University of California
Santa Barbara

Essays on Education and Comparative Advantage in Trade, Immigration and Gender Disparity in Film Industry

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Economics

by

Hongyuan Jin

Committee in charge:

Professor Javier Birchenall, Chair
Professor Peter Kuhn
Professor Peter Rupert

June 2022

The Dissertation of Hongyuan Jin is approved.

Professor Peter Kuhn

Professor Peter Rupert

Professor Javier Birchenall, Committee Chair

May 2022

Essays on Education and Comparative Advantage in Trade, Immigration and Gender Disparity in Film Industry

Copyright © 2022

by

Hongyuan Jin

I dedicate this paper to myself.

Acknowledgements

I am indebted to my committee members, Javier Birchenall, Peter Kuhn, and Peter Rupert. They provided me with a tremendous amount of help, support, and encouragement. I also thank Aashish Mehta, who is a great mentor, researcher, and inspiration. All the incompleteness is mine from constant procrastination and meandering.

Curriculum Vitæ

Hongyuan Jin

Education

2022	Ph.D. in Economics (Expected), University of California, Santa Barbara.
2017	M.A. in Economics, University of California, Santa Barbara.
2022	M.A. in Statistics (Expected), University of California, Santa Barbara.
2015	M.S. in Economics, University of Wisconsin-Madison.

Research Papers in Progress

"The Influence of Foreign-born Directors on the US Film Industry"

"Gender Gaps in Productivity and Labor Market Opportunities: The Celluloid Ceiling and Film Directors' Career Trajectories" (with Maxine J. Lee)

"Education and the Evolution of Comparative Advantage" (with Jesus Felipe and Aashish Mehta)

Abstract

Essays on Education and Comparative Advantage in Trade, Immigration and Gender Disparity in Film Industry

by

Hongyuan Jin

This dissertation consists of three separate papers on topics in the economics of labor markets and economic development. The first chapter is "Education and the Evolution of Comparative Advantage." We provide substantial evidence that education is helpful for workers in low-and middle-income countries to accumulate human capital and is helpful for the countries to develop comparative advantage in products that are unrelated to what they have been exporting. In contrast, controlling for the relatedness of target products to these countries' exports, education appears unimportant for developing comparative advantage in products that are intrinsically complex or education intensive.

The second chapter is "The Influence of Foreign-born Directors on the US Film Industry." This paper studies the effect of high-skilled immigrant labor on the production of cultural product in the U.S. With director-distributor matched data, I disentangle the director's effect from the distributor's effect by controlling for a full set of individual distributors' fixed effects. I further controlled for the estimated production budget to capture the value added by a director assignment that is additional to the quality of a film prior to director assignment. A foreign-born director result in a differential effect on the box office revenues in the domestic and international domain–29.9% international box office, and 16.4% lower domestic box office, and it suggests that the value of a foreign-born director is likely in mitigating the "foreignness liability" of the American films in the international market and generating a higher revenue overseas.

The final chapter is "Gender Gaps in Productivity and Labor Market Opportunities: The Celluloid Ceiling and Film Directors' Career Trajectories." We estimate the gender gap in employment outcomes and its interaction with productivity by following the careers of film directors. We proxy productivity with the audience and critical responses to the director's previous film. According to our analysis, there are no discernible gender gaps among high-productivity directors, but low-productivity women are much less likely to direct another film than comparable men.

Contents

Curriculum Vitae **vi**

Abstract **vii**

1 Education and the Evolution of Comparative Advantage **1**
1.1 Introduction 1
1.2 Specification 5
1.3 Data 10
1.4 Results 14
1.5 Conclusion 36

2 The Influence of Foreign-born Directors on the US Film Industry **39**
2.1 Introduction 39
2.2 Data 50
2.3 Empirical Strategy 61
2.4 Possible Mechanisms 76
2.5 Dynamics of the Absolute and Relative Advantages 86
2.6 Conclusion 91

3 Gender Gaps in Productivity and Labor Market Opportunities: The Celluloid Ceiling and Film Directors' Career Trajectories **93**
3.1 Introduction 93
3.2 Data and Empirical Analysis 99
3.3 Descriptive Results 102
3.4 Regression Results 105
3.5 Conclusion 114

A Appendix **115**
A.1 Deriving the Estimation Equation From a Simple Evolutionary Model of Education, Tacit and Book Knowledge, and Comparative Advantage . . . 115
A.2 Film Clustering with Plot Summary 119
A.3 Tables 121

A.4 Figures 138

Chapter 1

Education and the Evolution of Comparative Advantage

1.1 Introduction

Developing countries significantly improve their economic prospects by developing industries that produce complex, core products.[1] Unfortunately, these new industries require capabilities that must be acquired incrementally through practice, making industrial development difficult and path dependent.[2] Successful leaps into unrelated industries are rare, and the industrial policies responsible for many successes rely on institutional

[1]Economies tend to grow faster when they export more complex products (Hausmann et al., 2014, Hidalgo et al., 2007) and a larger variety of products (Saviotti and Frenken, 2008). A well-diversified and/or complex product mix has been linked to employment that grows faster and is more resilient to shocks (Frenken et al., 2007), to shorter recessions (Hausmann et al., 2006), and to lower inequality (Felipe and Hidalgo, 2015, Hartmann et al., 2017).

[2]We apply standard definitions and concepts: "Capabilities" is shorthand for productive knowledge and practices embedded in individuals, firms, industries, supply chains, and institutions (Hausmann et al., 2014). Industries relying on overlapping capabilities (or inputs) are said to be related (Hidalgo et al., 2018). Path dependency means a tendency to develop new industries that are related to existing industries (Bahar et al., 2019). More complex products require more capabilities. Core products, like chemicals, sophisticated machinery, and advanced scientific instruments, are complex and are therefore related to many other products. Peripheral products, including many agricultural and mined commodities, are less complex and therefore less related to other products.

conditions that are difficult to replicate in other countries.[3] Seeking more neutral alternatives to industrial policy, this paper therefore asks whether education can facilitate the development of new export industries, and it characterizes the types of industrial transitions that education enables. While several papers suggest that education might help develop and sustain a salubrious export mix, these questions have not received focused attention in the literature.[4]

This is a noteworthy omission for at least three reasons. First, theory suggests that education could be important for developing new comparative advantages. The complexity literature and related work rooted in evolutionary theories (Hidalgo et al., 2018, Nelson, 1985, Stiglitz and Greenwald, 2014) regard the acquisition of tacit knowledge through learning by doing and the translation of knowledge across related domains to be key processes by which new capabilities are developed. Education has the potential to speed up learning by doing, and to permit the translation of knowledge across less related domains. In addition, education is thought to enhance actors' abilities to respond to emerging opportunities Schultz (1975)—abilities that are crucial to the building of new industries. Second, while education, particularly high-quality education, has been shown to promote economic growth Hanushek and Woessmann (2008), Krueger and Lindahl (2001), the mechanism underlying this relationship has not been clearly established. It is therefore useful to examine whether helping diversify toward core products might be such a mechanism. Third, some authors credit human capital with facilitating export-

[3]See, for example, Johnson (1986); Amsden et al. (2001); Jomo (2003); Jomo (2013); and Studwell (2013)

[4]Previous studies show that: more rapid employment growth in skill-intensive industries in countries that had more highly educated workers and in those that expanded education faster (Ciccone and Papaioannou, 2009); more educated countries are better able to maintain a diverse export mix in the face of terms-of-trade shocks (Agosin et al., 2012); and primary education attainment is a strong Bayesian predictor of national export diversification (Jetter and Ramírez Hassan, 2015). None of this work examines the role of education in overcoming path dependence. Coniglio et al. (2021) come closest, showing that two crude proxies for education—scientific publication and educational expenditures—are associated with the development of unfamiliar industries in developing countries.

driven growth in East Asia, but do not provide comparative statistical evidence that it facilitates export transformation per se.[5]

This paper examines the role of education in the evolution of comparative advantage using export data for 1,240 different goods for 49 low- and middle-income countries between 1995 and 2015.[6] In particular, we test three hypotheses, each about a role that education could play in altering a country's export mix.

Two of these three hypotheses are motivated by the theory of economic complexity. First, we ask whether countries whose workforces were more educated in 1995 were more likely to develop comparative advantage by 2015 in products that were unrelated to those they exported with comparative advantage in 1995 (i.e., whether they developed strengths in "unfamiliar" products). Second, motivated by the same theory, we ask whether countries with high education levels in 1995 were more likely to develop comparative advantage in products that are intrinsically more complex, controlling for those products' initial relatedness to their export mix. The third hypothesis derives from the Heckscher–Ohlin–Vanek intuition that education expansions should shift the export mix toward more education-intensive products.

Our key finding is that countries whose workforces were more educated in 1995 were indeed more likely to move into unfamiliar products. We also provide some evidence that, as might be expected, good quality basic education and high primary attainment facilitate movements into unfamiliar peripheral products, but not into unfamiliar core products. There is at best weak evidence to support the second hypothesized role of education, and none at all to support the third. We apply a two-step procedure which

[5]See, for example, World Bank (1993), Hobday et al. (1995), and Stiglitz (1996). Others studying this history are more skeptical of education's role (Bank, 2007, Booth, 1999, Chang, 2012, Studwell, 2013)

[6]Industrial development is a qualitatively different phenomenon in rich countries, many of which are deindustrializing. See Figure 1.1 in section 1.3. The complete list of countries is found in Appendix Table A1

confirms that the lack of support for the latter two hypotheses is not driven by errors in the measurement of education or its change over time. We also demonstrate that our results are robust to errors in specification or operationalization; to biases owing to omitted variables related to institutional quality, infrastructure, foreign direct investment (FDI) receipt or regional specialization patterns; and to the fact that countries that underwent fast industrial development prior to 1995 tended to have both higher educational attainment in 1995, and more rapid industrial development between 1995 and 2015. While the instruments required to produce fully credible causal estimates of the effects of education on comparative advantage are not available, the robustness of our results to every alternative explanation suggests that they do provide a useful qualitative indication of the causal connections involved. The primary role of education in industrial development, at least among those we examine, is to help navigate the unfamiliar.

Although the most successful cases of industrialization in developing countries, with the exception of the People's Republic of China (PRC), predate 1995 (Felipe et al., 2019), we study the period 1995–2015 for three reasons. First, there is much more competition between nations for footholds in tradable industries now than in the past, so results from recent times are more relevant for policy. Second, the effectiveness of education should depend not only upon the quantity of schooling obtained, but also on its quality. We proxy for this using cognitive skills measures derived from international standardized tests that are only available beginning in the late 1990s. Third, trade policies vary less across countries after the structural adjustment era.

The paper is structured as follows. We introduce our specifications and hypothesis tests in section 1.2. We derive them in the appendix from a simple trade model that accounts for the connections between education, complexity, and path dependence.[7] We

[7]This model combines a dynamic Ricardian trade model in the spirit of Redding (1999) with some simple assumptions about how the acquisition of tacit knowledge from production produces relatedness, and by extension, how education can facilitate this process. As discussed in section 1.5, tacit knowledge

describe our data and variable definitions in section 1.3 and our results in section 1.4. Section 1.5 concludes.

1.2 Specification

1.2.1 The Single-Stage Specification

We examine the relationship between education, familiarity, product characteristics, and the evolution of comparative advantage using a linear probability model, estimated on a pooled sample of countries and products, respectively indexed by c and p. Our generic, baseline specification, is:

$$\begin{aligned} CA_{c,p,t_1} = \alpha_c + \alpha_p + f(RCA_{c,p,t_0}) + \beta_F F_{c,p,t_o} + \beta_{EF} E_{c,t_0} F_{c,p,t_o} + \gamma_{EC} E_{c,t_0} C_{p,t_o} \\ + \delta_{EE} el_{p,t_0} \Delta el_{c,t} + \delta_{KE} kl_{p,t_0} \Delta kl_{c,t} + e_{c,p,t_1} \end{aligned} \tag{1.1}$$

Our initial time period, t_0 is 1995, and the final time period, t_1, is 2015. Here $RCA_{c,p,t} = (X_{c,p,t}/X_{c,t})/(X_{p,t}/X_t)$ is Balassa's (1965) index of revealed comparative advantage (RCA) at time t, where X denotes exports. Let $CA_{c,p,t_1} \equiv I\{RCA_{c,p,t_1} \geq k\}$ indicate that country c has a comparative advantage in product p in 2015. Our baseline results use $k = 1$. Discretizing RCA sacrifices variation in the dependent variable, but is standard in the literature because it solves a range of econometric problems, and because varying the value of k allows us to check whether results are driven by information loss in particular RCA ranges (Bahar et al., 2014, 2019).[8]

is the most popular of several possible explanations of relatedness (Hausmann et al., 2014, Hidalgo et al., 2018)

[8]RCAs are nonnegative, often zero, and strongly right skewed, suggesting that a corner solution model is required if we are to treat them as continuous. Identification of these models relies on untestable distributional assumptions, and the product fixed effects required by theory also raise incidental parameter problems in a maximum likelihood context (Cameron and Trivedi, 2005). Beyond these consistency problems, corner solution models yield nonlinear conditional expectations functions, which complicates hypothesis testing (Wooldridge, 2010). Log-linearizing RCA, as required by the exponential Churdle

We control for RCA in 1995 to capture long-run drivers of trade patterns, such as history and geography, as well as the availability of human and physical capital prior to t_0. Controlling for lagged RCA also means that our coefficients capture the relationship between the independent variables and changes in comparative advantage. Country fixed effects allow for differences in the general level of development and diversification, while product fixed effects capture complexity, education intensity, and other traits that make it more difficult to evolve comparative advantage in some products than in others.[9] Other than familiarity, $F_{c,p}$, the remaining variables are normalized to have a mean of 0 and a standard deviation of 1. The independent variables and interaction terms are motivated by our two theoretical approaches to the evolution of comparative advantage.

As noted, the economic complexity approach argues that it is difficult to develop comparative advantages in products with which a country is unfamiliar (Hausmann et al., 2014, Hidalgo and Hausmann, 2009, Hidalgo et al., 2007). If this is true, then the evolution of comparative advantage will be path dependent, in the sense that whether a country's comparative advantage in a product grows over a time interval depends upon what else it exported at the start of that period. We will conclude that it is, on average, true for a country with initial education level E_{c,t_0} if $\beta_E + \beta_{EF} E_{c,t_0} > 0$. A positive β_E indicates that a country endowed with average education experiences path dependence. The complexity literature also argues that it is more difficult to develop comparative advantages in intrinsically complex products, because doing so requires more know-how. We therefore utilize product fixed effects to permit the probability of comparative advantage

model also results in findings being driven by differences close to RCA = 0, while using an inverse hyperbolic sin transformation would implicitly assume that starting to export a product poses similar challenges to increasing exports in an already exported product. In contrast, linear probability models are consistent and easy to interpret (Angrist and Pischke, 2008). To ensure that our findings are not specific to the dynamics of comparative advantage around $RCA_{c,p,t_1} = 1$, we follow Bahar, Hausmann, and Hidalgo (2014)Bahar et al. (2014) in estimating the model after discretizing around $RCA_{c,p,t_1} = 0.5, 0.8,$ and 2 (see Table 1.9).

[9]More educated countries export a larger number of products with comparative advantage (Mehta and Felipe, 2014).

to differ between products.

We turn next to our hypotheses regarding the role of education. First, high education levels may help countries to develop comparative advantage in unfamiliar products. This would be expected if education is useful for translating knowledge across domains, for identifying and acquiring required knowledge that was not already available from a country's initial product mix, or for responding to emerging business opportunities. This corresponds to a test of the alternative hypothesis that is negative. Second, holding familiarity constant, higher education levels may predispose countries to develop comparative advantage in intrinsically more complex products. This would be expected if larger amounts of knowledge are required to efficiently produce more complex products, and if education helps acquire this knowledge. If these two conditions hold, should be positive. Appendix 1 derives the first row of specification (1.1) from an N-good Ricardian model, augmented to capture the role of education in promoting the acquisition of tacit and book knowledge over time.

The second row of specification (1.1) captures standard factor abundance considerations. We denote the initial education and capital intensity of products by el_{p,t_0} and kl_{p,t_0}, and the growth in countries' per-worker endowments of education and capital by $\Delta el_{c,t}$ and $\Delta kl_{c,t}$. Our third hypothesis regarding education is that acquiring (rather than initially possessing) more education helps countries gain comparative advantage in education-intensive products. Producing education-intensive products involves more tasks in which educated workers have big productivity advantages over less educated workers. An increase in the number of educated workers is therefore predicted to increase production and export of education-intensive products to clear the market for educated workers Leamer (1984). This corresponds to a test of the alternative hypothesis that $\delta_{EE} > 0$. One would also expect countries accumulating more physical capital per worker to shift more strongly toward capital-intensive products (i.e., $\delta_{KE} > 0$).

We measure education as a vector whose dimensions include quality and quantity, with quantity decomposable into contributions from primary, secondary, and college attainment. This permits us to test hypotheses regarding the roles of these dimensions of education in the evolution of comparative advantage. To examine whether and how the role of education varies with the type of product, we also reestimate this specification on subsamples of core and peripheral products.[10]

The estimated interaction coefficients reflect differences in the characteristics of the target products in which RCA is most often developed between more and less educated countries. They provide causal estimates of education's effect on the character of the export mix only if those differences are not explained by omitted variables that vary across country–product dyads.[11] While causal identification is challenged by the unavailability of suitable instruments for education and familiarity, we will show that our findings are robust to the inclusion of a wide variety of omitted country-product-level variables. To ensure they are robust to the omission of variables capturing institutional or infrastructure quality, openness to FDI, or industrial dynamism pre-1995, we estimate specifications that interact proxies for these national characteristics with familiarity and with product complexity. Finding, as we do, that our results are robust to this, and to several other potential sources of error, suggests that they do provide insight into the role that education has played.

[10]Our parameter estimates are summary statistics capturing the key historical differences between the export diversification experiences of better- and worse-educated countries. Given country and product fixed effects, as well as a rich array of corrections for theorized drivers of change in comparative advantage, they are useful for examining the plausibility of hypotheses about the role education has played in export diversification.

[11]Reverse causation is unlikely because the dependent variable is measured 15–20 years after the independent variables.

1.2.2 A Two-Stage Approach

While the pooled, single-stage model produces efficient parameter estimates, the following two-stage approach is helpful for understanding why some interaction terms carry large, significant coefficients and others do not:

$$I\{RCA_{c,p,t_1} \geq 1\} = \alpha_c + f(RCA_{c,p,t_0}) + \beta_c F_{c,p,t_o} + \gamma_c C_{p,t_o} + \delta_c el_{p,t_0} + u_{c,p} \tag{1.2}$$

$$\hat{\beta}_c = b_0 + b_y y_{c,t_0} + b_E E_{c,t_0} + v_c \tag{1.3a}$$

$$\hat{\gamma}_c = c_0 + c_y y_{c,t_0} + c_E E_{c,t_0} + w_c \tag{1.3b}$$

$$\hat{\delta}_c = d_0 + d_y y_{c,t_0} + d_E E_{c,t_0} + z_c \tag{1.3c}$$

The first-stage regression (1.2) is run separately for each country, with products as observations. A positive, significant coefficient on familiarity for a country, β_c, implies that its trade patterns evolved in a path-dependent fashion. The sign of γ_c summarizes whether country c's comparative advantage tended to shift toward more or less complex products—i.e., whether it climbed up or down the product ladder. The sign of δ_c indicates whether it shifted toward more or less education-intensive products. Our two-stage estimates scale F_{c,p,t_o}, C_{p,t_o}, and el_{p,t_0} to have a mean of 0 and standard deviation of 1, so that the coefficients on each, are comparable to each other in magnitude.

The second-stage regressions (1.3a–1.3c) pool the estimated coefficients across countries and examine their relationship with initial per capita gross domestic product (GDP), y_{c,t_0}, and education. As in specification (1.1), we hypothesize that high education quantity and quality in 1995 promote movements into unfamiliar and complex products between 1995 and 2015 (regressions 1.3a and 1.3b), and that increases in education levels

between 1995 and 2015 promote movements into education-intensive products. Countries are weighted by the inverse of the standard errors of the relevant first-stage coefficients.

To see how this approach helps, suppose our estimates of (1.1) yield a negative result—that there is no tendency for more educated countries to develop comparative advantage in more complex products. In other words, suppose we fail to reject the null that $\gamma_{EC} = 0$. This could be because, controlling for the initial export mix, there is little variation between countries in their movements up or down the product ladder ($Var(\hat{\gamma}_c)$ is small), and therefore little in the historical record for differences in education across countries to explain. Or it could be because education does not correlate with these movements (i.e., $c_E = 0$ in regression 1.3b). Separating the stages will reveal which of these two explanations is relevant. This is important information if one is interested in economic history. Moreover, when the first explanation is relevant, the negative result cannot be attributed to measurement errors in our education variables.[12]

1.3 Data

We use data from The Atlas of Economic Complexity on most countries' exports of 1,240 Harmonized System products in 1995 and 2015 (The Growth Lab at Harvard University (2019b)). We exclude countries with per capita incomes in 1995 above $19,000. We do so because these countries had already developed comparative advantage in many core products by 1995 (Figure 1.1), making it difficult for them to establish a presence in many more core products during the period of our study. As advanced economies are among the most educated in the world, including them would lead us to underestimate the role of education among countries still attempting to move into the core of the global

[12]We are motivated to deal with the possibility that measurement error drives negative results by previous work showing that inability to link growth to educational accumulation is at least partially explained by errors in measuring educational accumulation Krueger and Lindahl (2001).

economy.

