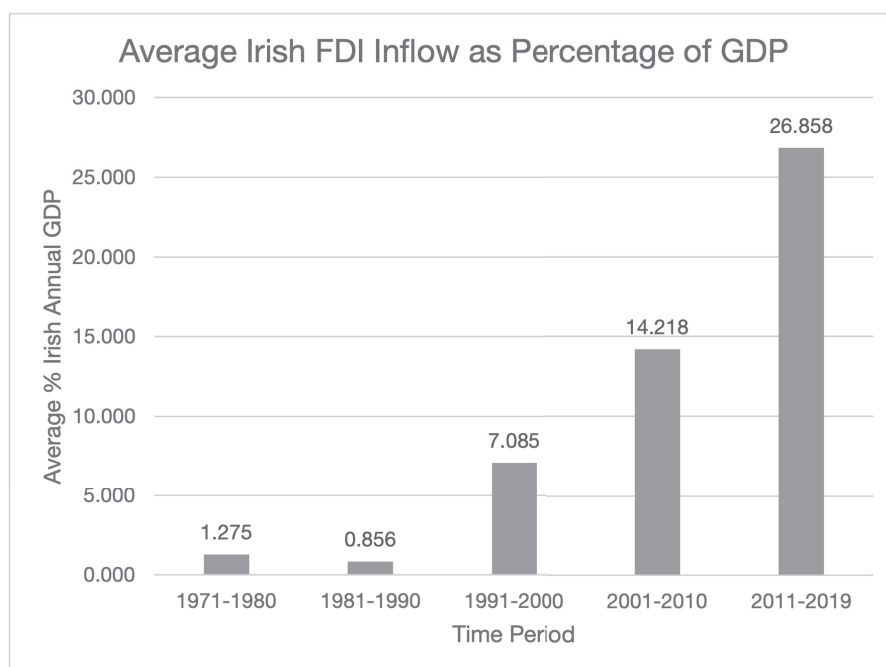
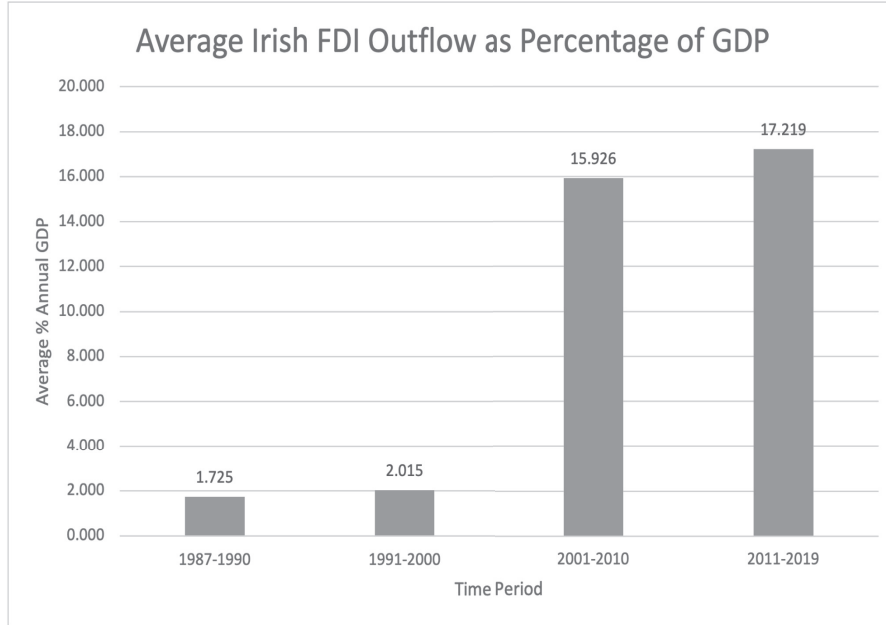


Figure 5: Ireland's Average FDI Outflows, Percent of GDP



3.2 Corporate Tax Structure & Intellectual Property

The driver of Ireland's massive increase in FDI inflows is the country's low corporate tax rate. In the mid 1990s, Ireland experienced a period of rapid real economic growth fuelled by FDI, earning the country the nickname "The Celtic Tiger" after similar periods of rapid economic expansion in East Asia. During this period, Ireland had an extremely low corporate tax rate of 10% (Bourke, 2004). Since 2003, Ireland has maintained a slightly higher corporate tax rate of 12.5%, still the second lowest corporate tax rate in the European Union (Taqui, 2023). Ireland's low corporate tax rate has encouraged foreign companies, especially American tech companies such as Intel and Microsoft, to invest in Irish markets (OECD, 1993; Taqui, 2023). The positive effects of FDI on Ireland's economy were seen in 2015 when Ireland's GDP increased 20% from 2014, an extraordinary growth rate, especially for a developed country in Western Europe. Much of this GDP growth was attributed to Ireland's low corporate tax rate, which prompted several large multinational corporations to relocate their economic activities, and their underlying intellectual property, to Ireland. With this shift, production from the use of intellectual

property now contributed to Ireland's GDP, causing a large boost in Ireland's overall GDP growth rate (OECD, 2016).

However, there is some degree of skepticism as to whether shifts in intellectual property should really be reflected in Ireland's GDP growth. Some have argued that Ireland's GDP no longer accurately reflects the country's economic activity. The United National System of National Accounts (SNA) defines production as "an activity, carried out under the responsibility, control, and management of an institutional unit that uses inputs of labor, capital, and goods and services to produce outputs of goods and services" (OECD, 2016). With globalization, the role of intangible assets in GDP has been increasingly brought into question, especially with multinational enterprises (MNEs) seeking to minimize their global tax burden by re-allocating their economic activities to low-corporate tax countries, such as Ireland. In the 2010s, when many MNEs relocated to Ireland, economic ownership of their Intellectual Property Products (IPPs) transferred to Ireland as well (OECD, 2016). Large MNEs shifted their assets to Ireland by having their Irish subsidiaries buy the intellectual property of their other offshore subsidiaries in higher tax countries (Setser, 2019). This was especially prevalent during the 2010s, when Ireland saw large-scale onshoring of IPPs by foreign-owned MNEs, sometimes coupled with the relocation of group headquarters, known as redomiciliation, to Ireland as well (Andersson et al., 2023). Often, the intellectual property is used in contract manufacturing arrangements where Irish enterprises, including Irish affiliates of foreign MNEs, contract with manufacturers, some of which are domiciled outside of Ireland, to produce final products using blueprints from the IPPs. The distribution and sale of the products created from the IPPs are then organized by Irish enterprises, resulting in the value from the products produced with the IPPs being counted towards the Irish economy and thus Irish GDP, even though much or all of the physical production occurred outside Ireland's borders (OECD, 2016). From a macro-level, this transfer of IPPs creates large year-over-year increases in Irish GDP and volatile non-construction "investment" (Andersson et al., 2023). As a result, Ireland has experienced periods of extremely high GDP growth far outside the norm of other Western European economies.

Ireland's tax structure in fact encourages this process. Ireland's Capital Allowance for Intangible Assets (CAIA) tax structure was designed to encourage firms with high amounts of intangible assets to settle in Ireland. By doing so, MNEs can effectively dodge a large portion of their taxes. MNEs can deduct the cost of a large investment in intangibles (for example, the purchase of a subsidiary's IP rights) from its Irish tax bill, and it can also deduct the funds it borrows (typically from a subsidiary) to pay for the purchase. In fact, Apple did just this in 2013. Apple Ireland borrowed funds from Apple Jersey to buy the IP rights that had been assigned to Apple Jersey. The new Apple Ireland was a tax resident of Ireland, and thus used Ireland's very low corporate tax rate. These Irish tax residents are now factored into Ireland's GDP. As a result, some have argued that Ireland's GDP now tells us less about the economic growth of Ireland and more about the growth and performance of large, typically American, MNEs and corporations (Andersson et al., 2023).

3.3 GNI and Modified GNI - A Better Measure of Irish Economic Growth and Welfare?

Considering that much of Ireland's TFP growth is from FDI and the economic activity of foreign-owned MNEs, does Ireland's high observed TFP actually reflect real domestic technological progress? In light of the ambiguity around GDP's ability to measure the true welfare and growth of the Irish economy, many organizations and researchers now rely on Gross National Income (GNI) and modified GNI as a more accurate measure of Ireland's economic development since the 1990s. GNI differs from GDP by the net amount of incomes sent to or received from abroad. In other words, GNI excludes the net profits of MNEs that may be operating in Ireland, but are not owned by Irish citizens or corporations. Thus GNI, rather than GDP, provides a better picture of Ireland's domestic economy. However, the Central Statistics Office of Ireland's Department of Finance has developed an even more accurate metric for understanding the domestic growth of the Irish economy - modified GNI. Modified GNI is GNI minus depreciation on intellectual property, depreciation on leased aircraft, and the net factor income of redomiciled public limited companies (PLCs) (Central Statistics Office, 2023). Modified GNI excludes depreciation on intellectual property to account for the fact that most IPPs belong to

foreign-owned MNEs and thus are not reflective of domestic production or development. Depreciation on leased aircraft is excluded because, due to Ireland's low corporate tax rate, most of the largest international aircraft lessors are based in Ireland. However, these aircraft lessors are foreign-owned. Even though the aircraft themselves may be owned in Ireland, their operations, profits, and manufacture occurs outside of Ireland's borders. Therefore, Ireland's Department of Finance chooses to exclude depreciation on leased aircraft from modified GNI. Finally, modified GNI is absent the net factor income of redomiciled PLCs to exclude the profits redomiciled companies receive from their subsidiaries abroad (Central Statistics Office, 2023).

Ireland's GNI and modified GNI suggest that Ireland's exceptional TFP and GDP growth may overstate the technological progress of the Irish economy. Figure 6 below plots Irish GNI (blue) and modified GNI (orange) from 1995-2019 in constant market prices, using data from the Irish Department of Finance's Central Statistics Office (Central Statistics Office, 2023). Figure 7 shows Irish GNI and modified GNI growth by plotting log GNI (blue) and log modified GNI (orange) from 1995 to 2019. Beginning in 2014, there is a clear deviation between Irish GNI and Irish modified GNI in both level and growth. Modified GNI, which is expected to better represent the domestic growth of the Irish economy, did indeed grow in the 2010s. However, in both level and growth, modified GNI has exhibited much smaller increases than Irish GNI or GDP. Figure 7 shows that since 1995, both Irish GNI growth and modified GNI growth have been fairly low and stable. Figure 1's notable increase in Irish GDP per capita growth since the mid 1990s is largely absent in Figure 7, where modified GNI growth has been relatively flat. Thus, when analyzing GNI and modified GNI, Ireland's high observed TFP growth appears to have a much more muted effect on the Irish economy and the welfare of the Irish people than the growth accounting results and analysis of Irish GDP would suggest. Figure 6 and Figure 7 provide evidence that Ireland's TFP growth may not truly reflect real domestic technological advances or sizable improvements in welfare.

Figure 6: Irish GNI and Modified GNI, 1995-2019

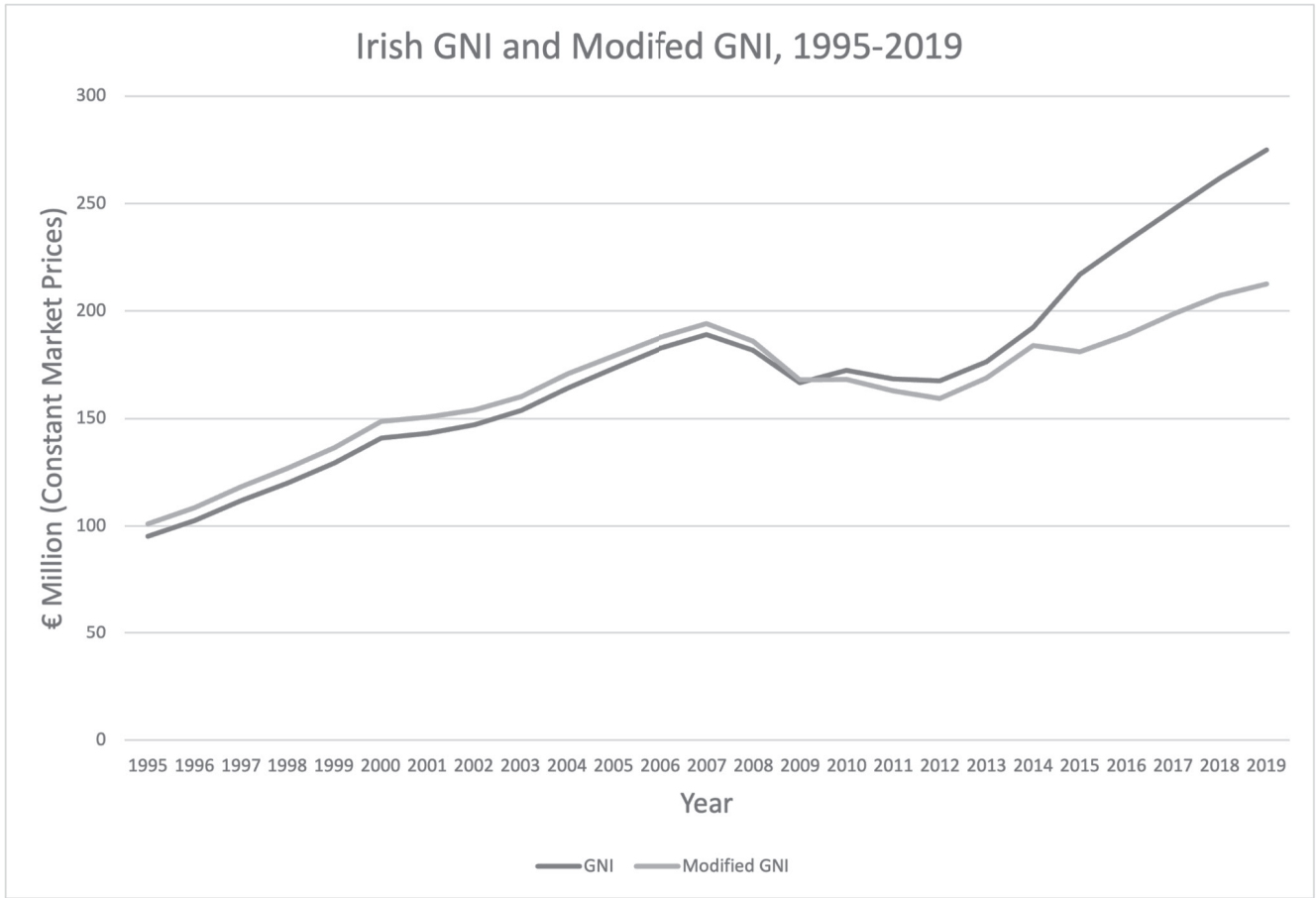
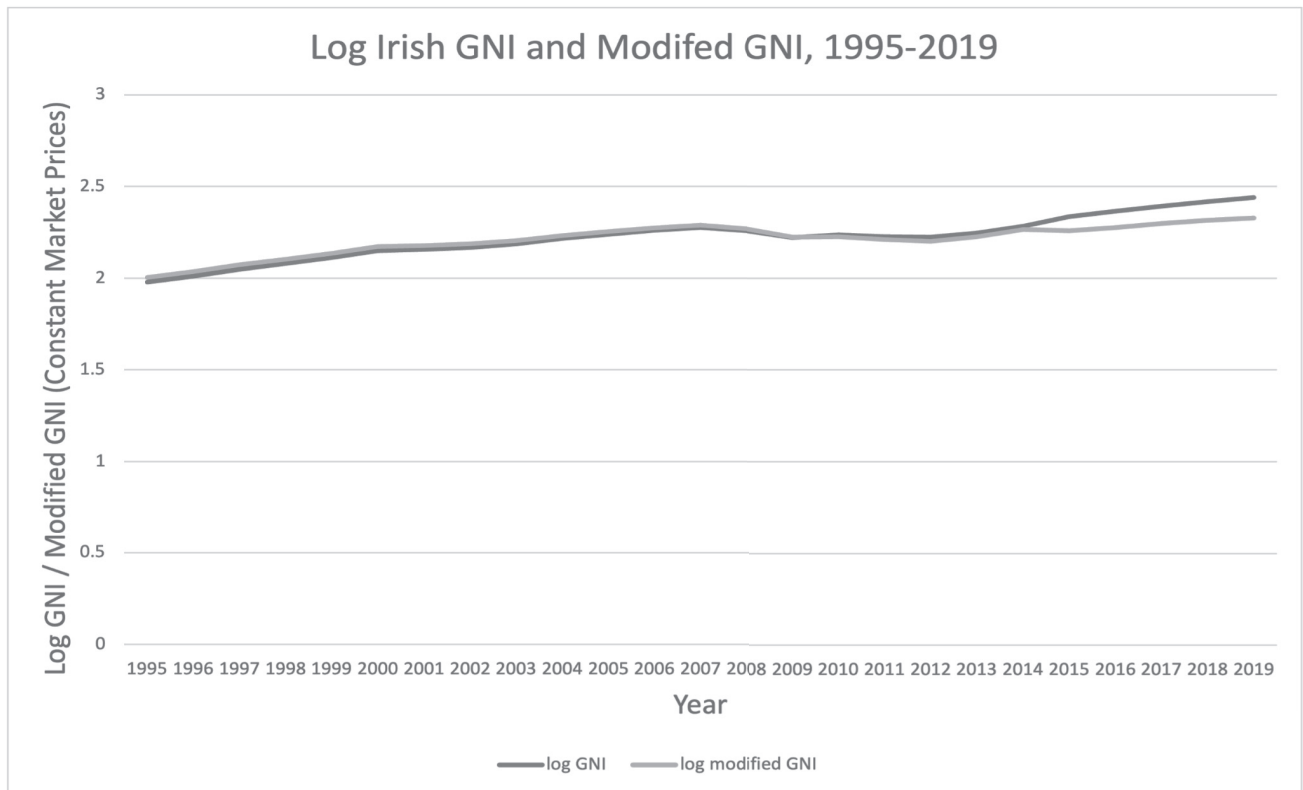


Figure 7: Irish Log GNI and Log Modified GNI, 1995-2019



Yet, despite ambiguity around the true scope of Ireland's technological progress, TFP plays a dominant role in Ireland's overall GDP growth. TFP growth in Egypt, on the other hand, has been consistently negative or near zero since the 1970s. Instead, an overwhelming majority of Egypt's GDP growth can be explained by capital accumulation.

3. Egypt's Capital Growth - Crude Oil Exports

In this section, I identify and explore the underlying fundamental causes of Egypt's capital growth, particularly the role of crude and petroleum exports, oil price volatility, public ownership of the petroleum industry, and the 1973 Open Door Policy - a transformative economic initiative that opened and privatized much of the Egyptian economy for the first time.

4.1 The Open Door Policy

The 1973 Open Door Policy is likely a significant underlying cause of Egypt's capital growth. As shown in the growth accounting results, Egypt's GDP is explained most strongly by capital accumulation. From the 1960s through the 1980s, capital accumulation's contribution to Egypt's overall GDP growth increased significantly. Changes in political leadership and economic policies largely explain Egypt's capital accumulation and growth over this time period, especially Egyptian President Anwar al-Sadat's Open Door Policy, which decentralized the Egyptian economy and encouraged investment following the Yom Kippur War (Alissa, 2007; Goldberg & Beinin, 1982; Handoussa, 2010; Kamaly, 2006).

In the 1960s, the Egyptian economy was government-controlled and highly-centralized with a state-led industrialization model under the rule of President Gamal Abdel Nasser (Goldberg & Beinin, 1982). The public sector was the main engine of growth and was responsible for most investment, capital accumulation, and employment (Kamaly, 2006). Large-scale nationalization was the norm and private-sector activity was severely restricted in most sectors, including agriculture, real estate, and finance. State-owned enterprises (SOEs) monopolized the banking, manufacturing, and transportation sectors, as well as foreign trade, which was limited to primarily the surrounding Arab nations. Thus, in the

1960s, Egypt's economic and political structure drastically limited private-sector capital accumulation and growth (Alissa, 2007).

However, in 1973 the Egyptian economy underwent a massive transformation that paved the way for greater capital accumulation and growth when Egyptian President Anwar al-Sadat launched *Infitah*, or the "Open Door Policy" (Alissa, 2007). The Open Door Policy opened the Egyptian economy to foreign investment as well as inter-Arab joint investment projects, both of which encouraged capital accumulation. Further, the policy promoted the role of the private sector, essentially privatizing many industries and transitioning away from SOEs for the first time. Notably, the Open Door Policy abolished government monopolies on finance and foreign trade and allowed private sector activity to enter all fields (except petroleum) without exception. One major economic shift within the Open Door Policy was the passage of Law No. 43 in 1974. Law No. 43 introduced foreign and Arab capital, granting investors access to all fields of economic activities, again with the exception of petroleum (Handoussa, 2010). To implement the Open Door Policy, Sadat created a new organization - the General Authority for Investment and Free Zones - which was tasked with overseeing a new series of regulations and incentives aimed at protecting private property and encouraging foreign trade. These regulations included a new law promising to refrain from re-nationalizing key industries or confiscating invested capital without judicial procedures, as well as incentives such as five-to-eight year exemptions from taxes on profits, deferment of payments on customs duties, and permissions to import without a license (Drummond, 2021).

4.2 New Private Investment

The new laws governing foreign trade and private investment had a direct, positive effect on Egypt's capital accumulation in the 1970s and 1980s, which is evidenced by a clear increase in Egypt's investment rate and the share of private investment. Egypt's Open Door Policy, and particularly Law No. 43, promoted investment. Because the country was in the process of transitioning from a centrally-planned economic system to a more open model, the majority of investment even into the mid-1980s was public. However, the Open Door Policy introduced private investment for the first time in

1973. Private investment began to grow and expand Egypt’s rate of capital accumulation, with private investment eventually surpassing public investment in the late 1980s. Figure 8 below, created from data provided by Egyptian authorities to the IMF, depicts Egypt’s public and private investment rates for fixed capital formation (excluding changes in stocks) as a percentage of GDP from 1969 to 1997. During Egypt’s period of high investment immediately following the Open Door Policy (approximately 1973 to 1982), the total investment rate (black line) more than doubled from just 12% of GDP to slightly under 25%. At its peak in the 1980s, public investment (dashed line) reached 22% of GDP, of which half was accounted for by public enterprise investment. However, even during the high-investment and high-growth period, private investment (purple line) remained low, averaging only 5% of GDP (Handy, 2010).

Figure 8: Egypt’s Investment Rate as Percentage of GDP, 1969-1997

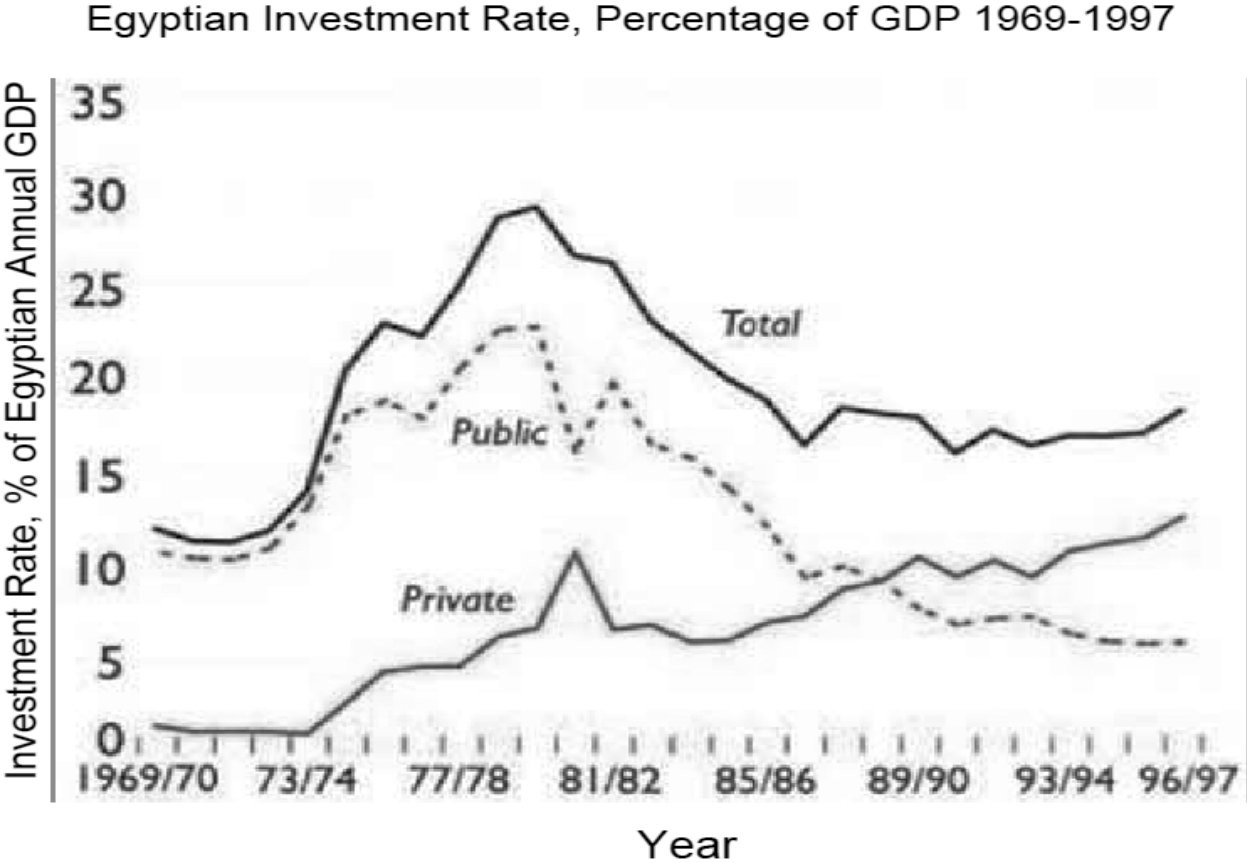


Figure 9: Private Investment in Egypt, Percent of Total Investment

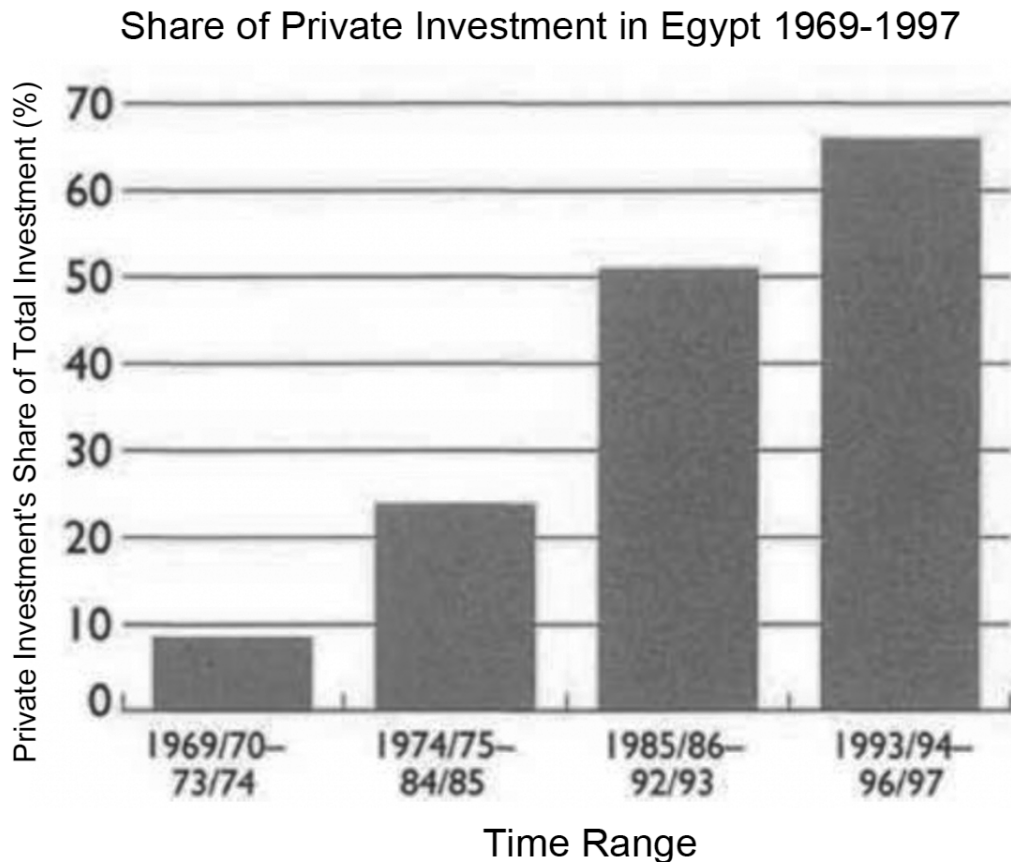


Figure 9 above, also created from data provided by Egyptian authorities to the IMF, sheds further light on the limited scale of private investment. The bar graph shows each time period's average private investment as a percentage of total investment. Although private investment significantly increased after the enactment of the Open Door Policy, the share of private investment still remained quite low even into the mid 1980s. Private investment did not reach half of total investment until the 1990s (Handy, 2010).

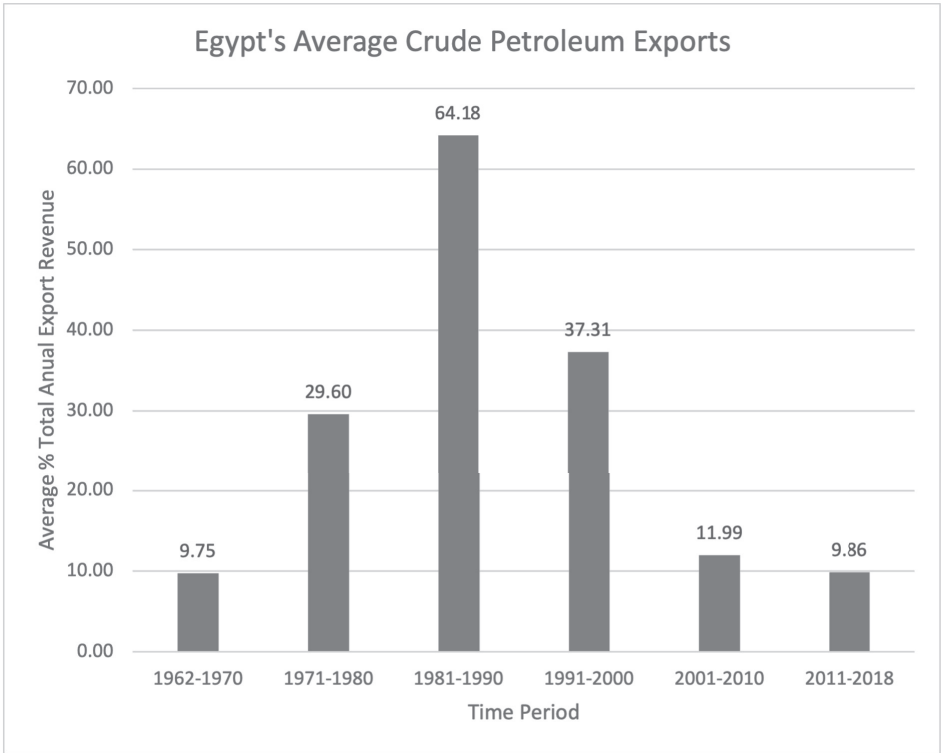
4.3 Crude Oil & Petroleum Exports

Although Egypt's Open Door Policy did indeed introduce private investment for the first time, given the small scale of private investment even into the mid 1980s, it's unlikely that private investment and private capital accumulation explains all of Egypt's capital growth. Figure 8 above makes clear that public investment grew dramatically after 1973, so where was all this public investment coming from?

The answer: crude oil extraction and petroleum production.

The opening of Egypt’s economy, coupled with the discovery of new, large oil and natural gas deposits in the late 1970s, resulted in a huge expansion of Egypt’s crude oil and petroleum exports. Oil production wasn’t a new phenomenon for Egypt. Oil was first found in Egypt almost 100 years earlier and oil operations began for the first time in 1885 (Khaled, 2014). However, the opening of Egypt’s economy to foreign nations and the transition away from import-substitution policies enabled Egypt to greatly expand its global trade not only with other Arab countries, but with Western and Asian nations as well. Now open to foreign trade, Egypt took advantage of the high demand for oil on the global market, expanding their crude oil and petroleum exports immensely. Figure 10 below uses Standard International Trade Classification (SITC) data from the Observatory of Economic Complexity (OEC) to show crude oil and petroleum products as a percentage of Egypt’s total annual export revenue (OEC, 2023). After the Open Door Policy, Egypt’s crude and petroleum exports shot up more than 200% between the 1970s and 1980s. While crude and petroleum averaged around a third of Egypt’s export revenue in the 1970s, by the 1980s, crude’s share had jumped up to nearly 65%. There were several years in the mid-1980s during which crude oil and petroleum made up nearly all of Egypt’s export revenue; crude and petroleum were 72% of Egypt’s total annual export revenue in 1983 and 1984, before peaking at 74.2% in 1985.