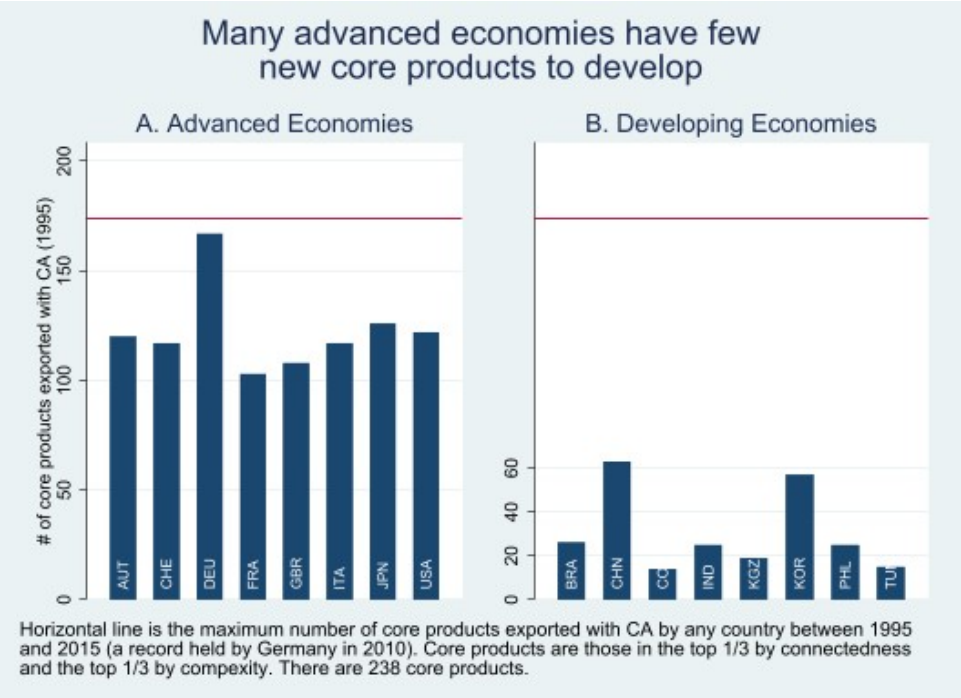
The main variable limiting our sample size is the quality of education. Hanushek and Woessmann (2008)(henceforth, H&W) carefully calibrate and splice together the results of several international standardized mathematics and science tests administered to 15-year-olds to produce a cross-sectional dataset of countries' average student cognitive skills by the late 1990s. This calibration is performed relative to a group of Organisation for Economic Co-operation and Development countries that took multiple tests over time. Altinok et al. (2014) use slightly different criteria and procedures to assemble not only cross-sectional, but also time-series estimates of student cognitive skills for these and other countries between 1995 and 2012. In addition to a wider country coverage, these estimates are arguably more reliable than the H&W estimates for countries whose standardized test performances are most unlike those of Organisation for Economic Co-operation and Development countries that H&W use in their calibrations. Our sample includes 49 countries appearing in Altinok et al. (2014) cross-sectional dataset, 35 of which also appear in H&W, and our results are robust to switching to the H&W measures.

We measure the quantity of education in 1995 and 2015 using data from Penn World Tables (PWT)(Feenstra et al., 2015) on countries' average years of schooling. Data on primary, secondary, and college attainment rates in 1995 in the population aged 15 and above come from Barro and Lee (2013).

Define the proximity between products p and q by $\varphi_{p,q} \equiv \min[Pr(CA_{c,p} = 1|CA_{c,q} = 1), Pr(CA_{c,q} = 1|CA_{c,p} = 1)]$. Proximate products are presumed to rely on similar capabilities. We proxy for familiarity using "density," which measures how close a product is to the country's export basket, is calculated as $F_{c,p} \equiv \sum q \neq p(CA_{c,q}\varphi_{p,q}) / \sum q \neq p(\varphi_{p,q})$ and must lie between 0 and 1.[13] Our measures of RCA and density are drawn from the

[13]This is in line with previous literature. Hidalgo et al. (2007) introduce density as a measure of the fraction of knowledge relevant to making product p that the country already possesses.

Figure 1.1: Advanced Economies and New Core Products Development



Atlas data, and were calculated using all countries in the dataset.

Let M be a $C \times P$ matrix with each element equal to $CA_{c,p}$. We measure CA_{p,t_0} by the product complexity index (PCI_p), which utilizes information in M and a recursive method to rate as more complex those products that are exported with comparative advantage by fewer countries (uniqueness), especially when those countries' exports are themselves diverse (Hausmann et al., 2014).

For any country-level factor endowment, Z_c, define $ProdZ_p \equiv \sum c(RCA_{c,q}Z_c)/\sum c(RCA_{c,p})$.[14] This infers, from the trade record, how intensive each product is in the use of this factor. We create three measures of the education content of products in this way. First, when Z_c is national average years of schooling, estimated from PWT, we obtain the measure $ProdYrs_p$. Second, when it is the share of a countries' population aged 15+ that com-

[14]These measures are analogous to PRODY, introduced to measure products "income content" by Hausmann et al. (2007)

pleted college, taken from Barro and Lee (2013) country data, we obtain $ProdColl_p$. Finally, when it is high schoolers' cognitive skill levels, using H&W's country data, we get $ProdCog_p$. We create each of these measures using data capturing conditions as close to 1995 as possible. We also use this procedure and PWT data to transform each country's ratio of capital to employment to produce our measure of products' relative capital intensities kl_{p,t_0}. The same PWT data are used directly to measure log-changes in countries' endowments of education and capital per worker between 1995 and 2015 ($\Delta el_{c,t}$ and $\Delta kl_{c,t}$).

Next, we define the connectedness of each product as the sum of its proximities to all other products: $C_q \equiv \sum_{n \neq q} \varphi_{n,q}$. We classify products as "core" if they are in the top tercile of the distributions of both connectedness and PCI; and "peripheral" if they are in the bottom tercile of both distributions. To illustrate: most unprocessed agricultural and mined commodities, human hair, jute fibers, and electric power are revealed to be peripheral; jet engines, x-ray machines, watch movements, optical devices, and machine tools are core products; and paper, electric shavers, hats, copper wire, and wine are in-between. In our dataset of 1,240 products, 230 are core and 232 are peripheral. The remaining 778 are in-between.

The control variables used in this paper include multiple measures of the quality of countries' institutions and infrastructure, as well as the average FDI:Exports and FDI:GDP ratios between 1995 and 2015, drawn from the World Development Indicators. Finally, we use three proxy measures of countries' industrial dynamism prior to 1995: real per capita GDP growth from the World Development Indicators, labor productivity growth from PWT, and estimates of β_c and γ_c for the periods 1975–1995 and 1985–1995 from specification (1.2) using data on exports by 4-digit Standard International Trade Classification products (The Growth Lab at Harvard University (2019a)).

Table 1.1 provides summary statistics. Usefully, the countries in our sample dif-

fer widely in educational attainment and quality, and RCAs in many industries (country–product dyads) demonstrate significant changes between 1995 and 2015.

Table 1.1: Summary Statistics

	Observations	Mean	SD	Minimum	Maximum
Country level variables					
Quantity (Average Years of Schooling)	49	7.11	2.35	2.16	11.39
Edu Quality A (Altinok et al., 2018)	49	494.39	75.07	282.24	652.62
Edu Quality B (Hanushek & Woessman, 2009)	35	4.27	.58	3.09	5.34
Primary (Primary Attainment, aged 15+)	48	73.13	17.08	36.21	98.03
Secondary (Secondary Attainment, aged 15+)	48	33.32	16.65	4.81	76.70
College (College Attainment, aged 15+)	48	5.88	4.19	.58	19.35
ΔYrsSch (Change in Average Years of Schooling)	49	2.31	.76	.50	4.55
ΔQuality A (Change in Edu Quality A)	22	9.81	30.02	-36.60	63.47
ΔK/L (Change in Capital Intensity)	44	.55	.44	-.51	2.02
Product level variables					
PCI (Product Complexity Index)	1,242	0.00	1.00	-2.93	2.84
ProdYrs (Average years of schooling across product exporters)	1,242	9.02	1.18	5.04	12.59
ProdCog (Average education quality across product exporters)	1242	4.75	.22	3.89	5.20
ProdColl (Average college attainment across product exporters)	1242	8.07	1.74	3.20	15.58
ProdKL (K/L averaged across product exporters)	1,242	-9.49	.74	-13.11	-6.66
Country-Product level variables					
RCA (1995)	60,858	1.20	9.24	0	868.77
RCA (2015)	60,760	1.18	9.19	0	1243.06
CA = I(RCA_{2015}>=1}	60,760	.17	.37	0	1
ΔRCA (Change in Export RCA, 1995-2015)	60,760	-.03	9.23	-866.65	476.23
Familiarity	60,858	.16	.09	.00	.70

Note: All country-level variables will be normalized to have a mean of zero and standard deviation of one when used in regressions.

1.4 Results

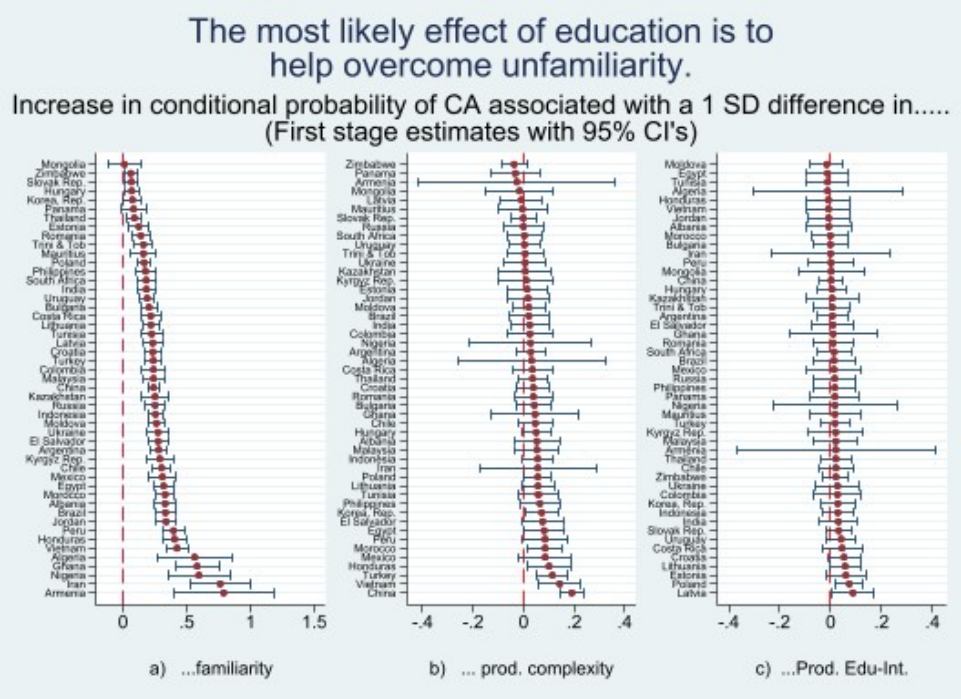
1.4.1 Two-Stage Analysis

Figure 1.2 shows each country's first-stage coefficient estimates.[15] Panel (a) indicates that countries differed significantly in terms of their tendency to keep exporting familiar

[15]The lagged RCA correction takes the form: $f(RCA) = f_0 \times I\{RCA = 0\} + f_1 \times [1 - I\{RCA = 0\}]g(RCA)$, where $g(RCA) = \ln(RCA)$ when $RCA > 0$ and $g(RCA) = m$ when $RCA = 0$. This specification makes allowance for the possibility that exporting any of a product has different effects on the likelihood of future comparative advantage than does having a high RCA in it. Our coefficient estimates are invariant to the value chosen for the constant m by construction. Log-linearizing the nonzero values is recommended by the q-q plot of $ln(RCA)$ (Appendix Figure A.1). We have also run our main regressions using a hyperbolic sine function in place of $g(\cdot)$, but this is restrictive and does not alter our main results.

products. In Iran and Armenia, for example, a 1 standard deviation difference in familiarity between two products in 1995 is associated with a roughly 75 percentage point difference between them in the conditional probability of comparative advantage by 2015 (after conditioning on lagged RCA, product complexity, and education intensity). In Mongolia, the Republic of Korea, and Zimbabwe, on the other hand, familiarity played almost no role. It follows that there is significant variation in the degree of this type of path dependence across countries.

Figure 1.2: Education and Unfamiliarity



In contrast, panel (b) provides much less evidence, controlling for product familiarity, that product complexity influenced the evolution of countries' comparative advantages, or that this influence varied across countries. Product complexity was significantly and positively associated with gains in comparative advantage in six countries (the Republic of Korea, Morocco, Honduras, Turkey, Viet Nam, and the PRC). But even in the PRC,

a 1 standard deviation difference in product complexity was associated with less than 20 percentage point difference in the probability of acquiring a comparative advantage by 2015. No tendency to move up or down the complexity ladder is statistically discernible in the remaining 43 countries.

Panel (c) shows even less variation in the relationship between products' education intensity and their conditional probability of acquiring a comparative advantage by 2015. Indeed, there is only one country (Poland), where this relationship is significantly different from zero.

These results show that countries differ much more with respect to their success overcoming unfamiliarity than with their success overcoming complexity or education intensity. Our second-stage regressions examine whether education can account for the cross-country differences just documented.

Figure 1.3 explores specification (1.3a) graphically. It shows that countries with more average years of schooling in 1995 and higher levels of cognitive skill by around 2000 were significantly less tethered to familiar products. Columns 1–3 of Table 1.2 show that these two relationships survive correcting for per capita GDP and switching measures of education quality—although in the case of the Altinok et al. (2014) measure, the relationship is not statistically significant. Columns 4 and 5 correct for both quantity, quality, and an interaction between the two. Whether quality or quantity have explanatory power is sensitive to the quality measure used. The quality–quantity interactions are statistically insignificant and their inclusion does not change these findings qualitatively.

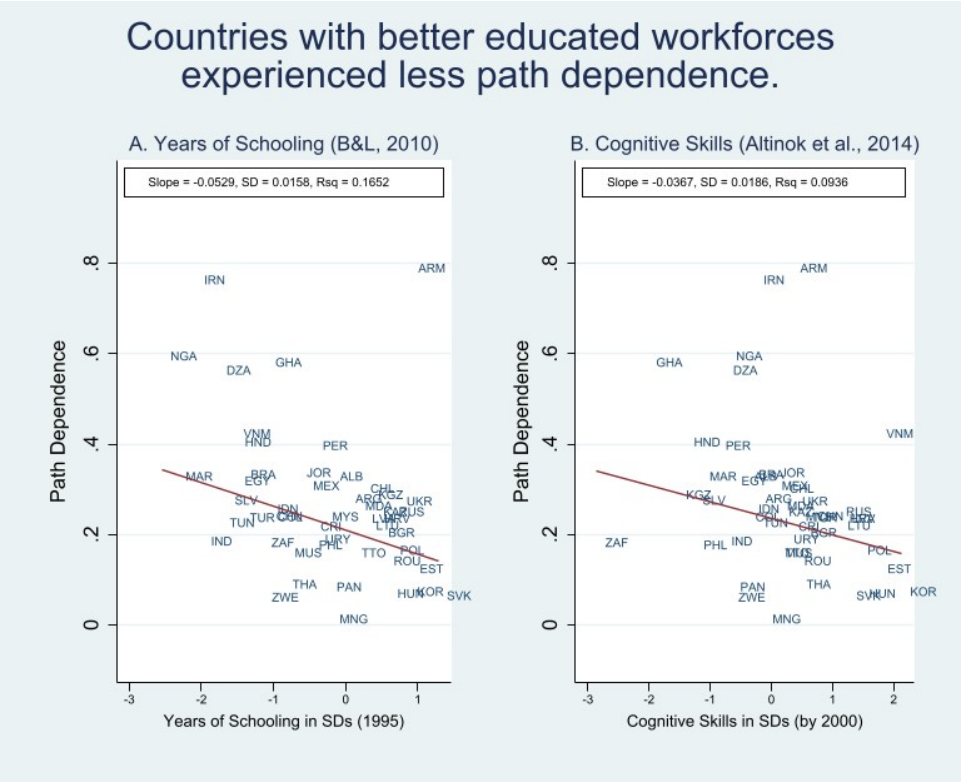
Figures 1.4 and 1.4 explore specifications (1.3b) and (1.3c), maintaining the same y-axis scale as Figure 1.3 for the sake of comparison. Neither provides strong support for the other two hypothesized roles of education. Figure 1.4(a) shows that higher initial years of schooling is associated with less movement toward complex products. Figure 1.4(b) shows a positive but extremely small relationship between the quality of education and

Table 1.2: Explaining Shifts into Unfamiliar Products

Dependent variable	First-Stage Coefficient on Familiarity/Familiarity				
	(1)	(2)	(3)	(4)	(5)
Per Capita GDP in 1995 (Constant 2010 US$)	-0.026	-0.037	-0.024	-0.017	-0.012
	(0.023)	(0.025)	(0.026)	(0.025)	(0.034)
Quantity (Average Years of Schooling)	-0.042**			-0.036*	-0.007
	(0.017)			(0.021)	(0.034)
Quality A (Altinok et al.)		-0.024		-0.017	
		(0.018)		(0.017)	
Quality B (H & W)			-0.048***		-0.053**
			(0.015)		(0.020)
Quantity x Quality A				-0.016	
				(0.014)	
Quantity x Quality B					-0.022
					(0.013)
Constant	0.423**	0.536**	0.433*	0.370*	0.345
	(0.196)	(0.207)	(0.218)	(0.205)	(0.295)
Observations	49	49	35	49	35
R-squared	0.186	0.142	0.223	0.202	0.248

Note: Weighted least squares estimates, per specification 1.3a. The dependent variable is each country's coefficient on familiarity from first stage specification 1.2. Countries are weighted by the inverse of the standard error of that coefficient. Abbreviated variable names appear in Table 1.1. All standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1.

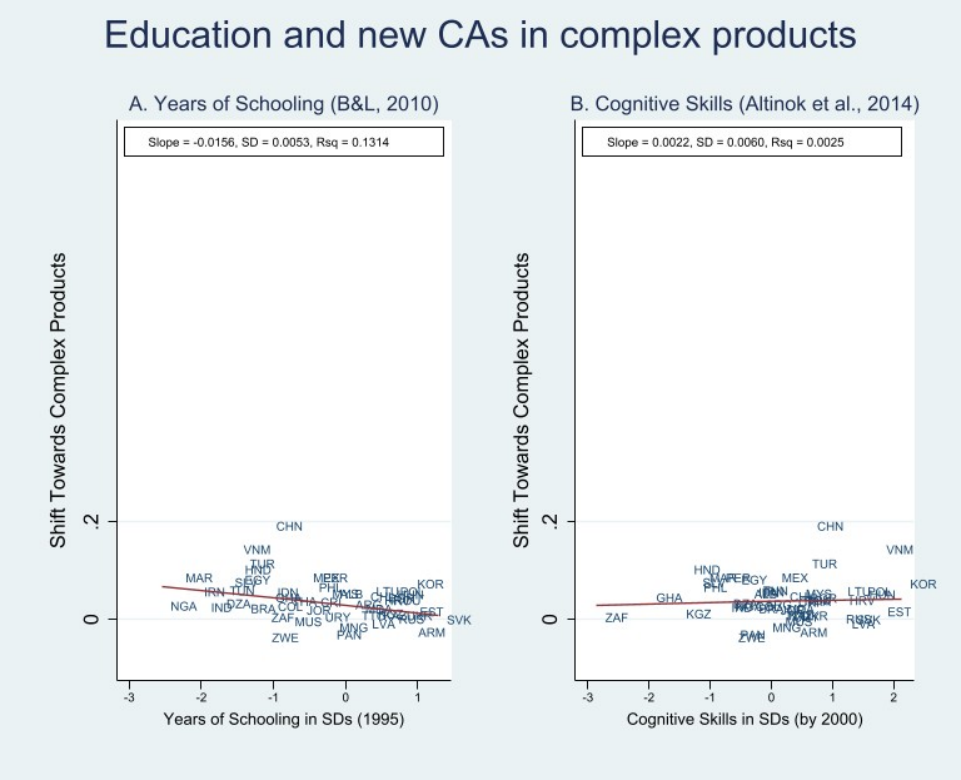
Figure 1.3: Education and Path Dependence



movement into complex products. Figure 1.5 shows that expansions in the quantity of education were associated with movements into less education-intensive products, and that there is no relationship between quality improvements and increased education intensity in the product mix. Tables 1.3 and 1.4, which again add per capita GDP corrections, confirm that most of these relationships have the wrong (negative) sign, or are small or statistically insignificant. The one partial exception (Table 1.3, column 4) is that when controlling for the quantity of education and the Altinok, Diebolt, and Demeulemeester (2014)Altinok et al. (2014) quality measure simultaneously, the latter is associated with a slightly higher tendency to shift toward complex products. Table 1.4 shows that neither increases in education quantity (column 1) nor quality (in a much smaller sample of countries, column 2) are associated with movement into education-intensive products,

and that the relationship with quantity expansions is not stronger if the education is of a higher quality (column 3).

Figure 1.4: Education and New Comparative Advantages in Complex Products



Together, these results suggest that: (i) the main effect of education on the evolution of comparative advantage is to help countries overcome unfamiliarity, (ii) the effects of quantity and quality in this regard are difficult to tease apart,[16] and (iii) it is unlikely that education helped counties move into more complex or education-intensive products. This is because while our education measures vary a lot across countries, countries did not vary much in their tendency to achieve comparative advantage in complex and education-intensive products (i.e., $Var(\hat{\gamma}_c)$ and $Var(\hat{\delta}_c)$ are relatively small).

[16]The raw correlation between Quality A (Quality B) and years of schooling in 1995 is 0.559 (0.588).

Table 1.3: Explaining Shifts into Complex Products

Dependent variable	First-Stage Coefficient on Product Complexity				
	(1)	(2)	(3)	(4)	(5)
Per Capita GDP in 1995 (Constant 2010 US$)	0.001	-0.009	0.000	-0.005	0.007
	(0.008)	(0.012)	(0.014)	(0.009)	(0.011)
Quantity (Average Years of Schooling)	-0.016***			-0.025***	-0.016
	(0.005)			(0.007)	(0.010)
Quality A (Altinok et al.)		0.006		0.017**	
		(0.008)		(0.008)	
Quality B (H & W)			-0.003		0.003
			(0.010)		(0.011)
Quantity x Quality A				-0.002	
				(0.006)	
Quantity x Quality B					-0.005
					(0.010)
Constant	0.019	0.112	0.037	0.064	-0.021
	(0.070)	(0.100)	(0.114)	(0.073)	(0.093)
Observations	49	49	35	49	35
R-squared	0.132	0.028	0.006	0.242	0.109

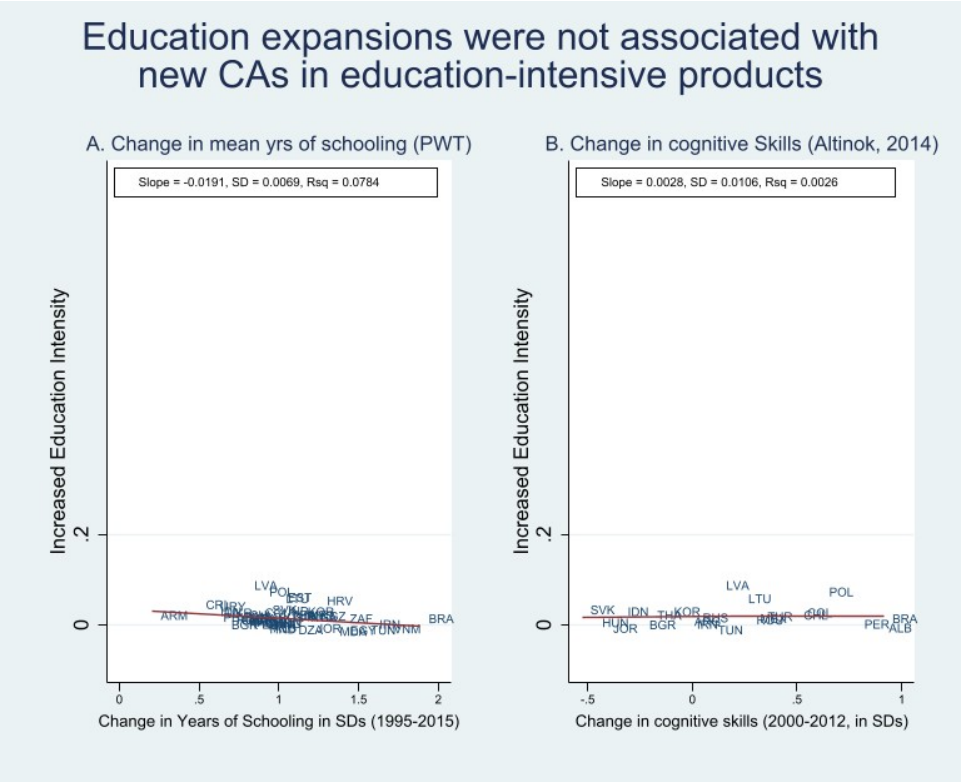
Note: Weighted least squares estimates, per specification 1.3b. The dependent variable is each country's coefficient on familiarity from first stage specification 1.2. Countries are weighted by the inverse of the standard error of that coefficient. Abbreviated variable names appear in Table 1.1. All standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.4: Explaining Shifts into Education-Intensive Products

Dependent variable	First-Stage Coefficient on Education Intensity		
	(1)	(2)	(3)
Per Capita GDP in 1995 (Constant 2010 US$)	0.009***	0.016**	0.003
	(0.003)	(0.007)	(0.004)
ΔQuantity	-0.018***		-0.019***
	(0.006)		(0.007)
ΔQuality A (Altinok et al., time-varying)		0.005	
		(0.009)	
Quality A (Altinok et al., time-varying)			0.020
			(0.013)
ΔQuantity x Quality A			-0.012
			(0.010)
Constant	-0.041*	-0.121*	0.009
	(0.024)	(0.061)	(0.031)
Observations	49	22	49
R-squared	0.200	0.128	0.303

Note: Weighted least squares estimates, per specification 1.3c. The dependent variable is each country's coefficient on familiarity from first stage specification 1.2. Countries are weighted by the inverse of the standard error of that coefficient. Abbreviated variable names appear in Table 1.1. All standard errors are robust. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1.5: Education and New Comparative Advantages in Education-Intensive Products



1.4.2 Single-Stage Analysis

Table 1.5 builds up our baseline estimates of specification (1). Columns (1) and (2) include terms suggested by complexity theory. Column (1) measures the quantity of education in 1995 by average years of schooling, while column (2) measures it by the proportions of the population aged 15+ that completed primary, secondary, and college education. Column (3) includes only the explanatory variables suggested by a factor abundance approach. As we do not have attainment rates by level in 2015, or time series for most countries on changes in education quality, we focus on the effects of changing average years of schooling. The three-way interaction allows that increasing years of schooling would be more supportive of the development of comparative advantage in

education- intensive products if that education is of a high quality. Columns (4) and (5) combine the two sets of coefficients. Other than familiarity and the terms derived from lagged RCA, every variable entering the table, whether on its own or interacted, is normalized to have a mean of 0 and standard deviation of 1.

As expected, the coefficients on the lagged dependent variable indicate positive relationships between having nonzero and larger RCAs in 1995 and the probability of comparative advantage in 2015.[17]

We use country-clustered standard errors throughout this paper to ensure conservative inferences. These standard errors are extremely conservative, given that we have not sampled a small number of countries from a large universe, but rather attempted to include every low- and middle-income country for which the relevant data are available (Abadie et al., 2017). We report in the text any instances in which using unclustered robust standard errors alters our qualitative inferences.

The results in Table 1.5 are consistent with those in the previous section.[18] The positive significant coefficient on familiarity indicates that countries of average education quality and quantity are much more likely to develop comparative advantage in products that are more familiar to them. The negative sign on the familiarity–quantity interaction is consistent with education helping to develop comparative advantages in relatively unfamiliar products. Evidence of this effect is strongest for primary school, and indeed when using unclustered standard errors, the p-value on the interaction between familiarity and primary attainment is effectively zero.

[17] Postestimation calculation of $E(CA_{p,t}|CA_{p,t})$ as well as a kernel-weighted local polynomial regression of the same (Appendix Figure A.2) confirm a nearly monotonic positive relationship.

[18] This was expected. Familiarity and product complexity are orthogonal to each other by definition (familiarity with a given product varies across countries, while the complexity of that product does not), practically guaranteeing that estimating familiarity and complexity interactions together will not alter the implied coefficients on either.

Table 1.5: Single-Stage Regression Results

	Economic Complexity				Heckscher-Ohlin-Vanek		Both			
	(1)		(2)		(3)		(4)		(5)	
Corrections for lagged RCA										
I{RCA in 1995 = 0}	-0.203***	(0.013)	-0.199***	(0.013)	-0.263***	(0.016)	-0.202***	(0.014)	-0.195***	(0.014)
[1 - I{RCA in 1995 = 0}] *ln(RCA in 1995)	0.049***	(0.002)	0.049***	(0.002)	0.066***	(0.003)	0.050***	(0.002)	0.050***	(0.002)
Familiarity	1.885***	(0.191)	2.132***	(0.156)			1.917***	(0.182)	2.163***	(0.156)
Familiarity x Quantity	-0.672***	(0.172)					-0.655***	(0.165)		
Familiarity x Quality A	0.181	(0.157)	0.004	(0.181)			0.114	(0.132)	-0.078	(0.162)
Familiarity x Primary			-0.332	(0.232)					-0.400	(0.240)
Familiarity x Secondary			-0.084	(0.183)					0.072	(0.204)
Familiarity x College			-0.134	(0.246)					-0.169	(0.257)
PCI x Quantity	-0.024**	(0.010)					-0.029**	(0.012)		
PCI x Quality A	0.028***	(0.008)	0.023***	(0.008)			0.021**	(0.010)	0.015*	(0.008)
PCI x Primary			-0.010	(0.009)					-0.013	(0.010)
PCI x Secondary			-0.022*	(0.012)					-0.018	(0.014)
PCI x College			0.008	(0.012)					0.002	(0.013)
ΔK/L x ProdKL					-0.005	(0.018)	0.015	(0.010)	0.014	(0.011)
ΔQuantity x ProdYrs					-0.005	(0.010)	-0.014	(0.013)	-0.013	(0.012)
Quality A x ProdYrs					0.031	(0.020)	0.002	(0.023)	0.007	(0.021)
ΔQuantity x Quality A x ProdYrs					-0.014	(0.019)	0.005	(0.023)	0.002	(0.021)
Country Fixed Effects	✓		✓		✓		✓		✓	
Product Fixed Effects	✓		✓		✓		✓		✓	
Observations	60,760		59,520		54,560		54,560		53,320	
R-squared	0.266		0.266		0.242		0.264		0.264	
Baseline without education variables										
Observations	60,760				54,560		54,560			
R-squared	0.261				0.239		0.259			

Note: Estimates based on linear probability specification 1.1. Country-clustered standard errors in parentheses: p<0.01, ** p<0.05, * p<0.1.

This effect is of a modest size. Consider two products and two countries. Assume, for each product, that its lagged RCA and familiarity are the same in both countries, but that one product is 0.20 points (roughly 2 standard deviations, Table 1.1) more familiar than the other product in both countries. We also assume that both countries have average education quality, that country A has average education quantity, and that country B's years of schooling are 1 standard deviation higher than the mean. The estimates in column (4) then indicate that in country A, the probability of comparative advantage in 2015 is 38 percentage points higher in the more familiar product than in the unfamiliar product. However, in country B, this probability will only be 25 percentage points higher in the more familiar product. The 13 percentage point difference attributable to education is sizable compared to the mean probability of comparative advantage of 17 percentage points.[19] On the other hand, despite this large effect of education quantity, there are no countries in our sample with enough years of schooling to eliminate path dependence (i.e., there are no countries for which $\beta_E + \beta_{EF} E_{c,t_0} \leq 0$).

As was the case with the second-stage results in Table 1.3, the single-stage results in Table 1.5 paint a mixed picture regarding our second hypothesis—that (controlling for familiarity) education helps develop comparative advantages in more complex products. Higher average years of schooling in 1995 are associated with developing comparative advantages in less complex products, but higher quality education is associated with developing comparative advantages in more complex products. The effect of school quality on comparative advantage in complex products is roughly one-third the size of the effect of school quantity on comparative advantage in unfamiliar products.[20]

[19] In country A, the difference in probability of comparative advantage is $0.2 \times 1.917 = 0.383$. In country B, it is $0.2 \times (1.917 - 0.655) = 0.252$.

[20] Consider two products 2 standard deviations apart in complexity, and two countries 1 standard deviation apart in education quality but with the same years of schooling. Holding all other variables constant across products and countries, the difference between the probabilities of comparative advantage in the more and less complex products would be 4.2 percentage points ($= 0.021 \times 2 \times 1$) greater in the better-educated country. Compare this to the 13 percentage point effect of school quantity in unfamiliar

Regressions (3)–(5) provide no significant evidence that growth in education shifts countries toward education-intensive products. The point estimates suggest that countries with larger increases in average years of schooling between 1995 and 2015 tended to develop comparative advantages in less education-intensive products. They also do not support the possibility that increasing years of schooling shifts comparative advantages toward education-intensive products more reliably in countries with higher quality education. In theory, this absence of measured effects of changes in education could be driven by attenuation biases arising from this differenced variable's low signal to noise ratio (Krueger and Lindahl, 2001). However, the fact that countries vary so little in the extent to which they shifted into education-intensive products (Figure 1.2(c)) suggests that our estimate of δ_{EE} would have been small even in the absence of measurement error.

Table 1.5 also reports R-squared statistics for models that restrict the coefficients on all education terms to be zero in order to assess education's explanatory power. Allowing that education could be useful for overcoming unfamiliarity, product complexity or education intensity adds very little to model R-squared.[21] In combination with the large and statistically significant coefficient on the interaction between familiarity and education quantity, this indicates that education is useful for overcoming unfamiliarity, but that past specialization patterns, product and country characteristics, are still the main determinants of subsequent specialization patterns.

products.

[21] For example, the R-squared rises from 0.2614 to 0.2664 when the four education-related terms in column (1) are added. In the linear probability context, R-squared captures the difference in the predicted probabilities of "success"—in our case, a comparative advantage in the product—between observed cases of success and failure (Gronau et al., 1998).

1.4.3 Robustness Tests

The previous two sections suggest that the primary beneficial effect of education when seeking to gain comparative advantage in new products is that it can help overcome a lack of familiarity with target products. The level of education most strongly associated with this effect is primary school. Conclusions about whether it is the quantity or the quality of education that matters depend upon the measure of education quality used. Next, we check whether these findings are robust.

Table 1.6 is analogous to Table 1.5, but replaces the Altinok et al. (2014) estimate of education quality with estimates on a smaller number of countries from Hanushek and Woessmann (2008). The results change very little. The most powerful role of education is still to overcome unfamiliarity.

Table 1.7 checks whether our inability to confirm a role for education accumulation in promoting the development of comparative advantage results from specification errors. Column 2 replicates Table 1.5, column 3 for comparison's sake. Column 3 strips it of quality effects, and column 5 checks to see whether quality accumulation over time is associated with moves into education-intensive products (note the sharply reduced sample size). Columns 1 and 4 provide restricted regressions that eliminate the education variables. Physical capital accumulation is associated with a movement into capital-intensive products in columns 4 and 5. However, the education effects are always statistically insignificant, usually of the wrong sign, and their inclusion adds little to R-squared. As Figure 1.2 demonstrates, their insignificance is unlikely to be driven by attenuation.

Table 1.8 examines the possibility that educated countries develop comparative advantage in products that are sophisticated in ways not captured by the Product Complexity Index (PCI). To do so, columns (2)–(4) replace the PCI with sophistication measures derived from the education endowments of those countries that successfully export them.

Table 1.6: Single-Stage Regression with Hanushek and Woessmann's (2009) Education Quality Measure

	Economic Complexity				Heckscher-Ohlin-Vanek		Both			
	(1)		(2)		(3)		(4)		(5)	
Corrections for lagged RCA										
I{RCA in 1995 = 0}	-0.220***	(0.016)	-0.211***	(0.016)	-0.288***	(0.019)	-0.219***	(0.018)	-0.208***	(0.018)
[1 - I{RCA in 1995 = 0}] *ln(RCA in 1995)	0.053***	(0.003)	0.053***	(0.002)	0.071***	(0.004)	0.053***	(0.003)	0.053***	(0.003)
Familiarity	1.914***	(0.250)	2.193***	(0.191)			1.929***	(0.249)	2.247***	(0.195)
Familiarity x Quantity	-0.485*	(0.258)					-0.566**	(0.252)		
Familiarity x Primary			-0.130	(0.222)					-0.233	(0.248)
Familiarity x Secondary			-0.010	(0.203)					0.020	(0.200)
Familiarity x College			-0.303	(0.341)					-0.316	(0.335)
Familiarity x Quality B	-0.036	(0.234)	-0.251	(0.224)			-0.004	(0.230)	-0.220	(0.238)
PCI x Quantity	-0.019	(0.015)					-0.024	(0.016)		
PCI x Quality B	0.023*	(0.013)	0.022*	(0.012)			0.018	(0.015)	0.015	(0.014)
PCI x Primary			-0.008	(0.012)					-0.010	(0.014)
PCI x Secondary			-0.023	(0.015)					-0.023	(0.016)
PCI x College			0.009	(0.019)					0.000	(0.019)
ΔK/L x ProdKL					-0.003	(0.021)	0.018	(0.012)	0.019	(0.013)
ΔQuantity x ProdYrs					-0.005	(0.014)	-0.018	(0.022)	-0.014	(0.019)
Quality A x ProdYrs					0.011	(0.016)	-0.007	(0.019)	-0.009	(0.019)
ΔQuantity x Quality A x ProdYrs					0.003	(0.016)	0.012	(0.020)	0.015	(0.019)
Country Fixed Effects	✓		✓		✓		✓		✓	
Product Fixed Effects	✓		✓		✓		✓		✓	
Observations	43,400		42,160		40,920		40,920		39,680	
R-squared	0.276		0.276		0.251		0.274		0.274	

Note: The estimation is based on the linear probability model, per specification 1.1. Country-clustered standard errors in parentheses: p<0.01, ** p<0.05, * p<0.1.

Table 1.7: Alternative Estimates of Heckscher-Ohlin-Vanek Effects

	(1)	(2)	(3)	(4)	(5)
ΔK/L x ProdKL	0.001	-0.005	-0.000	0.046**	0.044*
	(0.019)	(0.018)	(0.019)	(0.021)	(0.021)
ΔQuantity x ProdYrs		-0.005	-0.010		-0.009
		(0.010)	(0.010)		(0.016)
Quality A x ProdYrs		0.031			
		(0.020)			
ΔQuantity x Quality A x ProdYrs		-0.014			
		(0.019)			
ΔQuality A x ProdYrs					0.004
					(0.016)
Country Fixed Effects	✓	✓	✓	✓	✓
Product Fixed Effects	✓	✓	✓	✓	✓
Corrections for Lagged RCAs	✓	✓	✓	✓	✓
Observations	54,560	54,560	54,560	26,040	26,040
R-squared	0.239	0.242	0.239	0.259	0.259

Note: The estimation is based on linear probability model. Columns 1-3 include all countries in our sample; columns 4-5 are based on the sample with the time-varying Altinok quality data. Significance using country clustered standard errors: p<0.01, ** p<0.05, * p<0.1.

The interactions between these sophistication measures and education quantity continue to carry a negative coefficient, and making these substitutions does not qualitatively alter our findings on the interactions between education quantity and familiarity.

Table 1.9 shows that our results are not sensitive to most changes in the RCA cutoff used to decide whether a country exports a product with comparative advantage. The only discernable shift is that the interaction between familiarity and years of schooling declines as the bar for achieving comparative advantage is lowered.

Table 1.10 checks whether our education-related coefficients might be biased by the omission of controls for national institutional and infrastructure quality or for openness to FDI.[22] This is done in a separate regression for each control variable by adding an interaction between the control variable and both familiarity and the PCI. It also includes,

[22]FDI is widely considered to have been an important source of new productive knowledge for late industrializing countries (e.g., Jomo (2003), and Felker et al. (2013)). We have not presented all of these regressions here in the interests of parsimony. Regressions controlling for access to information and communication technology, corruption, the number of procedures needed to open a small business, and kilometers of road per square kilometer yield the same qualitative results as those shown here. Mehta and Felipe (2014) list the sources of these institutional variables, while FDI data are drawn from the World Bank's World Development Indicators database.

Table 1.8: Robustness to Alternative Measures of Product Sophistication

	Measure of Sophistication							
	PCI		*ProdYrs*		*ProdCog*		*ProdColl*	
	(1)		(2)		(3)		(4)	
Familiarity	1.917***	(0.182)	2.052***	(0.189)	2.005***	(0.181)	2.066***	(0.197)
Familiarity x Quantity	-0.655***	(0.165)	-0.404***	(0.121)	-0.366***	(0.124)	-0.312**	(0.127)
Familiarity x Quality A	0.114	(0.132)	-0.072	(0.112)	-0.070	(0.119)	-0.128	(0.120)
Soph x Quantity	-0.029***	(0.012)	-0.012**	(0.006)	-0.041*	(0.024)	-0.004*	(0.002)
Soph x Quality A	0.021**	(0.010)	0.007	(0.025)	0.043	(0.026)	-0.002	(0.002)
ΔK/L x ProdKL	0.015	(0.010)	0.014	(0.011)	0.013	(0.011)	0.013	(0.012)
ΔQuantity x ProdYrs	-0.014	(0.013)	-0.012	(0.014)	-0.007	(0.013)	-0.006	(0.013)
Quality A x ProdYrs	0.002	(0.023)			-0.005	(0.024)	0.000	(0.023)
ΔQuantity x Quality A x ProdYrs	0.005	(0.023)	0.009	(0.023)	0.011	(0.023)	0.014	(0.022)
Country Fixed Effects	✓		✓		✓		✓	
Product Fixed Effects	✓		✓		✓		✓	
Corrections for Lagged RCAs	✓		✓		✓		✓	
Observations	54,560		54,560		54,560		54,560	
R-squared	0.264		0.263		0.263		0.263	

Note: The estimation is based on linear probability model, per specification 1. Significance using country clustered standard errors: p<0.01, ** p<0.05, * p<0.1.