Figure 10: Egypt’s Average Petroleum Exports, Percent of Total Export Revenue



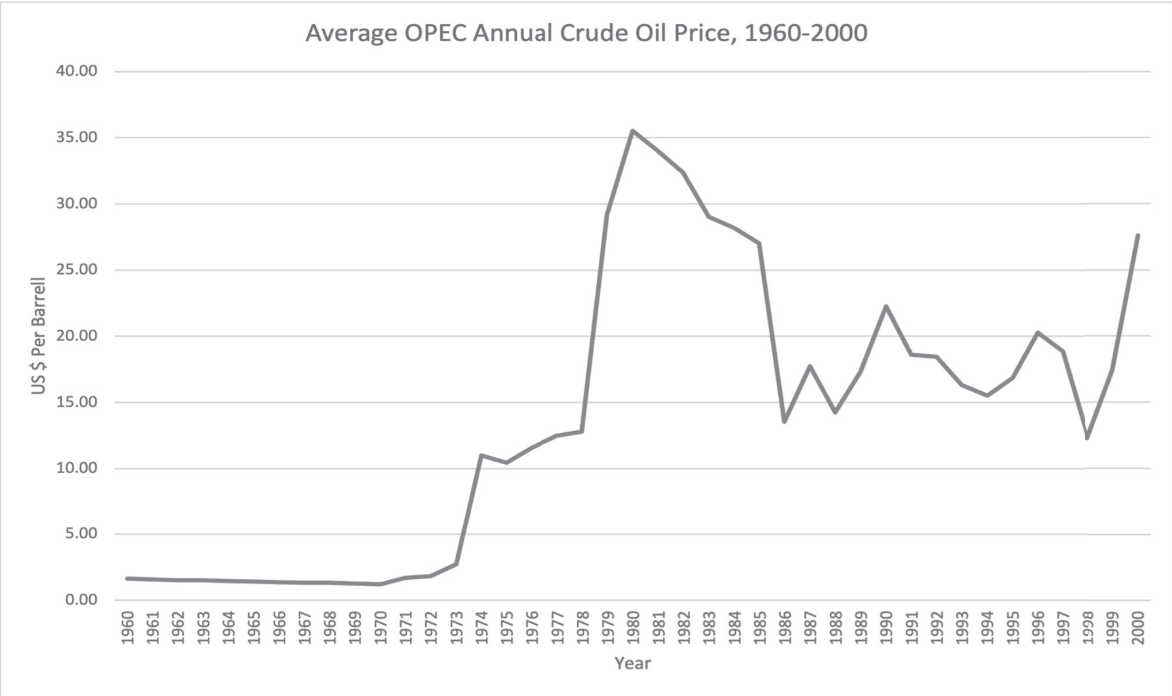
Egypt's increase in crude oil exports directly supported its high public investment, capital accumulation, and GDP growth in the 1970s and mid-1980s. These linkages are appropriate given that Egypt's petroleum industry remained nationally owned and crude oil and petroleum production are highly capital intensive processes. Oil and gas production is widely considered one of the most capital intensive industries. Crude oil and petroleum production requires substantial up-front investment costs, including oil rigs, offshore drilling apparatuses, chemical and heating machinery, and fracking infrastructure (Library of Congress, 2019). To access the high demand for oil on the world market, Egypt had to scale up its production by investing significantly in the expensive infrastructure needed to convert natural oil reserves into marketable crude and petroleum products. Although the Open Door Policy privatized much of the Egyptian economy and eliminated major SOEs from industrial and financial sectors, Egypt's petroleum industry remained under the control of the state-run Ministry of Petroleum and Mineral Resources. Under the Ministry, five state-owned oil companies continued to operate: The Egyptian General Petroleum Corporation (EGPC), The Egyptian Natural Gas Holding Company (EGAS), The Egyptian Petrochemicals Holding Company (ECHEM), The Ganoub El-Wadi Holding Company (GANOPE), and The Egyptian Geological Survey and Mining Authority (EMRA) (International Trade Administration, 2020). Thus, investment and capital accumulation to meet the demand for Egyptian oil on the world market would be considered a public, rather than private, initiative. The trends in Egypt's public investment, crude oil exports, and capital accumulation suggests that opening Egypt's economy to foreign trade facilitated a large expansion of public-sector capital investment aimed at supporting crude oil and petroleum production. This capital accumulation thus contributed to much of Egypt's capital and GDP growth in the 1970s and early 1980s.

4.4 The Natural Resource Curse & Economic Volatility

However, Egypt's oil exports did not sustain investment, capital accumulation, or GDP growth into the 1990s. Oil price volatility and Egypt's influx of resource wealth created macroeconomic imbalances that eventually began to curtail both investment and GDP growth. Egypt's oil exports were

subject to dramatic volatility in market oil prices, as shown in Figure 11 below. Figure 11 depicts the average annual OPEC crude oil price, measured in US dollars per barrel, from 1960 to 2000 (OPEC, 2023). In the 1970s and early 1980s, Egypt’s oil exports benefited from a period of especially high oil prices, including the twin oil shocks in 1973 and 1979 when OPEC doubled the price of oil (Cooper, 2012). However, between 1981 and 1985 oil prices slowly eroded, falling nearly 40%. In 1986, oil prices tanked, falling more than 50% in the first six months of the year (Gately, 1986). As a result, Egypt’s revenues from oil prices plummeted, curtailing both foreign and public investment.

Figure 11: Average Annual OPEC Crude Oil Price, 1960-2000



Because much of Egypt’s petroleum industry remained state-owned, private investment did not compensate for the contraction in public investment in the late 1980s, resulting in an overall decline in capital accumulation. At the same time, Egypt was struggling with the macroeconomic consequences of its resource wealth. Sudden increases in wealth from oil exports had resulted in high inflation, which was in excess of 20% a year in the mid 1980s, and a current account deficit that was more than 10% of GDP (Handy, 2010). The country harbored substantial fiscal deficits, rising external indebtedness, and

mounting public enterprise losses. Figure 12 below depicts Egypt's inflation rate and fiscal deficit (as a percentage of GDP) beginning in 1981, relying on data from the Egyptian Ministry of Finance (Abdu, 2022). Egypt's fiscal deficit increased sharply in the early 1980s, jumping from 15.8% of GDP in 1981 to over 25% of GDP in 1982. This high fiscal deficit was sustained for the first half of the 1980s; from 1981-1985, the country's fiscal deficit averaged over 21% of GDP. Although the fiscal deficit began to decline in the latter half of the decade, the deficit still remained high, resting at 15.1% of GDP in 1990. Egypt's external indebtedness was similarly high. Figure 13 below shows Egypt's external debt stocks as a percentage of Gross National Income (GNI), using data from the World Bank's International Debt Statistics (The World Bank, 2021). From the beginning of the Open Door Policy in 1973 to the late 1980s, Egypt's external debt stocks increased dramatically. External indebtedness rose from 20% of Egypt's GNI in 1973 to a peak of 132.7% of its GNI in 1988. These difficulties were exacerbated by remaining administrative restrictions left over from Egypt's centralized economy just a decade earlier, including administered prices, interest rate ceilings, multiple overvalued exchange rates, administrative allocations of foreign exchange, and surviving restrictions on private and foreign sectors (Handy, 2010). Although privatization was now more common, the private sector remained "crowded out" of many fields, including the petroleum industry, which remained dominated by large-scale public ownership.

Figure 12: Egypt's Fiscal Deficit and Inflation (Percentage of GDP), 1981-2020

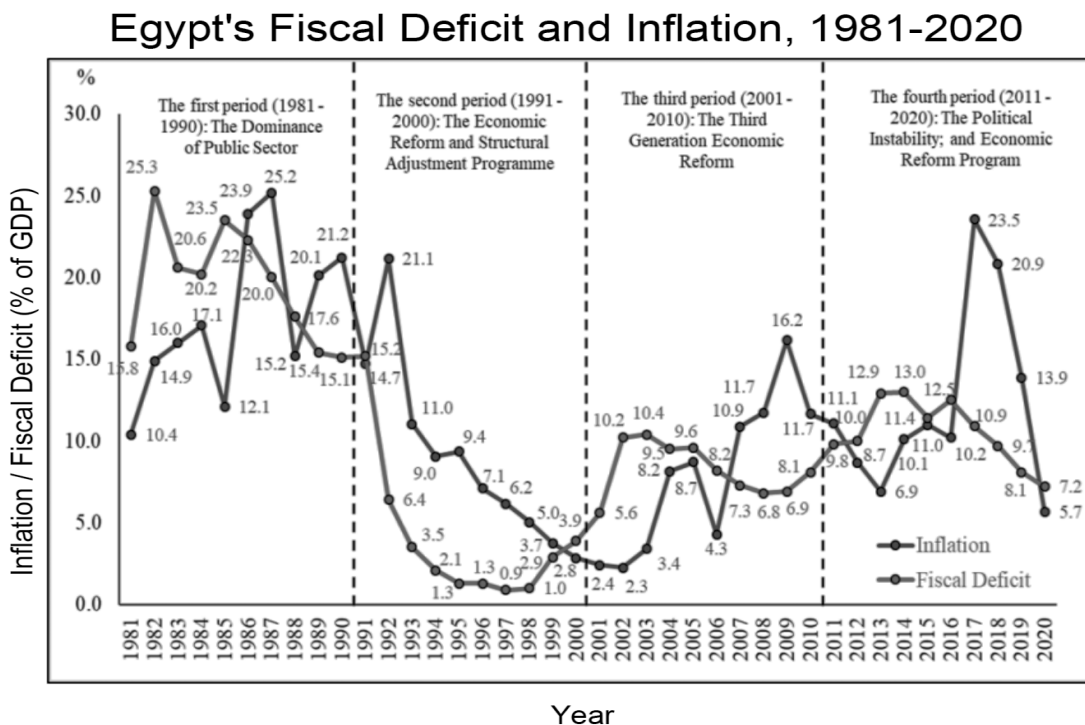


Figure 13: Egypt's External Debt Stocks (Percentage of GNI), 1970-2019



While capital still remains by far the largest factor of GDP growth in Egypt to the present day, the decline in oil prices in the late 1980s and the country's existing economic inefficiencies began to shrink public investment and slow the country's rate of capital growth beginning in the 1990s. In an attempt to stabilize and reform the Egyptian economy, President Hosni Mubarak initiated the Economic Reform and Structural Adjustment Programme (ERSAP) in 1991. Mubarak sought to tackle Egypt's financial crisis with a tight fiscal policy and contractionary monetary policy intended to curb inflation and subside final demand. However, ERSAP's contractionary policies decreased investment further. Public investment dropped to less than 10% of GDP, where it remains to this day, and Egypt's overall capital growth began to stall as well (Kamaly, 2006).

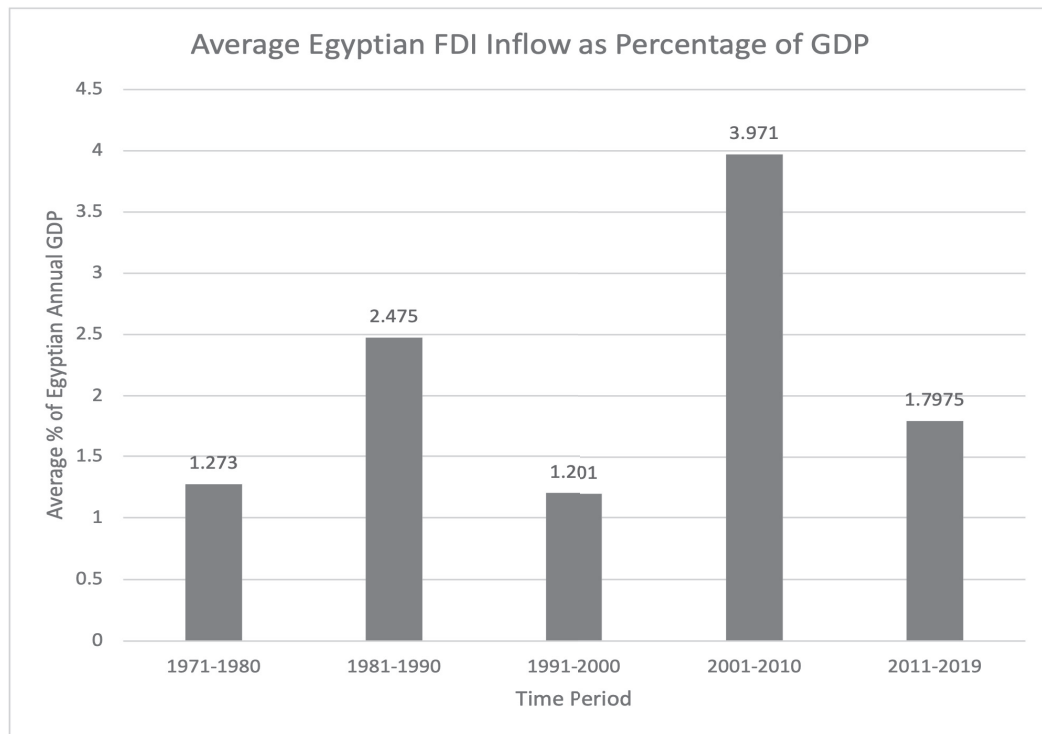
4. A Comparison of Egypt and Ireland's TFP Growth

Throughout the analysis above, we have seen that while capital formation explains the overwhelming majority of Egypt's GDP growth since the 1960s, TFP growth is the primary factor

responsible for Ireland's overall growth. In Egypt, on the other hand, TFP growth is consistently negative or near zero. Thus, the most noticeable difference in the determinants of GDP growth between Ireland and Egypt is the role of TFP. Although there is substantial skepticism as to whether Ireland's TFP growth truly reflects Irish technological progress, the difference in TFP growth between Egypt and Ireland remains large. What explains the vast differences in TFP growth between the two countries?

As explored earlier, Ireland's TFP growth is largely due its high level of FDI inflows, which are a result of the country's attractive low corporate tax rate. Absent such an advantageous tax rate, Egypt's FDI inflows are significantly lower than Ireland's. Figure 14 below shows Egypt's average FDI inflows as a percentage of GDP over each decade since 1970, using data from the World Bank's World Development Indicators database. In sharp contrast to Figure 4, which displays Ireland's average FDI inflows as a percentage of GDP, Egypt's FDI inflows are quite low and do not exhibit a clear trend. Egypt experienced its highest average FDI inflows in the 2010s, however even then, FDI inflows were still less than 4% of the country's GDP. Ireland's FDI inflows were much higher, averaging over one quarter of Irish GDP this past decade. Thus, it is likely that Egypt's comparatively low FDI inflows partially explains the difference in TFP growth between Egypt and Ireland. However, just as Ireland's TFP growth is exceptionally high, Egypt's TFP growth is exceptionally low. Egypt's TFP is consistently negative or near zero across the entire time period, 1960-2019. This introduces another question. Besides the lack of FDI, are there other reasons why Egypt's technological progress is particularly low?

Figure 14: Egypt's Average FDI Inflows, Percent of GDP



Egypt's low technological progress and TFP growth is also because the country faces a Dutch disease effect from its abundance of oil and natural gas. Named after the economic situation in the Netherlands in the 1960s, Dutch disease refers to the economic phenomenon under which development in one sector of a country's economy precipitates a decline in other sectors, and often a substantial appreciation of the national currency (Corporate Finance Institute, 2023). The abundance of crude oil, natural gas, and petroleum in Egypt gave the country a comparative advantage in oil extraction. However, oil extraction and production are highly capital-intensive processes. Although the industry requires significant capital investment, there is little room for "learning-by-doing," and thus the development of Egypt's oil and natural gas export industries were not conducive to its technological progress. Therefore, Egypt's natural resource wealth and its dominant oil export industry creates a Dutch disease effect - the oil industry thrives at the expense of technological progress, resulting in an extremely low level of TFP growth.

However, while the level and causes of Egypt and Ireland's TFP growth may be vastly different, it is possible they both share a similar Dutch disease effect - while Egypt's oil and natural gas abundance has led to low technological progress, Ireland's status as a tax haven poses risks to competitiveness and employment in its manufacturing industry. As discussed previously, Ireland's low corporate tax rate has incentivized many multinational enterprises to relocate their economic activities to Ireland, resulting in a large transfer of intellectual property, and the value of the products produced from such intellectual property, to Ireland's GDP. However, there are concerns that the sectors of the Irish economy dominated by these multinational corporations, and thus the sectors driving Ireland's GDP growth, represent a disproportionately smaller share of manufacturing employment. Ireland's competitiveness in employment-intensive sectors has been much weaker, posing a potential risk to employment in the future (Cerra & Soikkeli, 2002). Researchers Valeria Cerra and Jarkko Soikkeli of the IMF use unit labor costs to assess the competitiveness of employment-intensive sectors of the Irish economy, particularly manufacturing. Despite high production growth, they find that Irish unit labor costs remained broadly stable from 1995-2000, largely due to rapid wage increases. Increased FDI and economic activity from MNEs drove Irish wages up, and as a result, gains in competitiveness in high-employment sectors such as manufacturing have been relatively limited (Cerra & Soikkeli, 2002).

Yet, the overall competitiveness of Irish manufacturing has remained fairly stable and strong because several of the MNEs seeking out Ireland as a tax haven have been within the manufacturing sector. The especially strong performance of some of these firms in the late 1990s and early 2000s has propped up Irish manufacturing, ensuring that while gains in competitiveness have been relatively limited, the Irish manufacturing sector has not yet experienced a true contraction in output or employment (Cerra & Soikkeli, 2002). The danger, however, is that Ireland's role as a tax haven creates the conditions for Dutch disease in the future, similar to how Egypt's natural resource wealth created the conditions for its own Dutch disease. If FDI continues to drive increases in Irish wages, and if the MNEs seeking Ireland's corporate tax structure shift away from manufacturing towards less employment-intensive industries - as

has been the case this past decade as more technology-based MNEs have shifted their economic assets to Ireland, Irish manufacturing could see significant declines in competitiveness and employment. Thus, the very factor underlying Ireland's TFP growth - its corporate tax structure - could potentially inhibit the country's extraordinary GDP growth in the future.

5. Conclusion

In summary, while the two countries may have a somewhat similar colonial history, the economic growth of Ireland and Egypt is vastly different. Capital formation explains the overwhelming majority of Egypt's GDP growth since the 1960s while TFP growth is the primary factor responsible for Ireland's overall growth. In both cases, globalization and increasing contact with foreign nations - through FDI for Ireland and trade for Egypt - were the underlying fundamental causes of each country's growth. However, the countries' different economic and political structures led them along vastly different growth paths. Ireland is a privatized, open, and well-industrialized economy. As other countries in the European Union raised their corporate tax rate, Ireland's corporate tax rate remained low. Multinational corporations, especially American technology companies, sought to take advantage of Ireland's advantageous tax structure, relocating much of their economic activities, and thus their underlying intellectual property, to Ireland. Thus, increased FDI inflows are likely a fundamental cause of the high TFP growth that undergirds Ireland's dramatic growth rate (Setser, 2019; Taqui, 2023; OECD, 2016). However, there is substantial debate concerning if production from intellectual property should really be included in Ireland's GDP, and if such production is truly reflective of underlying *Irish* economic growth. Such concerns have led to the use of GNI and modified GNI as metrics better suited to capture Ireland's economic growth and welfare. Modified GNI suggests that, absent the effects of globalization and foreign-owned MNEs, Ireland's growth and welfare have experienced much more modest increases in recent years than the country's GDP would suggest. Despite some concerns that Ireland's high FDI may

decrease future employment and competitiveness in manufacturing, overall, Ireland's increased economic interactions with foreign nations have had positive effects on the domestic economy.

However, Egypt's foreign economic interactions have had a much more mixed effect on the domestic economy, chiefly due to the country's higher degree of centralization and its natural resource curse. In 1973, Egyptian President Anwar al-Sadat initiated the Open Door Policy, a series of economic reforms aimed at privatizing industry and opening the country up to foreign trade and investment. The policy did indeed privatize many economic sectors, such as agriculture and finance, thus initiating private investment and capital accumulation for the first time in Egyptian history. However, government bureaus and state-owned enterprises retained control of key industries, most notably petroleum and crude oil production (International Trade Administration, 2020). Now open to foreign trade, Egypt dramatically increased its natural resource exports, particularly in oil and natural gas, which initially spurred substantial public investment, revenue inflows, and GDP growth. In the mid 1970s, Egypt benefited from a period of extremely high oil prices, and its investment in oil production infrastructure facilitated the capital growth that drove Egypt's GDP growth. However, because the petroleum industry remained under the control of the state-run Ministry of Petroleum and Mineral Resources, investment and export revenues were directed towards the public sphere and private investment remained low. When oil prices tanked in the mid 1980s, existing macroeconomic imbalances from high inflation, a substantial current account deficit, and remaining administrative restrictions sent the country into a financial crisis. Political leaders responded by pursuing a contractionary fiscal policy that restricted investment and capital accumulation even further, ultimately slowing GDP growth (Handy, 2010).

This comparison of Ireland and Egypt indicates that increased economic interactions with foreign nations - whether through FDI or trade - can have vastly different effects on a country's long-term GDP growth, depending on the institutional, economic, and political environment. These results also clearly indicate the power of the resource curse and the Dutch disease effect, particularly in regards to crude oil and petroleum. The volatility of oil prices and state control of the petroleum industry, as is common in

many resource-rich African and Middle Eastern nations, limited Egypt's private sector investment and led to steep peaks and troughs in capital accumulation and GDP growth (Handy, 2010). Thus, a shared colonial history does not necessarily determine a country's economic path. Institutional structures, economic policies, and perhaps most importantly, natural resource wealth, have a much more direct effect on a country's growth trajectory. As more countries in Africa and the Middle East begin opening their borders to foreign trade and investment, their economic and political structures, rather than their colonial or tribal history, are likely to determine whether they follow the path of Egypt or Ireland in our increasingly global economy.

References:

1. Abdu, Mohammed, "Analyzing the relationship between Fiscal Deficit and Inflation: The Case of Egypt," *Ministry of Finance, Egypt*, December 2022.
https://www.researchgate.net/publication/365915676_Analyzing_the_relationship_between_Fiscal_Deficit_and_Inflation_The_Case_of_Egypt.
2. Alissa, Sufyan, "The Political Economy of Reform in Egypt: Understanding the Role of Institutions," *The Carnegie Endowment for International Peace*, October 2007.
https://carnegieendowment.org/files/cmec5_alissa_egypt_final.pdf
3. Allouche, Yasmina, "Remembering the 1952 Egyptian Revolution," *Middle East Monitor*, July 2017.
<https://www.middleeastmonitor.com/20170723-remembering-the-1952-egyptian-revolution/>.
4. Andersson et al. "Intangible assets of multinational enterprises in Ireland and their impact on euro area GDP," *European Central Bank*, 2023.
https://www.ecb.europa.eu/pub/economic-bulletin/focus/2023/html/ecb.ebbox202303_02~3404c284d0.en.html.
5. Beatty, Aidan, "Social Revolution Beyond the Volga: Egypt and Ireland," *Age of Revolutions*, February 2018.
<https://ageofrevolutions.com/2018/02/12/social-revolutions-beyond-the-volga-egypt-and-ireland/>.
6. Bourke, Cormac, "Low-tax policies created the Tiger," *Irish Independent*, October 2004.
<https://www.independent.ie/opinion/editorial/low-tax-policies-created-the-tiger/26225791.html>
7. Buckner, Elizabeth, "Access to Higher Education in Egypt: Examining Trends by University Sector," *The University of Chicago Press Journals*, August 2013.
<https://www.journals.uchicago.edu/doi/10.1086/670665>.
8. Cassidy, "Productivity in Ireland: Trends and Issues," *Central Bank of Ireland Economic Analysis, Research and Publications Department*, Spring 2004.
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=52993972b5f4f7681e1d3b0cc1024918b9d79083>
9. Centra Statistics Office, "Modified GNI," *Irish Department of Finance, Central Statistics Office*, November 2023.
<https://www.cso.ie/en/interactivezone/statisticsexplained/nationalaccountsexplained/modifiedgni/>
10. Cerra, Valerie and Jarkko Soikkeli, "How Competitive Is Irish Manufacturing?" *International Monetary Fund*, September 2002.

<https://www.imf.org/en/Publications/WP/Issues/2016/12/30/How-Competitive-Is-Irish-Manufacturing-16057>.

11. Cooper, Andrew Scott, "A Tale of Two Oil Shocks | Part 1: 1973-76," *PBS*, June 2012.
<https://www.pbs.org/wgbh/pages/frontline/tehranbureau/2012/06/a-tale-of-two-oil-shocks-part-1-1973-76.html>.
12. Corporate Finance Institute, "Dutch Disease," *Corporate Finance Institute*, November 2023.
<https://corporatefinanceinstitute.com/resources/economics/dutch-disease/>.
13. Department of Finance, "The Fastest Growing Economy in the EU," *Government of Ireland*, September 2023. <https://www.gov.ie/en/publication/86872-irish-economy/#>.
14. Drummond, James, "How Anwar Sadat's Open Door policy integrated Egypt with developed market economies," *Arab News*, October 2021.
<https://www.arabnews.com/node/1942181/middle-east>.
15. Gately, Dermot, "Lessons from the 1986 Oil Price Collapse," *Brookings Papers on Economic Activity*, 1986.
https://www.brookings.edu/wp-content/uploads/1986/06/1986b_bpea_gately_adelman_griffin.pdf.
16. Global Edge, "Egypt: Economy," *Global Edge*, 2021.
<https://globaledge.msu.edu/countries/egypt/economy>.
17. Goldberg, Ellis & Joel Beinin, "Egypt's Transition under Nasser," *Middle East Research and Information Project*, August 1982. <https://merip.org/1982/07/egypts-transition-under-nasser/>.
18. Handoussa, Heba, "Egypt's Investment Strategy, Policies, and Performance Since the Infitah," *IMF eLibrary*, 2010. <https://www.elibrary.imf.org/display/book/9781557751409/ch07.xml>
19. Handy, Howard, "A Historical Examination of Growth, Investment, and Saving," *IMF eLibrary*, 2010. <https://www.elibrary.imf.org/display/book/9781557757203/ch02.xml>.
20. IMF - European Department, "Firm-Level Productivity and its Determinants: The Irish Case," *IMF eLibrary*, July 2016.
<https://www.elibrary.imf.org/view/journals/002/2016/257/article-A003-en.xml>
21. IMF, "Ireland: From Tiger to Phoenix," *International Monetary Fund (IMF)*, August 2018.
<https://www.imf.org/en/Countries/IRL/ireland-from-tiger-to-phoenix>.
22. International Trade Administration, "Egypt - Oil & Gas," *International Trade Administration*, 2020.
<https://www.trade.gov/energy-resource-guide-egypt-oil-and-gas?anchor=content-node-t7-field-lp-region-2-1>.

23. John, Robert, "Gamal Abdel Nasser," *Britannica*, September 2013.
<https://www.britannica.com/biography/Gamal-Abdel-Nasser>.
24. Kamaly, Ahmed, "Economic Growth Before and After Reform: The Case of Egypt," *International Journal of Applied Econometrics and Quantitative Studies*, 2006.
<https://www.usc.es/economet/reviews/ijaeqs322.pdf>
25. Khaled, Nourallah, "More than a 100-year journey," *Egypt Oil & Gas*, December 2014.
<https://egyptoil-gas.com/features/more-than-a-100-year-journey/>.
26. Library of Congress, "Oil and Gas Industry: A Research Guide - Upstream: Production and Exploration," *Library of Congress Research Guides*, 2019.
<https://guides.loc.gov/oil-and-gas-industry/upstream>.
27. OEC, "What does Egypt Export?" *Observatory of Economic Complexity (OEC)*, 2023.
https://oec.world/en/visualize/tree_map/sitc/export/egy/all/show/2018/
28. OECD, "Irish GDP up by 26.3% in 2015?" *OECD*, Oct. 2016.
<https://www.oecd.org/sdd/na/Irish-GDP-up-in-2015-OECD.pdf>
29. OECD, "OECD Review of Foreign Direct Investment: Ireland," *OECD*, September 1993.
<https://www.oecd.org/ireland/34383945.pdf>
30. OPEC, "Data Download: Table 7.1: OPEC Reference Basket (ORB) and corresponding component's spot prices & Table 7.2: Selected spot crude oil prices," *Organization of the Petroleum Exporting Countries (OPEC)*, 2023. https://asb.opec.org/data/ASB_Data.php.
31. Schrader-King, Kristyn, "EU11 Growth Drops in 2012 and Remains Slow in 2013 Amid Weak External Confidence," *The World Bank*, June 13, 2013.
<https://www.worldbank.org/en/news/press-release/2013/06/13/eu11-regular-economic-report-june-13-2013>.
32. Setser, Brad W. "Ireland's Statistical Cry for Help," *Council on Foreign Relations*, November 2019. <https://www.cfr.org/blog/irelands-statistical-cry-help>.
33. Taqui, Anood, "2023 Investment Climate Statements: Ireland," *US Department of State, Embassy of Ireland*, 2023.
[https://www.state.gov/reports/2023-investment-climate-statements/ireland/#:~:text=One%20of%20Ireland%27s%20many%20attractive,the%20European%20Union%20\(EU\)](https://www.state.gov/reports/2023-investment-climate-statements/ireland/#:~:text=One%20of%20Ireland%27s%20many%20attractive,the%20European%20Union%20(EU))
34. The World Bank, "World Development Indicators Database," *The World Bank*, October 2023.
<https://datatopics.worldbank.org/world-development-indicators/>

35. The World Bank, "External debt stocks (% of GNI) - Egypt, Arab Rep." *The World Bank International Debt Statistics*, January 2021.
<https://data.worldbank.org/indicator/DT.DOD.DECT.GN.ZS?locations=EG&view=chart>.
36. Wallenfeldt, Jeff, "The Troubles," *Britannica*, Oct 2013.
<https://www.britannica.com/event/The-Troubles-Northern-Ireland-history>.

Right-to-Work Laws and Innovative Output

Paul Irvine

MIT Department of Economics

6 December 2023

ABSTRACT

This paper uses an event study approach to investigate the effect of right-to-work laws on states' innovative output, as measured by research and development (R&D) spending and utility patents issued in the state. We find little to no evidence of a causal effect of RTW laws on R&D intensity (total R&D spending / GDP), business R&D spending, nor total R&D spending per capita; on the other hand, we find some evidence that RTW laws increase utility patents issued per million state residents by around 50 to 110 within 5 years after passing the RTW law, with the effect increasing over time.

1 Introduction

Innovative output is an important consideration for policymakers, businesses, and workers alike. Not only is innovation needed to sustain long-term economic growth, but issues

that have or will have broader societal implications— such as in health- or climate-related industries— will require substantial technological innovation. Given the importance of innovation in these contexts, it is worthwhile to investigate different factors that can influence innovative output, either at the micro- (firm or facility) or macro- (aggregate economy) level. This paper decides to investigate the effect of right-to-work (RTW) laws on innovative output.

There are a few possible avenues by which unions can influence innovation. On the positive side, since innovation is a long and uncertain process, unionization may provide employees with “protection against dismissal in bad faith [that] is necessary to effectively motivate and nurture innovation,” which Bradley et. al (2014) dubs as the *employee protectionism hypothesis*; though, the authors suggest that for unionization to support this hypothesis, union contracts would have to also include long-term rewards for success, not just protect from short-term failure in innovation.

On the other hand, Bradley et. al (2014) finds that, on average, “passing a union election leads to an 8.7% decline in patent counts and a 12.5% decline in the number of citations per patent three years after the election.” By comparing the R&D expenditures, patent output and quality, and highly productive inventors’ firm departure rate for firms that barely vote to unionize to those that did not, the paper finds evidence for their *misaligned incentives hypothesis* via three ways unionization may decrease a firm’s innovation:

1. *Reduction of Research and Development (R&D) Expenditures* — After a significant investment in R&D is sunk, unionized employees may use the fact that the success of the project relies on them and therefore demand higher wages or other terms, known as “ex-post holdup” (Bradley et. al, 2014). Recognizing this ahead of time, firms may underinvest in R&D ex-ante (Grout, 1984; Malcomson, 1997). Bradley et. al (2014) estimates the causal effect of unionization being a 0.6% reduction in R&D / Assets.
2. *Increase in Shirking Reduces Productivity* — Unionization reduces the chances of dismissal, so unionized employees may supply less effort in their work which could lead to lower productivity among workers. Bradley et. al (2014) finds a significant reduction in both number of patents and number of citations per patent, for both inventors that stayed and those that were just hired at the firm compared to their previous productivity, generated within three years after a firm votes to unionize compared to those that do not.
3. *Departure of Highly-Productive Inventors* — Since unionization reduces wage inequality among workers (Frandsen, 2012), highly-productive inventors may leave the unionized firm to make more money as a non-unionized employee— exactly what Bradley et. al (2014) found.

There is some evidence that unionization can reduce a firm’s innovative output, but it is less clear if this impact noticeably scales up in state-wide aggregate innovative activity. Bradley

et. al (2014) may be treating all workers equally when it comes to innovation, underestimating unionization's impact: Intuitively, HVAC workers in firm A voting to unionize will likely have a much lower effect on innovation compared to if the scientists in firm B's research lab unionize, even if the paper classifies both firms as unionized. Furthermore, since the paper uses a regression discontinuity design from unionization elections, there is of course the possibility that its finding that unionization reduces patent output by nearly 10% cannot be generalized. It could be that the workplaces where union elections are decided closely (or perhaps even workplaces that hold an election in the first place) are somehow fundamentally different from ones where elections are won or lost decisively—and thus the paper's results do not apply to workplaces that are more on the tail ends of union election results or do not hold union elections.

This paper seeks to build on the previous literature on unionization and innovation by investigating whether the passing of Right-to-Work (RTW) laws impacts a state's innovative output. Other research has found that RTW laws significantly reduce unionization rates by around 4% within 5 years of passing (Fortin et. al, 2022), so we would expect that passing RTW laws would increase innovative activity. Taking Bradley et. al's (2014) results at face-value, and assuming RTW laws effects were roughly equally distributed among firms, we can estimate that passing RTW laws would increase patents-per-capita by around 0.348% and R&D spending / asset by around 0.024%. Although our actual estimated causal effect coefficients are indeed mostly positive, most of them are not surprisingly statistically significant, with the exception of utility patents issued per million state residents which yields (mostly) significant estimates. Our

results suggest that adopting RTW laws increases utility patents issued per million state residents by around 10 the first year and ramps up to over 120 by the fifth year— far off from our back-of-the-envelope calculation. The rest of the paper is as follows: Section 2 discusses some background around RTW laws and the datasets used for this paper’s empirical test, section 3 outlines the event study process used to estimate the causal coefficients, and the last section provides further discussion of the results, its limitations, and policy and academic implications.