Table 1.9: Robustness to Changes in Revealed Comparative Advantage Cutoffs

	Cutoff = 0.5 (1)		Cutoff = 0.8 (2)		Cutoff = 1 (3)		Cutoff = 2 (4)	
Familiarity	1.985***	(0.189)	2.017***	(0.184)	1.917***	(0.182)	1.376***	(0.166)
Familiarity x Quantity	-0.309*	(0.168)	-0.512***	(0.167)	-0.655***	(0.165)	-0.708***	(0.160)
Familiarity x Quality A	-0.068	(0.118)	0.019	(0.129)	0.114	(0.132)	0.099	(0.141)
PCI x Quantity	-0.026*	(0.014)	-0.028**	(0.013)	-0.029**	(0.012)	-0.023***	(0.008)
PCI x Quality A	0.033**	(0.012)	0.025**	(0.010)	0.021**	(0.010)	0.008	(0.006)
ΔK/L x ProdKL	0.018	(0.012)	0.017*	(0.010)	0.015	(0.010)	0.006	(0.008)
ΔQuantity x ProdYrs	-0.010	(0.016)	-0.014	(0.014)	-0.014	(0.013)	-0.016*	(0.009)
Quality A x ProdYrs	0.009	(0.029)	0.003	(0.024)	0.002	(0.023)	-0.004	(0.017)
ΔQuantity x Quality A x ProdYrs	-0.006	(0.029)	0.003	(0.024)	0.005	(0.023)	0.011	(0.017)
Country Fixed Effects	✓		✓		✓		✓	
Product Fixed Effects	✓		✓		✓		✓	
Corrections for Lagged RCAs	✓		✓		✓		✓	
Observations	54,560		54,560		54,560		54,560	
R-squared	0.317		0.283		0.208		0.264	

Note: The estimation is based on linear probability model, per specification 1.1. Significance using country clustered standard errors: p<0.01, ** p<0.05, * p<0.1.

at the end, a regression controlling for the first principal component of all 10 institutional and infrastructure quality measures. None of these changes makes any difference to our conclusions regarding the role of education. In contrast to education, no infrastructure controls predict movement into unfamiliar or complex products. Only one institutional control—an index of regulatory quality—predicts developing comparative advantage in complex products, and none predict developing comparative advantage in unfamiliar products. And, countries receiving more FDI were more likely to develop comparative advantage in products with which they were already familiar.[23]

The correlations we report between education and post-1995 export dynamism (movements into unfamiliar and complex products) could reflect a tendency for intrinsically dynamic countries to invest in education. To check whether this is plausible, we first calculate four different measures of prior industrial dynamism. We calculate each of these over two prior time intervals: 1975–1995 and 1985–1995. The four measures are: (a) real per capita GDP growth; (b) real labor productivity growth; and the coefficients

[23]This is consistent with FDI flowing mainly toward industries relying on well-established capabilities. While intuitive, this idea cannot be tested without data on FDI receipts disaggregated by industry.

Table 1.10: Robustness to Corrections for Institutions, Infrastructure, and Foreign Direct Investment

	Institutional Control								
	Contracts	Infrastructure	Days	Internet	RQ	RL	FDI/Exports	FDI/GDP	Principal Component
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Familiarity	1.961***	2.018***	1.897***	3.322***	1.913***	1.986***	2.031***	1.949***	2.148***
Familiarity x Institutional control	-0.026	0.044	0.042	2.371	0.309	-0.069	0.271**	0.065	-0.206
Familiarity x Quantity	-0.663***	-0.675***	-0.646***	-0.749***	-0.738***	-0.659***	-0.640***	-0.683***	-0.920***
Familiarity x Quality A	0.083	0.089	0.121	0.096	0.059	0.112	0.080	0.094	0.129
PCI x Institutional control	0.000	0.002	-0.001	0.377	0.019***	0.011	-0.003	-0.006	-0.001
PCI x Quantity	-0.031*	-0.033**	-0.029**	-0.045***	-0.035***	-0.029**	-0.026**	-0.025*	-0.085***
PCI x Quality A	0.023	0.023*	0.021**	0.026**	0.018*	0.017	0.019*	0.020**	0.064***
ΔK/L x ProdKL	0.013	0.014	0.015	0.014	0.013	0.016	0.017*	0.015	0.010
ΔQuantity x ProdYrs	-0.019	-0.019*	-0.015	-0.026**	-0.014	-0.014	-0.015	-0.015	-0.067**
Quality A x ProdYrs	0.007	0.005	0.001	-0.003	0.001	-0.001	0.004	0.002	-0.004
ΔQuantity x Quality A x ProdYrs	0.001	0.003	0.006	0.009	0.007	0.008	0.004	0.006	0.013
Country Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Product Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Corrections for Lagged RCAs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	43,400	47,120	53,320	48,360	54,560	54,560	54,560	54,560	22,320
R-squared	0.271	0.273	0.264	0.274	0.266	0.265	0.265	0.265	0.303
F-test	4.834	7.241	6.275	6.008	6.019	6.891	8.691	7.981	17.658
Prob>F	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: FDI, exports and GDP are nominal values, drawn from World Development Indicators. Sources of infrastructure and institutional variables appear in Mehta and Felipe (2014). The estimation is based on linear probability model, per specification 1. Significance using country clustered standard errors: p<0.01, ** p<0.05, * p<0.1.

Table 1.11: Correcting for Prior Industrial Dynamism

	Measure of Prior Industrial Dynamism Used					
	Real GDPPC Growth Rate		Labor Productivity Growth Rate		First stage coefficients on familiarity & PCI	
	(1)	(2)	(3)	(4)	(5)	(6)
Familiarity	1.753***	2.352***	1.983***	2.358***	1.710***	1.940***
Familiarity x Quantity	-0.830***		-0.691***		-0.702***	
Familiarity x Quality A	-0.107	-0.103	0.021	0.007	0.134*	-0.060
Familiarity x Primary		-0.605***		-0.438***		-0.331***
Familiarity x Secondary		0.144		-0.075		0.113
Familiarity x College		-0.163		-0.216**		-0.231**
Familiarity x Dynamism (1975-95)	0.857***	0.780***	0.583***	0.410**	-0.043	-0.650
Familiarity x Dynamism (1985-95)	-1.544***	-1.812***	-0.964**	-1.258***	-1.546***	-1.220***
PCI x Quantity	-0.032***		-0.020***		-0.019***	
PCI x Quality A	-0.009**	-0.019***	0.010**	0.004	0.019***	0.010***
PCI x Primary		-0.009**		-0.006*		0.004
PCI x Secondary		-0.033***		-0.017***		-0.009**
PCI x College		0.017***		-0.000		-0.009**
PCI x Dynamism (1975-95)	0.043***	0.057***	0.019***	0.021***	0.102***	0.112***
PCI x Dynamism (1985-95)	0.005	-0.019	0.040***	0.038***	0.669***	0.678***
ΔK/L x ProdKL	0.011***	0.008	0.011***	0.010*	0.010**	0.008
ΔQuantity x ProdYrs	-0.031***	-0.018***	-0.025***	-0.018***	-0.034***	-0.029***
Quality A x ProdYrs	0.003	0.006	0.010	0.005	0.020***	0.007
ΔQuantity x Quality A x ProdYrs	0.001	-0.000	-0.005	-0.001	-0.016***	-0.003
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Product Fixed Effects	✓	✓	✓	✓	✓	✓
Corrections for Lagged RCAs	✓	✓	✓	✓	✓	✓
Observations	33,480	32,240	38,440	37,200	42,160	40,920
R-squared	0.313	0.313	0.293	0.292	0.294	0.292

Note: GDP per capita data are from World Development Indicators, labor productivity from Penn World Tables, and first stage coefficients are explained in the text. Estimates from a linear probability model of comparative advantage in 2015, per specification 1.1. Significance using country-clustered standard errors: p<0.01, ** p<0.05, * p<0.1.

in country-by-country first-stage regressions (specification 1.2) using Standard International Trade Classification export data from the start and end of the time interval on (c) familiarity (i.e., prior $\hat{\beta}_c$) and (d) complexity (i.e., prior $\hat{\gamma}_c$).

The correlations between these years of schooling in 1995 and these four dynamism measures are higher in 1985–1995 than during 1975–1995. In 1985–1995 they are, in the same order: (a) 0.195, (b) 0.132, (c) 0.408, and (d) –0.237. Only the third of these correlations, between prior $\hat{\beta}_c$ and 1995 years of schooling, is statistically significant at even the 10% level.

Table 1.11 checks whether our finding that education helps to overcome familiarity reflects a spurious correlation owing to these relationships. Each regression interacts prior dynamism with familiarity and the PCI. Our preferred estimates are those appearing in columns 5 and 6, for two reasons: they use the same notion of dynamism in the earlier and later periods; and they include the dynamism measure that correlates most strongly with schooling (and so, a priori, is the likeliest source of omitted variables bias). With the limited exception of one of the four coefficients in the two less preferred specifications, the results do confirm that countries whose export mixes were more dynamic between prior to 1995 were more likely to move into unfamiliar and complex products between 1995 and 2015. However, perhaps because most measures of prior dynamism are not strongly correlated with education, these corrections do not alter our findings that countries with higher average years of schooling and higher primary attainment were more likely to develop comparative advantage in unfamiliar products. Moreover, the role of college and especially primary attainment in overcoming familiarity appears much stronger once these corrections are made.

Table 1.12 checks our results for robustness to corrections for neighborhood effects—spillovers of knowledge and supply chains to nearby countries (Bahar et al., 2014, 2019). We capture these effects by correcting for the weighted average familiarity in 1995 of every other

Table 1.12: Correcting for Neighbors' Specialization Patterns

	(1)	(2)	(3)	(4)	(5)	(6)
Familiarity	1.917***	1.916***	1.924***	1.918***	1.911***	1.921***
Familiarity x Quantity	-0.655***	-0.655***	-0.646***	-0.654***	-0.659***	-0.646***
Familiarity x Quality A	0.114	0.115	0.062	0.112	0.118	0.064
PCI x Quantity	-0.029**	-0.029**	-0.030**	-0.028**	-0.028**	-0.030**
PCI x Quality A	0.021**	0.021**	0.023**	0.021**	0.021**	0.023**
ΔK/L x ProdKL	0.015	0.015	0.013	0.015	0.015	0.013
ΔQuantity x ProdYrs	-0.014	-0.014	-0.016	-0.014	-0.014	-0.016
Quality A x ProdYrs	0.002	0.002	-0.004	0.002	0.002	-0.004
ΔQuantity x Quality A x ProdYrs	0.005	0.005	0.011	0.006	0.005	0.011
Neighbors' Familiarity with the Product		-0.001	-0.015			-0.010
Neighbors' Familiarity x Quantity			-0.018			-0.025
Neighbors' Familiarity x Quality A			0.068**			0.076**
Neighbor's RCA in the product				-0.002	-0.002	-0.003
Neighbor's RCA x Quantity					0.002	0.002
Neighbor's RCA x Quality A					-0.002	-0.003***
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Product Fixed Effects	✓	✓	✓	✓	✓	✓
Corrections for Lagged RCAs	✓	✓	✓	✓	✓	✓
Observations	54,560	54,560	54,560	54,560	54,560	54,560
R-squared	0.264	0.264	0.265	0.264	0.265	0.265

Note: Neighborhood familiarity is the average familiarity of all other countries in the world with product p, weighted by the inverse of each country's distance from country c. Neighborhood RCA is analogous. Estimates from a linear probability model of comparative advantage in 2015, per specification 1.1. Significance using country-clustered standard errors: p<0.01, ** p<0.05, * p<0.1.

country with the product in question—where the weight is simply the inverse of that country's geographic distance from country c, and an analogous measure of neighbors' RCA in 1995. Column (1) reprises our main results, which do not change when correcting separately for neighbors' familiarity or RCA (columns 2 and 4), for these variables interacted with education (columns 3 and 5), or all of the above (column 6).[24]

[24]Education quality is associated with developing comparative advantage in products with which neighbors are familiar—consistent with quality education facilitating knowledge flows across borders, but that neighbors do not actively produce—consistent with it helping to overcome competition.

Table 1.13: Regressions in Subsamples of Core and Peripheral Products

	Core Products			Peripheral Products		
	(1)	(2)	(3)	(4)	(5)	(6)
Familiarity	2.906***	2.523***	2.935***	2.101***	1.837***	2.118***
Familiarity x Quantity		-0.700**			-0.901***	
Familiarity x Quality A		0.536	0.136		-0.300	-0.498*
Familiarity x Primary			-0.075			-0.677**
Familiarity x Secondary			-0.328			0.206
Familiarity x College			0.148			-0.182
PCI x Quantity		-0.025*			-0.024**	
PCI x Quality A		0.016	0.014		-0.012*	-0.015**
PCI x Primary			-0.024**			-0.027***
PCI x Secondary			-0.017			-0.002
PCI x College			0.019			0.007
ΔK/L x ProdKL	0.014	0.022	0.012	0.027***	0.038***	0.037***
Quality A x ProdYrs		-0.018	-0.007		-0.000	0.005
ΔQuantity x Quality A x ProdYrs		0.029	0.016		0.008	0.003
ΔQuantity x ProdYrs		0.011	0.007		-0.010	-0.011
Country Fixed Effect	✓	✓	✓	✓	✓	✓
Product Fixed Effect	✓	✓	✓	✓	✓	✓
Corrections for Lagged RCAs	✓	✓	✓	✓	✓	✓
Observations	10,164	10,164	9,933	10,208	10,208	9,976
R-squared	0.2557	0.2583	0.2574	0.3189	0.3295	0.3300

Note: The estimation is based on linear probability model, per specification 1.1. Significance using country clustered standard errors: p<0.01, ** p<0.05, * p<0.1.

1.4.4 Core and Peripheral Products

Table 1.13 presents results on the subsets of core and peripheral products. Our main results, that countries with higher average years of schooling were substantially more likely to develop comparative advantage in unfamiliar products, holds for both subsets.

However, there are four differences between core and peripheral products. First, while countries with education of a higher quality exhibit a slight tendency to develop comparative advantages in unfamiliar peripheral products, this is reversed for core products.[25] Second, primary attainment was associated with learning to produce unfamiliar periph-

[25]When using unclustered standard errors, the coefficients on the familiarity–quality interaction for core products in columns (2) and (3) have p-values of 0.008 and 0.092. The analogous p-values in the peripheral regressions (5) and (6) are 0.018 and 0.000.

eral products, but this is not the case for core products. Third, capital accumulation is significantly associated with movements toward capital-intensive products among peripheral products, but not among core products. Fourth, the effects of familiarity are larger among core than among peripheral products. All of these results suggest that it is hard to develop comparative advantage in core products, and that general education is particularly important for moving up lower rungs of the product ladder.

1.5 Conclusion

We have analyzed the relationship between education and the evolution of comparative advantage among low- and middle-income countries. We found strong evidence consistent with education helping countries to develop comparative advantages in unfamiliar products—products that are unrelated to those in which they already have comparative advantages. In contrast, controlling for familiarity, more educated countries were not much more likely to develop comparative advantages in complex products, and those countries that experienced faster increases in education levels were not more likely to develop comparative advantages in education-intensive products. Taken together, these results are more obviously supportive of an approach to the development of comparative advantage that emphasizes relatedness between products than of a Heckscher–Ohlin–Vanek approach. The relatedness approach emphasizes that the process is evolutionary, with productive capabilities developing in path-dependent fashion. Education's core contribution to the process, it appears, is to reduce this path dependence by facilitating longer leaps into previously unfamiliar products.

While there are no plausible instruments for our key independent variables, several types of auxiliary evidence suggest that these results should be taken seriously. They are robust to many changes in specification, to changes in how product sophistication is

operationalized, and to corrections for national, institutional, infrastructural, and FDI-related variables. We also find that primary education and education of a higher quality is most strongly associated with overcoming unfamiliarity when developing comparative advantage in peripheral products, but not in core products—which is consistent with the widely accepted idea that core products require the acquisition of more capabilities, but also with the idea that quality basic education is important for amassing basic capabilities. Perhaps most importantly, our results are not explained by the fact that the most educated countries in 1995 tended to be a little more dynamic in the prior decades. Lacking the micro data needed to test alternative mechanisms, we remain agnostic regarding how education might facilitate shifts toward unfamiliar products. Three possibly complementary mechanisms appear worth exploring in light of related research. First, developing new capabilities requires the acquisition of tacit knowledge and its transfer across less related activities (Hidalgo et al., 2018). Education may well facilitate both a more rapid acquisition of tacit knowledge through learning by doing, and greater efficiency in assembling teams with the right knowledge mix. Second, a more educated workforce might be better placed to acquire the additional pieces of book knowledge required to make inroads into new industries. Third, education may permit actors to identify and take advantage of new disequilibrium opportunities (Schultz, 1975). There is evidence that new industrial strengths develop where downstream industries already exist (Bahar et al., 2019), which is consistent with changes in the organization of supply chains creating such disequilibrium opportunities.

These results do not mean that education is unimportant for developing countries seeking to move from peripheral products to complex, core products. They imply that education is important, but not because products in the core are complex. Education is important because core products are unfamiliar to these countries, given what they already produce. Put differently, education would be no less helpful in the unlikely sce-

nario that a developing country wished to develop comparative advantage in an unfamiliar peripheral product.

Finally, we caution that education differences account for rather little of the cross-country variation in export diversification. Education variables have limited explanatory power overall, and while education is associated with less path dependence, it does not eliminate it. Even the most educated countries tend to develop comparative advantages in products that are related to those they already produce. This path dependence suggests that industrial development needs to be partly a deliberate process, with governments facilitating the development of a series of stepping-stone industries that the economy can traverse on its way to the core. Our findings suggest that investments in education allow the stepping stones to be spaced a little further apart.

Chapter 2

The Influence of Foreign-born Directors on the US Film Industry

2.1 Introduction

This paper studies the effect of foreign-born directors on the quality of American films. I construct a unique dataset on the US-produced (or co-produced) films released between 1925 and 2019. The data allow me to document film attributes, including production budget, box office, Oscar nominations and awards, distributors, and director's name and identifier matched with the director's country of birth. There are many ways to quantify the value of films, and I use Oscar awards and box office revenue, where Oscar awards are a proxy for the cinematic value, and the box office is for the commercial value of the films.

Two main issues arise in identifying the effect of a foreign-born director on a film. One is to disentangle the director's effect from the distributor's effect, and the second is to control for the potential film quality that a foreign-born director may be disproportionately selected into. To deal with the first issue, I start with the standard fixed effect

analysis and control for a full set of individual distributors' fixed effects. To address the second issue, I control for the estimated production budget to capture the value added by a director assignment that is additional to the possibly endogenous quality of a film prior to director assignment.

My empirical analysis shows that the foreign-born directors' films are 1.87% more likely to be nominated into Oscars, and they receive 0.58 more nominations and 0.46 more awards than the native-born directors' films conditional on the nomination. I also find that a foreign-born director does not result in an overall higher gross box office (worldwide box office); however, a foreign-born director yields 29.9% higher international box office; correspondingly, a foreign-born director yields 16.4% lower domestic box office. In addition, I find that a foreign-born director is more likely to direct films co-produced by the US and another country (countries). This finding is utilized for instrumenting a foreign-born director assignment in a two-stage analysis. The results based on the IV analysis are consistent with the results from the main specification.

Foreign-born directors may induce a positive effect on the film outcome for several reasons. Foreign-born directors are subject to positive selection, where the positive selection applies to the foreign-born directors who migrated early with their parents to the U.S. and also to the directors who established careers overseas and were recognized by Hollywood for their film directing capability.[1] In addition, as Hirschman (2005) points out, foreign-born directors may be more successful in film directing due to their exposure to multiple cultures and bearing a higher level of cultural resources. Even though the positive selection or a higher cultural resource may help explain an overall higher film directing capability, neither one of the two factors may fully account for the differential box office outcomes of the foreign-born director in the domestic and international

[1]Studies show that immigrants are a self-selected group regarding the sense of achievement and individual capability. (Borjas, 1987, 2001).

markets. Moreover, the positive selection and cultural capital are embedded in the director's intrinsic capability and are difficult to pin down as precise mechanisms for film success. I provide additional empirical results in the following sections to help interpret the differential results.

Specifically, I test for whether the positive effect of a foreign-born director in the international market is realized through the director's home-country bias or the outsized market power of certain importing countries that a foreign-born director may have more ticket power in, if the a higher intrinsic capability does not fully account for the positive effect of a foreign-born director in the international market. To test for the effect of home bias, I leave out the revenue from the director's home country and estimate the foreign-born director effect in the rest of the international markets. The effect of a foreign-born director is attenuated, but the positive effect on international revenue persists and is significant. In other words, home bias does not fully determine the effect of a foreign-born director. I then conduct similar exercises to leave out the revenues from influential importing countries to test for the market size effect in confounding the effect of a foreign-born director. The results show that market size is limited in explaining the foreign-born director effect, which is suggestive of a more universal relative advantage of a foreign-born director in film directing catering to the international market.

My main identification strategy depends on using a set of film fixed effect controls to put films on the comparable common ground and using the production budget to account for additional endogeneity in film quality for director assignment. To verify the robustness of such strategy in accounting for the latent factors of films, I use a suite of state-of-the-art natural language processing tools to construct twin film "pairs" (pairs may contain more than two films) based on the film plot summaries, and test for the robustness of the foreign-born director effect aggregated from each film twin pair. The film plot summary is usually written in 300 to 700 words to deliver condensed information

of a film, and it provides information that other attributes may not carry, for example the film's topical issue. *United 93* and *Flight 93* both chronicle the events aboard the United Airlines Flight 93, and they form such film twin pairs. Testing of the foreign-born director effect with the additional metric on film similarity shows robustness of my main estimates. Overall, a foreign-born director results in a lower domestic box office, a higher international box office, and no difference in the worldwide box office.

Consumers' tastes for goods exhibit home bias, and like many other exporting goods, films experience a value discount in translating to a foreign market. Among films, the so called "foreignness liability" Zaheer (1995) appears more prominently for the films that are culturally relevant (Holloway, 2014). Miskell (2014) also finds that American comedy and romance films were significantly less popular with international audiences than domestic ones for the films in the 1940s. I conduct sub-sampling exercises with film genres that differ in cultural relevance and test for the effect of a foreign-born director in the domain of domestic and international markets. I find that the effect of a foreign-born director in the international market is more prominent for the films (genres) that are culturally relevant–comedy, drama, than for the films that are not–action and adventure. In other words, a foreign-born director's comedy films are more popular than a native-born director's comedy films in the international market. Domestically, however, consumers show preference for the culturally relevant films directed by a native-born director, where the foreignness liability does not apply and the demand for translation is low given the same level of expected film quality. The findings suggest that the value of a foreign-born director partially is in mitigating the cost of translating culturally relevant American films to the international market.

To enhance the argument that a foreign-born director's effect is partially in mitigating the translation cost of the American cultural products to a foreign market, I test for the hypothesis that a foreign-born director with a higher linguistic capability and hence

a better understanding of the American culture is better at mitigating the translation cost of the cultural products to the international market. I use the director's second occupation to construct a proxy for the foreign-born director's linguistic capability. I assume that a foreign-born director whose another proclaimed job title is *Actor*, and thus is Actor-Type, is an indicator for a higher linguistic capability. The assumption implicitly nests that native-born directors have relatively homogeneous linguistic capability. Interacting the indicator of a director being Actor-Type and her nationality, I find that the capability of being an actor contributes to a higher revenue for a foreign-born director, however, it does not affect the film outcomes for a native-born director. The differential effect of being an Actor-Type of director for the foreign- and native-born directors suggest that the linguistic capability and the associated cultural understanding play a role for the directing capability and are essential for a foreign-born director to mitigate the translation cost of the films to the international market.

Overall, the findings imply that foreign-born directors' films have relative advantage in international markets, native-born directors' have relative advantage domestically, and the advantage of a foreign-born director is more prominent in appealing culturally relevant American films to the international market. I then aggregate the effect of foreign-born directors in box office by decade and test for the evolution of the relative advantages of the two director groups. The results show that the disparity in relative advantage between the two director groups' films was highest in the 1970s and 1980s, which coincides with the New Hollywood movement–an organizational change in filmmaking initiated by creative independent filmmakers experimenting with radical perspectives in filmmaking. The disparity has then been decreasing over time; the performance of the two director groups has been converging in both markets, as could be a reflection of the cultural globalization and blockbuster films taken as a formula for commercial success.

Related Literature

This paper contributes to understanding the effect of immigrants on the US economy. There exists a large literature examining the relationship between immigration and labor market, they focus on the complementarity or substitutability of immigrants with the natives as production inputs (Altonji et al., 2012, Borjas, 1999, Borjas and Doran, 2012, Card, 2001, 2009, Ottaviano and Peri, 2006, Peri and Yasenov, 2015). My paper studies the foreign-born workers' impacts on producing cultural products, and the findings shed light on understanding the immigrants' role in producing the non-rival goods, and the making of American society and culture (Hirschman, 2005).

This paper is closely related to the studies of immigration and innovation using patent data (Burchardi et al., 2019, Hunt and Gauthier-Loiselle, 2010, Moser et al., 2014). Patents measure the potential contribution of high-skilled immigrants to economic growth. However, the monetary value of the patents is usually difficult to quantify. My paper studies the effect of foreign-born talents in the context of the film industry, where innovation is realized by the film directors' capability of managing creativity and creating higher-quality films. With films, the effect of foreign-born talents (directors) is easily monetized.

My study also contributes to the literature on predicting film outcomes, and it demonstrates why a foreign-born director's effect can be economically relevant. Film directors play essential roles in distinguishing films' financial and aesthetic success (John et al., 2003). Ravid (1999) shows that directors, among many important factors in the making of a film, matter significantly for film revenues after controlling for the effect of star actors/actresses. De Vany (2003) gives evidence about the uncertainty feature of the film industry and documents the effect of the market structure, distributive system, film rating, and the star actor effect on film outcomes with the films released between 1985 and

1996. My study complements the literature by quantifying the effect of a foreign-born director on films, and shows how a director's nationality may add prediction power for the film outcomes.

Studies on the film industry mostly focus on the industrial organization aspect. For example, Hui and Png (2002) look at the impact of economic incentives on the US movie production, Gil and Spiller (2007) study organizational implications of creativity, Gil and Lampe (2014) examine the impact of competition in the television industry on the number of movie theaters between 1993 and 2005. Hanson and Xiang (2009) develop a theoretical framework for examining international trade in information services, and apply the framework in analyzing trade in motion pictures. Galenson (2010), however, articulates the directors' filmmaking styles and the correlation of the directing styles with the directors' life-cycle creativity. My study also considers films as creative cultural products and study the value added by the film directors as creators based on their foreign-born background.

Background

2.1.1 Historical context

Hollywood is one of the world's oldest and largest national film industries. At its height of popularity in the mid-1940s, the studios produced about 400 movies a year, involving an audience of 90 million Americans per week (Cook and Sklar, 2019).[2] The Bureau of Economic Analysis built the Arts and Cultural Production Satellite Account (ACPSA) to recognize the development in the sector beginning 2012. According to the ACPSA, the film industry comprises the single most outstanding share of the US exports

[2]David A. Cook and Robert Sklar, *History of the motion picture*, Encyclopædia Britannica, January 08, 2019.

of arts and cultural goods and services. Motion picture and video services comprise 4 percent of the total U.S. trade surplus in services and generate $10.3 billion in trade surplus in 2018.

Besides the financial success of the film industry, films have substantial social implications; they affect public opinions and discourse. *Gone with the Wind* controversy is one of the famous examples that address the impact of films on cultural debate. The film has been denounced for a wrong depiction of slavery and stereotyping black people as naive and simple-minded. The racial tensions following the death of George Floyd brought up the film again as an important discussion point about race.[3] In addition, Science Fiction films are also known to inspire people's thinking about society and technology for the future (Maynard, 2018). Stanley Kubrick's *2001: A Space Odyssey* released in 1968 still inspires people's imagination about the future. *Ghost in the Shell* released in 1996 talks about the human-machine interfaces and explores the meaning of being a human, which seems to present a plausible discussion for now. Hirsch (1972) also stresses that the value of commercial cultural products is beyond a clearly utilitarian purpose. Hollywood influences the cultural identity of the Americans, and it showcases to the world what is like to be American.

Hollywood has been a globalized industry since the beginning of its establishment, and the talent flow to Hollywood has been reflecting the receptivity of American institutions to global talents. First-generation immigrants founded two of the major-five film studios, Paramount and Warner Bros., back in the 1910s and 1920s. In 1928, *Variety* magazine highlighted the diverse composition of foreign-born film workers in the industry: "A census of foreign-born studio workers, prominently and actively engaged in making pictures in Hollywood, shows a total of 189, of whom 15 are producers or executives; 36 in directorial positions; 14 writing for the screen; seven in various technical occupations;

[3]Explained: Why the Hollywood classic 'Gone with the Wind' is in controversy, again

78 actors and 39 actresses representing 27 countries foreign to America except Hawaii which is a territory of the United States."[4]

Foreign-born film directors make a substantial fraction of the foreign-born workforce in Hollywood, and there has been a steady flow of foreign-born directors to Hollywood over time. According to the American Community Survey 2017, about 25,000 foreign-born workers are working in the film industry for different positions, including actors, producers, and directors; and they make up about 12.5% of the workforce. During World War I, film directors from European countries arrived at Hollywood to seek career success. Hollywood, at the same time, searched actively for talented film directors from abroad who had proved their directing capability. Especially since, unlike acting or writing, directing did not require high English language proficiency. Directors who arrived during the World War I include Ernst Lubitsch (German Empire), Alfred Hitchcock (England), Fritz Lang (Austria-Hungary), and Jean Renoir (France). The rise of Nazism resulted in another wave of talented director migration from Germany, including directors like Billy Wilder and Robert Siodmak. By the late 1930s, 28.7 percent of film directors in Hollywood were foreign-born (Regev, 2018), and the fraction has been steady at the level over time.

Foreign-born directors and their films have been continuously recognized for cinematic achievement and commercial success. Hirschman (2005) traced the immigration background of the directors who have won multiple times the Best Director award by the end of the 20th century and found that nine out of 17 of the recipients were foreign-born. I update the list to the most recent records, and find that among 21 film directors with multiple awards, 12 are foreign-born, 4 are second-generation immigrants, and 5 are third- or higher-order immigrants (Table A4). The list also shows that the early recognized directors are from Europe, the later ones from Latin America and Asia. This

[4]Talent From 30 Nations, *Variety*, Aug. 22, 1928.

pattern coincides roughly with the US immigration waves. During the last ten years, eight of the Best Director Oscar awards are received by foreign-born directors (Table A2), and most of the directors are from Asian and Latin American countries. Among the top 30 grossing US films of all time, ten are directed by foreign-born directors (Table A3).

2.1.2 In making of a film

The making of a film involves numerous participants, joint activity, and cooperation; it is a collective activity like many art works as described in Becker (2008). Roussel (2017) documents the complex process and amount of energy it takes to match talents and resources for the films. The filmmaking process can be idiosyncratic, not fall on one single sequence of steps or set of formulas. However, a typical filmmaking takes the stages of development, pre-production, production, post-production, and lastly to the stage of distribution.

In the development stage, studios, who would usually be the distributors in the stage of film distribution, weigh on various filmmaking ideas and decide on the financing for different projects. Then in the pre-production stage, major cast and crew members, most importantly the director, are recruited to match up with the project given the budget. I assume that the recruiting of either a foreign- or a native-born director is to maximize the film outcomes given a certain level of production budget and film attributes, and the expected film outcome is independent of a foreign-born director assignment. The recruitment is usually project-based, which ensures there is enough movement of the talents to match with different distributors.[5] Then it goes to the production and post-

[5]The contracts for the early days Hollywood are term-based, which means during the contract terms, the talents can only work on the films produced by the contracted studios. But the system has been relinquished in the late 1950s. Contracts for the series films, for example the Lord of Rings, actors are contracted for multiple films at once.

production stages, where the directors get to visualize the screenplay and guide the film cast and crew members to fulfill the vision.[6] In the final stage of the film distribution, the studios (distributors) utilize their distributing networks to release the films. The film genre, the potential audience, and the past experience on similar films determine the releasing format (wide or limited; how many screens to allocate to the films), which in turn determines the film revenues.

Director

Film directors play essential roles in distinguishing films' financial and aesthetic success (John et al., 2003); they are known to be the main author of the film, and their roles in films are comparable to the ones of CEOs in firm operation. Directors guide the cast and crew members to fulfill their visions for the screenplay. Films with similar scripts could be visualized very differently depending on the directors' understanding of the scripts and their artistic visions for the films. *Cabin Fever (2016)* was almost a shot-by-shot remake of *Cabin Fever (2002)* with the same script. Despite the improved costumes and sets, the 2016 version was criticized for its director showing a lack of human comprehension and failing to portray the charms of the characters from the 2002 version. The 2016 version only received $39,065 in box office whereas the original received $30.6 million. Film outcomes are vastly difficult to predict. William Goldman once famously stated that the film industry is a field where "nobody knows anything" about how the products are going to turn out in the market, which addresses its feature of uncertainty.[7] Ravid (1999) shows that directors among many important factors in the making of a

[6] In the studio era (back in the 1940s), studios have a big control over the post-production, the cuts, but there is no reason to think the system may have affected foreign- and native-born directors differentially and thus to confound the effect of a foreign-born director.

[7] William Goldman was an American novelist, playwright, and screenwriter, who had won Academy Awards for his screenplays Butch Cassidy and the Sundance Kid (1969) and All the President's Men (1976).

film, however, do matter significantly for film revenues after controlling for the effect of star actors and other film attributes.

Distributor

Film distributors (studios) are essential for films' commercial success. They are usually the big major studios who are also involved in financing and making the films. Film sales depend on the distributor network–how many screens a distributor can allocate to the film, and whether the distributor has international releasing capability–even though the matching with a certain distributor is also endogenous to the film attributes. Distributors are different from production companies. Production companies can be formed for certain projects in a rather short period of time, and the goal is to maximize the film quality garnering financial and human resources, whose role is mostly reflected in the pre-production and production stages. Distributors on the contrary control the final stage of film distribution, they are usually the subsidiaries of the major studios, and have a more stable lifespan and limited in number than the production companies, and thus the estimation of the individual fixed effect is more tractable. Many independently made films are also distributed through the major distributors and their subsidiaries, and controlling for the distributor fixed effect instead of the production company fixed effect makes the estimation more stable.

2.2 Data

For the purpose of the study, I construct a unique director-distributor matched dataset from four main sources: OpusData, StudioSystem, Wikipedia, and Google search knowledge panels.

First, OpusData (OD) provides information on films released in the US every year.[8] I obtain data on the US-produced (co-produced) feature films released between 1925 and 2019.[9] For each film, OD reports technical credits, acting credits, distributor, release date, cumulative domestic and international box office, and estimated production budget. The domestic box office includes the ticket sales from the US and Canada. The international box office includes the ticket sales from the rest of the world that the film is officially released to. By taking the sum of the two, I get cumulative worldwide box office for each film. All the revenues are in nominal terms. In the regression analysis, I include the year fixed effect to address inflation. In addition, the data provide countries involved in making the film, which I later exploit to construct an instrumental variable for the director assignment. OD also provides film attributes like genre, MPA rating, release type, whether a film is a sequel (franchise), and the languages spoken in a film.

There are several features of the OD database useful for my study. It provides cumulative and also the country breakdowns of the international box office. The breakdowns are useful to test for director home country bias, cultural similarity, or importing country's market size. OD also records each entity–film or person involved in making the film–with a unique numerical identifier, which makes it possible to distinguish the entities with the same name in the data. One shortfall of OD is that they track the information on films for which there is a reported box office revenue (domestic or international) or a video sales estimate (DVD or Blu-ray). The non-commercial films or the ones with no recorded box office would likely be excluded from the sample. I get in total 12,941 film records from the OD database. To use fixed-effect analysis to identify the foreign-born director

[8]OpusData provides film industry data by tracking box office revenue on a daily basis and provides data services for research purposes for films.

[9]The whole set of records runs from 1902 to 2020. I exclude the data before 1925 because there was no consistent observation of films and most of the films released during those days are short films. The films released after 2019 are excluded because the market experienced an unprecedented shock due to the pandemic.

effect while disentangling the effect of the distributors, it is important to have moving of the foreign-born directors across different distributors. Each director on average directs 2.3 films, and the director who has the most recorded films in the data is Woody Allen with 38 films. A director's films are distributed on average by 3.129 distinct distributors, a distributor would on average release 352.4 unique directors' films.

StudioSystem (SS) is my second data source.[10] SS provides similar attributes on films as in OD, and it has film technical and acting credits, production and distributing companies, releasing date, genre, estimated budget, cumulative domestic and worldwide box office. SS complements OD in the awards information (*ten most recent awards*) for the available films, including whether a film has received or been nominated for an Academy Award. However, SS does not provide country breakdowns of the international box office. Both SS and OD tend to have more complete information on films released in the recent years (after the 1970s), but SS provides a larger dataset for films, as it includes those non-commercial films or films with zero box office. I get the records of 27,557 US-produced (or co-produced) films released between 1925 and 2018, which is about twice the size of OD. However, as mentioned, most of the films do not have recorded box office, and the summary statistics of the films with available box office information show comparability to the ones of OD in the box office distribution (Table 2.1).

In addition, SS allows me to access data on directors active between 2015 and 2020.[11] The data include directors' demographics–gender, birthplace, age (as of the data extraction date), and race. The data also include proclaimed job titles of the directors–some directors can also be actors, producers, writers, cinematographers and other technicians, and the director's last ten credits and last five film credits for directing.[12] Each entity

[10]StudioSystem is a subscription-based database that is owned by Gracenote, Nielsen. It is one of the most reliable entertainment industry data sources for TV, film, and digital data.

[11]Being active means that the directors are involved in at least one project in development, in pre-production, being scheduled on a release, or having been released.

[12]The current dataset is based on data extraction on date Sept. 27, 2020.

(film and director) has its unique numerical ID to be used for matching. However, the SS film data do not contain the director's ID number. For films associated with duplicated director names, I merge the director information by checking a match between the film title and the titles mentioned in last five film credits from the SS director data.

I then merge the films' box office and awards, and associated director country of birth information from SS to OD data based on the film title and release date information. I matched 7,718 out of 12,579 films in the OD data. Since SS does not provide the country breakdowns of the international box office revenue, which is essential for my study, I use the box office information from OD for my analysis of the commercial values, and I use SS to check for robustness. A director's birthplace is another essential piece of information for my study. SS provides information for the directors who were active in recent years, and the information for directors of the early films is mostly missing. I use my third and fourth data sources to fill in the missing data.

My third data source is Wikipedia. From Wikipedia film and film director-related pages, I mainly scrape the following information: film summary plots, Academy Awards information associated with the available films, and directors' birthplace information for those directors who are known to be American film directors. From the Academy Awards pages, I get from each year of the ceremony, the title of films, the number of nominations and awards, nominated or awarded Best Picture, the directors who have been nominated and/won Best Director. Then from the American film directors' list, I take the name and Wikipedia page URL for each director and then scrape the Infobox from each director's page. Wikipedia Infobox is used to provide essential information about the director, and it usually includes the director's birth date, birthplace, occupation, etc. This third source of data helped me identify 2,504 American directors' foreign-born status.

Lastly, I utilize SerpAPI service for Google search to scrape the knowledge panel (KP) of each director missing the birthplace information from both SS and Wikipedia. KP

is the information box that appears on the right-hand side of the browser screen when you search for an entity, for example, a person, place, or organization. It is similar to the Wikipedia Infobox, and it is meant for displaying a quick snapshot of information on a topic based on Google's understanding of available content on the web. For a search result with a director's name, it includes the director's birth date and birthplace, familial relationship, short filmography, etc. I scrape the information to fill in the missing from SS or Wikipedia for the directors.

Combining the above data sources, I construct a unique dataset that allows me to investigate the effect of having a foreign-born director on the film. A director is defined to be *Foreign-born* in the data if they were born outside of the US and have directed at least one US-produced (or co-produced) film. My data do not identify the directors' legal resident status. Some foreign-born directors are the ones who established careers outside of the US and migrated for working on the film projects, and some foreign-born directors are more home-grown who migrated in their early age with family. The latter is in a more strict sense immigrant, but my data do not distinguish the two.

2.2.1 Summary Statistics

With all the data sources combined, I identify 59.5 percent of the directors with birthplace information and hence their foreign-born status. The identified directors are associated with 62.7 percent of the films in the sample. They direct significantly higher budget films and yield higher box office revenues than the unidentified director group.

Table 2.1 provides the summary statistics on the film production budget and box office based on the director nationality.[13] Firstly, it shows that foreign-born directors'

[13]There are other commercial values can be considered for film performance depending on the distribution channels, for example DVD sales, and television or in-flight sales. In my study, I focus on the revenues generated from theater distribution, and the box office indicates the aggregate revenues from possibly multiple releases in different territories.

films have a higher average box office both domestically and internationally, and foreign-born directors also manage higher production budget films, which may suggest an overall positive selection of the foreign-born directors in the director population. In addition, directors whose birthplace information is missing tend to have a lower average box office and manage a lower production budget than the other two groups. This implies that the reason for the missing information is that the directors are less commercially successful and whose information is less known to the public. The unidentified director group can only bias the estimation if foreign- or native-born directors over-represent the unidentified group. Given that the native-born directors encounter a relatively lower entry cost to the industry, if there would be any bias, the estimate of the foreign-born effect is more likely to be downward biased. The table also shows the comparability of the two data sources–SS and OD–based on the domestic box office and production budget. I use OD as my main source of data for the box office analysis, and SS for robustness check. OD may be subject to positive selection in films as they collect data on the films that have commercial records. However, the selection does not show up obviously from the summary statistics, and even with the case of sample selection, it would unlikely to bias the estimate of foreign-born director effect as OD do not select the films based on the directors' foreign-born background.

Film commercial data are usually highly skewed. As is shown in the Table 2.1 for the box office and production budget. The variables being highly skewed may cause the estimation to be driven by extreme values. I take the logarithmic transformation of the box office and production budget to address the issue. As the production budget is determined based on the estimation of the film revenues, controlling for the production budget in the regressions largely addresses the issue and make it possible to study the effect of a director on the value-added basis.

Foreign-born directors are associated with about 27 percent of the Oscar nominated

films produced in the US (Table 2.2). This sample includes the Oscar nominated films that are matched with the commercial information in the main dataset. The table shows that foreign-born directors' films on average receive more nominations and they also receive more awards. However, it does not show significant difference in the probability of their films getting Best Picture award or the director getting Best Director award.