2 Background & Data

A Right-to-Work (RTW) state prohibits employees from being required to join or financially support a labor union (Right-to-work laws, 2023), even though they benefit equally from the union’s collective bargaining. This generates a “free-rider” issue that reduces union power, finances, and organizational ability— as suggested by the lower unionization rates in RTW states compared to non-RTW states (Fortin et. al, 2022). Twenty-seven states have RTW laws as of March 2023, making for a decent sample variation across the US states; however, nearly all of them adopted RTW laws in the 1940s or 1950s, before detailed data across many topics existed, making it difficult to run event studies. During a conservative wave between 2012 and 2017, five states— Michigan, Wisconsin, Indiana, Kentucky, and West Virginia— adopted RTW laws, offering an opportunity to run empirical tests, following the methodology of Fortin et. al (2022). Although that paper explores the effect of RTW laws and unionization on employee wages, I will use their methodology to instead investigate RTW laws’ effect on innovative output, attempting to build on the Bradley et. al (2014) paper discussed earlier.

Following Fortin et. al's (2022) lead, I restricted the years of interest from 2007 to 2019 to keep the number of years pre- and post-RTW adoption relatively balanced, and to avoid the data from being biased by the COVID-19 pandemic. From the National Science Foundation's Science and Engineering State Profiles dataset, I was able to get each state's business R&D performance (in millions of US dollars), total R&D performance (in millions of US dollars), and gross domestic product (in billions of US dollars). Since each profile was offered year-by-year, I downloaded all 13 years of spreadsheets, and put it into a master spreadsheet with all the data and made sure it was formatting correctly (e.g. some of the data points had footnotes that needed to be manually removed or else the regression would run 1544²² as 154422). Additionally, I double-checked that the data points seemed reasonable and manually patched the points with the National Science Foundation's interactive table whenever there obvious errors in reporting—say, for example, when the data set reported Vermont's business R&D as jumping from \$307m to \$4486m to \$248m, when the middle value was actually supposed to be \$247m.

In the context of this paper, utility patents that capture a new or improved product or process make sense to measure innovative activity, compared to plant and decorative patents, which do not quite have the usefulness one typically associates innovation with; however, we recognize that this is not a perfect measure, as it captures the quantity of patents but not necessarily their quality, which is arguably just as or perhaps more important for technological innovation. Unfortunately, the dataset that Bradley et. al (2014) used to quantify patent quality via number of citations does not go to 2019 as we need, and as far as I can tell there exists no

analogous dataset that does. Similarly, another limitation with this data is that it is aggregated and therefore cannot be broken down to the industry, firm, or facility level— making it impossible to cross-reference it with other, more granular information on unionization. The National Science Foundation’s dataset had gaps in utility patents granted to state residents each year, so I supplemented it with the United States’s Patent and Trademark Office’s Utility Patents by Country, State, and Year to find the number of utility patents generated by each state each year. Intuitively, the more workers a state has, the more innovative output they should be able to generate: The National Science Foundation again had gaps in these relevant variables, so I used instead the United States Census Bureau’s State Population by Year datasets to find each state’s population. Lastly, I hard-encoded each state’s RTW adoption year (if applicable) from Ballotpedia’s Right-to-work laws interactive map. Each dataset included Puerto Rico and the District of Columbia, and all units were standardized (e.g. all GDP data was converted to be in billions of dollars, and population in thousands). These four datasets were spliced together to form one larger one that includes each state and year’s GDP, population, business and total R&D, RTW adoption status and adoption year, and utility patents issued. From this dataset, which totaled 676 observations (50 states plus Puerto Rico and the District of Columbia, each over 13 years), I was able to run the event study, detailed in section 3.

**** Insert Figure 1 Here ****

We plot the breakdown of innovative activity by RTW law adoption status in figure 1

above. For these plots, we define “Never-Adopter” states as those that are not RTW states as of the end of 2019; “Early-Adopter” states as those that passed RTW laws 2001 and before (so as to include Oklahoma); and “Late-Adopter” states as those that passed RTW laws 2012 and after—so Michigan, Wisconsin, Kentucky, Indiana, and West Virginia. Immediately, it is apparent that for all three population-controlled measures of innovative activity, the states that never introduced RTW laws rank far higher— usually over double— than either Early- or Late-Adopter states. Late-Adopter states are usually sandwiched in-between the two other kinds of states in the innovative metrics, but typically trends far closer to the Early-Adopter states. Clearly, to little surprise, these observations show that there are other drivers of innovation that are not RTW laws and unionization: If the micro-effects observed in the Bradley et. al (2014) were to scale up on the aggregate state economy, and unionization rates were the only difference between the three kinds of RTW states, we would expect the Early-Adopter states to have higher rates of innovation compared to Never- and Late-Adopter states, with the Late-Adopter states accelerating innovative activity after passing RTW laws. Indeed, we observe just the opposite— suggesting far more potent influences at play in determining a state’s innovative output.

**** Insert Figure 2 Here ****

For this event-study to be valid, we need parallel trends assumption— that is, innovative activity in the RTW adopter states would have followed the same trends as the non-RTW states if it was not for adopting RTW. From figure 2, we can see that innovative activity for all

Late-Adopter states seem to trend almost perfectly parallel from 2007 to 2011, before any of the states adopt RTW laws. Some states like Michigan then seem to take-off in terms of innovative activity, while others like West Virginia never increase at all. From this plot, even if there is variation in the trends, one cannot deduce each state's RTW year switch, which would be the case if the RTW causal effect was strong. We conclude the plausibility of the parallel trends assumption, but will further verify in section 3 with the causal effect estimation coefficients.

3 Event Study Setup and Estimates

For this analysis, we will use an event-study that flexibly allows the causal effect to vary year-by-year after the event. In this context, the event refers to when a state passes a RTW law.

We estimate the effect of RTW laws on innovative activity using the following regression:

$$Y_{st} = \alpha_s + \delta_t + \sum_{k=-5}^{k=5} \pi_k 1\{K_{st} = k\} + X_{st}\phi + \varepsilon_{st}$$

where Y_{st} is the measurement of innovative activity; α_s is the state fixed effects; δ_t is the year fixed effects; $1\{\bullet\}$ is the indicator function; π_k is the estimated effect of RTW k years after adopting RTW, with π_k for $k < 0$ capturing possible pre-trends; $K_{st} = k = \text{year } t - \text{year of RTW adoption}$, for the states that adopted RTW; and X_{it} is the state-level covariates. There are a few different ways to measure a state's innovative output: I will use Y_{st} as the utility patents issued to state s residents in year t , business R&D performance (i.e. expenditures) in state s in year t , and R&D intensity (calculated by *Total R&D Spending in*

state s year t / GDP in state s year t). The covariates are a bit trickier to measure, as we need to identify factors that can influence a state's innovative activity as measured by Y_{st} that doesn't involve unionization— there are quite a few. To start, we need to control for the population of the state since the more laborers a state has available to innovate, the more patents it is likely to generate. The year and state fixed effects should luckily capture many of these differences that could arise. In all these regressions, I cluster by state and encode the years as time series data to obtain the appropriate standard errors for the resulting coefficient estimates. Summary statistics and variable labels of the regression results are included in the appendix.

For the causal effect of RTW laws on innovation, we are interested in the π_k coefficients. As is typical with event studies, after running the regression, I plot the π_k coefficient on the vertical axis, and the k coefficient as the horizontal axis (i.e. time before/after a state adopts a RTW law), as seen in figure 3 below. The standard errors on the π_k coefficients are plotted as well, clustered by states. Since the time period considered for this analysis only goes to 2019, but a few states adopted RTW laws after 2014, the post-event coefficients' standard errors increase further away from the event since the treatment state sample size decreases (e.g. since Kentucky adopted RTW in 2017, it only has two years of post-event data available, meaning it is not included for any estimate with $k > 2$, increasing the standard error for those post-event times).

** Insert Figure 3 Here **

We can check that before the event, π_k is statistically indistinguishable from 0. In every case with only a single exception, we have that $\pi_k \cong 0$, so we find little evidence of pre-trends. This provides further evidence that the parallel trends assumption holds well. Next, we look at the coefficient estimate after the event takes place: Even though most of the estimates have positive values as we would expect, since the 95% confidence error bars firmly include 0, we find little to no evidence of a causal effect of RTW laws on R&D intensity (total R&D spending / GDP), business R&D spending, nor total R&D spending per capita. Interestingly, the coefficients tend to get larger in magnitude the farther away from the event it is, providing sort of an upward slope on the RTW effect, even if the estimates themselves remain statistically insignificant.

On the other hand, the utility patents issued per million state residents are more debatable. Although some of the estimates' error bars still technically include 0, more often than not they are barely on the edge of the confidence interval, especially for the later values of k . As was mentioned previously, the error bars increase for later values after the event trigger because some of the Late-Adopter states drop out from the sample since they adopted so late in the relevant time period. It could be that RTW laws do have a statistically significant effect on utility patents issued to state residents, but there simply was not enough data on the post-event period that would have reduced the error bars to make the estimates more definitive. Furthermore, we notice that the coefficient estimate for the utility patents also increases in magnitude the farther away from the RTW adoption. If this causal effect is truly significant, this ramp-up behavior could be explained by the fact that the effects of RTW laws tends to be more corrosive over

destructive to union power— that is, it takes time for workers to opt out of paying union dues and unions' financial strain to be realized once contracts expire and need to be renegotiated, meaning the effects of RTW could be delayed. Should this be the case, we would expect that the ramp-up would begin to plateau if we were to extend the post-event period by several years—and the true RTW casual effect would be the value the coefficient estimate approaches.

Assuming year three after the event is the causal effect, then our results suggest that adopting RTW laws would increase the utility patents issued per million state residents by around 50.

To push back against these results, one could imagine the following scenario: The political climate that allowed states to adopt RTW laws also allowed them to simultaneously adopt other business-friendly legislation, which is the actual reason for the results we find. To test this possibility, we look at CNBC's top states for business rankings over the relevant time

**** Insert Figure 4 Here ****

period. If the adoption of RTW laws coincided with a host of other business-friendly legislation that truly made these Late Adopter states more innovative, we would expect to see these states to rise in ranking (and thus drop in figure 4) around the time of RTW adoption. As we can see, Kentucky, Indiana, and West Virginia's rankings stayed flat before and after RTW adoption, while Michigan and Wisconsin seemed to trend downward before adoption. Absent additional evidence, the argument that RTW adoption coincided with other legislation that made it more

attractive for business, increasing innovation, therefore seems relatively unconvincing.

4 Conclusion

The results in this paper provide more nuance to the relationship between unionization and innovative output, but leaves much to be discovered. Since unionization has been found to reduce firm R&D spending and patent quantity and quality, we hypothesized that RTW laws would increase such innovative activity. We find little to no evidence of a causal effect of RTW laws on R&D intensity (total R&D spending / GDP), business R&D spending, nor total R&D spending per capita. Recall that the Bradley et. al (2014) paper found a significant reduction in firm R&D spending as a result of unionization— which should imply that the passing of RTW laws that discourage unionization and reduce collective bargaining power should increase aggregate business R&D spending. One can conceive a few reasons why the micro-results did not exactly scale onto the aggregate economy. As Fortin et. al (2022) found, the reduction of unionization rates from the passing of RTW laws were concentrated in construction, education, and public administration— industries that are not particularly known for R&D spending. Union influence in industries that tend to be more innovative— software and technology, manufacturing of all kinds, and health care, for example— simply may be more-or-less unaffected by the passing of RTW laws, blunting the laws' impact on aggregate R&D spending. Furthermore, the percentage of the workforce of unions as a whole is relatively low, so even if RTW laws do reduce union power and organizing rates, its effect on the aggregate economy might be diluted.

However, these stories conflict with the relatively significant results of RTW laws increasing the utility patents issued to state residents. Although Bradley et. al (2014) points to a reduction of R&D spending decreasing issued patents, they also show evidence that R&D efforts might become less productive as shirking increases and highly productive inventors depart for better pay. If RTW laws did indeed reinforce these pathways, it is possible that R&D expenditures do not change but productivity decreases anyways. Perhaps the passing of RTW laws reduced bargaining power of already-unionized R&D facilities, such that their ability to secure protections from dismissal is reduced and therefore workers' innovative productivity increases. Perhaps more highly productive inventors moved to states that passed RTW laws because they thought they would be able to secure higher wages without the threat of unions, increasing the state's innovative productivity— though this seems quite unrealistic. These are scenarios by which R&D spending would remain unchanged as a result of RTW laws being passed, while patent output would increase. There is the added complication that some of RTW's effect could be coming from firms responding to a reduction in the threat of unionization, and not the unionization itself. Teasing out these effects will require further investigation.

We also flag the fact that the Late-Adopter states are clustered geographically in the Rust Belt, which could mean these paper's results are not easily generalizable. It is feasible to imagine a scenario where the passing of RTW laws coincided with other conservative policies that especially appeal to the Rust Belt states, and thus are not replicated elsewhere in the United States. Although I would argue this issue is unlikely because of the nationalization of party

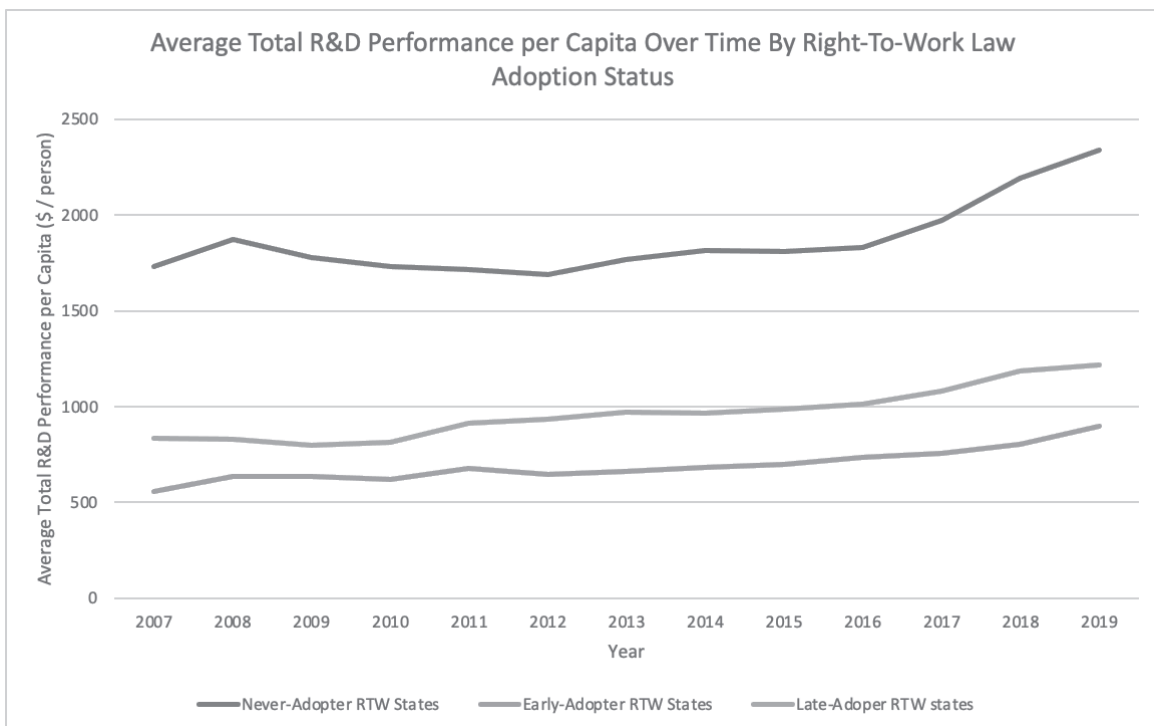
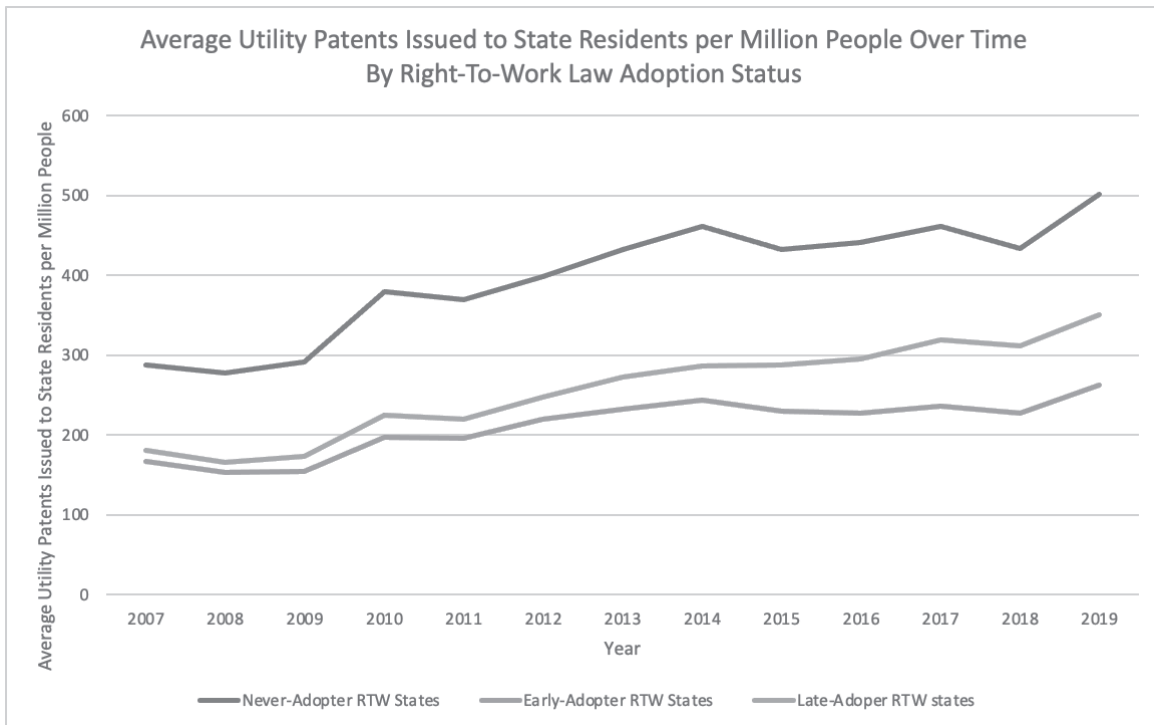
politics and the spread of years of RTW adoption, it is a risk that should be considered, especially when all the treatment states are clustered in roughly the same geographic region.

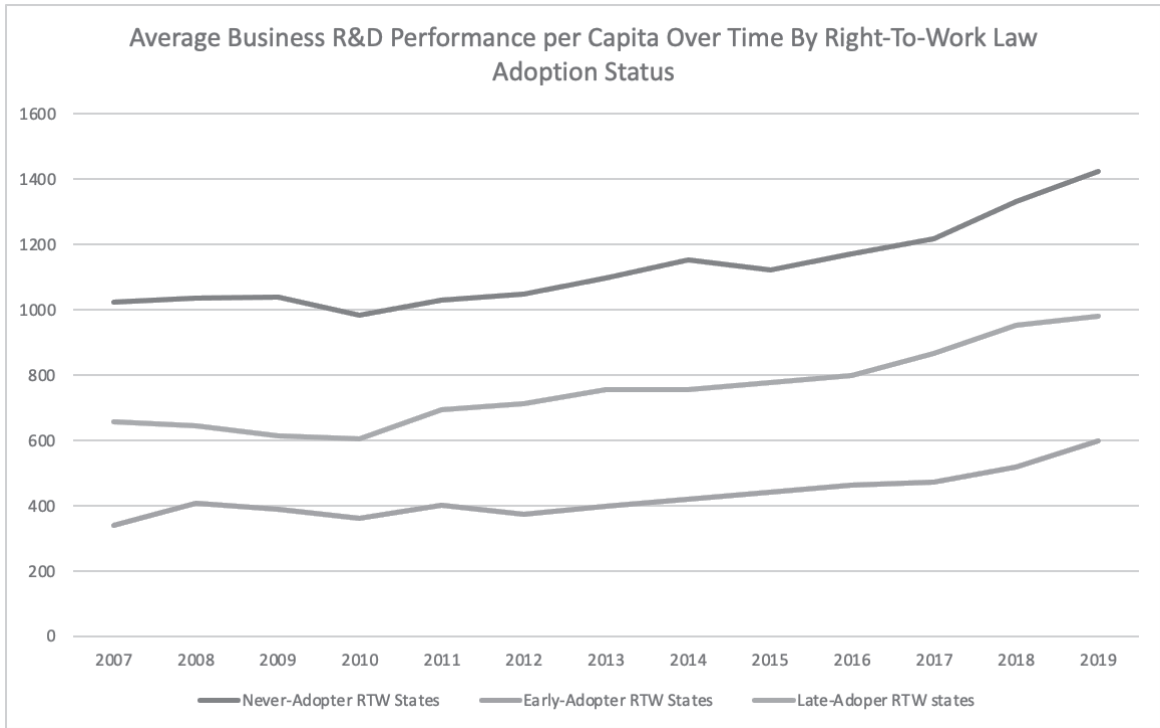
One last consideration is the possible spillover effects of RTW laws from multi-state firms. For example, say firm A had its R&D activity in New Jersey, which is a Never-Adopter of RTW, and all of its manufacturing activity in Indiana, which is a Late-Adopter of RTW. When Indiana passed RTW laws in 2012, it is feasible that firm A increased its cash flow (perhaps because the threat of its manufacturing workers unionizing decreases, so less money is spent on union busting) that can then be spent on increasing R&D activity, which is in New Jersey. These kinds of spillovers could complicate this paper's analysis, but since the dataset we used is not granular at the firm- or facility-level, we leave this issue to be explored in future work.

Our results shed some light on the relationship between unionization and innovation. Although much of this paper is built around the premise that unions typically are negative for innovative output, we want to point out that union contracts are assumed to be unable to balance short-term tolerance for failure (which is needed for the innovative process) with long-term rewards for success, such as in the form of royalties and stock options. Given the importance of innovation in solving society's most pressing challenges, as well as the need for workers to have liveable wages and working conditions, we hope this paper invites more discussion on how to best structure policy and other systems to further both mutual interests.

References

- Bradley, Kim, and Tian. Do Unions Affect Innovation? *Management Science*, 2017, vol. 63, no. 7, pp. 2251-2271. <https://doi.org/10.1287/mnsc.2015.2414>.
- CNBC. America's Top States for Business in 2019. 10 Jul. 2019.
<https://www.cnn.com/2019/07/10/americas-top-states-for-business-2019.html>
- Utility Patents Patent Counts By Country, State, and Year. *United States Patent and Trademark Office*. December 2015. https://www.uspto.gov/web/offices/ac/ido/oeip/taf/cst_utlh.htm.
- Fortin, Lemieux, and Lloyd. Right-to-work laws, unionization, and wage-setting. Working Paper. *National Bureau of Economic Research*. June 2022.
https://www.nber.org/system/files/working_papers/w30098/w30098.pdf
- Frandsen, B. 2012. Why unions still matter: The effects of unionization on the distribution of employee earnings. Working paper.
- Grout, P. A. 1984. Investment and wages in the absence of binding contracts: A Nash bargaining approach. *Econometrica* 52: 449-60.
- Malcomson, J. M. 1997. Contracts, hold-up, and labor markets. *Journal of Economic Literature* 35: 1916-57.
- Right-to-work laws. *Ballotpedia*. ballotpedia.org/Right-to-work_laws. Accessed 20 Nov. 2023.
- Science and Engineering State Profiles. *United States National Science Foundation*.
www.nsf.gov/statistics/states/#ui-tabs-4. Accessed 13 Nov. 2023.
- United States State Populations by Year. *United States Census Bureau*. Accessed 13 Nov. 2023.
<https://www2.census.gov/programs-surveys/popest/datasets/>





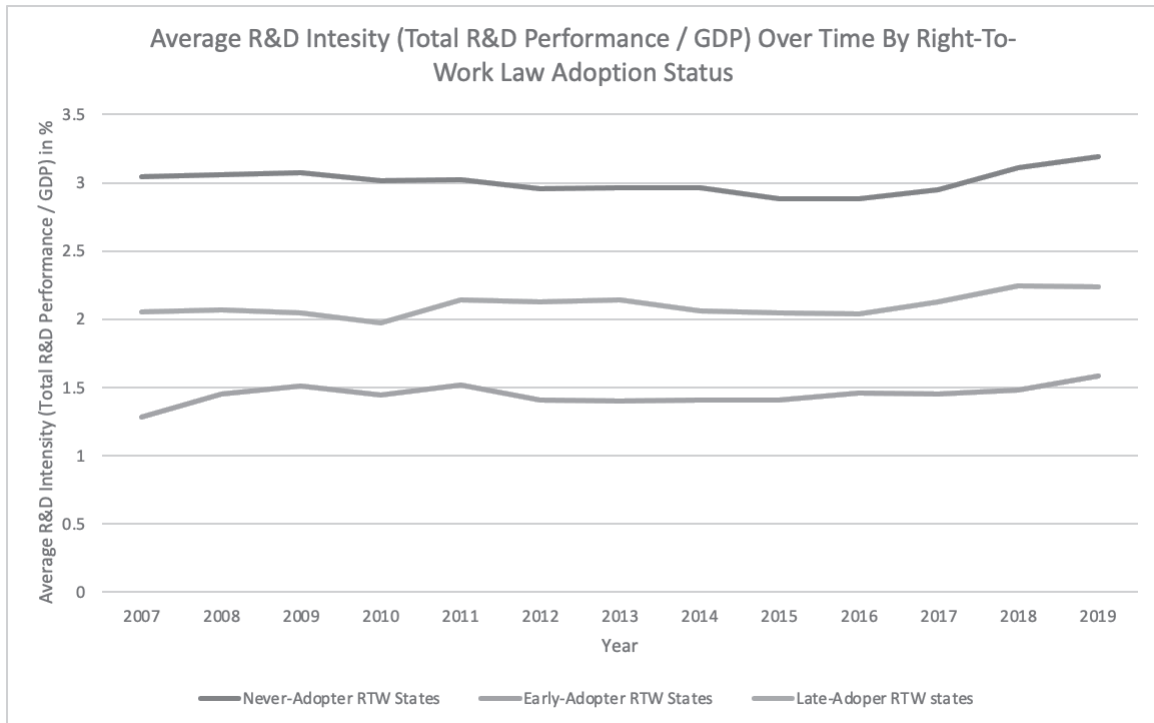


Figure 1. Average innovative activity over time broken down by state types.

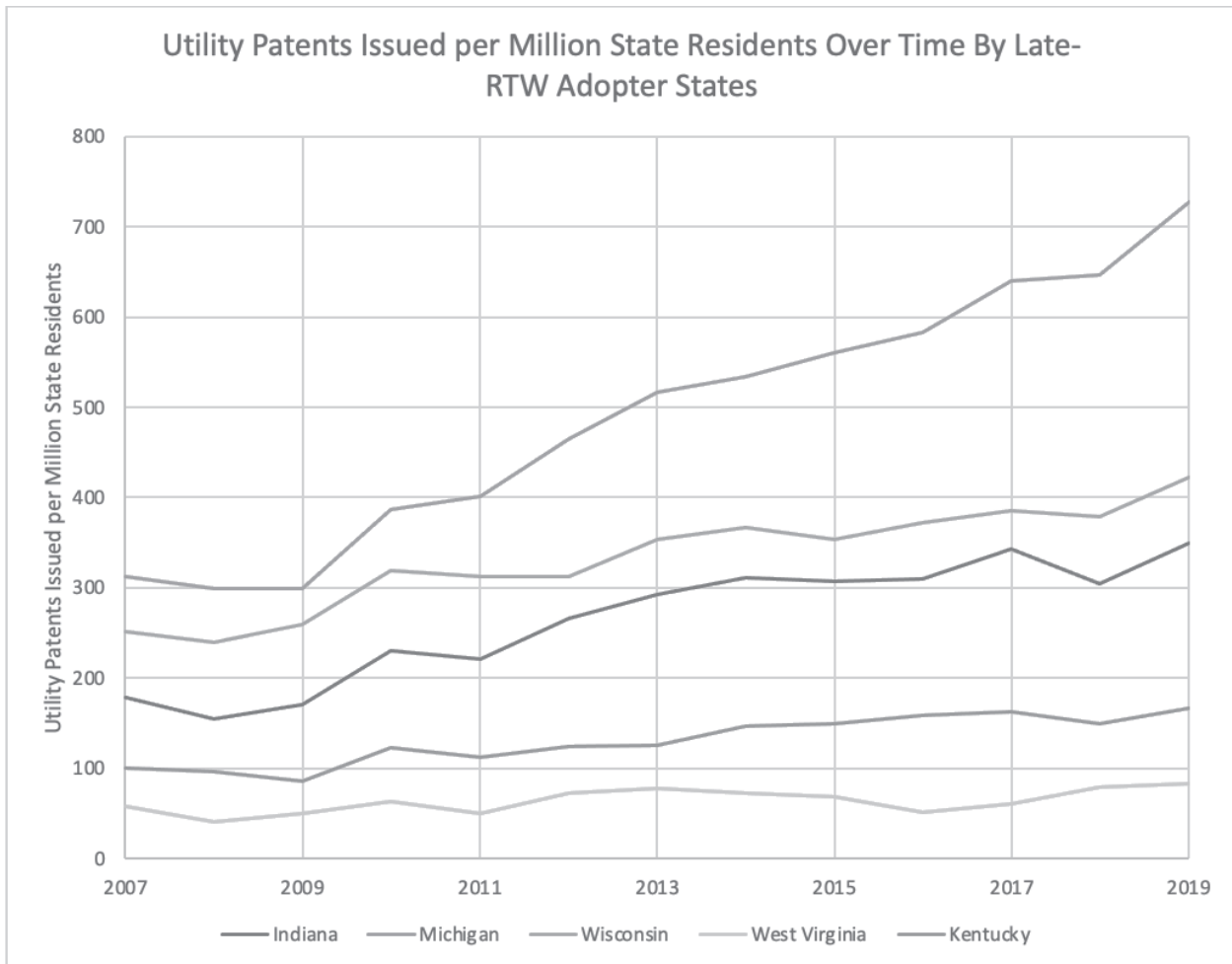


Figure 2. Utility patents issued per million state residents for late-adopter RTW states, 2007 - 2019

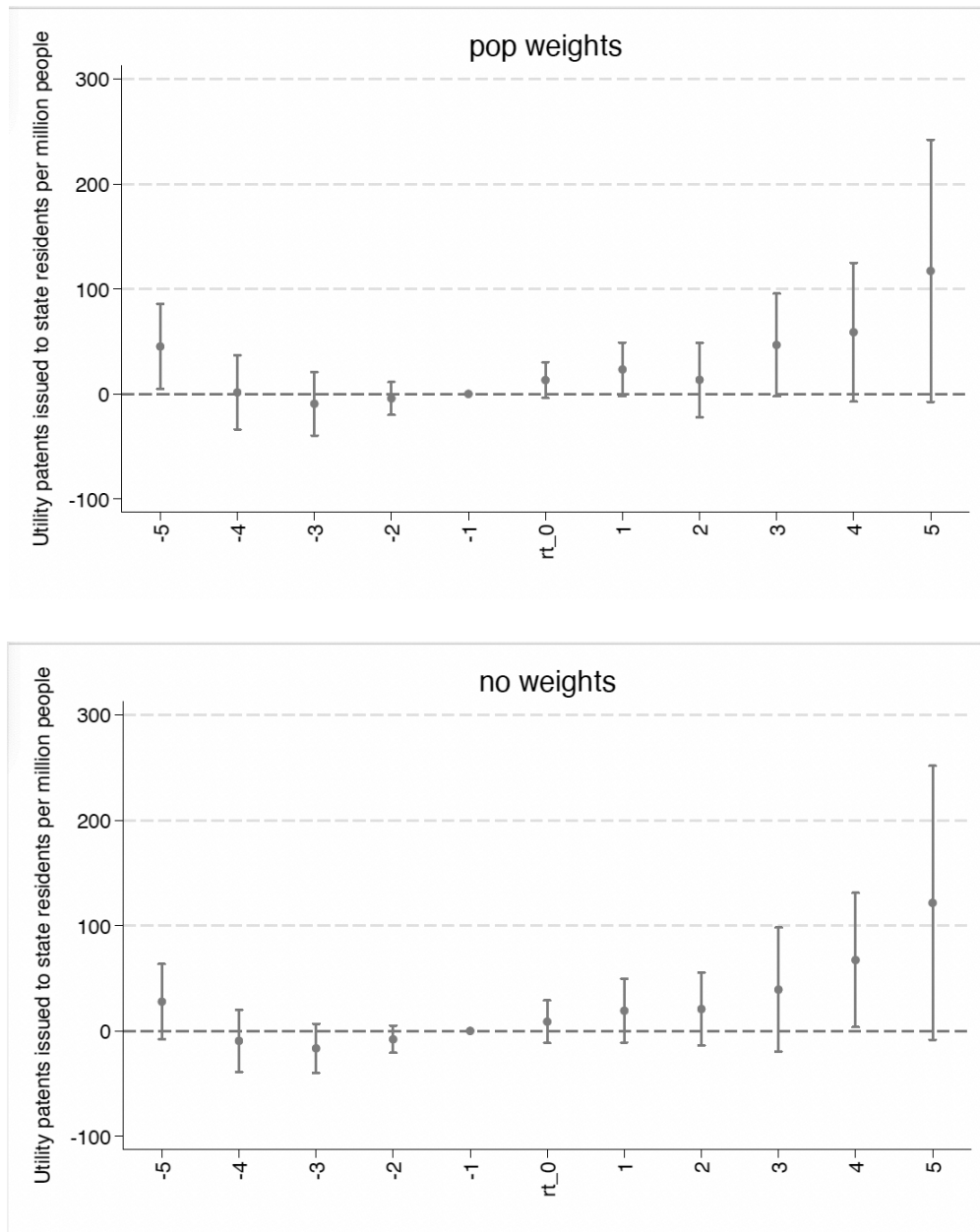


Figure 3a. The effect of passing RTW laws on the utility patents issued to state residents per million, π_k , is seen above. By the 5th year of passing a RTW law, utility patents issued to state residents per million by increases by 110. The top figure weighs the estimate by state population, and the bottom figure does not.