Table 2.1: Descriptive Statistics

	Foreign-born	Native-born	Unidentified	Overall
Domestic Box Office (SS)				
Count	1175	3189	1688	6052
Mean (SD)	46.41 (63.68)	42.29 (66.37)	8.89 (17.87)	33.77 (58.64)
Median (IQR)	25.12 (54.44)	17.95 (53.49)	0.46 (9.40)	11.02 (41.54)
Missing	979	2089	2630	5698
Domestic Box Office (OD)				
Count	1773	4596	3028	9397
Mean (SD)	41.07 (66.66)	40.23 (68.33)	8.32 (18.55)	30.10 (58.80)
Median (IQR)	17.06 (48.84)	15.95 (47.49)	0.47 (8.27)	7.71 (34.66)
Missing	381	682	1290	2353
International Box Office (OD)				
Count	1409	2967	1495	5871
Mean (SD)	71.57 (145.18)	59.61 (125.98)	9.44 (31.30)	49.71 (117.91)
Median (IQR)	17.00 (76.95)	10.80 (60.29)	0.24 (3.73)	4.74 (45.17)
Missing	745	2311	2823	5879
Worldwide Box Office (OD)				
Count	2007	4913	3645	10565
Mean (SD)	86.52 (185.06)	73.63 (163.45)	10.79 (34.84)	54.40 (142.72)
Median (IQR)	17.19 (84.77)	15.60 (66.78)	0.34 (6.53)	6.00 (41.31)
Missing	147	365	673	1185
Production Budget (SS)				
Count	1064	2734	1198	4996
Mean (SD)	43.28 (45.73)	34.77 (40.94)	12.86 (16.79)	31.33 (39.35)
Median (IQR)	26.00 (41.00)	20.00 (35.00)	7.32 (14.50)	18.00 (33.00)
Missing	1090	2544	3120	6754
Production Budget (OD)				
Count	1050	2545	786	4381
Mean (SD)	48.54 (47.99)	40.96 (47.44)	16.30 (22.52)	38.35 (45.45)
Median (IQR)	31.25 (52.50)	25.00 (45.00)	9.00 (17.63)	23.00 (41.50)
Missing	1104	2733	3532	7369

Notes: The table shows summary statistics for film box office and production budget in the three groups of directors–Foreign-born, Native-born, and Unidentified. All values are measured in millions of U.S. dollars without the adjustment for inflation. Directors with unidentified nationality are associated with about 37% of the films in the overall sample, and they direct lower level budget films in general and their films have box office that is at the lower tail of the distribution, compared to the other two groups of the directors. SS and OD indicate the data sources. The two data sources show comparability in the domestic box office and as well as the production budget. Based on the first two columns, the average box office is higher for the foreign-born directors' films both domestically and internationally, and foreign-born directors tend to manage a higher budget films according to the production budget provided at the lower panel of the table.

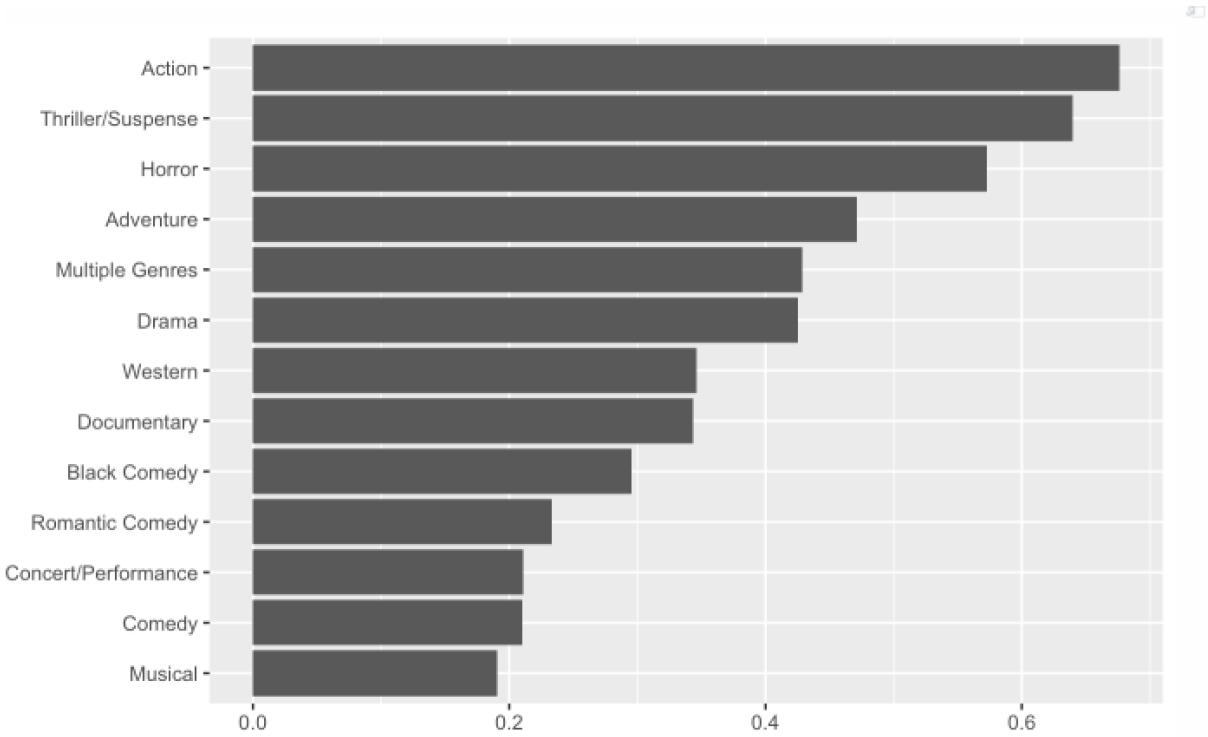
Table 2.2: Descriptive Statistics for Oscars

Variable	**Foreign-born**	**Native-born**	**Unidentified**	**Overall**
Number of Nominations				
Count	149	363	48	560
Mean (SD)	5.52 (3.45)	4.98 (3.21)	4.10 (3.14)	5.05 (3.28)
Missing	2012	4921	4271	11204
Number of Awards				
Count	143	352	45	540
Mean (SD)	2.28 (2.11)	1.97 (1.61)	1.67 (1.33)	2.03 (1.74)
Missing	2018	4932	4274	11224
Best Picture				
Count	149	363	47	559
Mean (SD)	0.13 (0.33)	0.12 (0.32)	0.06 (0.25)	0.12 (0.32)
Missing	2012	4921	4272	11205
Best Director				
Count	149	362	47	558
Mean (SD)	0.12 (0.33)	0.12 (0.32)	0.02 (0.15)	0.11 (0.31)
Missing	2012	4922	4272	11206

Notes: The table shows summary statistics for the Oscar awards by the director's nationality. The first two columns of the table indicate that about 150 films by foreign-born directors and 363 films by native-born directors are associated with the Oscars, and foreign-born directors have a slight mean advantage in their films being associated with a higher number of nominations, actual wins, probability of their film getting Best Picture, and getting a Best Director award, over the native-born directors. Comparing the first two columns with the third column shows that directors whose nationality is identified have a higher outcome in terms of all the Oscar measures. Foreign-born directors show on average higher number of nominations and actual awards than the native-born directors, however, they or their films do not show much difference in getting Best Director or Best Picture award.

Figure 2.1 shows the fraction of films directed by foreign-born directors in each genre. A foreign-born director is more likely to direct a film in Action, Thriller, Horror, and Adventure, and less likely to direct in Musical, Comedy, Documentary, and Western. The genre films with higher foreign-born director concentration are also the ones exported more to the international market encountering less cultural barriers. Unlike American comedy films which may require a foreign audience to understand the American sense of humor and catch it from the subtitles, action and thriller films are less demanding for the international audiences to follow. In contrast, the genres with lower foreign-born director concentration happened to be the films that are more culturally relevant. The films more often cater to the domestic preference and bear a higher cultural barriers when exporting to the international market. Foreign-born directors with birthplace origin of Canada show the closest distribution over genres to the one of the native-born directors.

Figure 2.1: Fraction of films by Foreign-born Directors



Notes: The figure shows the fraction of films directed by foreign-born directors in each genre. A foreign-born director is more likely to direct films in Action, Thriller/Suspense, Horror, Adventure, and Drama with larger than a half probability, while less likely to direct Musical, Comedy, Concert/Performance, Black Comedy, Documentary, and Western films. The genre categorization shown above is based on the default provided by OD.

2.3 Empirical Strategy

2.3.1 Fixed Effect Analysis

I first use the following standard fixed effect specification to estimate the value added by a foreign-born director to the film outcomes.

$$Y_{ijgtr} = \alpha I_i + \beta_1 X_{ijgtr} + R_{jt} + R_{gt} + \lambda_r + \lambda_m + \epsilon_{ijgtr}, \tag{2.1}$$

where Y_{ijgtr} is the outcome of a film by director i, distributor j, of genre g, released at time t, and with MPAA rating r. The variable of interest is I_i, which is an indicator for the film director being foreign-born. X_{ijgtr} is the film level characteristics controls, and it includes estimated production budget and the number of star cast members for each individual film. R_{jt} is a vector of a distributor by time fixed effect, which is to absorb both individual distributor and film release year specific distributor shocks. Similarly, I include R_{gt} to control for individual genre and release period specific genre fixed effect. λ_r controls for MPAA film rating fixed effect, which is to address the systematic differences in film outcomes due to the potential market demand for the films based on their ratings.[14] λ_m controls for the film release month effect to address for the seasonality in film outcomes. Lastly, ϵ_{ijgtr} is the residual.

The coefficient α captures the effect of a foreign-born director on the film outcome. Two main issues arise in the identification. One is to disentangle the director's effect from the distributor's effect. Film outcomes, especially the commercial outcomes, depend not only on the director capability, but also on the distributor influence. A major distributor maintains branch offices in critical regional markets both in domestic and international

[14]The MPAA rating determines the potential market size for films, and G- and PG-rated films have on average better performance in terms of film revenue than R-rated films (Ravid, 1999).

domains and provides sufficient publicity for films with massive marketing campaigns. This helps bring in the significant box office at least in the opening weekend (Scott, 2005), which distinguishes with the influence of the independent distributors on films. The key to tease apart the director (foreign-born director) and distributor effects is that filmmaking is mostly project-based and a sample of directors (foreign-born directors) having multiple films released by different distributors. These directors provide information to evaluate the individual distributor's effect on film outcomes. Suppose a director has two films, one is distributed by the Universal, and the other is distributed by Paramount Pictures. The difference between the two film outcomes can be attributed to the individual distributor's influence with all else equal. Second issue with the identification is that director capability can be endogenous to film quality prior to the director assignment. In other words, a director with higher capability may be selected to potentially better film projects and manage higher production budget. Foreign-born directors are usually positively selected to Hollywood, meaning that Hollywood is likely to hire foreign-born directors with proved capability, then the directors would be associated with potentially higher quality films. To address this issue, I control for the estimated production budget, and thus the identified effect of a foreign-born director is the value added by a director that is additional to the possibly endogenous quality of a film prior to director assignment.

The time-specific distributor fixed effect in equation (2.1) nests the linear additive control of distributor and time fixed effects and addresses additionally the time-specific shock to a certain distributor which may affect a director's film outcome differentially based on their nationality. For example, the effect of a distributor tapping into a new international market would be more of a positive shock to a foreign-born director if the new international market has a home country bias toward the foreign-born director's film. Such unbalanced shocks will be absorbed by the time-varying distributor fixed effect. Similarly, time-specific genre effect would absorb shocks that may influence a

director's performance differentially by their nationality.

The identification as specified helps to understand if there exists systematic differences in the film outcomes induced by the director nationality. A key question remains the extent to which the effects reflect unobserved differences in the market preferences for films or the beliefs about the capability of a foreign-born director. One potential concern is that a foreign-born director is selected into films that are more likely to appeal to the international market and thus receive a higher overall revenue as the international market makes an increasingly important part to the total revenue. I address this concern with twin film analysis and instrumental variable identification. Another concern could be that a foreign-born director is selected into films to appeal to the director's home country market. I address this issue and further test for the possible mechanisms of the foreign-born director effect in Section 2.4 with multiple sub-sampling exercises.

2.3.2 Findings

Oscars

Oscar awards represent the peer recognition of the film's cinematic value in a given year. I look at the number of nominations, awards, likelihood of receiving *Best Director* and *Best Picture* award to quantify cinematic achievement. Column (1) in Table 2.3 shows the effect of a foreign-born director on the likelihood of a film being nominated for an Oscar. With controls for the film attributes, a foreign-born director results in a film being 1.87% more likely to be nominated. Column (2) shows that a foreign-born director's film gets 0.58 more nominations conditional on being nominated, and the number receiving actual awards is 0.46 higher. The odds ratio of receiving an award conditional on the nomination is not significantly different within the sample. Contrary to the anecdotal observation of foreign-born director receiving more Best Director awards

or their films receiving more Best Picture, I do not observe the systematic differences in the US-produced films.

Box Office

In this section, I evaluate a foreign-born director's effect on box office revenues. Table 2.4 firstly shows that the production budget is positively correlated with the revenues; 1% increase in the budget is associated with 0.76% increase in the gross revenue as column (1) shows, 0.60% increase in the domestic revenue (column (2)), and a 1.02% in the international revenue (column (3)). The production budget shows a more positive correlation with the international box office than with the domestic box office. This agrees with the evidence that Hollywood films are better known for their higher production budget in the international market and the corresponding production quality than the competitors' films. It also suggests that the domestic and international consumers' preferences for American films do not perfectly align. More importantly, the table shows that the effect of a foreign-born director is not significant in the gross revenue (column (1)), however, a foreign-born director results in a lower domestic box office (column (2)) and a higher international box office (column (3)) than a native-born director counterpart, both as a value-added to the films controlling for the potential film quality measured by the production budget.

In Tables A5-A7, I use linearly additive fixed effects to evaluate the relative importance of the controls and the stability of the estimates. The findings agree with the baseline findings in Table 2.4, and the estimates are consistent with different set of the fixed effect controls. In addition, the table confirms that the revenues are heterogeneous across the distributors, and individual distributor fixed effect makes the most prominent part in distinguishing the revenues among the fixed effect controls according to the changes in the adjusted R^2. The inclusion of individual distributor effect also attenuates the coefficients of the production budget without altering the director's effect. This suggests a substantial influence of the individual distributors on the film revenues, and that

Table 2.3: Oscars

Dependent Variables:	Likelihood of Nomination	Number of Noms	Number of Awards	Odds of Awards	Best Director	Best Picture
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Foreign-born	0.0187*	0.5793*	0.4620**	-0.0021	0.0144	0.0008
	(0.0081)	(0.2273)	(0.1301)	(0.0284)	(0.0214)	(0.0155)
Log Budget	0.0308***	0.8213*	0.4285**	-0.0018	0.0678	-0.0123
	(0.0053)	(0.3535)	(0.1394)	(0.0083)	(0.0418)	(0.0239)
Fixed-effects						
Distributor	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Genre	Yes	Yes	Yes	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Adjusted R^2	0.11008	0.14020	0.14713	0.01370	0.10939	0.09527
Observations	3,452	336	322	320	336	337

Notes: This table analyzes the effect of a foreign-born director on the film quality measured by Oscars. Column (1) shows the likelihood of a film getting nominated explained by the director being foreign-born and the production budget after controlling for the set of fixed effects defined in equation **??** within the overall sample. A film directed by a foreign-born director is 1.87% more likely to be nominated. Columns (2)-(6) analyze the outcome variables within the sample of Oscar nominated films. Column (2) explains the number of nominations a film gets by the director being foreign-born and production budget given the film is being nominated with at least one Oscar. It shows that a foreign-born director's film gets 0.58 more nominations than a native-born director's film. In column (3), the dependent variable is the number of awards a film gets given the film is being nominated, and the coefficient shows that a foreign-born director results in 0.46 more award. However, the odds of awards given nomination does not vary according to the director being foreign-born (column (4)). In column (5), I regress an indicator for the film recognized with the director receiving a Best Director award on the director's foreign-born background and production budget. In column (6), I regress an indicator for the film receiving a Best Picture award on the director being foreign-born and production budget. Contrary to the anecdotal observations, a director being foreign-born does not affect the probability of the director getting a Best Director award or the film gets a Best Picture award. Columns (1)-(6) show that the production budget as a proxy for film quality is positively correlated with a film getting nominated (column (1)), getting more nominations (column (2)) and awards (column (3)). However, the production budget does not get a film higher odds of win, or the likelihood for the Best Director or Best Picture award. All the standard errors are clustered by Distributor, (Release) Year, Genre, MPAA rating, and (Release) Month. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

the production budget level partly reflects the distributor capability a film gets associated with.

Table 2.4: Box Office Baseline

Dependent Variables:	Log Worldwide	Log Domestic	Log International
Model:	(1)	(2)	(3)
Variables			
Foreign-born	0.0580	-0.1639*	0.2994**
	(0.0607)	(0.0763)	(0.0853)
Log Budget	0.7634***	0.6014***	1.020***
	(0.0656)	(0.0438)	(0.1234)
Fixed-effects			
Distributor-Year	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes
Month	Yes	Yes	Yes
Fit statistics			
Adjusted R^2	0.72706	0.68115	0.62053
Observations	3,443	3,405	3,001

Notes: This table analyzes the effect of a foreign-born director on the film outcome measured by the worldwide, domestic, and international box office. Column (1) shows that a foreign-born director does not result in a difference in the gross revenue controlling for the production budget; column (2) shows that a foreign-born director results in a 16.4% lower domestic box office; and, column (3) shows that a foreign-born director results in a 29.9% higher international box office. The production budget is positively correlated with the revenues; 1% increase in the budget is associated with 0.76% increase in the gross revenue (column (1)), 0.60% increase in the domestic revenue (column (2)), and a 1.02% in the international revenue (column (3)). A higher production budget is more positively correlated with a higher box office revenue in the international market than in the domestic market. It agrees with the evidence that Hollywood films are better known for their higher production budget and the corresponding production quality in the international market than the competitors' films. It also suggests a disparity in the domestic and international consumers' preferences for American films. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAA Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

2.3.3 Validity with Twin Films

In this section, I exploit film plot summaries to construct quasi-experimental film groups to evaluate the effect of a foreign-born director. Fixed effect specification in the previous section assumes that the selection of a foreign-born director is random given the fixed film characteristics such as distributor, genre and release time period, and especially given the pre-determined estimated production budget. To verify the robustness of the specification and address the concerns for omitted variable biases, I construct "twin" film groups to evaluate the effect of a foreign-born director based on the sample of films with more homogeneous characteristics.

Twin films are the films with similar plots.[15] Films dealing with the same topical issues may end up forming twin pairs, for example *United 93* and *Flight 93* both chronicle the events aboard *United Airlines Flight 93* that leading up to 9/11 event. Twin films are usually made as production companies tend to invest in similar scripts to avoid risk-taking. They provide a context to study the foreign-born director effect with less endogeneity issue of director assignment. I cannot find full records of twin films in any database, except for notable pairs provided by film critics.[16] To detect such "twins" among a big set of films, I apply machine learning tools to analyze film plot summaries. A film plot summary, also known as synopsis, introduces the main characters and the setting of a film with less than 700 words and accessible format; it provides condensed information of a film. I consider the film plot summary as a proxy for film plot and use plot summaries to construct "twin" films for the director effect analysis.

Specifically, I use a pre-trained BERT model to convert text data into vectors, then

[15]Twin films are different from mockbusters. The latter is created to exploit the publicity of the major film they "mock" with a similar title or subject. Mockbusters are often made with a significantly lower budget than the major films.

[16]Twin Films (Wikipedia) provides an integrated list of twin films based on several film critics' judgement.

use cosine similarity metric to measure the distance of the vectorized plot summaries, and lastly, I use K-means–an unsupervised machine learning method–to cluster films.[17] Details about the construction of the plot similarity metric and the clustering of the films based on the metric are included in Section A.2. I then use the constructed twin film sample to run the following two-stage analysis. In the first stage, I estimate:

$$Y_{ijgtr} = \beta X_{ijgtr} + R_{jt} + R_{gt} + \lambda_r + \lambda_m + \epsilon_{ijgtr}. \tag{2.2}$$

Then in the second stage, I evaluate the following equation with the residuals from the first stage:

$$Y_{ijgtr} - \hat{Y}_{ijgtr} = \alpha I_i + \mu_{ijgtr}, \text{ for } i \in C, \tag{2.3}$$

where C is a cluster of films.

Findings with Twin Films

Table 2.5 analyzes the effect of a foreign-born director using the sample of twin films based on equations (2.2)-(2.3). Columns (1) and (3) are the baseline results as in Table 2.4, and columns (2) and (4) are the findings based on twin films. The effects of a foreign-born director on both the domestic and international box office attenuate but the coefficients show consistency with the baseline results–a foreign-born director has an advantage in appealing films to the international market and a disadvantage in appealing to the domestic market. The attenuation may suggest that the plot summary is picking

[17]Bidirectional Encoder Representations from Transformers (BERT) is a transformer-based machine learning technique for natural language processing (NLP) pre-training developed by Google (Devlin et al., 2018). BERT outperforms other language models in interpreting contextual meanings of words and paragraphs and is proved to be efficient in many language processing tasks without substantial task-specific architecture modifications.

up additional heterogeneity of the films with director assignment, and more specifically their nationality.

Table 2.5: Film Plot Summary Clusters

Dependent Variables:	Log Domestic		Log International	
Model:	Baseline	Twin Films	Baseline	Twin Films
Variables				
Foreign-born	-0.1639*	-0.1002*	0.2994**	0.2012***
	(0.0763)	(0.0466)	(0.0853)	(0.0444)
Log Budget	0.6014***	0.5923***	1.020***	1.025***
	(0.0438)	(0.0476)	(0.1234)	(0.1109)
Fixed-effects				
Distributor-Year	Yes	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
Fit statistics				
Adjusted R^2	0.68115	0.51933	0.62053	0.58273
Observations	3,405	2,323	3,001	2,040

Notes: The table analyzes the effect of a foreign-born director within the matched film "twin" pairs based on the similarity in the plot summaries. Columns (1) and (3) show the baseline results and columns (2) and (4) show the results based on twin film sample. The estimates based on the twin films show consistency with the baseline results for both the domestic and international box office.

2.3.4 Validity with Instrument Variable for Foreign-born Director Assignment

To use fixed effect analysis to identify the foreign-born director effect, it requires the director assignment to be random, with the control of the fixed effect and production budget. fixed effect controls take care of the systematic differences in film outcomes due to the film characteristics, and production budget controls for the possible endogeneity in film expected outcome and director level. however, it is possible that the selection of foreign-born director is associated with unobserved factors including the film targeting different markets, film story lines are more likely to be popular in the overseas market. films may have different advantage to the markets if some films are co-produced with another country and have easier access to the overseas market. To account for such endogeneity, I conduct IV analysis using co-production as an instrumental variable for foreign-born director assignment.

Table 2.6: Coproduction

Variable	US-only	Co-produced	Overall
Worldwide Box Office			
Count	10553	1651	12204
Mean (SD)	48.17 (135.18)	43.28 (128.67)	47.51 (134.32)
Median (IQR)	2.79 (33.00)	1.71 (24.39)	2.51 (31.92)
Domestic Box Office			
Count	8434	1198	9632
Mean (SD)	30.44 (59.40)	23.19 (49.31)	29.53 (58.28)
Median (IQR)	7.93 (35.12)	3.49 (23.33)	7.20 (33.83)
International Box Office			
Count	5066	1104	6170
Mean (SD)	49.68 (117.88)	39.56 (105.84)	47.87 (115.87)
Median (IQR)	4.40 (45.26)	2.14 (28.36)	3.91 (42.50)
Production Budget			
Count	3866	576	4442
Mean (SD)	38.11 (45.70)	40.75 (44.92)	38.45 (45.60)
Median (IQR)	22.00 (42.00)	26.00 (44.00)	23.00 (41.50)

Table 2.6 shows the summary statistics for the films by their co-production status. Even though the co-produced films have a higher mean production budget level than the US-only produced films, a t-test result shows that the difference between the two is not statistically significant (t-stat is 1.2964, p-value is 0.1949). Similarly, the difference in the gross revenues from the two samples is insignificant as well (t-stat is -1.3769, p-value is 0.1686). However, the probability of having a foreign-born director is 35.49% higher for the co-produced films than for the US-only produced films.[18] I utilize this fact to construct an instrumental variable with the dummy *MultiCountry* indicating a film being produced by multiple countries (including the US), and also with the dummy *MultiLanguage* for film having lines in different languages to account for the possibility that a foreign-born director is selected for certain cultural ties. Based on the statistical evidence, I assume that the co-production affects the film outcomes only through the likelihood of the director being foreign-born. Then the identification based on IV specification should identify the effect of a foreign-born director. Specifically, in the first stage, I estimate the following equation:

$$I_{ijgtr} = \gamma_1 \text{MultiCountry} + \gamma_2 \text{MultiLanguage} + \gamma_3 X_{ijgtr} + R_{jt} + R_{gt} + \lambda_r + u_{ijgtr}. \quad (2.4)$$

Then in the second stage, I substitute the estimated $\hat{I}_i$ in equation 2.1 to estimate the effect of a foreign-born director on films.

Findings with IV Specification

Table 2.7 first shows the result with instrumenting the director being foreign-born director for the films' international box office. Tables 2.7 presents the results with the dummy for a film being produced by multiple countries, the dummy for a film having

[18]The fraction of films directed by a foreign-born director is 59.82% for the co-produced films, and the fraction is 24.33% for the US-only produced films.

lines in multiple languages, and the production budget as instrumental variables, also controlling for the same set of fixed effects. The first stage (column (2)) shows that a foreign-born director assignment is only positively correlated with *MultiCountry*. Specifically, the production budget is not correlated with the director being foreign-born, which suggests that foreign-born directors are not systematically selected into the films with potentially higher film quality. Column (3) shows the IV specification. Foreign-born directors result in 39.4% higher international box office than native-born directors, with the magnitude being larger than the baseline result. Table A8 shows qualitatively similar result with the number of production countries and languages as instrumental variables for foreign-born director assignment.

Table 2.7: International Box Office (IV)

Dependent Variables:		Foreign-born	Log International
IV stages		First	Second
Model:	Baseline	(2)	(3)
Variables			
Foreign-born	0.2994**		0.3938***
	(0.0853)		(0.0439)
Log Budget	1.020***	0.0301	1.017***
	(0.1234)	(0.0185)	(0.1320)
MultiCountry		0.3175***	
		(0.0369)	
MultiLanguage		-0.0785	
		(0.0532)	
Fixed-effects			
Distributor-Year	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes
Month	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes
Fit statistics			
Adjusted R^2	0.62053	0.05522	0.62022
Observations	3,001	3,001	3,001
F-test (IV only)		87.038	2.2989
Wald (IV only), p-value		5.25×10^{-42}	5.35×10^{-19}

Notes: This table analyzes the effect of foreign-born director using instrumental variable specification. A film is more likely to be directed by a foreign-born director when it is produced by multiple countries. Column (2) shows the first-stage estimation using the production budget, an indicator for the film produced by multiple countries, and an indicator for a film having lines in multiple languages as instrumental variables. In the second stage, the coefficient has an implication that for the films produced by multiple countries a foreign-born director yields 39.4% higher international box office than a native-born director. The magnitude is greater than the coefficient in the baseline specification (column (1)), whereas the coefficient for the production budget is consistent. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAAs Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Then Table A9 shows the director effect for the films' domestic box office revenue. With a similar setup, column (3) shows that a foreign-born director results in 121.6% lower domestic box office with a magnitude that is larger than the baseline result. Table A10 shows qualitatively similar result with the number of production countries and languages as instrumental variables for foreign-born director assignment.

The implications from the results are two-fold. First of all, the IV specification confirms the main findings that a foreign-born director's effect is mostly in generating positive revenue in the international market, while not in the domestic market. Secondly, it shows that even though co-productions are a solution to pool the financial and human resources, it may lead to a higher risk in producing relevant products to target different consumers than the US-only production–generating a larger disparity in the effect of a director between the domestic and international market.

2.4 Possible Mechanisms

In the previous sections, my findings suggest that a foreign-born director has a differing effect on the box office revenues in the domestic and international domain, and that a foreign-born director's films are recognized with cinematic value, which partly verifies the director capability. A foreign-born director may induce a positive effect on the film outcome for several reasons. Foreign-born directors are subject to positive selection, where the positive selection applies to the foreign-born directors who migrated early with their parents to the U.S. and also to the directors who established careers overseas and were recognized by Hollywood for their film directing capability. As Hirschman (2005) points out, foreign-born directors may be more successful in film directing due to their exposure to multiple cultures and bearing a higher level of cultural resources (cultural capital). The positive selection and cultural capital are embedded in the director's in-

trinsic capability and are difficult to pin down as precise mechanisms for film success. Nonetheless, either the selection or cultural capital does not fully address the differing effects of a foreign-born director on the domestic and international box office revenues. Even though it is difficult to pin down the precise mechanisms, I conduct sub-sampling exercises and robustness tests in the following sections to help interpret the findings.

Cultural Relevance and Director Effect

In this section, I test for the idea that a foreign-born director's effect is in mitigating translation cost of films to the international market and thus generating a higher revenue overseas. Films like many other products suffer from translation cost when exported to a foreign market. Studies show that more culturally relevant films (genres) may suffer more from such costs (Holloway, 2014, Miskell, 2014). I conduct a sub-sampling exercise with film genres heterogeneous in cultural relevancy and test for the effect of a foreign-born director in the domain of both domestic and international markets.

Table 2.8 shows that the effect of a foreign-born director in the international market is more prominent for the films (genres) that are culturally relevant–comedy, drama, than for the films that are not–action and adventure. In other words, a foreign-born director's comedy films are more popular than a native-born director's comedy films in the international market. Domestically, however, a foreign-born director's culturally relevant films are not as preferred as a native-born director's culturally relevant films, where the foreignness liability of American culture does not apply and the demand for translation is low. The findings are consistent with the general theory in trade, and suggest that the value of a foreign-born director is partially in mitigating the cost of translating culturally relevant American films to the international market.

Table 2.8: Cultural Relevance

	Irrelevant		Relevant	
Dependent Variables:	Log International	Log Domestic	Log International	Log Domestic
Model:	(1)	(2)	(3)	(4)
Variables				
Foreign-born	0.1129	-0.1245*	0.4673*	-0.3072**
	(0.0583)	(0.0561)	(0.1874)	(0.1101)
Log Budget	1.268***	0.7973***	0.9834***	0.5915***
	(0.0775)	(0.0742)	(0.0540)	(0.0431)
Fixed-effects				
Distributor-Year	Yes	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
Fit statistics				
Adjusted R^2	0.70072	0.65692	0.47890	0.61846
Observations	1,246	1,342	1,342	1,585

Notes: The first two columns evaluate the effect of foreign-born director on the relatively culturally irrelevant film genres–Action, Thriller/Suspense, and Adventure; and the last two columns evaluate the effect of foreign-born director on the culturally relevant film genres–Drama, Comedy, and Western. The cultural irrelevant film genres are also the ones that have higher foreign-born director concentration than the culturally relevant genres. Columns (1) and (3) suggest that a foreign-born director generates a higher margin in the international market for the films that are culturally relevant, i.e. a foreign-born director is better at mitigating the translation cost of the films that are culturally relevant; however, for the culturally relevant films, a foreign-born director generates further negative effect on domestic revenue than the culturally irrelevant films and also the overall sample as shown in Table 2.4.

Director Type

To support the argument that a foreign-born director's effect is in mitigating the translation cost, I check whether a foreign-born director has heterogeneous effect on films based on their American cultural understanding. I use the director's linguistic capability as a proxy for the cultural understanding, as studies show that the acquisition of linguistic capability and cultural understanding complement each other. There is no data documenting a director's linguistic capability, one can probably use as a proxy whether a director has received a degree in the US, the age at which a director migrated to the U.S., or whether a director is from an English-speaking country, etc. With the constraint of my current dataset, I use a director's second occupation to construct a proxy for the foreign-born director's linguistic capability. Specifically, a foreign-born director whose another proclaimed job title is *Actor* is assumed to have a higher linguistic capability and hence a better cultural understanding than the other foreign-born directors among many factors that distinguish such Actor-type director with other directors; and native-born directors are assumed to have relatively homogeneous linguistic capability.

I interact the director type with director nationality in investigating the variation of the film revenues. Table 2.9 firstly shows consistent results with the baseline specification (Table 2.4) that the effect of a foreign-born director is positive in the international market and negative in the domestic market, while no effect on the gross revenue. Column (1) exhibits that an actor-type of foreign-born director, who bears a higher level of linguistic skill, yields a higher gross outcome than the other foreign-born directors. However, a director being an actor-type does not contribute to a revenue gain for the native-born directors in either the domestic or the international market. This implies that it is the differed linguistic capability and associated cultural understanding that distinguishes a foreign-born director among the foreign-born directors, instead of other latent factors

that actor-type directors may carry. Comparison of the coefficients from columns (2) and (3) shows that the effect of having a higher linguistic capability shows up mostly in the international box office, which suggests that a higher linguistic skill is associated with the director's understanding of the American culture and lowering the translation cost of the films to the overseas. Table A15 shows similar analyses with a more detailed director-type categorization, and the results are consistent. Translating cultural products means is a rather abstract way of saying the director reinterpreting the American culture with new cultural context and elements. The reinterpretation may resonate with a wider audience as it may provide more diverse cultural references or deliver a more universal message to the audience.

Table 2.9: Director Type

Dependent Variables:	Log Worldwide	Log Domestic	Log International
Model:	(1)	(2)	(3)
Variables			
Foreign-born	0.0338	-0.1768*	0.2591**
	(0.0621)	(0.0759)	(0.0830)
ActorType	0.0784	0.0890	0.1974
	(0.0943)	(0.0796)	(0.1190)
Foreign-born × ActorType	0.3312*	0.2122	0.4101*
	(0.1525)	(0.1859)	(0.1651)
Log Budget	0.7617***	0.6009***	1.017***
	(0.0654)	(0.0450)	(0.1227)
Fixed-effects			
Distributor-Year	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes
Month	Yes	Yes	Yes
Fit statistics			
Adjusted R^2	0.72678	0.67972	0.62255
Observations	3,483	3,444	3,039

Notes: Column (1) exhibits that an actor-type of foreign-born director, however, yields a higher gross outcome than the other foreign-born directors, comparing the results from column (2) and (3) but the indicator of being an actor-type does not contribute to a revenue gain for the native-born directors in either the domestic or the international market. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAA Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Home Bias

In this section, I test whether a foreign-born director's capability in mitigating the translation cost only works for the director's home country or if it helps deliver the film more universally to the overseas market. In other words, I test for whether foreign-born directors' home country bias is the sole driver for their films to have a systematically higher box office in the international market. To conduct such analysis, I leave out the revenue from the director home country and estimate the foreign-born director effect in the rest of the international market. I also subtract the box office recorded with spurious categorization as "Rest of World" in the data, which may bias the result if the revenue of a director's home country is included in that category.

With the same set of controls as in the baseline results, Table 2.10 shows that the effect of foreign-born decreases from 29.9% to 19.6% after excluding the directors' home country revenue, and the effect is significant at 10% level (column (2)). This exercise shows that about 34.4% ((29.9-19.6)/29.9) of the effect of foreign-born directors on the international revenue could be due to home bias. However, home bias does not solely explain the effect of foreign-born directors on international revenues.

Market Power

Now I test for whether the foreign-born director effect is only limited to several markets which have substantial impact on the revenues. For example, China is one of the largest film markets globally, where Hollywood has become increasingly relying on to make profits. One might be interested in seeing whether the effect of foreign-born directors in the international market is mostly due to the films catering to the big markets like China. *The Great Wall (2016)* by Chinese-born director Yimou Zhang starring Matt Damon and with the production budget of $150 million grossed only $45 million in the

Table 2.10: Director Home bias

Dependent Variables:	Baseline	Excl. Director Origin & *ROW*
Model:	(1)	(2)
Variables		
Foreign-born	0.2994**	0.1964*
	(0.0853)	(0.0874)
ln(*Budget*)	1.020***	0.9878***
	(0.1234)	(0.1005)
Fixed-effects		
Distributor-Year	Yes	Yes
Genre-Year	Yes	Yes
MPAA Rating	Yes	Yes
Month	Yes	Yes
Fit statistics		
Adjusted R^2	0.62099	0.66718
Observations	3,039	1,778

Notes: This table analyzes whether the effect of a foreign-born director on the international box office is solely driven by the director's home country bias. A director's home country is where the director was born. A home country bias exists when the director's films are disproportionately favored by the consumers in the director's home country than the other directors' films. Column (1) shows the baseline result with the total international box office as dependent variable. In column (2), the dependent variable is the international box office excluding the box office generated in a director's home country for the associated film and the box office recorded with spurious categorization as "Rest of World", which may bias the result if the revenue of a director's home country is included in that category but not explicitly specified. This exclusion does not affect the international box office for the films directed by native-born directors. The result indicates that a foreign-born director still generate positive effect on the film's international box office, even though the magnitude of the coefficient gets smaller after excluding the home country revenue, about 34.4% ((29.9-19.6)/29.9) of the effect of foreign-born directors on the international revenue could be due to home bias. In other words, the advantage of a foreign-born director in the international market is not solely a consequence of the director's home country bias. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAA Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

US, while $289 million internationally, and about 60% ($171 million) of the international revenue came from China. If by excluding the revenue from China the effect of foreign-born directors diminishes, then it implies that the effect is only local catering to certain markets rather than being universal across the different countries, and the capability of foreign-born directors is realized only through the connection to the big markets. With similar leave-out exercises, I have the following findings.

Table A11 shows the comparison between baseline result (column (1)) and the ones by excluding the revenue from China (column (2)), further excluding the revenue from its submarkets–Hong Kong and Taiwan (column (3)), and subsampling by excluding the films directed by Chinese-born directors (column (4)). In column (5), I again subtract the revenues from the spurious category "Rest of World". The effect of foreign-born directors persists throughout the exercises. It suggests that the effect of foreign-born directors in the international market is not solely generated by the directors' ties to China and its associated market power.[19]

Cultural Distance

In this section, I test the foreign-born director's effect using the sample of films exported to the countries that have a closer cultural distance to the U.S, and this helps to partly interpret the foreign-born director's effect in the domestic market. To conduct the analysis, I subsample the films exported to at least one of the English speaking countries.[20] The subsample reflects a selection of films that are exported to the countries that have a closer linguistic and cultural distance to the U.S. than the other countries; and it also reflects a selection of films from the right tail of the box office distribution, as those

[19]I conducted the exercises with other countries, including the UK, Australia, South Korea, and Japan, and the results are qualitatively the same.

[20]English-speaking countries considered in this analysis are: Australia, Ireland, New Zealand, South Africa, and United Kingdom.

films are likely to have a more complete records of international box office breakdowns at the country level.

Table A12 shows the effect of a foreign-born director for the films exported to at least one of the English-speaking countries in comparison to the overall sample. Firstly, comparing columns (1) and (3), I find that the foreign-born directors' effect in the domestic market is not negative, though it is not significantly positive either. The dependent variable in column (2) is the sum of the box office generated within those English-speaking countries. Comparing the coefficient from column (2) to the one in column (4), it shows that the magnitude of the foreign-born directors' effect on the international box office is larger for the top grossing films. Comparison of columns (2) and (3) shows that, even comparing with the other English-speaking countries, the U.S. consumers present an unaligned taste for films directed by a foreign-born director.

In addition to the exercises above, I also test for the star effect on films that may confound with the director effect on films, as the a foreign-born director's films show with a higher probability of associating with a star actor (Table A13). The findings with the star actor controls are shown in A14, and even though the star actor effect attenuate the correlation of production budget with the box office outcomes, it does not attenuate the effect of a foreign-born director, in other words, a foreign-born director effect is robust to the star actor control, and the findings also suggest that the production budget as an estimate for film quality picks up many latent factors that may affect a film outcome including the star effect.

In summary, I find that one of the mechanisms for a foreign-born director to generate differing effect on the American films is in mitigating the translation cost of the films to the international market. The translator effect is more prominent for the films that are considerably carrying a higher level of American culture (culturally relevant) than the culturally irrelevant films like the Action or Adventure genre films to be exported

to a foreign market. The translation capability also reconciles the disparity in the main findings that a foreign-born director results in a higher international but a lower domestic box office revenue, as the demands for cultural translation differ domestically and internationally.

2.5 Dynamics of the Absolute and Relative Advantages

In this section, I evaluate the long-term dynamics of the effect of a foreign-born director. As mentioned, Hollywood's development intertwines with the history of immigration. As the composition of the immigrants changes over time, and more relatedly, the composition of the foreign-born directors active in Hollywood changes over time, the effect of a foreign-born director may also differ. I aggregate the effect of a foreign-born director in decades to document the evolution of the absolute and relative advantages in the domestic and international box office revenues. Specifically, I estimate the following equation:

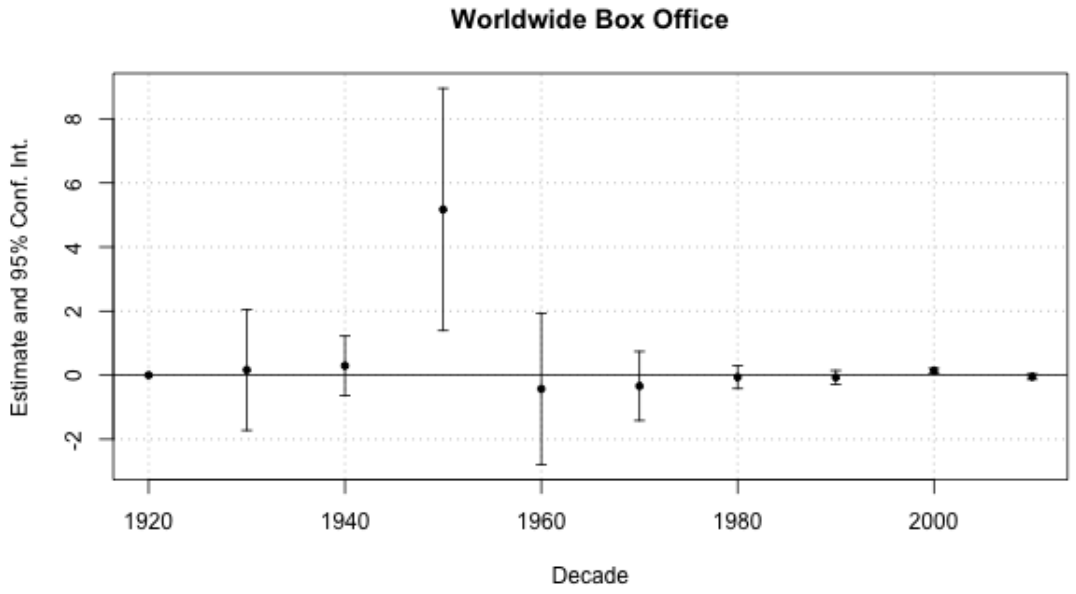
$$\mathrm{Y}_i = \alpha_t I_i + \beta X_i + \Phi + \epsilon_i, \tag{2.5}$$

where Y_i is individual film's outcome variable, X_i is the film's characteristics variable, Φ is a set of linearly additive fixed effect controls as specified in equation, and α_t is the coefficient of interest for director's foreign-born effect aggregated for each decade t. The evolution of α_t helps to understand the effect of foreign-born directors changing over time for the film outcomes.