The results are borderline statistically significant.

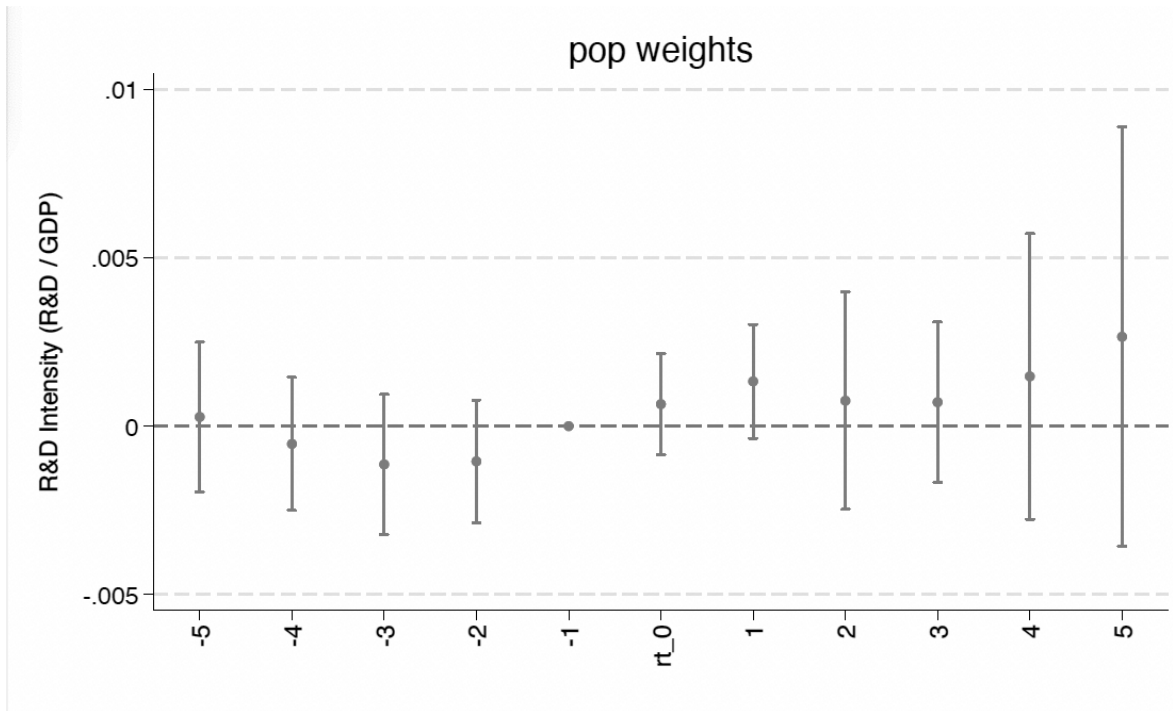


Figure 3b. The effect of passing RTW laws on R&D intensity (Total R&D Spending / GDP), π_k , is seen above. According to these estimates, by the 5th year of passing a RTW law, a state will increase its R&D intensity by around 0.003, though the results are not statistically significant.

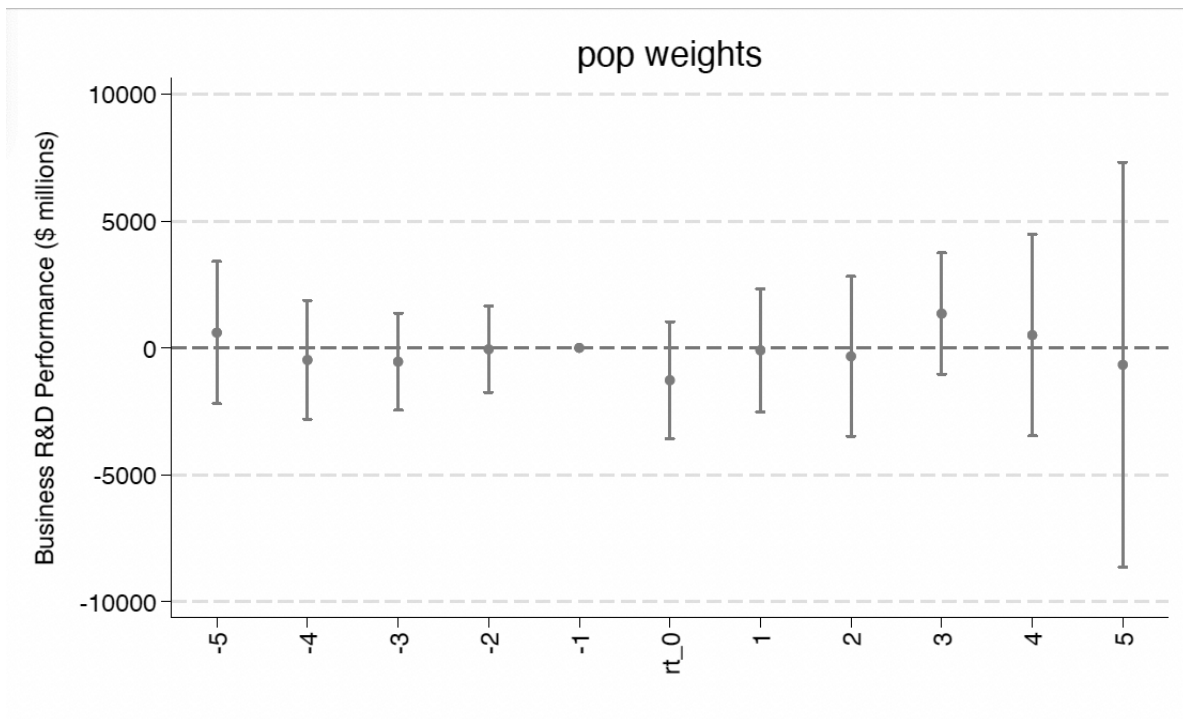


Figure 3c. The effect of passing RTW laws on total business R&D spending, π_k , is seen above. It is clear that the estimates are not statistically significant— in fact, they are not even confident in direction. Thus, we can comfortably say that passing RTW laws does not impact business R&D performance from this model.

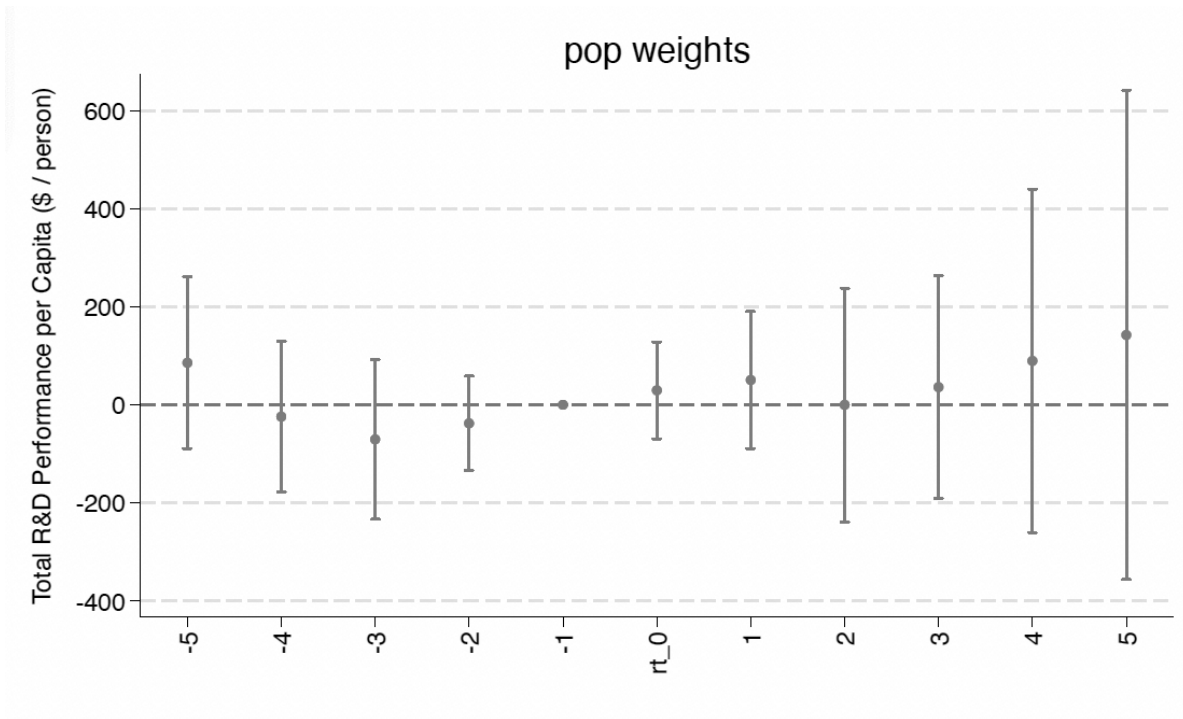


Figure 3d. The effect of passing RTW laws on R&D per capita (Total R&D Spending / population), π_k , is seen above. According to these estimates, by the 5th year of passing a RTW law, a state will increase its R&D spending per capita (\$ / person) by around 180, though the results are not statistically significant.

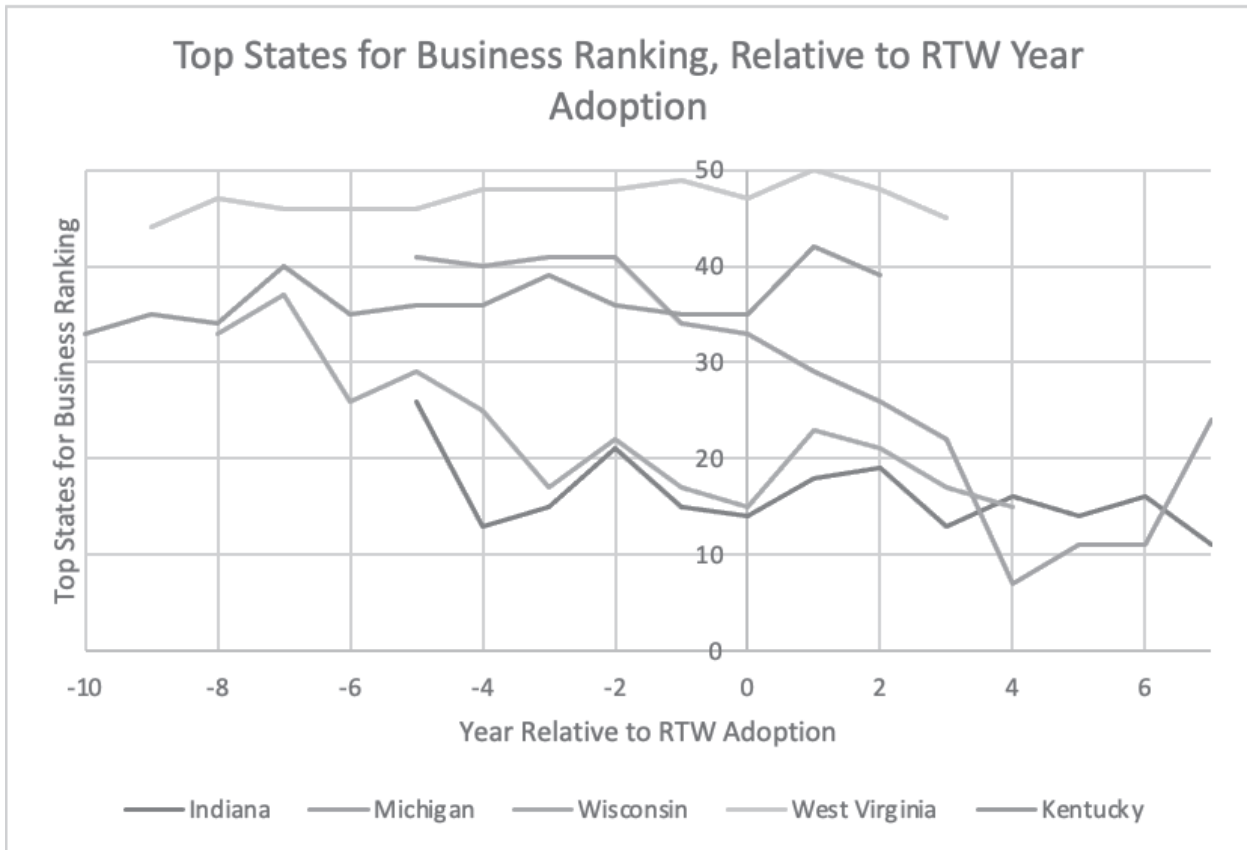


Figure 4. Top states for business ranking relative to RTW year adoption for Late Adopter states, according to CNBC. Note that ranking #1 represents the best state to do business in for that year, while #50 is the worst.

APPENDIX A

Variable name	Storage type	Display format	Value label	Variable label
st	str2	%9s		st
year	int	%10.0g		year
YearSwitch	int	%10.0g		Year Switch
RTWAdoption	byte	%10.0g		RTW Adoption
GDP	double	%10.0g		GDP (\$billions)
pop	double	%10.0g		Population (in thousands)
busRD	long	%10.0g		Business R&D Performance (\$ millions)
totalRD	long	%10.0g		Total R&D Performance ()
totalPats	long	%10.0g		Utility patents issued to state residents
RDperCapita	double	%10.0g		Total R&D Performance per Capita (\$ / person)
patsPerPeople	double	%10.0g		Utility patents issued to state residents per million people
RDIntensityRD~P	double	%10.0g		R&D Intensity (R&D / GDP)
state	long	%8.0g	state	st
relative_year	float	%9.0g		
rt_0	float	%9.0g		
rt_lag1	float	%9.0g		1
rt_lag2	float	%9.0g		2
rt_lag3	float	%9.0g		3
rt_lag4	float	%9.0g		4
rt_lag5_plus	float	%9.0g		5
rt_lead5_minus	float	%9.0g		-5
rt_lead4	float	%9.0g		-4
rt_lead3	float	%9.0g		-3
rt_lead2	float	%9.0g		-2
rt_lead1	float	%9.0g		-1

APPENDIX B

Variable	Obs	Mean	Std. dev.	Min	Max
st	0				
year	676	2013	3.744428	2007	2019
YearSwitch	676	973.1538	1032.621	-99	2017
RTWAdoption	676	.464497	.4991072	0	1
GDP	676	329.8889	417.5552	24.7977	3042.694
pop	676	6143.365	6910.347	523.414	39437.61
busRD	618	5837.15	12654.47	25	132473
totalRD	577	7782.049	14166.98	104	135129
totalPats	655	2257.095	4729.045	14	42012
RDperCapita	671	1278.518	1210	35.92773	10076.7
patsPerPeo~e	676	304.9234	235.1162	3.528859	1226.127
RDIntensit~P	671	.0222037	.0152641	.0015134	.0746269
state	676	26.5	15.01944	1	52
relative_y~r	676	-23.13462	75.55706	-99	75
rt_0	676	.0073964	.0857474	0	1
rt_lag1	676	.0073964	.0857474	0	1
rt_lag2	676	.0073964	.0857474	0	1
rt_lag3	676	.0059172	.0767519	0	1
rt_lag4	676	.0044379	.0665186	0	1
rt_lag5_plus	676	.4319527	.4957147	0	1
rt_lead5_m~s	676	.5059172	.5003352	0	1
rt_lead4	676	.0073964	.0857474	0	1
rt_lead3	676	.0073964	.0857474	0	1
rt_lead2	676	.0073964	.0857474	0	1
rt_lead1	676	0	0	0	0

From Guac to Glock: Exploring the Consequences of Avocado Production on Organized Crime in Mexico

Brian Liu

April 23, 2023

1 Introduction

In recent years, the popularity of avocados has soared to new heights. The humble fruit has gone from a relatively unknown ingredient to a ubiquitous presence in many households and restaurants around the world. Despite being a traditional ingredient in Mexican cuisine for centuries, avocados have recently gained a reputation as a “superfood,” with various health benefits touted by bloggers, social media influencers, and nutritionists. From avocado toast to guacamole, the versatility of this fruit has made it a beloved ingredient among foodies and health enthusiasts alike. However, the impact of the avocado craze extends far beyond social media and brunch menus. In particular, Mexico, the largest exporter of avocados to the United States, has seen dramatic effects from the increased demand on its agricultural sector and perhaps more surprisingly, potential impacts on violent crime rates.

On February 11, 2022, the USDA temporarily banned avocado imports from Mexico after a sanitary inspector was “verbally threatened” in Michoacan, choking the supply of avocados before the fruit’s biggest event: the Super Bowl (Creswell, 2022). While the ban lasted only nine days, it was enough to disgruntle unhappy football fans when stores and restaurants ran out of guacamole.¹ But the ban also represented one of many issues in the long history of Mexican cartel influence on licit industries. In an attempt to expand and diversify their

¹Approximately 90% of U.S. avocado consumption is supplied from Mexico

portfolios, cartels have latched on to lucrative crop markets, siphoning money from farmers most commonly through extortion and homicides/homicide threats. Between 2009 and 2013, cartels expropriated \$770 million from avocado farmers, approximately 13% of total revenue. In extreme cases, cartels have even induced supply shocks to jack up crop prices, allowing them to expropriate higher margins. In 2022, lime prices skyrocketed over 300% after cartels instructed lime farmers to only work two days a week (Simon 2022).

While cartels have carved out spheres of influence in the avocado industry, it is less clear whether their entries into the market have increased crime rates. A quick google search provides no shortage of articles claiming that avocados, also known as “green gold,” have become the next big conflict commodity in Mexico. The link between avocado producers and cartels is undeniable, but can we say that the explosion in avocado production has increased crime compared to the counterfactual, a world where avocado farmers would have been farming other crops or potentially working for cartels? There are a few competing theories. The U.S. has had a long history of restricting imports from Mexico due to pest and disease concerns and only recently allowed heavily-monitored importing via law changes in 1997 and 2016. These law changes provided opportunities for farmers to switch from illicit crop markets to the avocado market, potentially decreasing the intra-regional influence of cartels². In addition, historical evidence suggests that cartels have various strategies when entering new markets, ranging from territorial fortification and defense (which would lead to decreased violence) to increased aggression in an attempt to stamp out competing gangs (which would lead to increased violence).

Thus, the purpose of this paper is to determine whether the increase in demand for avocados led to an increase in cartel-related violence in Mexico. Using a difference-in-differences (DID) and fixed-effects design centered around the 2016 USDA lifting of import restrictions, we demonstrate that treated municipalities³ experienced statistically significantly lower rates

²While cartels extort licit markets, they have less relative control over their production and transportation compared to illicit crops since licit crops can be legally exported to the U.S.

³Municipalities that started producing avocados for export to the U.S. in 2016 because they were not allowed to previously.

of cartel-related crime and homicide than control municipalities. As the demand for avocados continues to increase year on year, our results have important policy implications for the well-being and safety of avocado producers and other agricultural workers in Mexico.

The structure of the paper is as follows. We first provide background on the history of avocado imports to the U.S. and the influence of cartels on avocado-producing regions. Then, we discuss related literature on cartel presence and impact on licit markets. Next, we explain our data and design a DID model to investigate the difference in crime rate between treated and untreated municipalities. Finally, we analyze our results and conclude with broader implications and future steps.

2 Background

2.1 Avocado production and imports

The North American Free Trade Agreement (NAFTA) in 1994 promised to lift an import ban on Mexican avocados dating back to 1914. The purpose of the ban was to protect American domestic producers from invasive pests.⁴ While the blanket ban was lifted, the USDA's Animal and Plant Health Inspection Service (APHIS) continued to demand strict regulations on all steps in the avocado production process and only allowed imports from Mexico's most developed avocado production state, Michoacan. In partnership with Mexico's Secretariat of Agriculture, Livestock, Rural Development, Fisheries, and Food (SAGARPA), APHIS imposed requirements for inspection and cleaning at packing centers, criteria for shipment rejection, detailed documentation and storage before transportation, and timely pest inspections at numerous supply chain choke points (SAGARPA, 2005).

As of a result of these guidelines, until 2015, only 24 municipalities in the state of Michoacan were approved to ship avocados to the U.S.⁵ These were the municipalities that were

⁴Five specific pests mentioned in USDA literature are: stem weevils (*Copturus aguacatae*); seed weevils (*Conotrachelus aguacatae*, *Conotrachelus perseae*, and *Heilipus lauri*); and seed moths (*Stenomoma catenifer*).

⁵Acutzio, Tancitaro, Uruapan, Tinguindin, Salvador Escalante, Nuevo Parangaricutiro, Periban de Ramos, Ario, Los Reyes, Apatzingan, Taretan, Tacambaro, Tingambato, Madero, Cotija de la Paz, Eron-

both located in the Trans-Mexican Volcanic Belt⁶ and able to develop avocado collection, packing, and transportation infrastructure in accordance with APHIS guidelines. Nevertheless, the U.S. and Mexico had become each other's largest trading partner with respect to avocados. 80% of Mexican avocado exports were sent to the U.S., and this accounted for approximately 90% of U.S. avocado consumption (USAID, 2014).

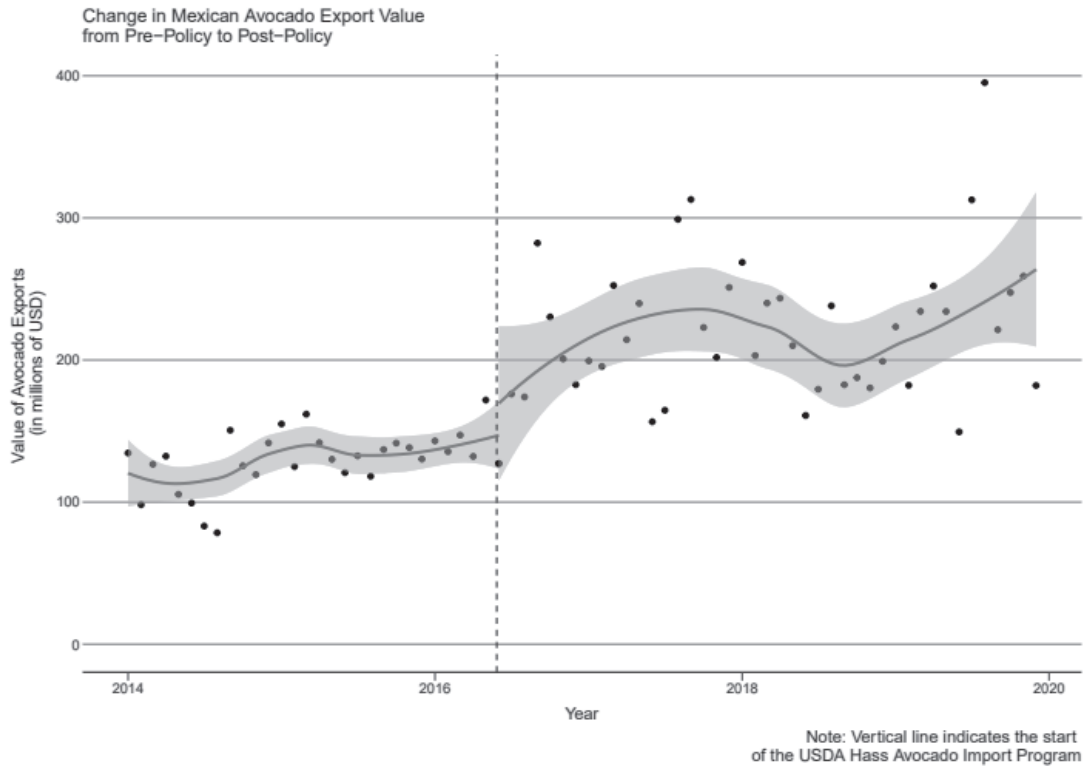


Figure 1: Mexican avocado exports over time

The landscape of production and export changed in 2015 when APHIS announced a revision in avocado policy that would allow all municipalities to export avocados to the U.S. as long as they met the requirements established post-NAFTA (USDA, 2016). The final policy was signed and enacted in June 2016, and the Volcanic Belt states of Jalisco, Mexico, Michoacan, Morelos, and Nayarit consequently saw massive increases in production of avocados, primarily aimed at export to the U.S, over the next few years. Figure 1 shows

⁶Ideal soil and weather conditions for avocado farming.

the impact of the U.S. import law change on Mexican avocado exports.⁷

2.2 Cartels in Michoacan and nearby states

Organized crime groups in Mexico, commonly referred to as cartels, function as highly efficient criminal enterprises that are involved in various illegal activities, including drug trafficking, extortion, and violence. Cartels operate as a decentralized network of independent factions that share common goals and collaborate in the drug trade. They are typically led by a few powerful individuals known as “capos,” who are responsible for coordinating the group’s activities and maintaining control over their territory. Cartels often rely on a complex system of alliances and rivalries with each other, which can shift over time as new factions emerge or existing ones are weakened. They frequently engage in violent conflicts with each other over control of lucrative drug trafficking routes and territory and use violence to intimidate and control local communities and government officials.

Cartels have increasingly entered licit markets as a way to diversify their revenue streams, launder money, and legitimize their businesses. This allows them to operate more openly and avoid the scrutiny of law enforcement agencies. Regarding agricultural markets and avocados in particular, cartels have infiltrated the industry through various points in the supply chain and rely on threats of violence to ensure local cooperation with their illegal activities. The presence of cartels in Michoacan dates back to the 1980s. A common theme of violent displacement of established cartels by new ones proceeded until 2006. Los Valencias maintained a drug empire involved in trafficking cocaine, marijuana, and methamphetamines in the 1980s and 1990s until violently expelled by Los Zetas in 1999. Los Zetas was the first cartel to expand extortion to licit markets, including avocados (Ornelas, 2018). They were ultimately uprooted by La Familia Michoacana who provided temporary protection to agricultural producers until defeating Los Zetas.⁸

La Familia Michoacana was ultimately destroyed when Felipe Calderon became president

⁷Created from U.S. Global Agricultural Trade System query.

⁸Afterwards, La Familia Michoacana charged farmers for their “protective services.”

in 2006 and started the Mexican Drug War. Pressured and bolstered by the U.S. Federal Government, Calderon sought to disrupt and dismantle the control of cartels through military force. Supplied with U.S. equipment, Calderon sent tens of thousands of federal police and militia into cartel-controlled territories with the intent of a blanket crackdown on criminal activity. Ultimately, the impact of the Mexican Drug War is controversial due to the widespread violence between cartels and government forces that spilled over to residential areas, catching civilians in crossfire and destabilizing political infrastructure.⁹

The Mexican Drug War exacerbated the influence of cartels on licit agricultural markets. Since the pathways and supply chains in drug production were destabilized, cartels sought to establish themselves in legal businesses. Avocado producers provided an attractive market because of various reasons according to Ornelas (2018):

“1) businesses in traditional sectors of the economy with a high degree of territorial specificity; 2) a relative small size of firms; 3) a relatively low technological level; and 4) a region where the public sector is relatively large and legal institutions are weak”

Indeed, the avocado industry in Michoacan checked all four boxes, and following the fragmentation of cartels during the Drug War, various groups fought for control of the avocado market. The Knights Templar and Los Viagras emerged as the dominant cartels in Michoacan during the 2010s, and proceeded to engage in maximally predatory behaviors towards the avocado industry as well as inter-cartel violence. Moncada (2021) notes that the Knights Templar efficiently extorted almost every checkpoint in the supply chain of avocado exports, including plant nurseries, orchards, packing houses, transport checkpoints. In addition, farmers were charged for simply owning land within the Templar’s territory and, in some cases, forced to sign ownership of land away to cartel members. Such demands on the supply chain and at the individual farmer level were enforced primarily through extortion, kidnapping, and murder.

⁹By the end of Calderon’s term in 2012, over 120,000 people were killed as a result of the War on Drugs (Booth 2012).

Most recently, the lifting of the restriction of the U.S. import ban has brought new actors to the avocado producing states, including Jalisco New Generation who is thought to be responsible for the threatening of the U.S. inspector in February 2022 (that led to the 9 day ban) when he rejected a truck of cartel-controlled avocados. The presence of multiple cartels within these regions set the stage for our analysis into how the increase in avocado exports has affected crime and violence. Figure 2 shows a visual history of cartel influence in Michoacan and neighboring states.¹⁰



Figure 2: History of Michoacan cartels

3 Related Literature

As mentioned in Ornelas (2018), the relative weakness of Mexican agricultural institutions is conducive to the large influence of organized crime. Two main theories exist on the effect of market shocks on violence due to organized crime. The first is the rapacity hypothesis, which claims that positive market shocks increase violence because rising commodity demand/prices increases potential gain from exploitation of said market, leading to increased conflict over its control. In simpler terms, the pie is bigger, so people are more willing to fight over control of the pie. The second is the opportunity cost hypothesis, which states that positive market shocks decrease violence because higher commodity demand/prices generates more employment opportunity, raising the opportunity cost of engaging in violent appropriation. In simpler terms, because the pie is bigger, more people are satisfied with their share

¹⁰Obtained from Erickson 2020.

of the pie and feel less inclined to fight over the rest of the pie.¹¹

In the context of agricultural industry, more research leans towards the opportunity cost hypothesis because of the labor-intensive nature of agricultural production. Dube and Vargas (2013) supported this hypothesis by studying the difference in crime rates in Colombia during the Civil War in response to exogenous price shocks to labor-intensive commodities and capital-intensive commodities. They found that market shocks were negatively correlated to violence in labor-intensive commodities such as coffee, sugar, banana, palm, and tobacco, while market shocks were positively correlated to violence in capital-intensive commodities such as oil, coal, and gold. Blair (2021) corroborates this claim through meta-analysis of 46 natural experiments that estimate the causal effect of commodity price changes on armed civil conflict. Dal Bo (2011) further supports this hypothesis and provides an equilibrium model that determines how commodity price shocks would influence violence. These papers demonstrate that the two hypotheses may both be applicable depending on the type of market, where the opportunity cost effect trumps the rapacity effect in labor-intensive markets and vice-versa. Therefore, we would expect that the crime rates in Mexican avocado-producing regions would decrease relative to other regions after the positive exogenous shock induced by the U.S. import restriction lifting.

Mejia and Restrepo (2015) suggest other reasons for why licit market shocks are negatively correlated with violence. Specifically, they argue the state plays a larger role in reducing violence in licit markets because licit crop farmers can use state authority to enforce property contracts. In addition, the state has more incentive to regulate organized criminal activity since it can collect taxes on licit crop production. Thus, if more actors enter the licit market from an illicit market as a result of a positive shock, organized crime loses influence in the form of increased state intervention. Such a scenario could feasibly play out in regions that were allowed to produce avocados starting in 2016.

¹¹Note that this analogy is not perfectly correct due to the complexity of the issues surrounding market shocks and cartels.

Regarding the Mexican avocado industry, Roett (2020) argues that positive price shocks in avocado imports to the U.S. caused an increase in violence in avocado-producing regions. However, we believe the model used can be improved since Roett’s difference-in-difference treatment indicator is staggered and proportional to the log import value of avocados per month. Roett does not use the 2016 USDA import restriction lifting as a treatment, rather indicating treatment as whenever a municipality started producing avocados. As Baker (2022) states, it is difficult to properly assess potential biases in staggered difference-in-differences estimates. For example, there were municipalities that started producing avocados between 2010 and 2016 for domestic Mexican consumption or export to countries other than the U.S. However, it is unlikely that these farmers experienced the same extortion as exporters to the U.S. since their avocados do not pass through the various supply chain check points (collection, transportation, border inspection, etc.) that avocados bound for the U.S. pass through. Thus, the extortion that occurs at these checkpoints by cartels such as the Knights Templar applies unevenly to the data Roett sampled.

4 Data

Data was taken from a variety of national agencies of the United States and Mexico. All cleaning, scraping, and preparation was done in python. Robust regressions are calculated using Stata. We include two primary outcome variables as proxies for cartel violence: crimes per capita (`crimes_pc`) and homicides per capita (`homicides_pc`). Homicides per capita is the number of intentional homicides per 100,000 people and is recorded at the municipal-monthly level. Crime per capita is homicides per capita plus the number of extortions per 100,000 people and is recorded at the municipal-monthly level. While we would have liked to include additional types of crime related with cartel violence such as torture and kidnappings, this data was not available in the designated time frame of 2011-2019. This data was collected from Mexico’s Executive Secretariat of the National Public Security System and (SESSNP) the National Institute of Statistics and Geography (INEGI).¹²

¹²Crime data available at <https://www.gob.mx/sesnsp>, population data obtained from <https://www.inegi.org.mx/app/scitel/>.

The primary regressor is the treatment variable in the difference-in-differences model. For control municipalities, this variable is always 0 since no avocados are produced between 2011 and 2019. If a municipality starts producing avocados after the import change or starts producing significantly more after 2016 than before, we deem it a treated municipality. Significantly more is defined as the average production post-2016 being 2.2 times more than the average production pre-2016. This threshold was estimated through the USDA GATS monthly avocado import data from 2011-2019. The threshold ensures that the treated municipalities are indeed increasing avocado production because of the U.S. import law change, not just due to global increased demand (which is growing albeit at a slower pace than U.S. consumption demand). As a sanity check, we confirm that all 24 municipalities that produced avocados for export to the U.S. before 2016 are not included in the treatment group. The treatment municipalities have treatment variable equal to 1 times the log mean production value from 2016-2019. We believe this gives us a more accurate treatment effect since the increase or decrease in crime should vary more depending on the potential productivity of the municipality with respect to avocados. All data is taken from the Mexican Agricultural and Fisheries Information Service (SIAP) and the USDA Global Agricultural Trade System (GATS).¹³

With these variables, we create panel data from 2011-2019 at the municipality-month level, dropping municipalities with more than 12 missing values (1 year of data). In the end, we have 189 municipalities from the five states, Jalisco, Mexico, Michoacan, Morelos, and Nayarit, that produced large amounts of avocado after 2016, 57 treated and 132 control. A heat map of avocado production by state is available in the appendix. All municipalities lie within the Trans-Mexican Volcanic Belt and have similar weather patterns and approximately Gaussian distributed log populations. 189 municipalities over 9 years, or 108 months, creates 20412 observations.

¹³Data available at <https://www.gob.mx/siap/acciones-y-programas/produccion-agricola-33119> and <https://apps.fas.usda.gov/gats/>.

Table 1: Variable Descriptions and Statistics

variable name	mean	std dev.	min	max
homicides	1.706	4.04	0	56
homicides_pc	1.312	2.804	0	52.514
crimes	2.603	6.096	0	73
crimes_pc	1.793	3.225	0	57.511
treat	1.856	3.860	0	13.167
pop	137924	250760	4862	1645352
avo_prod	16971	42223	5	226850
N	20412			

5 Method

We employ a difference-in-differences model to determine the effect of the U.S. import law change on cartel homicides and crimes:

$$(1) \quad \text{crimes_pc}_{it} = \beta(\text{treat}_{it} \cdot v_i) + \delta_t + \alpha_i + \epsilon_{it}$$

$$(2) \quad \text{homicides_pc}_{it} = \beta(\text{treat}_{it} \cdot v_i) + \delta_t + \alpha_i + \epsilon_{it}$$

where homicides_pc_{it} and crimes_pc_{it} are the number of homicides and cartel-related crimes (homicides plus extortions) per 100,000 people for municipality i and month t , respectively. treat_{it} is the DID treatment indicator, v_i is log average production, and β estimates the treatment effect of the lifting of the U.S. import restrictions in 2016 on homicides and cartel-related crimes. δ_t represents time fixed-effects, α_i represents municipality fixed-effects, and ϵ_{it} is the error term.

Using a difference-in-differences model allows us to control for unobservable factors that may bias our regression had we simply done fixed-effects. Because we are concerned with municipalities all from similar regions within 200 miles of each other that exhibit similar

relevant characteristics such as weather patterns, presence of agricultural, cartel presence, etc., we argue that the municipalities provide balanced treatment and control groups for the DID model.

Table 2: DID regression on crime per capita

	(1)	(2)
	crimes_pc	homicides_pc
treat	-0.0485*	-0.0408*
	(0.0197)	(0.0172)
constant	1.052***	0.976***
	(0.185)	(0.179)
Time fixed-effects?	Yes	Yes
Location fixed-effects?	Yes	Yes
R^2	0.0293	0.0192
N	20412	20412

Robust standard errors in parentheses

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

The coefficients in the row treat indicate the average treatment effect of the import law change on crime and homicide rates.

6 Results

Table 2 presents the results of our DID regression model on equations (1) and (2). The DID yields the subtraction between the difference in homicide/crime rate in treated municipalities and the difference in homicide/crime rate in control municipalities before and after the import law change in 2016. This eliminates potential selection bias among municipalities and allows us to obtain a quasi-experimental estimate of the average treatment effect of the law change on cartel-related crime. On average, a 1% increase in mean avocado production value led to a decrease of 0.05 crimes per 100,000 people and 0.04 intentional homicides per 100,000 people, significant at the $\alpha = 0.05$ level. For the median municipality whose production value increased by 28%, this accounts for a decrease in 1.36 crimes per 100,000 people and 1.14 homicides per 100,000 people. This is significant relative to the average number of crimes and homicides per 100,000 people, which are 1.79 and 1.71 respectively.

Figures 3 and 4 in the appendix give us a visualization of the treatment effect of the import

law change on crimes per capita and homicides per capita. Four data points are plotted on each graph, illustrating the average crimes/homicides per capita among treated and untreated municipalities before and after the law change in 2016. The dashed line represents the counterfactual where the law change never happened and treatment regions continued to not produce avocados. We can clearly see that the treatment had a significant effect on reducing crimes/homicides in treated municipalities. Had the law not been implemented, under the parallel trends assumption, these municipalities would have experienced more incidents of crime.

7 Conclusion

Through a twoway fixed-effects difference-in-differences model, we demonstrate that the increase in avocado production caused by the 2016 U.S. import law change had a statistically significant decrease on homicides and cartel-related crimes in treated municipalities. This result makes sense in the broader context of relevant literature. Since avocado production in Mexico is a labor-intensive commodity that was subject to an exogenous positive market shock in 2016, we expected homicide and crime rates to decrease as a result of the opportunity cost hypothesis. When profits are high from lucrative licit crops such as avocados, it makes sense for farmers to switch from producing illicit crops to avocados, weakening the presence of cartels and decreasing the amount of crimes. Moreover, the boom of the avocado industry only further incentivizes the Mexican government to crackdown on cartel influence within the industry. As the value of avocado imports nears \$3 billion, protecting farmers and workers along the export supply chain provides more tax revenue for the government and will hopefully prevent any supply shocks such as the February 2022 inspector threat incident.

Additionally, from the perspective of cartels, we argue that the avocado boom encouraged cartels to spend resources on fortifying their existing territory, rather than fighting other cartels to expand their control. Given the long history of cartels in regions such as Michoacan and the presence of at least 3 major cartels in the region during the 2010s, we argue that the cartels were long entrenched in these municipalities before the import law

change occurred in 2016. Erickson (2020) corroborates this statement. Given that municipalities were under the control of various cartels before the law change, it made more sense for cartels to secure the avocado production within their own territories in the years following the change. While media and anecdotes from farmers suggest that avocados have become the next conflict commodity, our analysis demonstrates that, on average, the policy caused a reduction in crime.

Nonetheless, future steps may include investigation of these anecdotes. It may be the case that violence from cartels against avocado farmers has increased (matching anecdotal evidence), yet has been outweighed in our research by the decrease in inter-cartel violence. We hope that the Mexican government in the future may be able to provide more detailed breakdowns of crimes specific to agricultural industries. We also hope to explore how the COVID-19 pandemic affected the relative strengths of cartels and government institutions in avocado-producing regions, though this may be a difficult task due to the missing data reports during 2020 and 2021. Given that the import law change occurred 7 years ago, cartels may have resumed violent attacks on each other after building sufficient fortifications on their territories. In this way, the effect of the rapacity hypothesis may override the opportunity cost effect because of the vast array of exploitations the cartels employ on all aspects of the avocado export industry. Finally, we hope to explore the environmental consequences of the explosive growth in demand for avocados. As a water and nutrient intensive crop, the never-ending increase in avocado production has led to a strain on natural resources such as freshwater sources. The importance of regulating violence and crime related to cartels matters only insofar as people's basic needs are met.

8 References

Animal and Plant Health Inspection Service, USDA (2016). Mexican Hass Avocado Import Program. Federal Register 81 (103), 33581–33588.

Asmann, Parker. “Powerful Mexico Crime Groups Grew by Extorting Avocado Trade: Report.” *InSight Crime*, Nov 19, 2017. <https://www.insightcrime.org/news/brief/powerful-mexico-crime-groups-got-their-startextorting-avocado-trade-report/>

Baker, Andrew, Larcker, David, Wang, Charles. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, Volume 144, Issue 2, 2022, Pages 370-395, ISSN 0304-405X, <https://doi.org/10.1016/j.jfineco.2022.01.004>.

Blair, G., Christensen, D., Rudkin, A. (2021). Do Commodity Price Shocks Cause Armed Conflict? A Meta-Analysis of Natural Experiments. *American Political Science Review*, 115(2), 709-716. doi:10.1017/S0003055420000957

Bó, E. D., Bó, P. D. (2011). Workers, Warriors, and Criminals: Social Conflict in General Equilibrium. *Journal of the European Economic Association*, 9(4), 646–677. <http://www.jstor.org/stable/25836084>

Booth, William. “Mexico’s crime wave has left about 25,000 missing, government documents show”. *The Washington Post*, Nov 30, 2012.

Creswell, Julie. “U.S. Temporarily Bans Avocados From Mexico, Citing Threat.” *New York Times*, Feb 15, 2022, <https://www.nytimes.com/2022/02/15/business/us-mexico-avocado-ban.html>.

Dube, Oeindrila, Vargas, Juan. Commodity Price Shocks and Civil Conflict: Evidence from Colombia, *The Review of Economic Studies*, Volume 80, Issue 4, October 2013, Pages

1384–1421, <https://doi.org/10.1093/restud/rdt009>

Erickson and Owen (2020). Cartel Violence over Licit Industries in Mexico. University of Washington Political Economy Forum.

Moncada, E. (2021). *Resisting Extortion: Victims, Criminals and States in Latin America*. New York, NY: Cambridge University Press.

Ornelas, R. G. (2018). *Organized Crime in Michoacan: Rent-Seeking Activities in the Avocado Export Market: Organized Crime in Michoacan*. *Politics Policy* 46 (5), 759–789.

Roett, Katie, "Green Gold: Avocado price shocks and violence in Mexico" (2020). Master's Theses. 1316. <https://repository.usfca.edu/thes/1316>

SAGARPA, E. U. M. (2005). NORMA Oficial Mexicana NOM-066-FITO-2002, Especificaciones para el manejo fitosanitario y movilizaci´on del aguacate. *Diario Oficial de la Federaci´on 2002* (66), 34–46.

Simon, Scott. "Mexican drug cartels are getting into the avocado and lime business." *NPR*, Feb 19, 2022, <https://www.npr.org/2022/02/19/1081948884/mexican-drug-cartels-are-getting-into-the-avocado-and-lime-business>.

USAID (2014). The U.S. Market for Avocado. *U.S. Agency for International Development*, Dec 2014, https://pdf.usaid.gov/pdf_docs/PA00KP28.pdf.

USDA (2023). U.S. Monthly Avocado Exports from Mexico. *Global Agricultural Trade System*, Mar 2023.

9 Appendix

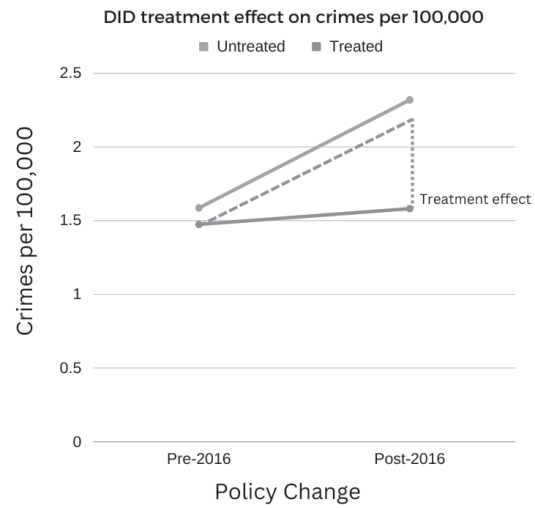


Figure 3: Crimes per capita visualization

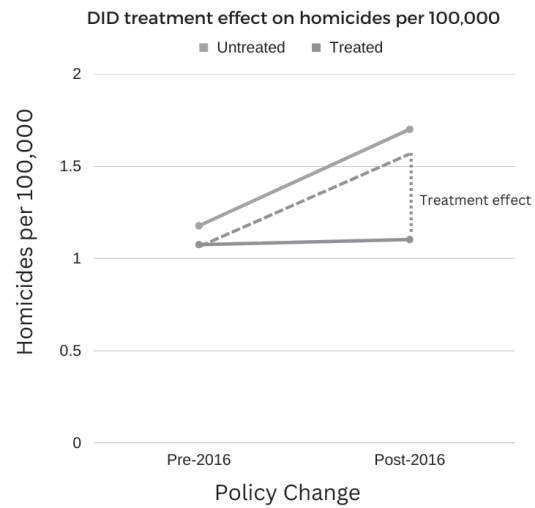


Figure 4: Homicides per capita visualization

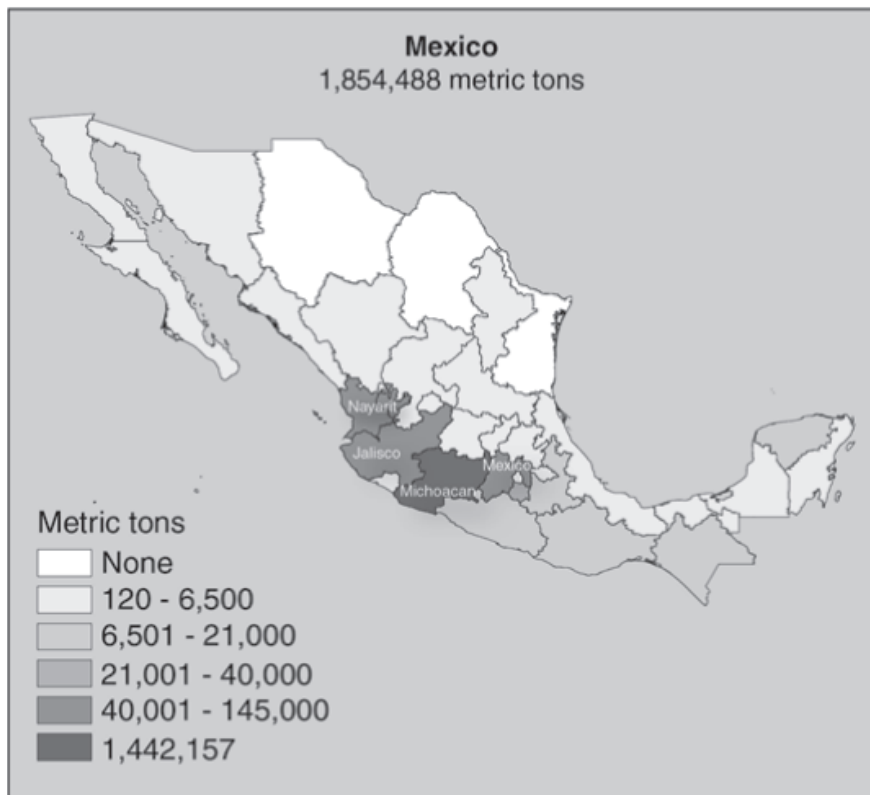


Figure 5: Avocado production in 2018 by state

Assessing the impact of police officer line-of-duty deaths and force size on civilian arrests

Srinidhi Narayanan

March 28, 2024

Abstract

The death of a police officer in the line of duty can be traumatic for the fallen officer's colleagues, and may elicit a response with downstream consequences — does the death of an officer in the line of duty affect how the force polices in the aftermath? This paper investigates this question by assessing the impact of an officer death on arrest numbers in the officer's municipality for up to nine months post-death with a difference-in-differences approach. Furthermore, the paper investigates the role of police force size in this impact with a triple difference estimator. I find that force size has no significant effect on arrests. I find that the death of an officer in the line of duty has a significant effect in the month of the death, and no significant effects in the nine months post-death. I find that the line of duty death also plays a significant effect as a term in the triple difference regression. This suggests that police force size may be correlated with the death of an officer in the line of duty in subtle ways not reflected in the dataset.

1 Introduction

Policing is a critical social and political issue globally. This is even more the case in the United States, a country which incarcerates disproportionate to its size and in which, compared to other democracies, civilian-officer relationships are unusually strained [1] [2]. The issue of policing intersects with broader themes of civil liberties, state authority, and public trust, and recent high-profile cases of police brutality have sparked nationwide calls for comprehensive police reform. It is therefore imperative that we develop a robust understanding of policing mechanisms and the drivers of police behavior.

The death of a police officer in the line of duty (LODD - Line Of Duty Death) is a shock that is oftentimes traumatic for the fallen officer's colleagues [8], and which may elicit a response in the form of changed policing behavior [4]. If an officer is killed in the line of duty, it is possible that other officers may engage in risk-mitigating behavior in the aftermath, being more reluctant to police or unwilling to interact with the public, causing arrest numbers to decrease. On the other hand, officers might become hyper-vigilant, over-policing and possibly driving arrest numbers upward. A body of work studying this question already exists — most notably [3] — but studies actually vary in their findings, even up to the sign of the impact.

Additionally, it is possible that the size of the affected police force is a contributor to this impact. In a smaller police force, the death of an officer may feel more personal to colleagues and may have an outsize impact on the change in arrest numbers; the size of the force may also indicate the level of support or resources the department experiences from the city, with a larger force benefitting from more support. Support and resources could reduce the impact of the officer death [5].

In summary, my paper will investigate the two-fold question “How does the death of an officer in the line of duty affect future policing behavior? And, does this effect vary based on police force size?” Though the former question has been answered before with the same data source I use in [3], I introduce a data sample restriction that I hypothesize will provide more accurate estimates of the coefficients derived in that paper. My paper also adds a (to my knowledge) previously-unexamined heterogeneity study with the investigation of the effect of police force size.

To this end, I examine FBI datasets with samples from 20,000 municipalities over a five year period from 2013-2017, and my empirical strategy uses a difference-in-differences design that exploits the staggered occurrence of line-of-duty deaths across municipalities and time. The initial analysis finds a significant effect in the exact month of the killing, and no significant treatment effects beyond month 0. I then use a triple-difference estimator [9] to examine the additional impact of force size, also finding no significant effect.

Section 2 of this paper highlights the datasets and data preparation. Section 3 describes the empirical framework and key assumptions of the difference-in-differences analysis; and summarizes and contextualizes key results. Section 4 examines possible future steps.

2 Data

2.1 Datasets

I use FBI datasets, further cleaned, concatenated, and available with more granularity via openICPSR [6] [7]. All data is available by municipality by month for the years 2013 to 2017. I use the Law Enforcement Officers Killed and Assaulted (LEOKA) dataset, and for the arrest numbers, the Uniform Crime Reporting (UCR) dataset. I also make use of employment data that is contained in the LEOKA dataset.

Each dataset initially contains data on around 20,000 municipalities. LEOKA contains columns on the breakdowns of various types of assault; demographic information on the assailant(s) and victim(s); the populations of each municipality in each year; and as stated above, information on the numbers of male and female federal and civilian officers in each year. UCR contains columns pertaining to arrests of different types, weapons involved in those arrests, etc.

2.2 Sample Restriction

It is evident from the range of populations represented that municipalities vary significantly in size; it is reasonable to assume they would vary in many other ways as well. Adding place fixed effects to account for demographic and other differences is one option, but an additional sample restriction could act to further control the analysis. I choose to restrict the data to only municipalities where an officer was at least shot at some point in the five-year period, and consider the treatment to be an officer being killed. With this control, the officer's being killed becomes a quasi-random occurrence, and is thus a natural variation I make use of. After adding the sample restriction, the size of the dataset reduces by about 50 percent to around 10,000 municipalities.

An overview of quantities of interest after applying the sample restriction, is here:

Table 1: Summary Statistics

	Variables			
	Mean	S.D.	Min	Max
Population	33419	100625	0	4007905
Total Employee Officers	61.5	252.3	0	9988
Officers Killed	.0005	.03	0	5
Total Officers Assaulted	.62	4.05	0	566
Total Arrests	21.7	80.9	-55	20000

For all quantities of interest represented above, the standard deviation is much larger than the mean. This is due to some very large municipalities' skewing the data — the presence of these large municipalities is evidenced by the "Max" column.

2.3 Preprocessing

After retrieving the datasets from the openICPSR repositories, I apply some preprocessing before performing the analysis. I add a total assaults variable and an assault indicator variable; and encode the months from January of the first year as a 1 to December of the last year as a 60.

Since I am interested in the time trajectory of officer behavior, I want to view indicators for several months post-treatment, where the treatment is an officer killing in a municipality in a month. I would also like to verify that there is no pre-treatment effect, so I wish to consider pre-treatment months as well. So, I create nine indicator variables to indicate “officer killed 1 month prior”, “officer killed 2 months prior”, . . . , “officer killed nine months prior”. When an officer is killed in a municipality j in month t , the dummy column for “1 month prior” is updated with a 1 in the row corresponding to municipality j , month $t + 1$; and so forth. Likewise, I create 1, 2, and 3 month pre-treatment dummies. The assumption in this creation is that the effect of the death, if any exists, will be apparent in nine months, and that there is no treatment effect after nine months.

I transform the arrests (dependent variable) column to a log of the arrests, since the interpretation of the regressions will then be with respect to percent changes; using raw arrest numbers could yield problems where results are skewed by some municipalities with very large or very small populations.

Finally, since I am interested in police force size, I calculate sizes, measured two ways: the first is a notion of “absolute” force size, and is a 1/0 indicator for large/small police force based on whether the force is larger or smaller than the median force size across all municipalities. I also calculate a measure that represents force size relative to the municipality’s population with the metric $\frac{\text{force employment}}{\text{population of municipality}}$.

The resultant data is a panel dataset with time variable being the numerically encoded month, and place variable being the municipality.

3 Empirical Strategy

I employ a difference in differences strategy for this analysis, making some key assumptions in the process.

The foremost assumption is that of parallel trends: in the absence of a police killing, the arrest trends in untreated and treated places are similar. An earlier paper [3] investigating the question of trends post police-killing conducts several robustness and balance tests that appear to verify this assumption. In the results described below, it also becomes clear that the pre-treatment coefficients do not exhibit trends, which, while not equivalent, is consistent with this assumption.

A simplifying assumption I make is that police force size does not change in response to arrest numbers (no reverse causation). As a simplification to the regression calculations, I use notions of size calculated with data from the first month of 2013. Thus, it is imperative that this assumption be close to true so that the assumed values of the variables are still valid

in the later months of the study.

Related to the previous assumption, it is also important the municipality populations don't change in response to arrest numbers, since one view of force size incorporates the municipality population. Since I study a relatively short time-frame of five years — not long enough for citizens to move in or out in response to crime rates — this appears to be a reasonable assumption.

I then break the overall analysis into three sub-stages, outlined below.

3.1 First stage - No Time Horizon

In the first stage, I investigate the effect of a police killing only on the arrests in that same month. I seek to estimate

$$Y_{it} = \beta_0 + \beta_1 \text{shoot}_{it} + \beta_2 (\text{shoot}_{it} \times \text{killed}_{it}) + I + T + \varepsilon_{it}.$$

A traditional difference-in-differences formulation might look like

$$\beta_0 + \beta_1 \text{shoot}_{it} + \beta_2 \text{Post}_{it} + \beta_3 (\text{Treat}_{it} \times \text{Post}_{it}) + I + T + \varepsilon_{it},$$

however, the Post column and the “shoot” column are equal since we consider the Post column to be simply an indicator for time $t = 0$ in the no-time-horizon case, so the traditional formulation and my estimating equation are equivalent.

Here, I and T are year and municipality fixed effects to account for intrinsic differences in explanatory or response variables by municipality or year.

The results of this first regression are shown below:

Table 2: Impact of LODD on arrests, same month

	Response Variable log(Total Arrests)
Shoot _{it}	0.0512*** (23.33)
Shoot _{it} × Killed _{it}	0.0289 (0.83)
Constant	2.180*** (2447.66)
Observations	330456

t statistics in parentheses

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

The main coefficient of interest from this regression is an insignificant $\beta_2 = 0.0289$. However, the table as a whole suggests that even an officer's being *shot* in the line of duty

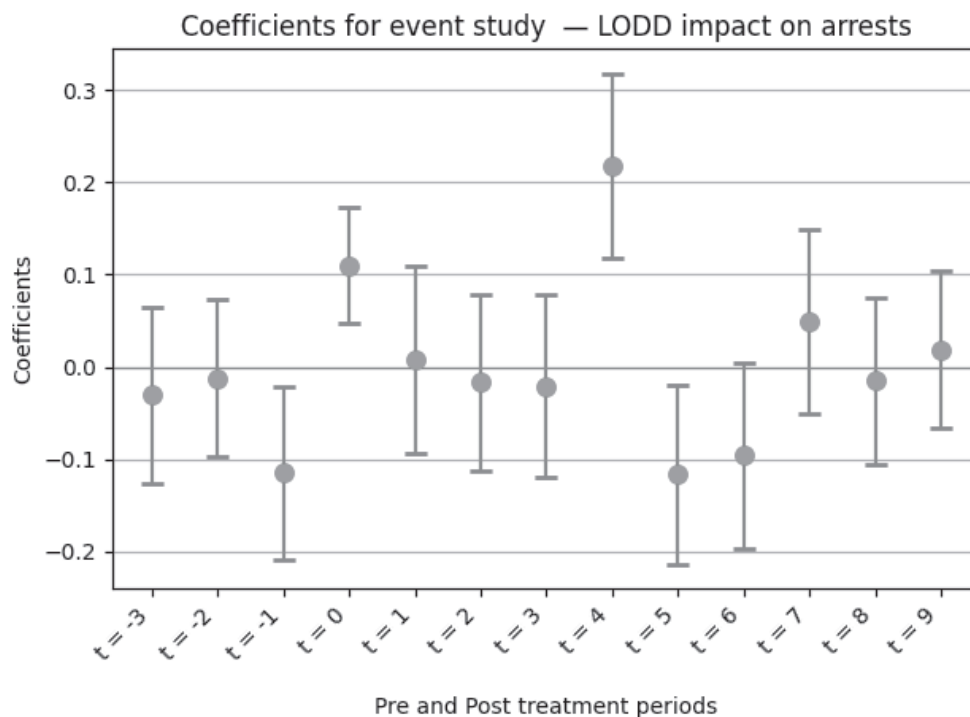
has a significant impact on total arrests in the municipality (in particular, prompting a percent increase of 5.07 in arrest numbers), though additionally being killed may not make as much of an impact.

3.2 Second stage - Event Study

In the second stage, I add post treatment indicators for 1, 2, . . . , 9 months after a police shooting and killing since it is possible that the effect of an officer death is not restricted to the month of treatment. In this stage, if there was a shooting in month t in municipality i , the dummy $Post_k$ takes on value 1 in month $t + k$ in municipality i ; I also add pre-treatment indicators to confirm the absence of pre-trends or an anticipation-like effect. Note that the interaction term coefficients represent the effects relative to untreated (shooting but no killing) baselines, and are thus the coefficients of interest. Taken together, the estimating equation becomes

$$Y_{it} = \beta_0 + \sum_{j=-3}^9 \beta_{Post,j} Post_{j,it} + \sum_{j=-3}^9 \beta_{PT,j} (killed_{it} \times Post_{j,it}) + I + T + \varepsilon_{it}.$$

The results of this event study are shown below, with bars indicating confidence intervals for coefficient values; the complete table of values, and graph of $\beta_{Post,j}$ baseline coefficients is given in the appendix. The coefficients are not normalized in this graph.



Firstly, the graph confirms an absence of pre-trends, since there are no patterns in pre-treatment coefficients; this affirms the validity of the parallel trends assumption, as there appears to be no systemic pattern of difference in the response variable among the treated municipalities prior to the introduction of the treatment. There is a significant coefficient at $t = -1$, but this can be reasonably attributed to random variation. The apparent lack of pre-trends is also a positive sign for the robustness of the treatment effect estimation, as pre-trends may bias the coefficient estimates in the post-treatment months, though it is not a guarantee of unbiasedness. Additionally, it suggests that the treatment and control groups were similar with respect to other unobserved factors, reducing the risk of the estimated coefficients' being affected by treatment-correlated unobserved variables.

Regarding the actual treatment effect, I find significance in month 0: arrests are recorded at the end of the month, so the response reflects arrests within 4 weeks of the treatment. There is a positive effect in this immediate aftermath, suggesting an uptick in arrests reflective of the hyper-vigilance theory — the coefficient $\beta_{PT,0} = 0.110$ suggests an 11.6 percent increase in arrests in month $t = 0$. I would hope that the uptick in arrests does not occur before the treatment — due to the absence of positive coefficients in the months prior to treatment, this seems to be a reasonable assumption: there appears to be no anticipation-like effect.

There is little significance in the nine post-treatment months. There is also no discernible pattern in the signs or magnitudes, so the significant coefficient in month 4 is likely to be random. This lack of pattern also makes it unlikely that there could be realized effects beyond nine months. While [3] finds significant negative effects in months 6-12 beyond the treatment, my lack of effects could be due in part to the sample restriction's removing places with profiles of little-to-no crime, which if included (as they are in [3]) would magnify the effect of the treatment.

3.3 Third stage - Triple Difference

Finally, I add the size of the police force as a regressor, estimating a triple difference equation with the following interaction terms:

$$Y_{it} = \beta_0 + \beta_1 \text{Post}_{it} + \beta_2 (\text{killed}_{it} \times \text{Post}_{it}) + \beta_3 (\text{ForceSize}_{it} \times \text{Post}_{it}) + \beta_4 (\text{killed}_{it} \times \text{ForceSize}_{it} \times \text{Post}_{it}) + I + T + \varepsilon_{it}.$$

The Post_{it} variable here is a consolidated Post indicator variable, which is assigned a 1 if any of the Post_i variables from the previous stage were set to 1. I use the same window as the previous stages. The coefficients capture an average treatment effect.

I run this regression twice, with two differently calculated force size metrics; each represents a different interpretation that might affect the response differently.

The first is a measure of size relative to other police force sizes, which can alternatively be thought of as a measure of “absolute size.” This can be interpreted as a measure of familiarity or fraternity within the police force. I predict that a smaller police force according

to this metric will be tied to higher-magnitude coefficients since the impact of a death might be felt more strongly in more absolutely small police forces.

The second metric is a measure of force size relative to municipality population. This can be interpreted as an indicator of how much cities are willing to invest in their police forces or the value of a police force to a city. Cities with a high ratio of force size to population might be more demonstrably supportive of the police or might offer other support resources to the police in the aftermath of police killings. I would expect that by this metric also, a larger force size is tied to smaller-magnitude coefficients. I specify only a magnitude and not a sign since the sign of the treatment effect is unclear.

The results of the regressions are below:

Table 3: Impact of LODD and “absolute” force size on arrests

	Response Variable log(Total Arrests)
Post-Treatment _{it} (PT _{it})	0.00761 (1.70)
PT _{it} × Killed _{it}	0.178* (2.13)
PT _{it} × Force Size _{it}	-0.0100 (-1.65)
PT _{it} × Force Size _{it} × Killed _{it}	-0.169 (-1.88)
Constant	2.457*** (823.10)

t statistics in parentheses

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

Table 4: Impact of LODD and population-relative force size on arrests

	Response Variable log(Total Arrests)
Post-Treatment _{it} (PT _{it})	-0.00257 (-0.61)
PT _{it} × Killed _{it}	0.0826 (1.83)
PT _{it} × Force Size _{it}	0.00981 (1.62)
PT _{it} × Force Size _{it} × Killed _{it}	-0.0962 (-1.55)
Constant	2.455*** (861.14)

t statistics in parentheses

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

I again find no significant effect due to either force size in the treated cases, or the treatment itself. Interestingly, when comparing the magnitudes of the coefficients in each notion of force size, it seems that there is a more significant reduction in arrest numbers when considering absolute force size, giving credence to the theory that familiarity with the officers may be the more important driver compared to resources in the police departments.

Another observation is that the $PT_i \times Killed_{it}$ term, which was the coefficient of interest in the event study analysis, is significant in this analysis in the absolute force size case. This is possible if police force size is correlated with the treatment effect in a way that is not reflected in the original event study.

3.4 A comment about the null effects

Most computed coefficients in the study appear to be nulls — there are few β significantly different from 0. The precision of this result is difficult to assess with certainty, but some contributing factors are the imbalance of the dataset, large dataset size, and variability in the dataset.

The imbalance in the dataset — that is, the very small ratio of positive examples to total examples — makes it likely that the results are noisy nulls. In each year of data, there were around 40 - 50 officers killed, constituting a very small percent of the examples in the 10,000 municipalities and 12 months per municipality of data. This lack of positive examples could be resulting in high variability and standard errors in the estimates; it could also result in increased likelihood of a Type 2 error, underestimating the significance of the results. While in general, large dataset size increases precision, in this case, the large dataset size when taken together with the small number of positive examples may increase the noise in the result. The effect of this imbalance is present across all regressions and results.

4 Conclusions and Future Work

This paper contributes to work in policing and studying drivers of police behavioral response. I provide new estimates with a sample restriction for the effect of a line of duty death on arrest numbers in the relevant municipality; I also introduce a heterogeneity, police force size, and study its effect when combined with the line of duty death treatment. I find a significant positive treatment effect in the exact month of treatment, and no significant effects in the following months or with the addition of the force size. There is a possible significant effect of the original treatment in the triple-difference case, which could be due to unobserved correlation between the treatment effect and the force size. I find that familiarity with the fallen officer is a bigger driver of downstream behavior change than perhaps the resourcedness of the police department. While the results were found to be non-significant, the relatively small number of officers killed in the dataset contributes to a possibly noisy null, and may result in underestimations of the significance of the treatment effect.

Additional work on the possible effect of the arrests or policing behavior on the population should investigate potential reverse causality. It could also be worth investigating the types of arrest stops made to lend credence to the $t = 0$ hypervigilance theory suggested by the

coefficients — the number of stops for trivial offences' increasing would affirm that officers act hypervigilant. I would also like to re-impose the sample restriction via matching: using a model to identify municipalities with high propensity for assaulted police officers could be another way to restrict the sample to municipalities with crime profiles similar to those where officers might be killed.

References

- [1] Highest to lowest - prison population rate, 2023.
- [2] Amelia Cheatham and Lindsay Maizland. How police compare in different democracies, 2022.
- [3] Sungwoo Cho, Felipe Goncalves, and Emily Weisburst. The impact of fear on police behavior and public safety. *NBER Working Paper 31392*, 2023.
- [4] Oeindrila Dube, Sandy Jo MacArthur, and Anuj K. Shah. A cognitive view of policing. *NBER Working Paper 31651*, 2023.
- [5] Will Johnson. Community policing: Much more than walking a beat, 2018.
- [6] Jacob Kaplan. Uniform crime reporting program data: Law enforcement officers killed and assaulted (leoka) 1960-2021, 2020.
- [7] Jacob Kaplan. Uniform crime reporting (ucr) program data: Arrests by age, sex, and race, 1974-2021, 2020.
- [8] Laurence Miller. Line-of-duty death: Psychological treatment of traumatic bereavement in law enforcement. *International Journal of Emergency Mental Health*, 9:13–23, 2007.
- [9] Andreas Olden and Jarle Moen. The triple difference estimator.

Appendix

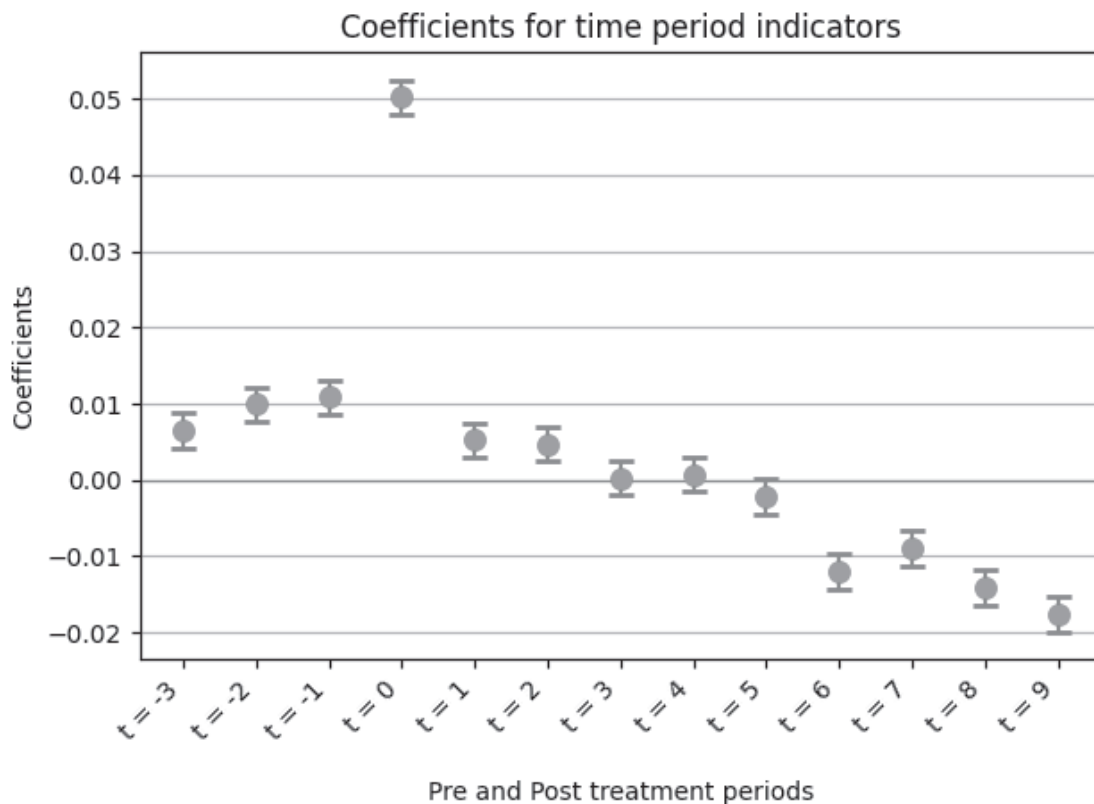
Here is the table of baseline time trajectory coefficients from the second-stage event study, along with a plot of the coefficients — these are the coefficients denoted $\beta_{\text{Post},j}$ in the estimating equation.

	log(total arrests)	t-stat	$p \geq t $	95-percent confidence interval
t = -3 (Post _{-3,it})	0.0065** (.00224)	2.90	0.004	[.0021, .0109]
t = -2	0.0100*** (.00224)	4.44	0.000	[.0056, .0143]
t = -1	0.0109*** (.00224)	4.87	0.000	[.0065, .01527]
t = 0	0.0503*** (.00221)	22.77	0.000	[.0459, .0546]
t = 1	0.0053* (.00225)	2.35	0.019	[.0009, .0097]
t = 2	0.0047* (.00226)	2.08	0.038	[.0003, .0091]
t = 3	0.0003 (.00227)	0.11	0.910	[-.0042, .0047]
t = 4	0.0007* (.00229)	0.31	0.757	[-.0038, .0052]
t = 5	-0.0022 (.00231)	-0.93	0.351	[-.0067, .0024]
t = 6	-0.0120*** (.00232)	-5.17	0.000	[-.0166, -.0075]
t = 7	-0.0090*** (.00234)	-3.85	0.000	[-.0136, -.0044]
t = 8	-0.0141*** (.00235)	-6.01	0.000	[-.0187, -.0095]
t = 9	-0.0176*** (.00235)	-7.47	0.000	[-.0221, -.0129]
Observations	330456			

95% confidence intervals in brackets

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

The corresponding plot of coefficients, which reflect the baseline effect of a shooting for many months, is shown. Coefficients have not been normalized:



Like the single coefficient from the first stage, we see that the effect of a shooting itself is significant in many months post-shooting. The effect in month $t = 0$, which is captured between 0 and 4 weeks post-shooting, appears most significant with a strong positive effect on arrests; in the remaining months, there is a clear downward trend in the impact of the shooting, starting with smaller positive effects for up to 3 months post shooting, and ending with significant negative effects in months 6 to 9.

Here is the table of interaction term coefficients from the second stage event study; these are the coefficients corresponding to the graph shown in the paper.

	log(total arrests)	t-stat	$p \geq t $	95-percent confidence interval
t = -3 (killed _{it} × Post _{-3,it})	-0.0307 (.095)	-0.32	0.748	[-0.218,0.156]
t = -2	-0.0125 (.086)	-0.15	0.884	[-0.180,0.155]
t = -1	-0.115 (.093)	-1.24	0.216	[-0.298,0.0674]
t = 0	0.110 (.063)	1.74	0.082	[-0.0140,0.233]
t = 1	0.00746 (.102)	0.07	0.942	[-0.192,0.207]
t = 2	-0.0167 (.095)	-0.18	0.861	[-0.204,0.170]
t = 3	-0.0209 (.099)	-0.21	0.833	[-0.215,0.173]
t = 4	0.218* (.100)	2.18	0.029	[0.0219,0.414]
t = 5	-0.116 (.097)	-1.20	0.231	[-0.307,0.0741]
t = 6	-0.0957 (.101)	-0.95	0.343	[-0.293,0.102]
t = 7	0.0490 (.100)	0.49	0.624	[-0.147,0.245]
t = 8	-0.0151 (.090)	-0.17	0.867	[-0.192,0.162]
t = 9	0.0190 (.085)	0.22	0.823	[-0.147,0.185]
Constant	2.182*** (.002)	1449.31	0.000	[2.179,2.185]

95% confidence intervals in brackets

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

All code for this paper can be found at <https://github.com/NSrinidhi/14.33-policing-project>.

Dropping the SAT Subject Tests

Quentin Smith

ABSTRACT

The SAT Subject Tests are no longer offered, but institutions with selective admissions used to strictly or implicitly require them. The disappearance of the Subject Tests provides an avenue for studying the effects of changing admissions policies on student behavior, and this paper uses a difference-in-differences approach to study the effect of dropping the subject tests on minority completions in different majors. Although this paper does not find a clear significant effect on completions, it does take the opportunity to look at several race and ethnicity categories independently and suggests that additional application and enrollment data along with further studies of the completions of Asian and Hispanic students may yield more fruitful results.

INTRODUCTION

University policies for standardized testing are a great source of contention. The spread of COVID-19 led many schools to adjust how they considered standardized testing in the admissions process, and universities had to decide whether or not to bring back the SAT and ACT and come up with reasonable explanations for either decision. MIT Dean of Admissions and Student Financial Services Stuart Schmill explains the reasoning behind bringing back the requirement at MIT in a 2022 Q&A. He mentions that equity is a concern in the admissions process, and testing is one of many factors potentially impacted by the race, ethnicity, and socioeconomic status of applicants. He argues that the tests provide a way to learn about student readiness and actually may benefit students with less access to challenging coursework and other means of demonstrating ability. At the same time, he provides the caveat that the tests are imperfect; he addresses opposition to these testing policies which asserts that the SAT and ACT may actually hurt underrepresented students.

In this paper, we investigate the impact of changing school testing policies on student behavior.

Student behavior regarding postsecondary admissions is a broad category that encompasses application decisions, enrollment behavior, and completions, but we specifically look at how testing policies affect the on-time graduation of students in different racial/ethnic categories.

We focus not on the SAT and ACT requirements, but on the SAT Subject Tests. The Subject Tests lost popularity throughout the 2000s and 2010s before being discontinued in 2021, and the decline in their popularity allows them to be a useful area of study because there is variation in regards to whether the Tests were required or not considered at universities in the years leading up to their termination. Besides having a source of variation, the Subject Tests also had the benefit of testing particular areas of study. There is existing literature that finds connections between AP Exam scores and major choice, another test that makes appearances in the college admissions process. Avery et al (2018) is an example of the utilization of tests beyond the SAT/ACT to study student behavior, and that paper investigates how an increased score on an AP Exam may influence major choice by comparing students of similar aptitude who received different scores on AP Exams in the mid to late 2000s. The paper does not find major differences between various demographic groups but suggests that students may treat a higher score as a positive signal of what they should major in. As another subject-specific test maybe the SAT Subject Tests also led to similar student behavior while they existed and were more prevalent in the 2000s. If there is an interest in leading students toward specific majors, research that reveals a relationship between the Subject Tests and completions in particular majors may lend support to using subject-specific testing policy as a means of influencing student decisions.

The choice to look at the Subject Tests while differentiating racial and ethnic categories follows from papers like Black et al (2020), a study of how some students' guaranteed acceptance into Texas public universities may impact student decisions. Black et al (2020) looks at application behavior in an attempt to study the causes of racial/ethnic differences in college enrollment, and one potential cause is differences in information accessibility. The paper finds that Black and Hispanic students who are eligible for automatic admission are more likely to apply to more selective public institutions, a display of how school policy can impact minority student decisions. In this paper, we carry over a feature from Black et al (2020): we isolate Black and Hispanic students instead of grouping them together as underrepresented minorities.

Black et al (2020)'s investigation is an example of how access to information may impact college outcomes differently depending on race and ethnicity, and several other papers study how students react to information about postsecondary education and related costs. Hoxby and Turner (2015) study how the cost of postsecondary education may lead students away from their best academic fit and find that clear information may assist students with choosing schools. Goodman et al (2020)'s study of retaking the SAT also utilizes an alternative to studying first-time SAT/ACT takers and mentions that students may struggle to make accurate judgements about the benefits and costs of taking tests. With the inclusion of Pallais (2015)'s suggestion that students may use testing fees as signals of reasonable behavior in the admissions process, the information barrier regarding postsecondary education not only accounts for the differences in information that students have but also that students may not know exactly what to do with the information that they are given. The costs associated with testing along with the complexity of the admissions process makes it worthwhile to study how testing impacts student behavior. Maybe testing is a means of steering students in a particular academic direction, but if testing increases socioeconomic or racial gaps in education because particular groups of students are misunderstanding the costs, it may not be worthwhile to maintain policy that enforces them.

This paper begins with a brief description of our school sample and the fields of study that we investigate. We run a simple regression with one variable to represent the changing Subject Test requirements and then utilize a staggered difference-in-differences design in order to study how changes in SAT Subject Tests impacted student completions over multiple years. We ultimately find that there is not a clear suggestion that the SAT Subject Tests had an effect on on-time-graduation. We generally estimate the smallest effect of changing Subject Test requirements for Black students and the largest for Asian and Hispanic students, negative for Asian students and positive for Hispanic, but potential issues with differences between completions at schools that dropped or kept the Subject Tests leads us to avoid making any strong conclusions about the effect of changing the Subject Test requirements on completions. More accessibility to application and enrollment data may be informative due to issues with the completion data along with the distance of completions from initial application decisions.

DATA

Sample

We investigate the effect of dropping the SAT Subject Test requirement on thirty-three schools in the United States that either required or recommended the SAT Subject Tests at some point between the 2006-2007 and 2017-2018 application cycles. We choose to treat schools as if they required the Subject Tests as long as they at least encouraged students to take them. The wording of "recommended" or "encouraged" may have been understood as an implicit requirement by students or admissions offices, and STEM-oriented schools seemed more inclined to strictly require the Subject Tests while selective liberal arts schools tended to heavily recommend the Tests. Grouping those policy categories with strict requirements is an attempt to avoid accidentally including schools that effectively required the Subject Tests for the entire time period in the treatment group while avoiding major differences in academic characteristics between the treatment and control schools.

Table 1 lists the schools and some of their characteristics. The majority of the schools in the sample are considered more selective and are located in Northeastern states and California.¹ There are 14 schools in the treatment group and 19 in the control group. A school was treated once it no longer mentioned that the Subject Tests were recommended or required. Confirmation for the treated schools comes from emails with admissions officers in addition to college catalogs and bulletins. Control school confirmation comes from the same means in addition to common data sets if the schools at least claimed to recommend the Subject Tests through their 2017-2018 application cycles.²

Our sample does not include all schools that recommended or required the Subject Tests. We exclude schools that listed the Subject Tests as an alternative to the SAT and ACT because a lack of knowledge about the Subject Tests or concerns about costs associated with additional testing would not prevent students from still applying to that school. We also exclude schools

¹Selectivity of admissions is according to these schools' Carnegie Institutions of Higher Education Undergraduate Profiles. Admissions for the University of California - Merced are considered inclusive. Table 1 lists the locations of each schools.

²Information about SAT Subject Test requirements is not centralized. The majority of schools do compile common data sets, but we did not solely use these to confirm testing requirement changes for treated schools because some schools predict admissions policies for future cycles while others go back and update their requirements.

that only required or recommended the Tests for specific departments and those that became entirely test optional and dropped the SAT/ACT along with the Subject Tests.

Something to note is that schools that only required the Subject Tests from students who took the SAT are included. We consider this case to be different from the case where the Subject Tests are listed as an alternative to other forms of testing because the Tests were tied to a more popular element of the admissions process. According to school 2018-2019 common data sets, the majority of schools in the sample had a greater percentage of students submitting the SAT than the ACT.³ We acquiesce that information about SAT and ACT popularity for one year is not enough to state that the SAT was always more popular for this sample of schools, but this finding suggests that the ACT was not the unilateral choice, and students were affected by Subject Test requirements that were paired with the SAT.

Race/Ethnicity

The race and ethnicity categories used in this paper come from the National Center for Education Statistics's Integrated Postsecondary Education Data System. The racial categories are American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White and then there is the ethnicity category Hispanic or Latino. Individuals are categorized as Two or more Races if they list more than one of the races and Unknown if their race/ethnicity are not known.

These categories were new in the 2000s. Some survey components required these categories starting with the 2010-2011 school year, and the component that this paper is concerned with, completions, was required to use the new categories beginning in 2011-2012. The schools in our sample all switched to the new categories by 2010-2011.

In this paper, we do not consider the racial categories of American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Two or more races, and Unknown because they tend to have small numbers of completions, or their results may be difficult to interpret because they can draw from multiple race/ethnicity categories. We also exclude U.S. nonresidents in order to focus on the completions of U.S. residents and citizens. We refer to the remaining categories as

³These schools include Barnard, Boston College, Brown, Caltech, Clarkson, Dartmouth, Georgetown, Harvard, Harvey Mudd, Lafayette, MIT, Pomona, Princeton, Scripps, Stanford, UPenn, Vassar, Wellesley, and Yale. Again, the common data sets were found individually.

Hispanic, Asian, Black, and White throughout this paper.

Fields of Study

Our sample of schools determined the possible fields of study. The National Center for Education Statistics provides the Classification for Instructional Programs (CIP), and we include eleven fields of study that are each offered by at least thirty of the thirty-three schools in the sample. Table 2 lists each of the two-digit CIP codes and the corresponding field of study. We group these eleven fields of study into three broad major categories: humanities, social sciences, and STEM.⁴ The humanities majors include English, Foreign Languages, Philosophy and Religious Studies, Visual and Performing Arts, and History. The social sciences include Social Sciences and Psychology, and the STEM majors include Computer Science, Biological Sciences, Mathematics and Statistics, and Physical Sciences.⁵

Completions and Time Period

Although application behavior is of interest, this paper focuses on completions due to data availability. We estimate the effects of dropping the SAT Subject Tests on on-time graduation, so we look at students who complete a degree in four years. This paper further specifies that the student receives a bachelor's degree in their first major. Completions are at the school level and are from the Integrated Postsecondary Education Data System. We download completions by race and two-digit CIP codes for field of study. We consider a school to have complete data on completions for a particular major if that school offers that field of study and collects data on completions for every year in the time period of interest.

We define shares of completions as the fraction of total bachelor's degrees awarded to students of a particular race and levels of completions as the number of students of a particular race who receive a bachelor's degree. The share of on-time completions that are from White students is the ratio of White students receiving a bachelor's degree in four years to the total number of

⁴One field of study: Multi/Interdisciplinary Studies was offered by thirty-one of these schools, but the specific majors within that field sometimes fit into vastly different categories. Mathematics and Computer Science from MIT is an example of a STEM major while Cultural Studies/Critical Theory and Analysis at Harvard is in either the humanities or social sciences category.

⁵Full names for these majors are in Table 2. There is some debate about whether history is in the humanities or social sciences and whether psychology is in the social sciences or STEM. We assigned history to the humanities and psychology to the social sciences for this paper but acknowledge that this categorization may not be shared by all.

students receiving a bachelor's degree in four years. We use on-time graduation because there are low graduation rates, at least through 2014. On-time graduation rates only crossed above 50% for White students at all institutions who entered college in fall 2014 and for Asian students who entered in fall 2010 or later.⁶ As a result, we are not very concerned that on-time graduation misses major completions of the first students affected by changing Subject Test policies.

On-time graduation links the fall entrance of students to completions four years in the future. The earliest entering class of students that we study in this paper graduates in 2011, so we link their entrance to fall 2007, and the last entering class enters in 2018 and graduates in 2022. This means that we investigate changes in admissions policies for the 2006-2007 application cycle through the 2017-2018 application cycle and spring 2011 completions through spring 2022. We begin with spring 2011 because of the new Integrated Postsecondary Education Data System race/ethnicity categories, and the time period ends with 2022 because that was the most recent year of completions collected at the time of this study.

METHODS

We begin with a simple regression that just accounts for whether a school changed their Subject Test requirement at any point.

$$Y_{st} = \alpha_s + \beta_t + State_s + \gamma(Post_t \times Treated_s) + \varepsilon_{st} \quad (1)$$

For a given race/ethnicity and field of study, the dependent variable Y_{st} is the share of bachelor's degrees in a student's first major awarded in four years at school s in year t . The interaction term $Post_t \times Treated_s$ is the independent variable of interest; $Post_t$ is equal to 1 if the first on-time graduations impacted by dropping the Subject Test requirement occurred in year t , and $Treated_s$ is equal to 1 if the school dropped its Subject Test requirement within our time period of interest. We also include school fixed effects, year fixed effects, and one control for the state that school s is located in.⁷

⁶Data about graduation rates from Table 326.10 in NCES Digest of Education Statistics

⁷Some other possible controls are school qualities which include whether the school is private or public, campus setting, graduate coexistence, and religious affiliation. With our sample of schools, most of these variables were omitted, so we exclude them from the final regression.

The second regression is a staggered difference-in-differences design that accounts for potential changes in the effect of removing the Subject Test requirement over time.

$$Y_{st} = \alpha_s + \beta_t + State_s + \sum_{\tau=-11}^{-2} \lambda_{\tau} lead_{\tau} + \sum_{\tau=0}^9 \lambda_{\tau} lag_{\tau} + \varepsilon_{st} \quad (2)$$

The dependent variable, fixed effects, and control are the same as in regression 1, but there are now several event time variables. The variable $lead_{\tau} = 1$ if in year t it is τ years before school s is treated. The variable $lag_{\tau} = 1$ if in year t it is τ years after school s is treated. The variable $\tau = 0$ when school s is treated. For control schools, the event time variables always equal 0.

Event Time

We calculate event time based on when schools changed their Subject Test requirements. Brandeis is the earliest treated school; students who applied to enter in fall 2010 were no longer required to take the Subject Tests, and 2022 is nine years after on-time graduations for those students in 2013. As a result, the maximum possible event time is 9. The minimum possible event time -11 arises because the earliest completions in 2011 occur eleven years before the last schools have their first treated on-time graduations in 2022.

We include all lag and $lead$ variables in the regression but only include the variables with event time between -4 and 4 in the event study figures in an attempt to avoid having a small number of schools greatly influence our results. Figure 1 plots the frequency of treated schools that have each event time for a certain type of major. Except for $lead_{11}$, the $lead$ variables are nonzero for at least half of the schools, with all of the schools having nonzero $lead_0$ through $lead_2$ variables because the first completions that we have data for occurred in 2011 two years before the first on-time graduations at Brandeis were impacted by dropping the Subject Tests. We do not have as many nonzero lag variables, but at least 6 of the treated schools have nonzero lag variables through lag_4 .

ESTIMATION

In order to learn a little more about the data before diving into the effect of dropping the Subject Tests, let us begin with completions that are differentiated by major but not race. We do not run

regressions 1 and 2, but we can investigate changes in completion levels graphically. Figure 2 captures completion levels by field of study.

The 14 schools in our sample that dropped the Subject Tests before the 2018-2019 application cycle have a greater number of completions in the humanities than schools that kept the Tests for the entire time period of interest, and completions generally decreased over time. The number of STEM completions looked pretty similar for both the schools that kept and dropped the Subject Tests, and these completions increased overall. There was the most variation between schools that kept and dropped the Tests for the social sciences. The smallest gap between the schools occurred in 2013 when the difference in completions was only four students; schools that at some point dropped the Subject Tests awarded 6,535 bachelor's degrees to first majors in the social sciences and schools that kept the Tests awarded 6,539. There was the greatest difference between these schools, 1,037 students, in 2021 when there were 6,569 completions from schools that dropped the Subject Tests and 5,532 from those that kept the Tests. Completions mostly decreased in the social sciences for schools that kept the Subject Tests while there was more fluctuation and an overall increase from schools that dropped the Tests.

Simple Regression

There is also data on bachelor's degrees awarded by race but not major. There are completion levels for this data which we discuss in the next section, but because we are now specifying race and can calculate completion shares, we can discuss the estimates we produce from running regressions 1 and 2 for this data along with data that specifies race and major.

We run regression 1 to look at the effect of dropping the Subject Test requirement on completions before specifying field of study. The coefficients for $Post_t \times Treated_s$ are in Table 3. Looking at race without specifying anything about major besides the restriction to eleven fields of study, the regression estimates that schools that dropped the Subject Tests awarded lower shares of bachelor's degrees to Asian, Black, and White students than schools that kept the Subject Tests and higher shares to Hispanic students, but none of these coefficients are statistically significant at the .05 level.

Moving on to look at completions for particular majors, we run regression 1 on the humanities, social sciences, and STEM majors. For the humanities, we estimate that schools that dropped

the Subject Tests awarded lower completion shares to Asian and Black students and more to White and Hispanic students than schools that kept the subject tests. The coefficients for Asian and Hispanic students are significant and estimate that the share of completions was 0.0444 or 4.44 percentage points lower for Asian students and 3.28 percentage points higher for Hispanic students at schools that dropped the Subject Tests than at schools that kept the Tests.

The regression estimates that there were lower completion shares of Asian and White students and higher shares of Black and Hispanic students in the social sciences at schools that dropped the Subject Tests. The coefficients for Asian and Hispanic students are significant; the regression estimates that schools that dropped the Tests had completion shares for Asian students that were 3.26 percentage points lower and shares for Hispanic students that were 3.45 higher than at schools that kept the Tests.

None of the coefficients are significant for STEM majors, but the regression estimates that there were lower completion shares for Asian, Black, and White students and higher shares for Hispanic students at schools that dropped the Subject Tests.

Event Study

Next, we run regression 2 in order to do an event study by race. Looking at completions by race but not major, Figure 4 only restricts majors to the eleven fields of study. The majority of the coefficients are insignificant, but some trends are that the event study estimates lower completion shares of Asian students and higher shares of Hispanic students at schools that drop the Subject Tests for each of the years after the first on-time graduations of students impacted by dropping the Subject Tests. Looking at the scale of the event study, Black students generally have coefficients of the smallest magnitude while Hispanic and Asian students have the largest. The coefficients become more negative for White and Asian students and more positive for Hispanic. Only one coefficient, lag_3 for Hispanic students is significant and estimates that schools that dropped the subject tests awarded shares that were higher by 3.90 percentage points for Hispanic students than shares at schools that kept the Tests three years after the first on-time graduations of students impacted by the changing requirements.

Looking specifically at humanities majors, there is potentially an issue with concluding anything about Hispanic students because the regression estimates that schools that drop the Subject Tests

have completions shares for Hispanic students that are lower by 2.76 percentage points than at schools that kept the Tests three years before the first on-time graduations should have been affected by dropping the Tests.

Interpreting other significant coefficients for the humanities majors, the coefficients for Hispanic students mostly increase, and the regression estimates that Hispanic students at schools that dropped the Tests had completions 3.31 percentage points higher four years after dropping the Tests than those at schools that kept the Tests. The coefficients for Asian students mostly decrease, and several coefficients are significant; the event study estimates that completion shares of Asian students at schools that dropped the Tests were 4.12 lower one year after the first on-time graduations impacted by changing test requirements, 4.99 lower three years after, and 5.13 four years after. White students at schools that dropped the subject tests are estimated to have had completion shares 4.81 higher than those at schools that kept the tests at the year that on-time graduations should have been first affected. Black students have the smallest coefficients, and none of them are significant. Comparing majors, the humanities have some of the noisiest coefficients, some examples are the coefficients for White students three and four years after the first impacted on-time graduations.

For the social sciences and STEM, none of the coefficients before the first impacted on-time graduations are significant. The social sciences have the largest estimated coefficients. The coefficients for Hispanic students mostly increased, and it is estimated that shares of Hispanic students were 5.35 percentage points higher at schools that dropped the Subject Tests and 7.37 lower for White students four years after the first on-time graduations of students affected by dropping the Subject Tests. The coefficients for Black students are generally small relative to the coefficients for other races and positive, but none are significant. The coefficients for Asian and White students mostly decrease over time, but none are significant.

None of the coefficients are significant when looking at STEM majors, and this group of majors has some of the smallest coefficients. Even though none of these are significantly different from zero, some trends for the STEM majors are that Hispanic students at schools that dropped the Subject Tests are estimated to have had higher completion shares than those at schools that kept the tests, and White students are estimated to have had lower shares for the first impacted

on-time graduations and for each year after.

Assumptions

One of the assumptions for a difference-in-differences approach is that there is no anticipation. Students should not have been able to react to changing subject test requirements ahead of time. If students chose not to apply to schools because of the subject tests, advanced knowledge about changing requirements could result in students waiting to apply to college until the tests were no longer required or recommended. If that was the case, we might see that there was a drop in completions immediately before treatment and an increase after. If for some reason students really wanted to take the subject tests, maybe they worried that other factors of the admissions process would hurt their chances, we might see a jump in completion shares immediately before the first on-time graduations should have been impacted and lower shares after.

The event study figures for Hispanic and Asian students cause the most concern for this assumption. Again, these are mostly insignificant coefficients, but the leading coefficients for Hispanic students are negative and the lagging are positive, with the opposite being true for Asian students. We do not suspect that there was anticipation though. Wesleyan University, Massachusetts Institute of Technology, and California Institute of Technology's announcements of their changing test policies exemplify that schools may have timed their decisions so that previous application cohorts are unable to adjust their application behavior. Caltech released their announcement about removing the Subject Tests for the 2020-2021 application cycle in late January after regular decision applications were likely due for the 2019-2020 application cycle. Wesleyan's announcement about becoming test optional and two announcements that MIT released, removing the Subject Tests and bringing back the SAT and ACT post-COVID, all occurred after regular decisions were released for students from the previous application cycle.⁸ Another assumption is parallel trends, that completion patterns would not have differed between control and treatment schools had the Subject Tests not been dropped. In order to get an idea of whether there were preexisting differences between schools that kept and dropped the Subject Tests, let us look at completion levels by race. There were more completions of Asian and

⁸We exclude Wesleyan from our sample because it became entirely test optional when it removed its Subject Test requirement. We mention changes that Cal Tech and MIT made to their testing policies; those all occurred after our time period of interest ends.

Hispanic students at schools that dropped the Subject Tests for the entire time period than those that did not and fewer Black students. Completion levels only crossed between schools that dropped and kept the Subject Tests for White students. The greatest number of completions came from White students and the least came from Black students.

Something that is concerning is that the increase in Hispanic student completions for schools that kept the Subject Tests was more drastic than at schools that did not in the total, humanities, and social sciences categories. There was also an increase in the gap between the two types of schools for STEM, but it was not as drastic and also occurred later in the time period. For the total, humanities, and social sciences categories, this gap began to increase even by 2015 when only one school was treated, so the changing difference between schools that kept and dropped the Tests did not seem to be the result of changing Subject Test requirements.

Completions between the two types of school seemed to converge for Asian students overall, though the change in the gap between the two groups was not as extreme as for Hispanic students, and more visibly converged for the humanities. Even in the humanities, this convergence did not occur as early on as it did for Hispanic students. The greatest difference for Asian students in the humanities was in 2013 with a difference of 226 bachelor's degrees awarded, but the difference in completions fluctuated between 170 and 230 through 2016 and then shrank from 146 completions in 2017 down to 73 in 2022. That contrasts with Hispanic students where the gap in completions between schools that kept and dropped the Subject Test more than doubled from 121 completions in 2011 to 261 in 2015 and then went up to 335 in 2022. Another concern for this assumption is with White students in the total, humanities, and STEM categories because the completion levels cross for the two groups of schools throughout the time period.

Going back to the event study graphs, the coefficients generally increase for Hispanic students and decrease for Asian students. Just focusing on Hispanic students, maybe there was some factor besides the changing Subject Test requirements that simultaneously increased the completion levels of Hispanic students at treated schools and led us to underestimate the effects of changing Subject Test requirements before on-time graduations were actually impacted and overestimate them afterward. Determining the exact causes requires further study, but we find that one potential difference between schools in the treated and completion groups that may

specifically impact Hispanic students outcomes is their status as Hispanic-Serving Institutions (HSIs). HSIs are federally-recognized eligible schools where at least 25% of undergraduate full-time equivalent students are Hispanic. From 2010-2011, University of California - Merced was the only HSI in our sample, and by 2021-2022 University of California - Santa Cruz, another school in our treatment group, had also achieved HSI status.⁹ University of California - Santa Cruz received a grant from the Developing Hispanic-Serving Institutions program in 2020, and the 2019 grant for the HSI Antelope Valley College's project objectives included increasing transfers to University of California. We suspect that schools in this category may make greater efforts to increase the completions Hispanic students, and if those programs include things like increasing transfers, we are not purely looking at the effect of the Subject Tests on completions. The discussion of HSIs provides a glimpse into how the multitude of initiatives employed by a school with diversity concerns may provide a challenge to singling the effect of any one policy or program.

The last major assumption is that nothing else changed when these subject test requirements were dropped. We just mentioned that if schools are making these changes due to diversity concerns, we may worry that other policies are also impacting the completions of students in a particular racial/ethnic group. Another concern that comes up with this assumption is changes that were made to the SAT during this period. One major change occurs right before our time period of interest begins; SAT scoring shifted in 2005, and the writing section was added. Another major shift occurred in 2016, and SAT sections and scoring changed again. This is potentially an issue if changes in test format also somehow impacted major choice and student behavior regarding completions. That seems less likely than Subject Tests impacting major choices unless specific sections of the SAT/ACT like the math section also had the ability to signal students toward certain majors. These changes may also be a concern just because of the ties between the SAT and the Subject Tests. If these changes to the SAT drove people toward or away from the SAT, that might also impact who applied and completed majors at schools that kept the Subject Tests. The investigation of common data set information about SAT and

⁹There are also schools that are suspected to reach HSI status in the coming years that are called Emerging HSIs; those schools included Santa Cruz and Stanford in 2010-2011 and Amherst College, Cal Tech, Harvey Mudd, MIT, Pomona, Rice, Stanford, and Davis from 2021-2022. We only focus on the HSIs though, because the emerging HSIs are not federally recognized.

ACT popularity at least at years surrounding those SAT changes might also be informative for that question. In the case that changes to the SAT did contribute to changing completion levels, something that we did do to attempt to address this assumption was excluding schools that went entirely test optional when they dropped the SAT and ACT.¹⁰

CONCLUSION

The difference-in-differences approach we utilize in this paper did not reach a clear conclusion. Asian and Hispanic students often had the most significant estimated differences in completions in shares between schools that dropped and kept the Subject Test requirements, but they might have also had issues with their samples and the parallel trends assumption. At this point, we make no definitive statements about how changing testing policies impact student completions but suggest that it may be of interest to continue to study the Subject Tests, especially with Hispanic and Asian students.

One difficulty that arises with this topic comes from creating the sample itself, schools have varying priorities when it comes to institutional data, and there is not a centralized mandatory source regarding the Subject Tests and all admissions policies. Future areas of interest are to more directly investigate the impact of testing policies on admissions and enrollment. The previous grievance about differing priorities is the reason why this is unlikely across many schools, but a focus on a smaller sample maybe over a longer time period may lead to more detailed data on enrollments or declared majors by race.

¹⁰Something that we do not keep track of is what schools did with their testing policies in the years following dropping the SAT Subject Tests, and that is potentially a concern for making conclusions once we get very far from the treatment year. Again, we only looked up to four years after treatment, and the hope is that graduation rates are low enough we do not catch too many effects from later changes in testing policy on completions.

REFERENCES

- Avery, C., Gurantz, O., Hurwitz, M., & Smith J. (2018). Shifting College Majors in Response to Advanced Placement Exam Scores. *Journal of Human Resources*, 53(4), 918-956.
- Black, S. E., Cortes, K. E., & Lincove, J. A. (2020). Apply Yourself: Racial and Ethnic Differences in College Application. *Education Finance and Policy*, 15(2), 209–240.
https://doi.org/10.1162/edfp_a_00273.
- Browse CIP Codes*. IPEDS - National Center for Education Statistics.
<https://nces.ed.gov/ipeds/cipcode/browse.aspx?y=55>.
- Caltech to Eliminate Requirement for SAT Subject Test Scores in Admissions*. (2020). Caltech.
<https://www.caltech.edu/about/news/caltech-eliminate-requirement-sat-subject-test-scores-admissions>.
- Carlisle, K. (2014). *Wesleyan Makes Tests Optional in Admissions*. The Wesleyan Connection.
<https://newsletter.blogs.wesleyan.edu/2014/05/07/admissionstests/>.
- College Navigator. National Center for Education Statistics.
<https://nces.ed.gov/collegenavigator/>.
- Compare Institutions. IPEDS - National Center for Education Statistics.
<https://nces.ed.gov/ipeds/datacenter/InstitutionByName.aspx?stepId=1&sid=05418cb7-8dfc-4274-942b-41e370a25ba3&rtid=1>.
- Digest of Education Statistics* (2021). IPEDS - National Center for Education Statistics.
https://nces.ed.gov/programs/digest/d21/tables/dt21_326.10.asp?current=yes
- Goodman, J., Gurantz, O., & Smith, J. (2020). Take Two! SAT Retaking and College Enrollment Gaps. *American Economic Journal: Economic Policy*, 12(2), 115–158.
<https://doi.org/10.1257/pol.20170503>.

Hoxby, C. M., & Turner, S. (2015). What High-Achieving Low-Income Students Know About College. *American Economic Review*, 105(5), 514–517.

<https://doi.org/10.1257/aer.p20151027>.

Institutional Responsibility For Reporting Race And Ethnicity To IPEDS. IPEDS - National Center for Education Statistics.

<https://nces.ed.gov/ipeds/report-your-data/race-ethnicity-reporting-changes>

Institution Search. Carnegie Classification of Institutions of Higher Education.

<https://carnegieclassifications.acenet.edu/institutions/>

Pallais, A. (n.d.). *Small Differences That Matter: Mistakes in Applying to College*.

Schmill, S. (2020). *A special announcement about SAT Subject Tests*. MIT Admissions.

<https://mitadmissions.org/blogs/entry/a-special-announcement-about-sat-subject-tests/>.

Wren, K. (2022). *Q&A: Stuart Schmill on MIT's decision to reinstate the SAT/ACT requirement*.

MIT News. <https://news.mit.edu/2022/stuart-schmill-sat-act-requirement-0328>

Table 1. Basic School Information

School	Type	State	Treated Year
Amherst College	Private, not-for-profit	MA	2022
Barnard College	Private, not-for-profit	NY	2021
Boston College	Private, not-for-profit	MA	2020
Brandeis University	Private, not-for-profit	MA	2013
Brown University	Private, not-for-profit	RI	-
California Institute of Technology	Private, not-for-profit	CA	-
Carleton College	Private, not-for-profit	MN	-
Clarkson University	Private, not-for-profit	NY	-
Dartmouth College	Private, not-for-profit	NH	-
Davidson College	Private, not-for-profit	NC	-
Duke University	Private, not-for-profit	NC	-
Georgetown University	Private, not-for-profit	DC	-
Harvard University	Private, not-for-profit	MA	-
Harvey Mudd College	Private, not-for-profit	CA	-
Lafayette College	Private, not-for-profit	PA	-
Massachusetts Institute of Technology	Private, not-for-profit	MA	-
Northwestern University	Private, not-for-profit	IL	-
Oberlin College	Private, not-for-profit	OH	2018
Pomona College	Private, not-for-profit	CA	2022
Princeton University	Private, not-for-profit	NJ	-
Rice University	Private, not-for-profit	TX	-
Scripps College	Private, not-for-profit	CA	2018
Stanford University	Private, not-for-profit	CA	-
Tufts University	Private, not-for-profit	MA	-
University of California - Davis	Public	CA	2016
University of California - Merced	Public	CA	2016
University of California - Santa Cruz	Public	CA	2016
University of Pennsylvania	Private, not-for-profit	PA	-
University of Virginia - Main Campus	Public	VA	2021
Vassar College	Private, not-for-profit	NY	2022
Washington and Lee University	Private, not-for-profit	VA	2021
Wellesley College	Private, not-for-profit	MA	2022
Yale University	Private, not-for-profit	CT	-

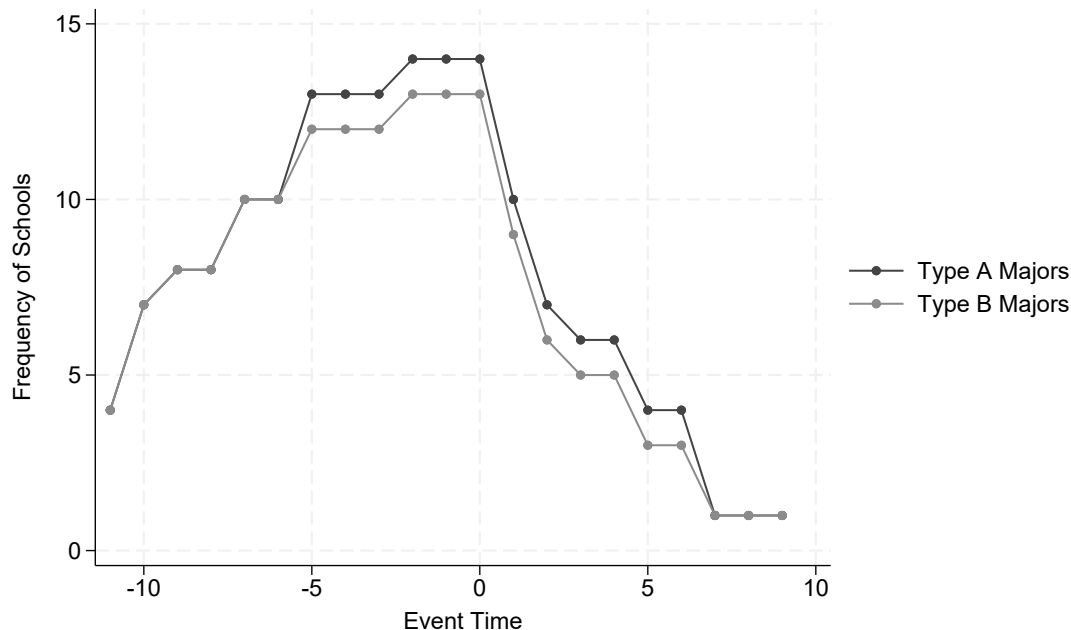
Table includes full institution name, type, state, and year of treatment (if applicable). School type and state are from NCES College Navigator.

Table 2. Number of Schools With Data on Completions

Field of Study	Schools With Completions Data
11 - Computer and Information Sciences and Support Services	30
16 - Foreign Languages, Literatures, and Linguistics	30
23 - English Language and Literature/Letters	30
26 - Biological and Biomedical Sciences	33
27 - Mathematics and Statistics	33
38 - Philosophy and Religious Studies	30
40 - Physical Sciences	33
42 - Psychology	30
45 - Social Sciences	33
50 - Visual and Performing Arts	30
54 - History	31

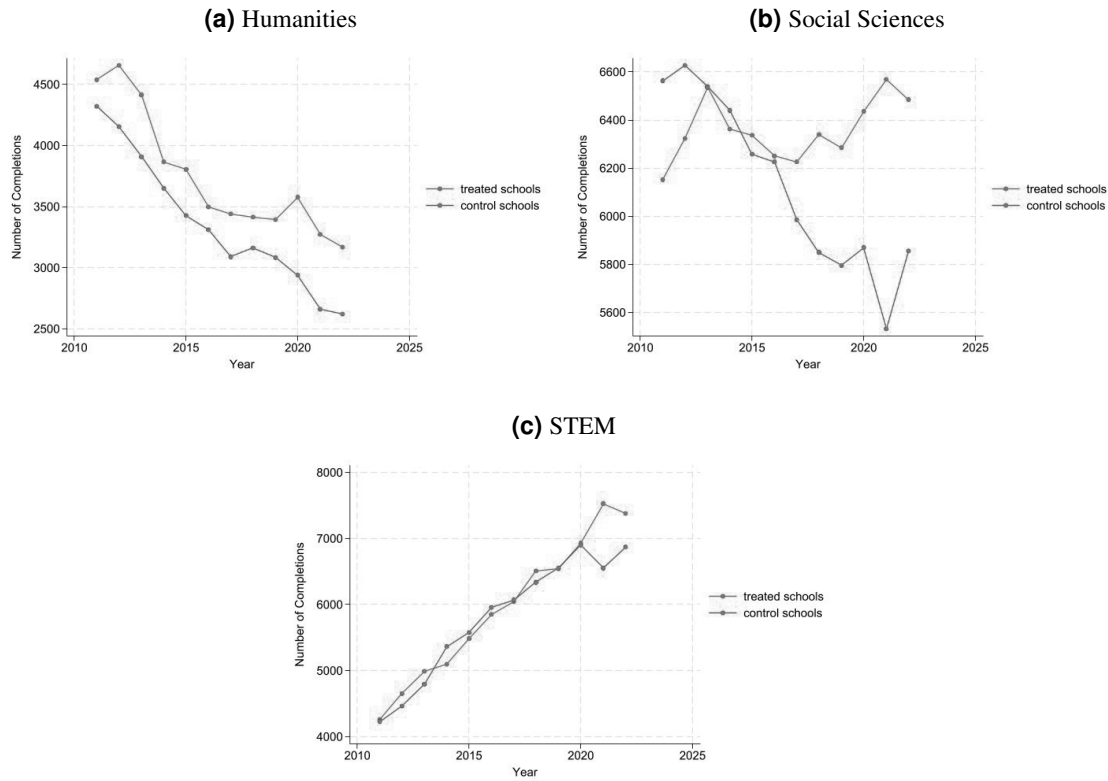
Number of schools is out of 33 schools in the sample. Fields of Study from NCES Classification of Instructional Programs (CIP). We exclude schools that either did not offer that field of study or did not collect completion data for the entire time period 2011-2022.

Figure 1. Schools Per Event Time



We graph the number of treatment schools that have each event time for each type of major. The maximum possible number is fourteen, and we exclude control schools because they do not receive treatment within our time frame. We group majors into two types. Type A majors (with CIP codes 16, 26, 27, 40, 42, 45, and 54) have data for all of the treated schools, Type B (11, 23, 38, and 50) are missing the treated school University of California - Merced that has its first treated on-time graduations in 2016. Type A schools do not necessarily have completion data for all of the schools in our sample, just schools that drop the Subject Test requirement

Figure 2. Completions Levels by Field of Study



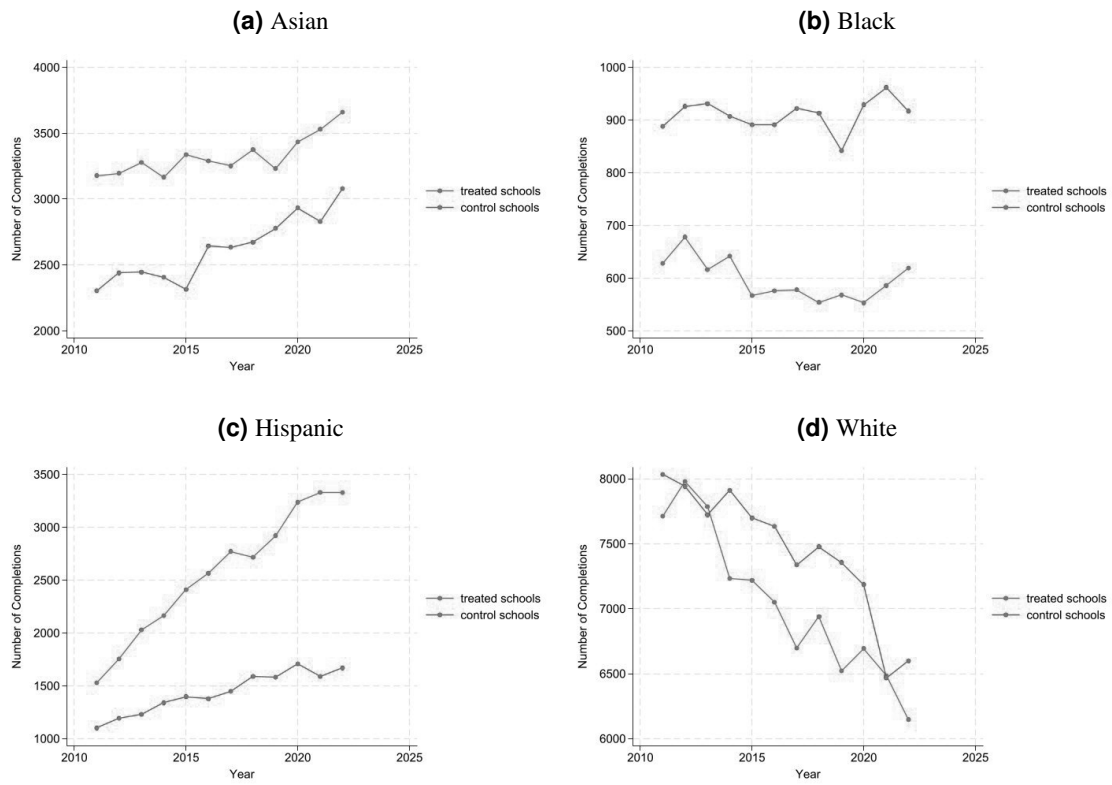
Given a particular major, the number of students of any race receiving bachelor’s degrees in their first major in four years. Fields of study include the humanities, social sciences, and STEM.

Table 3. Effect of Dropping the Subject Tests on Fraction of Completions by Race

Field of Study	Asian	Black	Hispanic	White
Total				
$Post_t \times Treated_s$	-0.0233 (0.0124)	-0.00175 (0.00283)	0.0224 (0.0148)	-0.00958 (0.0129)
Humanities				
$Post_t \times Treated_s$	-0.0444 (0.0122)	-0.00665 (0.00582)	0.0328 (0.0157)	0.000279 (0.0215)
Social Sciences				
$Post_t \times Treated_s$	-0.0326 (0.0161)	0.00133 (0.00725)	0.0345 (0.0160)	-0.0213 (0.0163)
STEM				
$Post_t \times Treated_s$	-0.0155 (0.0156)	-0.00161 (0.00458)	0.0172 (0.0167)	-0.0151 (0.0125)

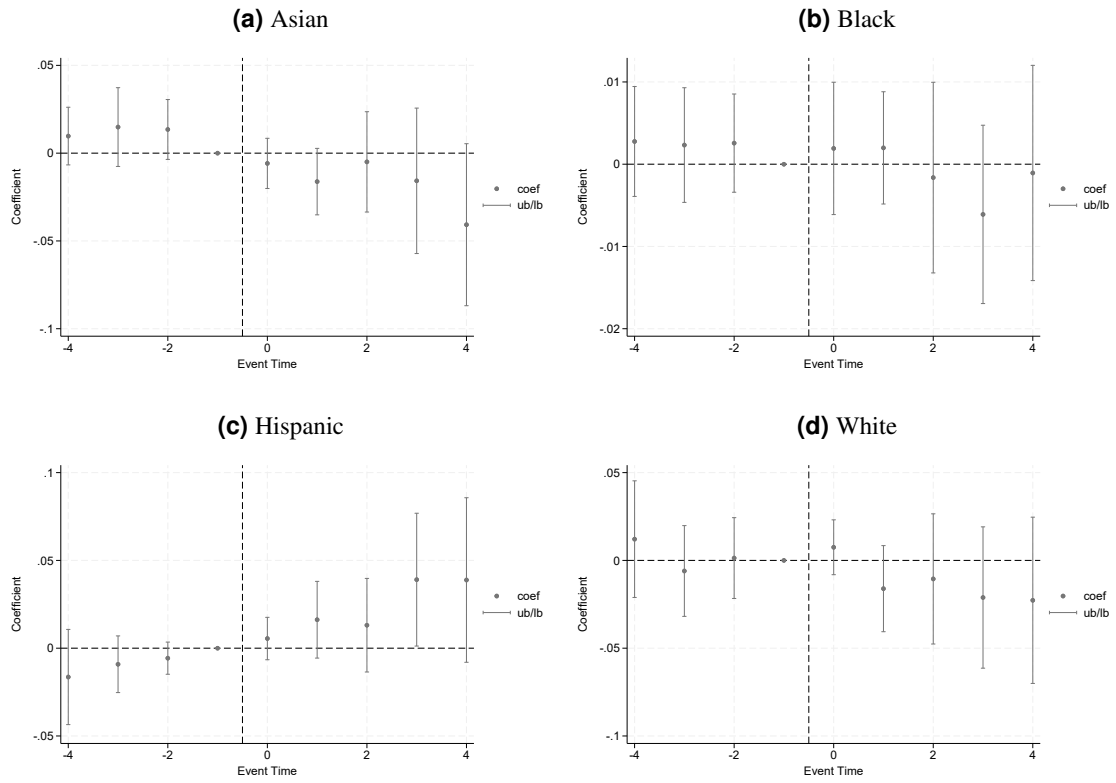
Standard errors in parentheses. Coefficients for variable $Post_{st}$ in regression 1. Includes school and year fixed effects along with an indicator variable for state that school is located in. Each aggregated major draws from the eleven fields of study in Table 2.

Figure 3. Total Completions Levels by Race



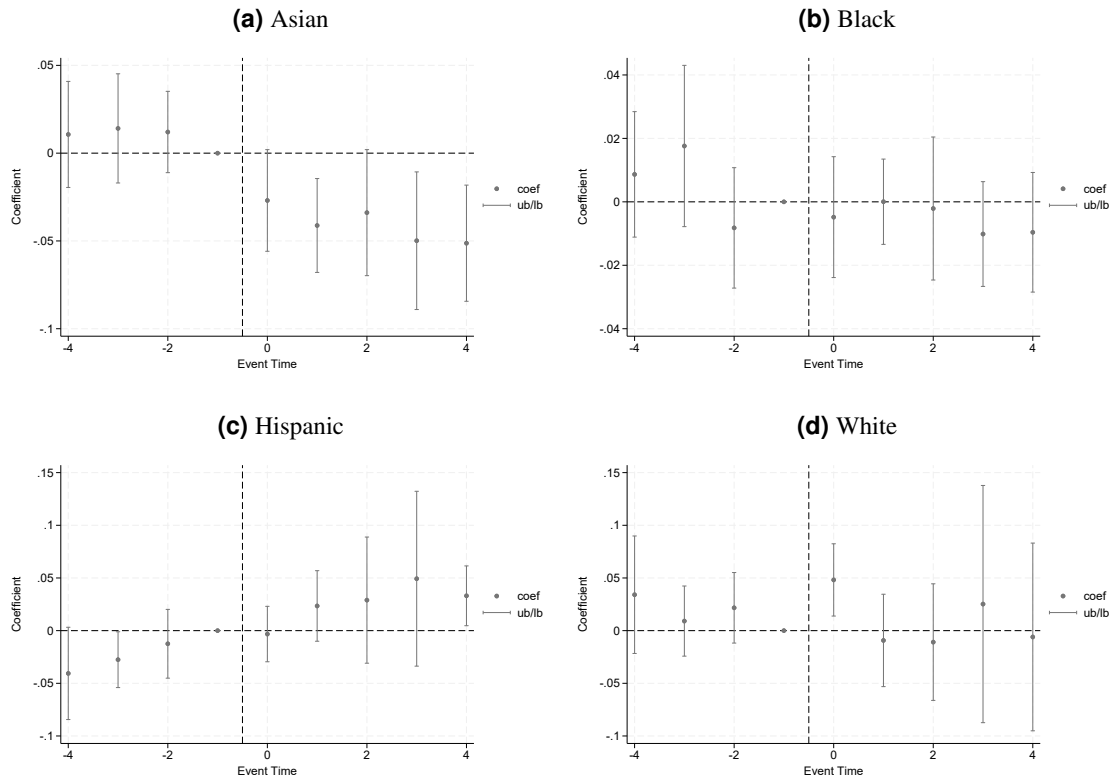
Completion levels by race but not major for eleven fields of study in Table 2

Figure 4. Event Study by Race but not Major



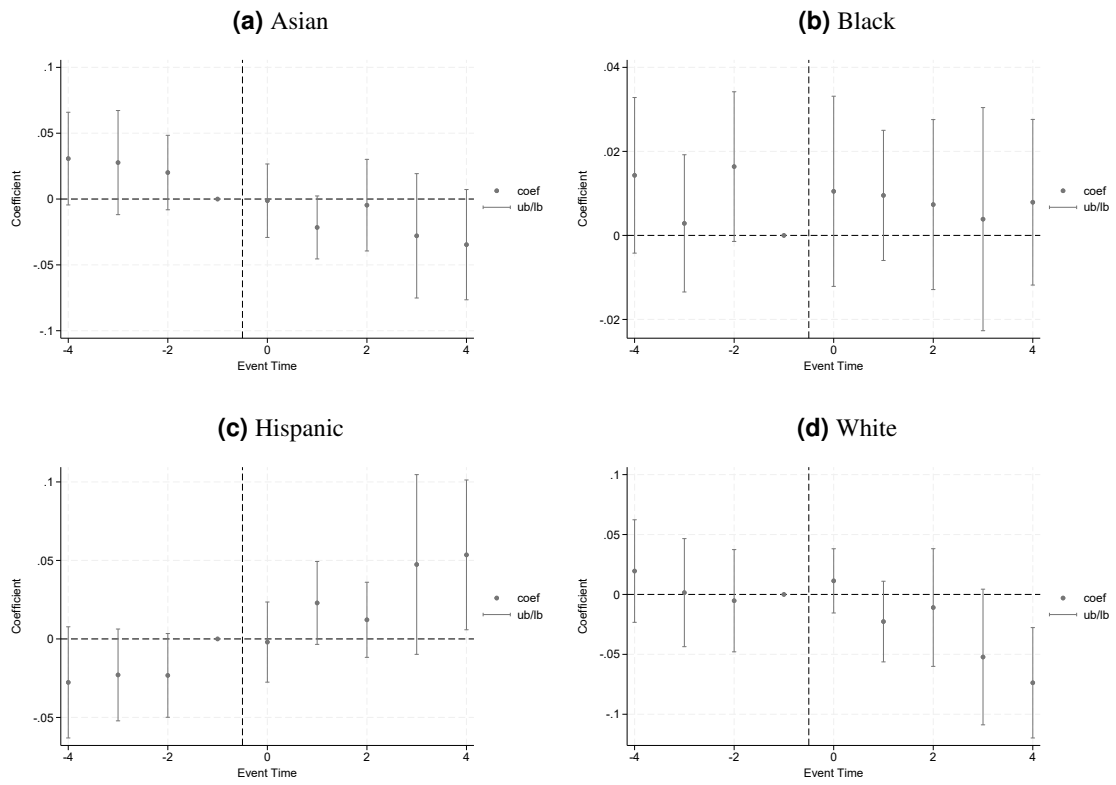
Results from regression 2. Includes 11 fields of study from Table 2 and displays coefficients for event time variables along with the 95% confidence intervals.

Figure 5. Event Study by Race. Humanities Majors



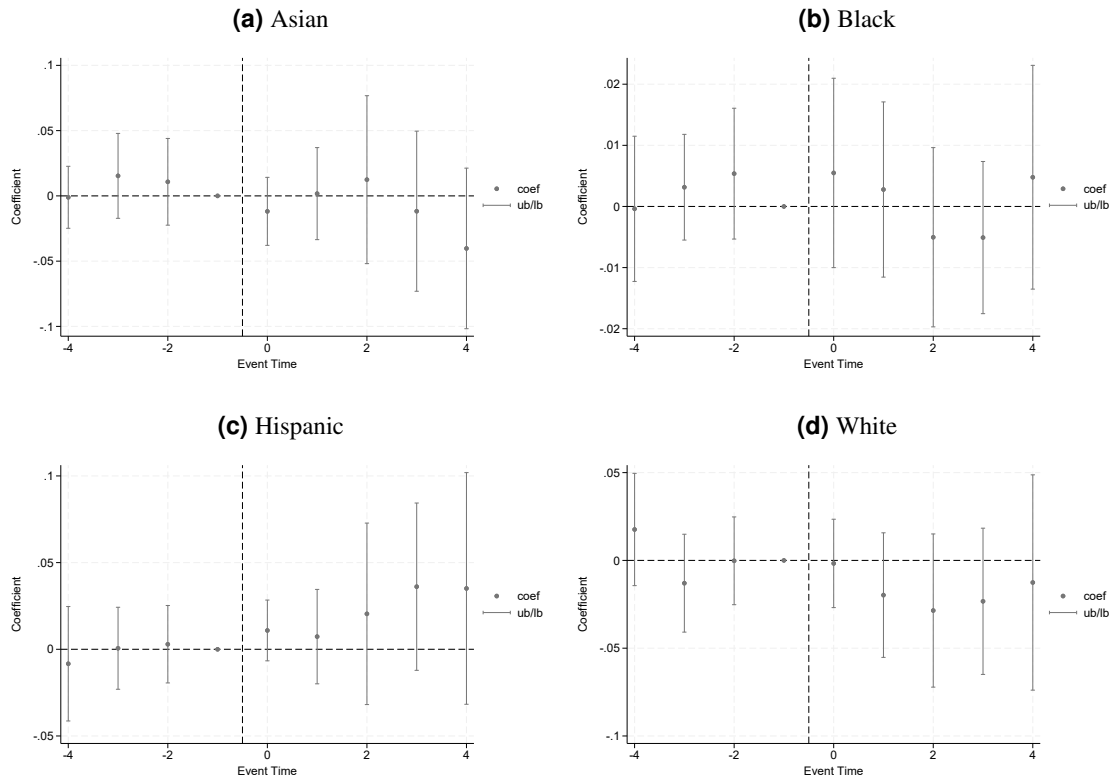
Results from regression 2 that specifies that completions are for students in humanities majors. Displays coefficients for event time variables along with the 95% confidence intervals.

Figure 6. Event Study by Race. Social Sciences Majors



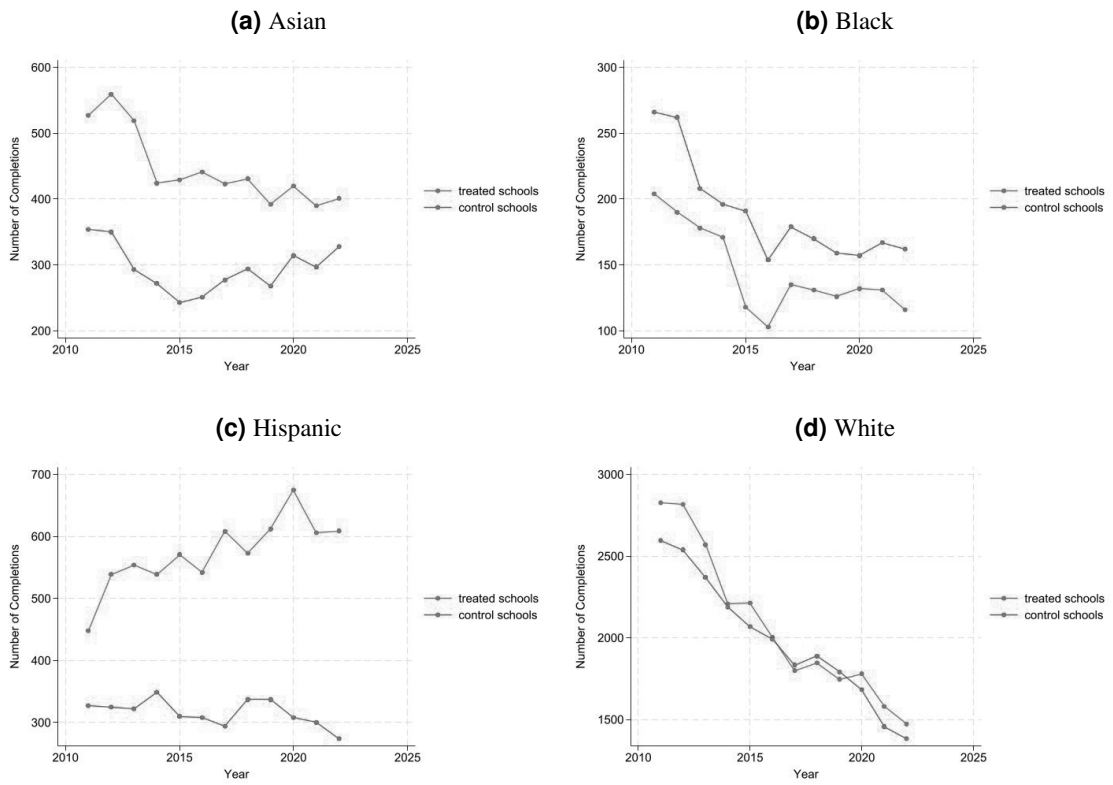
Results from regression 2 that specifies that completions are for students in social sciences majors. Displays coefficients for event time variables along with the 95% confidence intervals.

Figure 7. Event Study by Race. STEM Majors



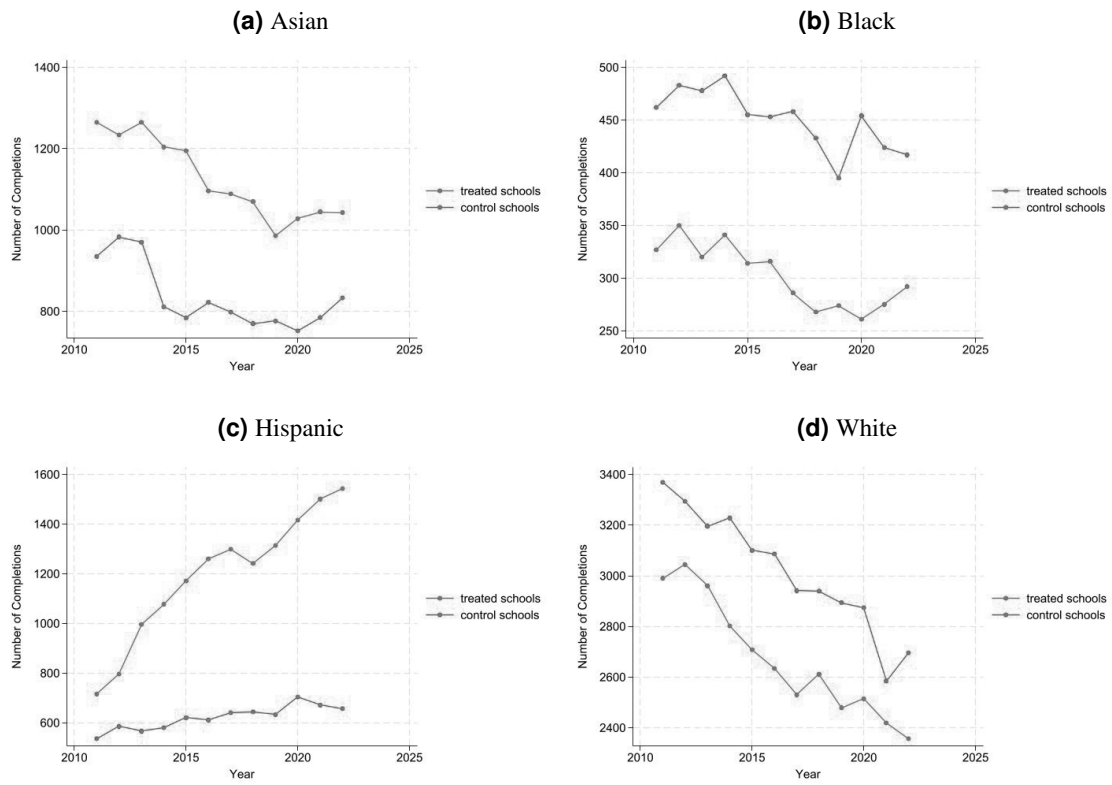
Results from regression 2 that specifies that completions are for students in STEM majors. Displays coefficients for event time variables along with the 95% confidence intervals.

Figure 8. Humanities Completions Levels



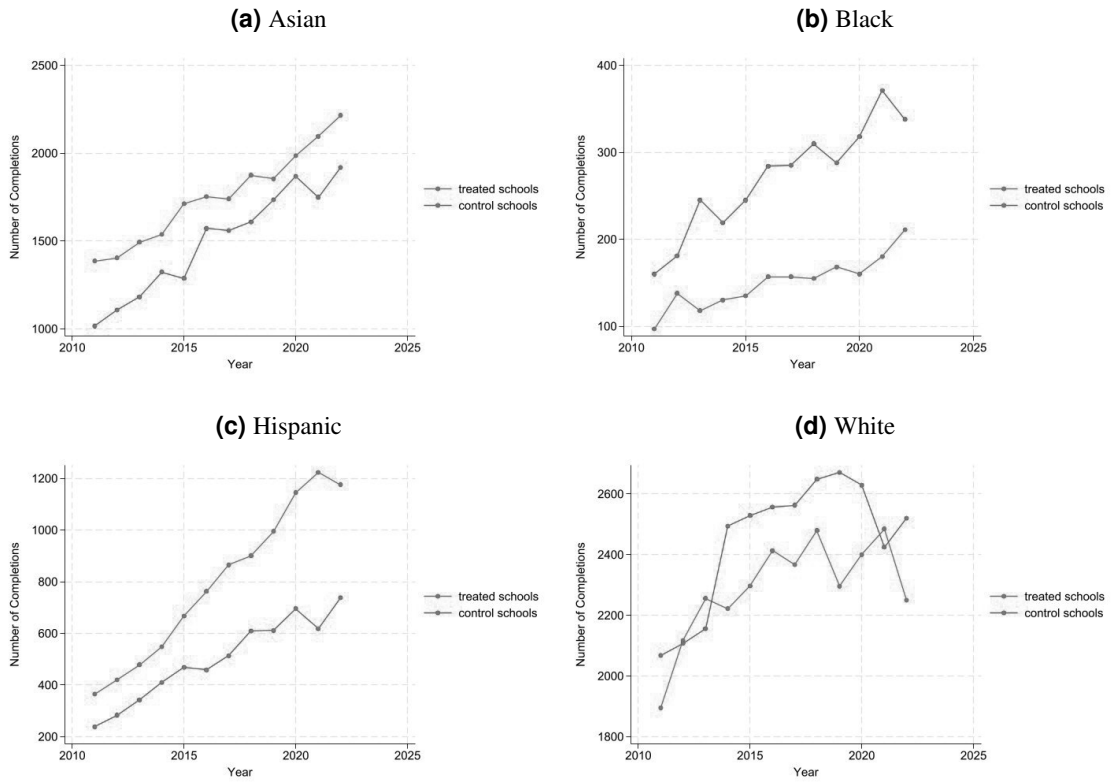
Completion levels by race for students in the humanities.

Figure 9. Social Sciences Completions Levels



Completion levels by race for students in the social sciences.

Figure 10. STEM Completions Levels



Completion levels by race for students in STEM.

An Empirical Analysis of the 3-Points-for-a-Win Rule in the English Premier League

Jeremy Smithline

14.33: Research and Communication in Economics

I. Abstract

Did the introduction of the 3-points-for-a-win rule in the English Premier League (EPL) cause more goals to be scored and less games to end in a tie? Using data from the seven years before and after the introduction of the rule, this paper estimates the causal effect of the rule change on goals scored and tie frequency in the EPL. Using a difference-in-differences approach and an event study approach, this paper finds nearly no evidence that the rule change had an effect on the outcomes of interest. These results suggest that the 3-points-for-a-win rule was largely a failure in the EPL.

II. Introduction

Prior to 1981, all soccer (association football) leagues awarded 2 points for each win, 1 point for each tie, and 0 points for each loss. Starting with the 1981-82 season, the EPL began awarding 3 points for each win. The EPL is played as a round-robin style competition in which every team plays every other team twice and the teams are ranked by the number of points they have come the end of the season. There is no play-off. In the decade or so following the EPL's adoption of the rule change, other leagues around the world adopted the 3-points-for-a-win rule. Eventually FIFA adopted the rule in 1995, meaning every soccer league was to abide by it. The goal of the rule change was to encourage teams to push harder for wins, as opposed to being

content with ties, leading to more exciting and attack-oriented soccer. This, in turn, would in theory increase the entertainment value of soccer matches and increase spectatorship.

The two most obvious ways to measure the success of the rule change are in goals scored per match and frequency of ties. If the rule change worked as intended, there should be more goals per match and a lower tie frequency. The aim of this paper is to assess the success of the 3-points-for-a-win rule in the EPL using those two metrics. Specifically, this paper asks whether the rule change causes more goals to be scored and less games to end in a tie. Answering these questions would contribute to our understanding of incentives in professional sports. This would lead to better decision-making when considering future potential rule changes and better the long-term entertainment value of professional sports. The EPL is particularly important to investigate because it is the most watched sports league in the world.

Previous work on this topic has led to mixed results. Brocas and Carillo (2004) analyze the rule change from a theoretical game theory perspective, finding that under the new rule teams may play more defensively, contrary to the intuition of those who spearheaded the rule change. Dilger and Geyer (2009) use a difference-in-differences approach and find empirical evidence that the rule change increased goal scoring and decreased tie frequency in the German soccer league. However, Hon and Parinduri (2014) use a regression discontinuity design on German soccer league data and find no evidence that the number of goals scored went up or that tie frequency went down as a result of the rule change.

This paper tackles the problem via a difference-in-differences regression approach and via an event study. Games from the FA Cup, an elimination-style English soccer tournament, are used as controls because the rule change does not affect that competition but many of the same teams compete in it and its rules are otherwise identical to the EPL. This paper finds no evidence

that the rule change worked as intended, neither causing an increase in goal scoring nor a decrease in tie frequency. Section III presents the data used, section IV discusses the methods in detail, section V estimates the coefficients of the models, section VI offers concluding remarks, section VII displays tables, and section VIII displays figures.

III. Data

EPL data was obtained from FBref.com at <https://fbref.com/en/comps/9/history/Premier-League-Seasons>. The data covers the seven years before and after the implementation of the rule change, which include the 1974-75 through 1987-88 seasons. In all, 6,426 EPL games are included in the dataset. The EPL is the highest level of the English football pyramid. For all seasons included, except for the 1987-88 season, 22 teams competed. In the 1987-88 season, 21 teams competed. The dataset includes game-level data in which we have the final score of each game included. All EPL games are 90 minutes regardless of if the game is tied at the end of regulation and ties are allowed. More granular game events, such as the time goals are scored, substitutions, and the issuing of red and yellow cards, are not available. Summary statistics for games both before and after the rule change are provided in Table 1.

FA Cup data was obtained from Worldfootball.net at <https://www.worldfootball.net/schedule/eng-fa-cup/>. This is also game-level score data for the seasons 1974-75 through 1987-88. The FA Cup is a single-elimination tournament in which teams that win in a given round advance to the next round. Teams from many levels of the English football pyramid enter the tournament at various stages of the competition. Teams from the EPL enter during the round of 64. For this reason, the dataset includes matches from the

round of 64 onward for all seasons considered. Notably, this means many of the matches include teams from levels of the English football pyramid below the EPL. For all seasons considered, if a match prior to the semi-finals was tied at the end of 90 minutes of regulation, the match was deemed a draw and a replay match was played at a later date to determine which team would advance. This process was repeated until a team won in regulation. Starting with the 1980-81 season and continuing to the end of the seasons included in the dataset, semi-final and final matches tied at the end of regulation went to a 30-minute extra time period, followed by a replay match if tied after extra time. For the sake of consistency, the dataset includes only scores after 90 minutes of regulation. In addition, no replay matches are included in the dataset so as not to introduce any biases from the disproportionate representation of matches between teams even in skill level. In all, 883 FA Cup matches are included in the dataset. Summary statistics for games both before and after the rule change are provided in Table 2. GOALS represents the number of goals scored in a given game and TIE is a 1 if the game ended in a tie and 0 otherwise.

In addition to the summary statistics provided in Table 1 and Table 2, Figure 1 is a histogram of goals scored across all EPL and all FA Cup games considered. It demonstrates the exogenous variation in goals scored necessary for the study.

IV. Methods

We will employ difference-in-differences and event study approaches in order to estimate the effect of the rule change on EPL games. Specifically, we aim to determine the effect on tie frequency and goal scoring, so for each game considered the response variables of interest are whether or not the game ended in a tie and the total number of goals scored by both teams. The treatment in question is the application of the 3-points-for-a-win rule. The control group will be

the set of FA Cup games for which the rule does not apply because it is an elimination-style tournament. Other than the distinction between a round-robin and elimination-style tournament, the EPL and FA Cup games are played using the exact same set of rules.

The basic difference-in-differences equation to be estimated is

$$Y_{it} = c + \alpha \cdot TREAT_i + \beta \cdot POST_t + \delta \cdot (TREAT_i \times POST_t) + \varepsilon_{it}$$

where Y_{it} is the response variable (either binary to indicate if there was a tie or number of goals scored), $TREAT_i$ indicates whether the data point i is in the treated group (EPL games), $POST_t$ indicates whether data point i is in the post period (1981-82 onward), and ε_{it} is an error term.

The coefficient of interest is δ , which is the treatment effect. Thus, the value obtained for the estimator $\hat{\delta}$ of δ represents our estimate of the average increase in the response variable caused by the rule change. The hypothesis that the rule change caused fewer ties to occur corresponds to an estimator $\hat{\delta}$ that is less than zero and statistically different from zero when Y_{it} indicates a tie.

When Y_{it} is total goals scored, $\hat{\delta}$ that is statistically greater than zero corresponds to the hypothesis that the rule change caused more goals to be scored. The first season in which the rule change took effect, the 1981-82 season, is omitted from the estimation. This is to allow an acclimation period to the rule change. While teams can change some aspects of their tactics instantly, they cannot, for example, make personnel changes instantly. The acclimation period allows for a more full set of tactical changes to take effect.

The equation to be estimated for an event study is

$$Y_{it} = c + \alpha \cdot TREAT_i + \beta_t + \sum_{t \neq 74} \delta_t \cdot (TREAT_i \times \beta_t) + \varepsilon_{it}$$

where Y_{it} is the response variable (either binary to indicate if there was a tie or number of goals scored), $TREAT_i$ indicates whether the data point i is in the treated group (EPL games), β_t indicates whether data point i is from season t , and ε_{it} is an error term. We denote a season by the year in which it started, so that the 1974-75 season is denoted by $t = 74$. The first season where the rule change was in place corresponds to $t = 81$. The term with $t = 74$ is excluded from the sum for multicollinearity reasons. The coefficients of interest are the δ_t for $t \geq 81$, which are the treatment effect. Thus, the value obtained for the estimator $\hat{\delta}_t$ of δ_t represents our estimate of the average increase in the response variable in season t caused by the rule change. The hypothesis that the rule change caused fewer ties to occur corresponds to estimatorz $\hat{\delta}_t$ that are less than zero for $t \geq 81$ and statistically different from zero when Y_{it} indicates a tie. When Y_{it} is total goals scored, $\hat{\delta}_t$ that are statistically greater than zero for $t \geq 81$ correspond to the hypothesis that the rule change caused more goals to be scored.

There are two key identifying assumptions necessary for the use of a difference-in-differences approach, no anticipation and parallel trends. No anticipation means that the impending rule change did not cause EPL teams to behave differently prior to the rule change's implementation relative to how they would have behaved in the absence of the rule change. This is entirely reasonable. EPL teams would gain no strategic advantage from implementing tactical adaptations a year early because any tactical adaptations could be implemented with just as much effectiveness once the rule change took effect. Parallel trends means that in the absence of the rule change, the expected difference between each response variable in the pre and post periods is equal for EPL and FA Cup games. This is less self-evident

than no anticipation. It is possible that fundamental differences exist between the EPL and FA Cup games, causing this assumption to not hold. One possible source of difference is the fact that an expanded set of teams participate in the FA Cup, so games may be more lopsided on average. Another source of difference may be the fact that teams approach FA Cup games differently, for example benching top players, as a result of taking those games less seriously.

One way to test the parallel trends assumption visually is to plot the outcome variables over time for both competitions and fit a regression line for each. The lines prior to the rule change should be parallel. Figure 2 displays GOALS regressed on Season for EPL games prior to treatment. The estimated slope of the regression line is $\hat{\beta}_{EPL} = -.00487$ with a standard error of $\hat{s}_{EPL} = .0146$. Figure 3 displays the same for FA Cup games prior to treatment. The estimated slope of the regression line is $\hat{\beta}_{FA} = .0743$ with a standard error of $\hat{s}_{FA} = .0407$. Neither slope is statistically different from 0. If we assume that the two slopes are independent, we may test against the null hypothesis that their difference is 0 by estimating the standard error of the

difference $\hat{s}_{Dif} = \sqrt{\hat{s}_{EPL}^2 + \hat{s}_{FA}^2} = .0432$. Given that $\hat{\beta}_{EPL} - \hat{\beta}_{FA} = -.0456$, we do not reject the null hypothesis that the difference of the slopes is 0. One reason the seemingly positive slope displayed in Figure 3 is not concerning is because each season of FA Cup data only includes 63 games, leading to a high degree of random variation. This is reflected in the slope not being significantly different from 0 due to its high standard error.

The above process is repeated for the outcome variable TIE in Figures 4 and 5. For EPL games the estimate of the slope is .000232 with a standard error of .00396. For FA Cup games the estimate of the slope is .0102 with a standard error of .0110. Just as with goals, neither slope is statistically different from 0. Assuming the two slopes are independent, we can again calculate

the standard error of the difference of the slopes in order to test against the null hypothesis that their difference is 0. The estimate for the difference comes out to $-.00997$ with a standard error of $.0117$, meaning we do not reject the null hypothesis that the two slopes are identical.

Another way to test the parallel trends assumption would be to use the event study model. We would hope to see coefficients δ_t on the interactions between $TREAT_i$ and the season fixed effects that are not statistically significant for seasons prior to treatment, which are the seasons when $t \leq 80$.

Other than a breakdown of the parallel trends assumption, there are other potential issues the study may encounter. It is possible that the effect of the rule change differs depending on the relative strength of the teams. For example, teams may be more enticed to push for wins aggressively when they are much better than their opponent. In addition, if one team becomes more attack-minded, their opponent may respond by becoming more defensive, obscuring the effect of the rule change detectable by looking at tie frequency or goals scored.

V. Estimation

We first estimate the basic difference-in-differences equation without season fixed effects for goals scored and tie frequency. The results for goals scored are displayed in Table 3. The estimate $\hat{\delta}$ of the treatment effect is $.225$. That means we estimate an additional $.225$ goals scored per game as a result of the rule change. That corresponds to approximately a 8.5% increase in goals scored per game. Despite this seemingly notable effect, none of the coefficients, except for the constant, are statistically different from 0 at the 5% level. The p-value corresponding to the coefficient $\hat{\delta}$ is $.071$. This means that at the 5% level we do not reject the null hypothesis that the introduction of the 3-points-for-a-win rule had no effect on goals scored in EPL games. There is

mild evidence, if any, for the conclusion that the rule worked as intended to push teams to score more goals.

The results for ties are displayed in Table 4. In this case, the estimate $\hat{\delta}$ of the treatment effect is -.002. This corresponds to an decrease in tie frequency of .2 percentage points as a result of the rule change. Again, none of the coefficients, except for the constant, are statistically different from 0 at the 5% level. The p-value corresponding to the coefficient $\hat{\delta}$ is .947, which is extremely high. This means that we do not reject the null hypothesis that the introduction of the 3-points-for-a-win rule had no effect on tie frequency in EPL games. According to this specification, there is no evidence for the conclusion that the rule worked as intended to motivate teams to push harder for wins.

Now, we estimate the event study specification with season fixed effects for goals scored and tie frequency. The results for goals scored are displayed in Table 5. None of the interactions between the season fixed effects and the $TREAT_i$ indicator for the post period are statistically different from 0 at the 5% level. This means that at the 5% level we do not reject the null hypothesis that the introduction of the 3-points-for-a-win rule had no effect on goals scored in EPL games. This contributes to our mounting evidence that the rule did not work as intended to push teams to score more goals. The results for ties are displayed in Table 6. None of the interactions between the season fixed effects and the $TREAT_i$ indicator for the post period are statistically different from 0 at the 5% level except for that of the 1987-88 season. This means that for the most part at the 5% level we do not reject the null hypothesis that the introduction of the 3-points-for-a-win rule had no effect on tie frequency in EPL games. This adds to the evidence displayed in the basic difference-in-differences regression for tie frequency demonstrating that the rule did not work as intended to motivate teams to push more for wins.

Upon a first glance at the summary statistics in Table 1, the lack of significant effect of the rule change appears surprising. The seven years following the rule change saw increased goals scored and decreased tie frequency. In the case of tie frequency, making conclusions based on those preliminary summary statistics is complicated by the fact that the FA Cup, where the rule change was irrelevant, saw an even greater decrease in tie frequency than the EPL did. However, the FA Cup also saw a decrease in goals scored. This is reflected in the fact that the estimated $\hat{\delta}$ for the goals scored outcome variable has a p-value of .102, which is much smaller than the p-value for the $\hat{\delta}$ corresponding to the tie outcome variable.

One interpretation of the negative results is that the limited variation in goals scored per game in soccer means that an extremely large sample of games would be needed in order to detect the rule change having an effect on game strategy. To illustrate this point, we consider a power calculation. Assuming independence of EPL and FA Cup games, we may calculate the standard deviation of the difference between the mean goals per game in the EPL after the rule change and the mean goals per game in the FA Cup after the rule change using the second rows of Tables 1 and 2. The standard deviation comes out to

$\sqrt{1.691^2/3192 + 1.582^2/441} = .0811$. This implies that in order to be statistically different at the 5% level, the means in the second rows of Tables 1 and 2 would need to have a difference of at least $1.96 * .0811 = .159$, which they do not. A similar calculation for ties using the fourth rows of Tables 1 and 2 finds the standard deviation of the difference of their means to be .0225. This implies the means in the fourth rows of the tables would need to have a difference of at least .0441 in order to be statistically different at the 5% level, which they do not.

VI. Conclusion

The analysis of this paper finds nearly no evidence that the introduction of the 3-points-for-a-win rule in the EPL caused teams to score more goals or push harder for wins. This was done via a difference-in-differences approach and an event study and results are consistent across specifications. However, it is possible that a more nuanced study is required to ascertain the full effect of the 3-points-for-a-win rule. Perhaps, a robust set of control variables could be added in order to boost statistical confidence given, for example, the limited variation in goals scored per game in soccer. Variables to consider include the progression of spending by teams, the globalization of the pool from which players are drawn, or the progression of financial incentives that the league system gives teams to place highly or avoid relegation. For example, increased spending and an increasingly global pool of players would increase team quality. If such an increase in team quality occurred unevenly across teams, competitiveness would go down, leading to less ties and more goals as the gap between top and bottom teams grows. Also, it is possible that the rule has an effect but only in particular scenarios. Factors that might impact the effect of the rule are the skill gap between teams and the current league standings. Moreover, it's possible that the rule impacts team behavior, but does so in a way that is more complicated than can be detected by the methods used in this paper. For example, as one team becomes more attack-minded, the other may become more defense-minded. Data that could be used to analyze the presence of such a dynamic are team formations and lineups and the location of ball possession (one very offensive team and one very defensive team might lead to increased possession in the defensive team's defensive third of the field).

The results of this paper call into question the intuition behind the 3-points-for-a-win rule. This serves to warn sports rule-makers of the complexity of incentives in sports, especially in games as free-flowing as soccer. Given that the rule change does not seem to offer the benefits

it promised, it's unclear why it should remain in world soccer. It can be reasonably argued that a 2-points-for-a-win rule is more fair because it does not disproportionately reward wins over ties. If a reversion back to 2-points-for-a-win will not lead to less exciting soccer, it's unclear why it should not be implemented.

VII. Tables

Table 1–Summary Statistics for EPL Games

Variable	After rule change?	Mean	Standard deviation	Minimum	Maximum	n
GOALS	No	2.616	1.661	0	10	3234
GOALS	Yes	2.671	1.691	0	11	3192
TIE	No	.283	.451	0	1	3234
TIE	Yes	.258	.437	0	1	3192

GOALS represents the number of goals scored in a given game. TIE is a 1 if the game ended in a tie and 0 otherwise.

Table 2–Summary Statistics for FA Cup Games

Variable	After rule change?	Mean	Standard deviation	Minimum	Maximum	n
GOALS	No	2.674	1.714	0	8	442
GOALS	Yes	2.533	1.582	0	8	441

TIE	No	.310	.463	0	1	442
TIE	Yes	.270	.444	0	1	441

GOALS represents the number of goals scored in a given game. TIE is a 1 if the game ended in a tie and 0 otherwise.

Table 3—Basic Difference-in-Differences Regression Results for Goals Scored

total_goals	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
1.treat	-.0579435	.084633	-0.68	0.494	-.2238508	.1079639
1.post	-.1477531	.1169185	-1.26	0.206	-.3769501	.0814438
treat#post						
1 1	.2252613	.1247053	1.81	0.071	-.0192003	.4697229
_cons	2.674208	.079382	33.69	0.000	2.518594	2.829822

n=7,309. 1.treat represents the $TREAT_i$ indicator for EPL games. 1.post represents the $POST_t$ indicator for games played in the post period. No coefficients are statistically significant at the 5% level. We have $R^2 = .0008$.

Table 4—Basic Difference-in-Differences Regression Results for Tie Frequency

tie	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
1.treat	-.0267142	.0226222	-1.18	0.238	-.0710608	.0176324
1.post	-.0242405	.031252	-0.78	0.438	-.0855042	.0370233
treat#post						
1 1	-.0022235	.0333334	-0.07	0.947	-.0675675	.0631204
_cons	.3099548	.0212186	14.61	0.000	.2683596	.3515499

n=7,309. 1.treat represents the $TREAT_i$ indicator for EPL games. 1.post represents the $POST_t$ indicator for games played in the post period. No coefficients are statistically significant at the 5% level. We have $R^2 = .0013$.

Table 5–Event Study Regression Results for Goals Scored

total_goals	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
1.treat	.450938	.2244918	2.01	0.045	.0108689	.891007
season						
75	.1904762	.2978221	0.64	0.522	-.3933414	.7742938
76	.6984127	.2978221	2.35	0.019	.1145951	1.28223
77	1.012897	.2966564	3.41	0.001	.4313642	1.594429
78	.5555556	.2978221	1.87	0.062	-.028262	1.139373
79	.4920635	.2978221	1.65	0.099	-.0917541	1.075881
80	.5396825	.2978221	1.81	0.070	-.0441351	1.1235
81	.3968254	.2978221	1.33	0.183	-.1869922	.980643
82	.2063492	.2978221	0.69	0.488	-.3774684	.7901668
83	.2063492	.2978221	0.69	0.488	-.3774684	.7901668
84	.4285714	.2978221	1.44	0.150	-.1552462	1.012389
85	.5079365	.2978221	1.71	0.088	-.0758811	1.091754
86	.3333333	.2978221	1.12	0.263	-.2504843	.9171509
87	.4285714	.2978221	1.44	0.150	-.1552462	1.012389
treat#season						
1 75	-.1536797	.3174794	-0.48	0.628	-.7760313	.468672
1 76	-.7633478	.3174794	-2.40	0.016	-1.385699	-.1409961
1 77	-.9739358	.3163862	-3.08	0.002	-1.594144	-.3537272
1 78	-.5468975	.3174794	-1.72	0.085	-1.169249	.0754541
1 79	-.6089466	.3174794	-1.92	0.055	-1.231298	.013405
1 80	-.507215	.3174794	-1.60	0.110	-1.129567	.1151366
1 81	-.4834055	.3174794	-1.52	0.128	-1.105757	.1389462
1 82	-.0959596	.3174794	-0.30	0.762	-.7183112	.526392
1 83	-.1262626	.3174794	-0.40	0.691	-.7486143	.496089
1 84	-.2662338	.3174794	-0.84	0.402	-.8885854	.3561179
1 85	-.3455988	.3174794	-1.09	0.276	-.9679505	.2767528
1 86	-.3290043	.3174794	-1.04	0.300	-.951356	.2933473
1 87	-.5564935	.3184304	-1.75	0.081	-1.180709	.0677224
_cons	2.174603	.210592	10.33	0.000	1.761782	2.587425

n=7,309. 1.treat represents the $TREAT_i$ indicator for EPL games. The season variables represent an indicator for the corresponding season. We have $R^2 = .0051$.

Table 6–Event Study Regression Results for Tie Frequency

tie	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
1.treat	-.049062	.0597671	-0.82	0.412	-.1662228	.0680987
season						
75	-.031746	.07929	-0.40	0.689	-.1871773	.1236853
76	-.015873	.07929	-0.20	0.841	-.1713043	.1395583
77	-.0674603	.0789796	-0.85	0.393	-.2222833	.0873627
78	.015873	.07929	0.20	0.841	-.1395583	.1713043
79	-.047619	.07929	-0.60	0.548	-.2030504	.1078123
80	.0952381	.07929	1.20	0.230	-.0601932	.2506694
81	-.1428571	.07929	-1.80	0.072	-.2982885	.0125742
82	-.031746	.07929	-0.40	0.689	-.1871773	.1236853
83	-.047619	.07929	-0.60	0.548	-.2030504	.1078123
84	6.06e-14	.07929	0.00	1.000	-.1554313	.1554313
85	.0793651	.07929	1.00	0.317	-.0760662	.2347964
86	-.047619	.07929	-0.60	0.548	-.2030504	.1078123
87	-.1428571	.07929	-1.80	0.072	-.2982885	.0125742
treat#season						
1 75	.0382395	.0845234	0.45	0.651	-.1274508	.2039299
1 76	.0440115	.0845234	0.52	0.603	-.1216788	.2097019
1 77	.0847763	.0842323	1.01	0.314	-.0803435	.2498961
1 78	.027417	.0845234	0.32	0.746	-.1382733	.1931074
1 79	.0692641	.0845234	0.82	0.413	-.0964263	.2349544
1 80	-.1082251	.0845234	-1.28	0.200	-.2739154	.0574652
1 81	.1363636	.0845234	1.61	0.107	-.0293267	.302054
1 82	.0036075	.0845234	0.04	0.966	-.1620828	.1692978
1 83	.034632	.0845234	0.41	0.682	-.1310583	.2003224
1 84	-.0367965	.0845234	-0.44	0.663	-.2024869	.1288938
1 85	-.1096681	.0845234	-1.30	0.195	-.2753585	.0560222
1 86	.0454545	.0845234	0.54	0.591	-.1202358	.2111449
1 87	.1887446	.0847766	2.23	0.026	.0225579	.3549313
_cons	.3174603	.0560665	5.66	0.000	.2075538	.4273669

n=7,309. 1.treat represents the $TREAT_i$ indicator for EPL games. The season variables represent an indicator for the corresponding season. We have $R^2 = .0056$.

VIII. Figures

Figure 1–Histogram of Goals Scored in all EPL and FA Cup Games

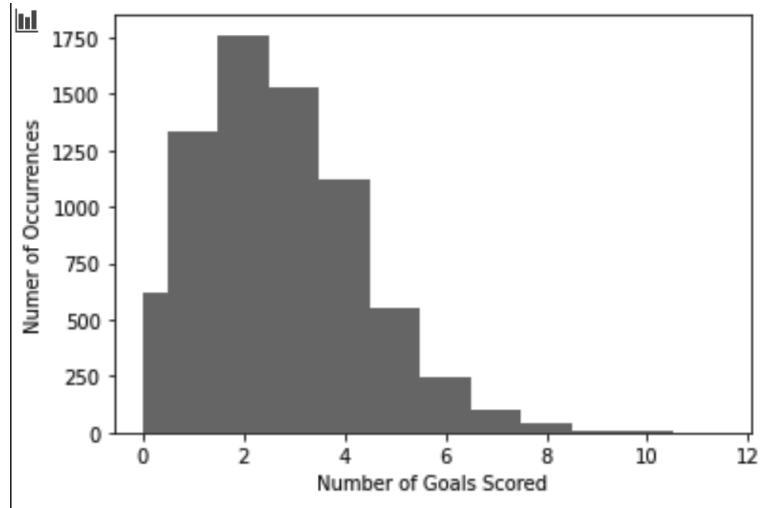
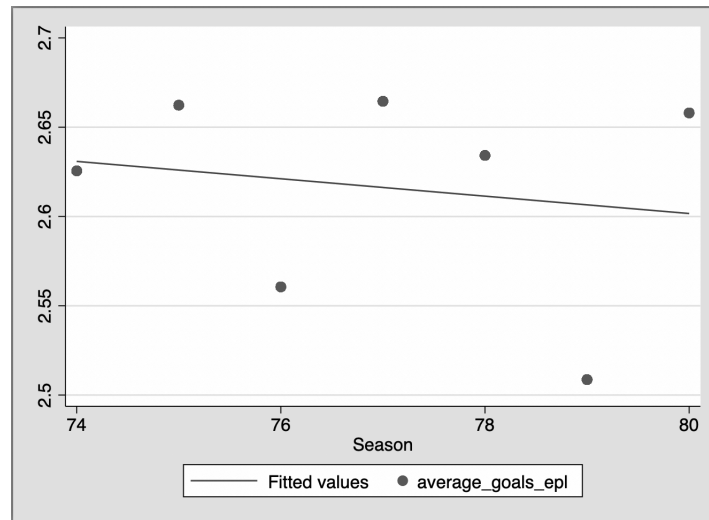
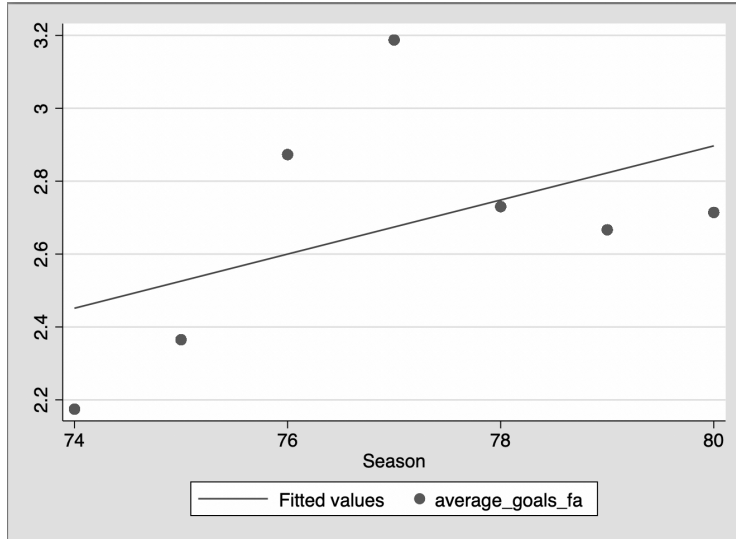


Figure 2–Relationship Between Goals Scored and Season Prior to Treatment in EPL



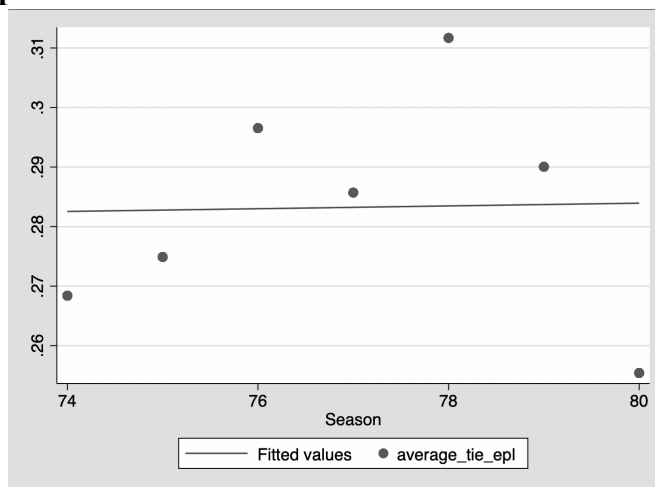
GOALS is regressed on Season for EPL games before the rule change and the line of best fit is shown. Also plotted are the means of GOALS for each season. The estimated slope of the line is $-.00487$ with a standard error of $.0146$.

Figure 3–Relationship Between Goals Scored and Season Prior to Treatment in FA Cup



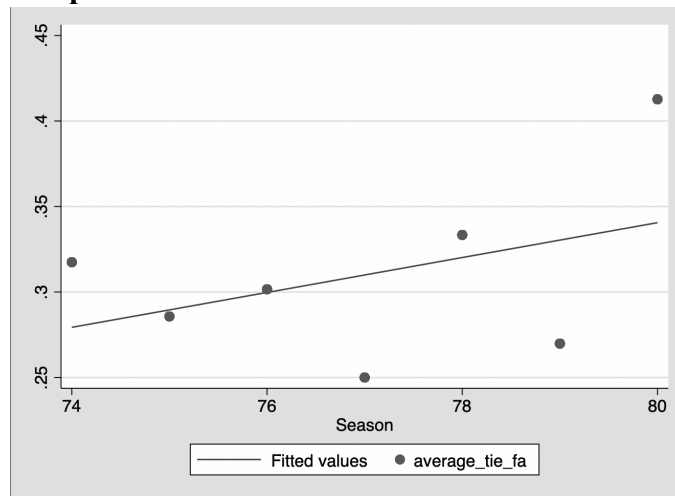
GOALS is regressed on Season for FA Cup games before the rule change and the line of best fit is shown. Also plotted are the means of GOALS for each season. The estimated slope of the line is .0743 with a standard error of .407.

Figure 4—Relationship Between Tie Rate and Season Prior to Treatment in EPL



TIE is regressed on Season for EPL games before the rule change and the line of best fit is shown. Also plotted are the means of TIE for each season. The estimated slope of the line is .000232 with a standard error of .00396.

Figure 5—Relationship Between Tie Rate and Season Prior to Treatment in FA Cup



TIE is regressed on Season for FA Cup games before the rule change and the line of best fit is shown. Also plotted are the means of TIE for each season. The estimated slope of the line is .0102 with a standard error of .0110.

Works Cited

Brocas, Isabelle, and Juan D. Carrillo. “Do the ‘three-point victory’ and ‘golden goal’ rules make soccer more exciting?” *Journal of Sports Economics*, vol. 5, no. 2, 2004, pp. 169–185, <https://doi.org/10.1177/1527002503257207>.

Dilger, Alexander, and Hannah Geyer. “Are three points for a win really better than two?” *Journal of Sports Economics*, vol. 10, no. 3, 2009, pp. 305–318, <https://doi.org/10.1177/1527002508327521>.

Hon, Lee Yoong, and Rasyad A. Parinduri. “Does the three-point rule make soccer more exciting? evidence from a regression discontinuity design.” *Journal of Sports Economics*, vol. 17, no. 4, 2014, pp. 377–395, <https://doi.org/10.1177/1527002514531790>.