Findings for the Dynamics

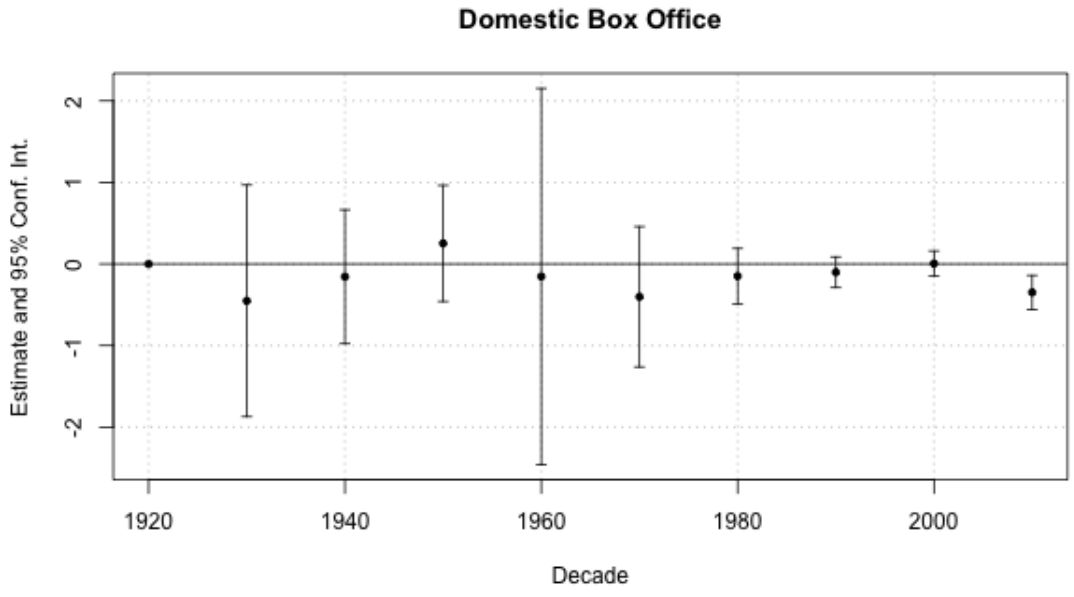
Figures 2.2-2.4 show the estimate of foreign-born director's effect on film revenues in each decade, α_t. The effects in the early decades are spurious with large standard deviation due to lack of effective observations, while the estimates show an improved precision for the later decades since the 1970s. Figure 2.2 shows that a foreign-born director does not present an absolute advantage in terms of the gross revenues over the decades, with the effects mostly centered around zero. Figure 2.3 shows that a foreign-born director's effect is most of the time below zero, which indicates that a foreign-born director has a relative disadvantage domestically in the film revenues. Figure 2.4 shows that a foreign-born director has positive effects over the decades in terms of international revenue, and their relative advantage in the market has been highest during the 1970s and 1980s. The period corresponds to the New Hollywood movement led by many creative and independent filmmakers experimenting with radical perspectives in filmmaking. It was also a period of Hollywood influenced by world cinema where similar new movements were underway. The effect of a foreign-born director may have been magnified as the innovative minds meet the capital support from Hollywood. The disparity has then been decreasing over time; the performance of the two director groups converges in both markets. The convergence could be a reflection of cultural globalization, and also could be a result of blockbuster films taken as a formula for commercial success by Hollywood.

Figure 2.2: Absolute Advantage



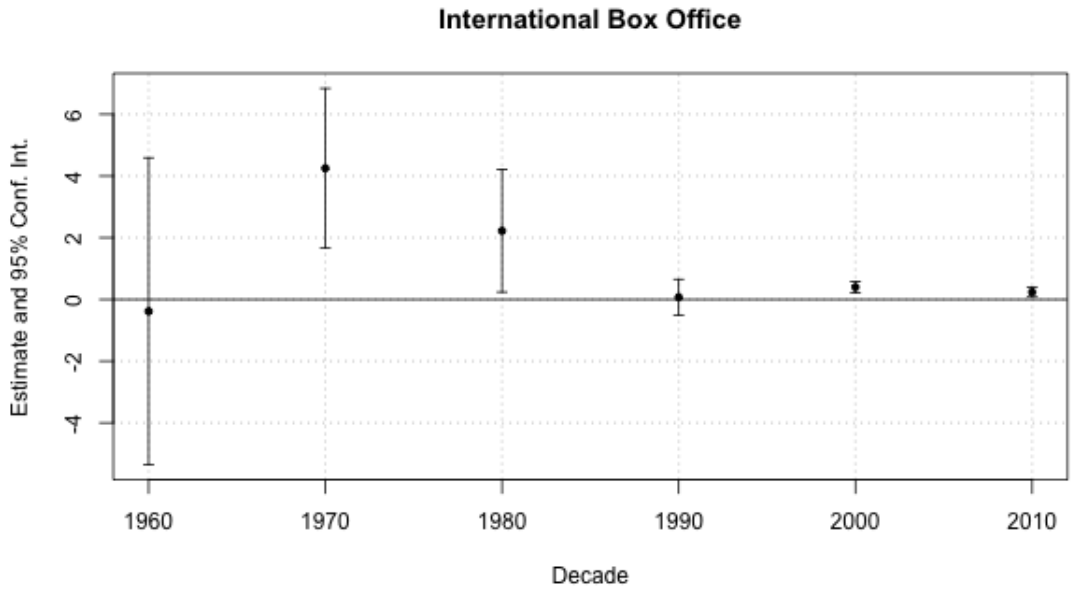
Notes: The figure shows the changes in the absolute advantage of foreign-born directors in filmmaking across the decades. The estimation results for the early decades are spurious with large confidence intervals as the box office information is limited. The box office records become more available for the films released after the 1970s, and the figure reports that foreign-born directors do not yield a significantly higher or lower gross box office than the native-born directors with the confidence intervals mostly containing zero.

Figure 2.3: Relative Advantage in the Domestic Market



Notes: The figure shows the changes in the relative advantage of foreign-born directors in the domestic market across the decades. Data scarcity on the box office information for the early day's films makes the estimation spurious. The estimation of the coefficient for the 1970s and onward shows that a foreign-born director yields lower revenue than a native-born director in the domestic market. The figure suggests a trend that the negativity of the effect of foreign-born directors has been decreasing over time except in the 2010s.

Figure 2.4: Relative Advantage in the International Market



Notes: The figure shows the changes in the relative advantage of foreign-born directors in the international market across the decades. International box office information is scarce for the early day's films, and hence the identification is mostly based on the released after the 1960s. It shows with more observations that a foreign-born director does yield a higher international box office almost in every decade except for the 1990s, and the foreign-born effect was largest in the 1970s and 1980s when the cinema was expanding with diverse impacts from the world cinema.

2.6 Conclusion

This paper studies the effect of foreign-born directors on film outcomes. Hirschman (2005) (H2005) opens up the discussion on the role of immigrants and foreign-born talents in shaping of the American culture with multiple intriguing observations. H2005 suggests to look at the role of immigrants and foreign-born talents beyond the scope of labor market, recapping the famous quote by Oscar Handlin that "Once I thought to write a history of immigrants in America. Then I discovered the immigrants were American history." Inspired by H2005, my study focuses on the role of foreign-born directors in making of the American films, and it provides a parsimonious way of looking at the foreign-born talents' role in making of the quintessential cultural products in the U.S.

With the impossibility of running a natural experiment with director assignment, I utilize a unique director-distributor matched dataset and construct film comparison groups based on the fixed effect controls and film production budget. Assuming the effects are linearly additive, I find that film distributor explains the variation in film outcomes the most, and film outcomes are also heterogeneous across genres, film ratings, and release time period. The findings indicate that foreign-born directors result in a higher international revenue and a lower domestic revenue, overall they provide no difference in the gross (worldwide) revenue of the films. I also integrate a natural language processing tool for film summary plots to construct pseudo-experimental twin film groups based on the similarity in film plots to control for the film heterogeneity. The findings confirm the foreign-born director's relative advantage in the international market and disadvantage in the domestic market.

The importance of foreign-born directors in film outcomes arises as Hollywood expands in the international market. Foreign-born directors, due to their bearing of a higher level of cultural resources, translate the American culture into a more universal language

to the international audience. As studies show that audiences prefer films with familiarity, foreign-born directors help increase the receptivity of the films by the international audience, and it explains why their films are perceived better in the international market. With multiple subsampling analyses, I address the concerns for director assignment endogeneity, and I exclude the possibility that foreign-born directors' effect is solely a result of the directors' home country bias or the related market power effect. This suggests that the foreign-born directors probably have a higher intrinsic capability regardless of their country of origins, due to their exposure to different cultures and positive selection to the film industry in the U.S.

Chapter 3

Gender Gaps in Productivity and Labor Market Opportunities: The Celluloid Ceiling and Film Directors' Career Trajectories

3.1 Introduction

Since Kathryn Bigelow became the first woman to win the Academy Award for the Best Director in 2009, only two more women joined her rank as the Best Director award recipients - Chloe Zhao in 2020 and Jane Campion in 2021. During its 94-year history celebrating films screened in the US, the Academy of Motion Picture Arts and Sciences crowned the Best Director three times to women and ninety-two times to men (multiple awards were given in the first year). The status of female directors in the film industry is similar elsewhere. At Cannes, women won Prix de la mise en scene (the award given to the best director at the festival) only twice (Yuliya Solntseva in 1961 and Sofia Coppola in

2017) in the 74-year history of the prestigious film festival. The lack of women receiving these awards is unsurprising when we consider the gender disparity in the number of directors. According to the Center for the Study of Women in Television Film's Celluloid Ceiling project, between 7 and 11 percent of 250 top-grossing films were directed by women during the mid-2010s before increasing to reach only 17 percent by 2021 (Lauzen, 2022).

We investigate the underrepresentation of women in film directing by providing the first estimates of the gender gaps in the labor market outcomes for film directors. There is a myriad of reasons why one might expect women to do worse than men in the film industry. Vanity Fair's Joy Press interviewed female directors who recounted the prejudice and discrimination they faced in the industry during the 1970s. During this period, studio executives barred women from the directing role, suggesting that there is an inherently higher risk of hiring women as a director, with one executive quoted as saying that "guys might think 'she didn't shoot well that day because she had her period'" (Press, 2021). The hostile climate for female directors may have pressured women into early retirement or occupation changes that led to fewer women directing multiple films, if they had a chance to direct a film at all. Alternatively, it could be that the lack of opportunities for women led to poorer quality in films directed by women, which may have affected the decision to hire a female director. Lastly, women may leave the labor market earlier than men due to the gender inequality in child rearing and housework, as Bertrand et al. (2015) suggest.

Understanding the gender disparity in film directing is of interest to economists for several reasons. The film industry provides a setting in which labor market discrimination is well-documented via numerous accounts from women. As noted above, female directors are woefully underrepresented in the industry, and their achievements often go unrecognized. Hence, examining how these gaps emerge is an important contribution to

the industry and to the economic literature on gender gaps. The paper is also relevant for other fields in which workers frequently switch jobs or most workers are freelancers. In today's labor market climate, where the job market is growing increasingly fluid, there is a growing importance to document gender gaps that are unique to this type of market. Similarly, the labor markets in the film industries are sometimes reputation- and relationship-based, which makes it easier for personal biases to influence the hiring decision (Doyle, 2012). At the same time, the industry's output, the films, is often widely available, and its quality is thoroughly scrutinized and critiqued by film experts and the public. The worker's labor market outcomes, such as the probability of directing another film or the budget given for the next film, are easy to observe as well. The combination of these factors makes the film industry a unique setting in which it is relatively easy to test economic hypotheses on the returns to productivity, gender-based biases and discrimination, and their interactions in the labor market.

We contribute to the literature by exploring the gender gaps in a labor market where workers are largely freelance, and the incentive to reach the upper end of the job market ladder is very high, similar to many other industries with an increasingly sizeable number of outsourced workers. The industry's hiring practice is potentially heavily dependent on their previous performance, yet the hiring process is rarely transparent. The combination of these factors creates a competitive and discrimination-prone environment, in which both the director's gender and their ability and productivity can influence the director's career success. To further investigate these factors, we focus on the gender gaps in the probability of directing another film. Specifically, we combine two datasets: detailed information about movies released to the public from OpusData and a collection of critical responses and audience ratings from RottenTomatoes. Then, we measure the impact of the director's gender, the quality of their previous film, and the interaction between these two factors on the probability of directing another film. The estimates

obtained from this regression inform us how the opportunity to continue one's career in the film industry differs between women and men, and whether the gender gap differs between directors with positive and negative track records. In other words, we estimate the gender gap in labor market opportunities and whether the returns to productivity, as measured by the quality of previous films, differs by gender.

Our primary finding is that men are more likely than women to direct another film when the previous film's critical responses and box office returns are not favorable, but the gender disparity largely disappears among directors with favorable critical responses and box office returns. This gap can also be observed when we compare the director's tenure when their last film is released. Only 6 percent of women in our sample had career spans longer than 10 years, compared to 12 percent of men. Similarly, the time elapsed between two subsequently directed films is longer in women than in men, suggesting that men tend to get opportunities to direct another film more frequently compared to women. The gap in the opportunity to direct another film could be interpreted in two ways. It could be that male directors are more persistent than female directors. However, it could also be that the studio executives give men more room for failure. To further explore this question, we examine whether the studio allows directors different degrees of independence in filmmaking, holding their productivity measures constant for the previous film. Women in our sample are less likely to hold multiple roles in a film than men. The implication is that women are more restricted in their role as a director, whereas men are able to influence the film from multiple roles. Women also receive lower budgets than comparable men. Hence, the studios are more hesitant to give female directors creative freedom and financial support compared to male directors with similar track records. Lastly, we are interested in whether there is a self-sorting behavior among women in the industry. According to Joy Press's Vanity Fair article, female directors, who became jaded from the industry's treatment of women, sorted into

directing documentaries, where the studio executives are less likely to infringe on the creative freedom of the director and put less weight on box office numbers in re-hiring decisions (Press, 2021). We see this sorting behavior in our data. Women are heavily overrepresented in documentaries relative to other genres, especially the male-oriented genres like action.

Previous papers in the literature studying the gender gaps in the film industry have uncovered the gender wage gaps and opportunity gaps among film and television writers. Among television writers, women's lifetime earnings were between 11 and 25 percent lower than similarly qualified men (Bielby and Bielby, 1992). Similarly, female writers in the film industry were subject to a 21 to 25 percent penalty in earnings compared to male writers, and the gap grew from 4 to 6 percent early career to 40 percent by 15 years after the first writing credit (Bielby and Bielby, 1996). Sociologists explain that workplace discrimination is more severe in industries in which productivity is evaluated arbitrarily or subjectively, especially when hiring managers rely on whether someone is a "good fit" with the pre-existing mostly-male workforce (Bielby and Bielby, 2002). In filmmaking, directors often have more authority and independence compared to writers, and they are more often in the spotlight. These differences could potentially lead to larger gender gaps in directing compared to writing, as pushing against the existing biases becomes riskier as workers move up the rank. Hence, we expect directors' experience of the gender gap differs from writers' experience.

Others have focused on the #MeToo movement, which originated in Hollywood from the Harvey Weinstein scandal. Luo and Zhang (2021, 2022) document these changes using the sample of writers, directors, and actors working on films released between 2014 and 2019. The authors find that Weinstein-associated producers became more likely to work with female writers and female directors and became less likely to portray women in a traditionally feminine manner (Luo and Zhang, 2021). In addition, Weinstein-associated

producers became more likely to allow female writers to work on male-oriented genres (Luo and Zhang, 2022). Hence, the economic and sociological literature focusing on large samples of film industry workers suggests that there are substantial gender disparities, and changing social climate and attitudes have somewhat reduced some of these gaps. However, it is still not well known whether these findings extend to the gender gaps in labor market opportunities for directors where we contribute.

In Economics, researchers have extensively studied how and why the gender wage gaps arise. We discuss some of the most notable and relevant examples here. One strand of the literature focuses on gender-based sorting. Dohmen and Falk (2011) use an experimental setting to find that women less often choose risky options for their payment scheme compared to similarly skilled men. Buser et al. (2014) find that gender gaps in the affinity to competition explain a substantial portion of the difference in students' choice of academic track, where girls were less likely to choose prestigious academic track compared to boys. More recently, Sarsons et al. (2021) find that female economists receive less credit than men when coauthoring a paper, which results in a lower likelihood of receiving tenure. Taken together, women tend to avoid risks due to either some innate differences in risk preferences or to avoid situations in which the evaluation of their work or their contribution to projects is swayed by subjective bias. Hence, one could hypothesize that women's risk preferences and experiences of discrimination encourage them to choose a particular line of work with lower risks in the case of failure and less systemic gender-based biases from the decision-makers of the industry. Our finding that female directors tend to sort themselves into specific genres, such as documentaries, where the perceived risk is smaller due to the lower budget and creative freedom from studio executives, confirms the hypothesis that women sort into specific fields to avoid discrimination.

The existing work that is closest to our article is John et al. (2017). The authors

follow the career paths of film directors from 1987 to 2006 and show that directors with more films under their belt are more likely to receive another opportunity (John et al., 2017). They also show that the probability of directing another film is highly dependent on the average ratings of previous films, and the number of films that directors worked on during their career is a good indicator of ability (John et al., 2017). The article, however, does not explore gender differences in returns to previous performances. We expand on their work by focusing on how the gender disparities in the workers' labor market opportunities emerge in a highly competitive, high turnover industry.

3.2 Data and Empirical Analysis

Our article's primary purpose is to accurately capture the differences in labor market opportunities between female and male directors over their careers. While descriptive results can capture the existing gender gaps and labor market disparities clearly, it is important to assess whether those gender gaps arise due to the gender differences in productivity or ability as far as they could be observed at the time when the relevant parties decide on whether to hire a director for another film. Hence, we need detailed longitudinal data to capture the quality of the director's previous works and their likelihood of directing another film. Our main sample, provided by OpusData, is a large database of films that were either theatrically released or through other modes such as a VHR, DVD, Blu-ray, or streaming services in the US. The dataset contains the title of the film, release date, name and gender of directors, producers, and actors, as well as the name of publishers. Some entries also include the box office receipt and production budget. We organize the films by the same director to construct a panel of directors' career paths. Next, we link this dataset to the collection of critical responses and audience ratings of each film retrieved from RottenTomatoes. Then, we use these measures in addition to

the box office receipt as the measures of quality (i.e., productivity).

We measure the gender differences in labor market opportunities and returns to productivity with a simple OLS regression model. We capture the raw gender gap using a binary variable of gender we obtain from OpusData. Then, we observe whether the returns to productivity differ between the two binary genders by interacting the gender dummy with continuous variables of worker's productivity. The resulting econometric model is:

$$Y_{idgt} = \beta_0 + \beta_1 \text{female}_{idgt} + \beta_2 \text{productivity}_{idgt} + \beta_3 \text{female}_{idgt} \times \text{productivity}_{idgt} + X_{idgt} + \phi_d + \phi_g + \phi_t + \epsilon_{idgt}, \quad (3.1)$$

where Y_{idgt} is the labor market outcomes related to the film directed by person i, published by distributor d, in genre g, released on year t. The vector of control variables capturing relevant information about the director and film is X_{idgt}. The main regression model also includes fixed effects for distributor, genre, and release year. The errors are clustered at the distributor level.

The model aims to capture the gender differences in outcomes that are unexplained by available covariates with β_1 and the gender differences in the returns to productivity with β_3. In other words, we expect the estimate of β_1 to be negative if women's labor market outcomes are worse than men's labor market outcomes on average holding the critical reception or revenue of their previous film constant. We expect the estimate of β_3 to capture whether this gender gap improves for more productive women.

The key limitation of the model is that we are unable to separately capture the role of gender from the role of other individual characteristics that may correlate with, but are unrelated to, gender. Realistically, we do not expect many confounders that fit the above description, and it is unlikely that these factors explain a large portion

of the observed gender gap. The greater potential concern is the gender bias in the measures of productivity. In the paper, we use three distinct measures of productivity: the critical reception, the audience rating, and the box office receipt of the previous film. All variables could be affected by the director's gender if critics and the audience hold some stereotypes or biases about women in the director role. It is also possible that some films do better among women while other films do better among men based on the gender differences in taste. When we estimate the potential bias in the critical response using the director's gender and the critic's gender, we find a small bias in which female critics tend to rate women-directed films higher than male critics do. However, as previously noted, our findings are consistent with both the gender-based bias in film critics and the taste difference.

The box office receipt may be subject to bias if the audience chooses films based on the director's gender. It is unclear whether the director's gender is an important contributor to the decision to watch and rate a certain film or how they rate the film. If there is any bias in these measures, we expect the bias to work against female directors, based on the studio executives' tendency to view male directors as less risky than female directors. Our expectation comes from the assumptions about executives' incentives. Suppose the executives form these biases based on the differences in the revenues by the director's gender, for example. In that case, we expect female-directed films to do worse than male-directed films in the market. The direction of this statistical discrimination may have shifted in the last few years with the #MeToo movement and general trends toward gender equality, but it is unlikely to apply to most of the observations we include in our study. Hence, the bias in audience ratings or box office receipt, if it exists, could lead us to underestimate women's ability and productivity. The underestimation of productivity means that when we compare women and men holding productivity constant, the true value of productivity is higher for women. Then, we would expect the average gender

gap, β_1, to be biased upward and the gender differences in the returns to productivity, β_3, to be biased downward. If β_1 is negative and β_3 is positive, as we hypothesized, the biases will attenuate the effect, working against finding significant estimates.

Lastly, it must be noted that we are unable to discern the director's performance in the previous film if the film is the director's debut work. Also, we do not observe the directors who are never given a chance to work on a film. Hence, our regression model is unable to estimate the effect of gender on the director's debut work. We will focus on the probability of directing the second, third, and fourth film for the sample of directors with at least one, two, and three films, respectively.

3.3 Descriptive Results

Before we introduce productivity into the discussion, we first list notable gender gaps in the analytic sample. We construct the analytic sample, described in Table 3.1, by first dropping all observations for which we do not observe and cannot discern the gender of the director. We further restrict the sample to only those with nonmissing productivity measures, such as the critical and audience responses or worldwide box office receipts. Overall, our analytic sample contains 6,505 films by 2,967 directors.

As expected, women are underrepresented in film directing. There are only 371 women in our data compared to 2,596 men, and men direct more films than women on average. However, we do observe an increase in the number of women directors debut in the recent decades as shown in Figure 3.1. Nearly three-quarters of films directed by women were documentaries, dramas, or comedies, whereas the genre distribution was more uniform for men. Notably, women were 15pp more likely to direct a documentary film than men. The large gender gaps in genre imply that women and men tend to direct different types of films, where women's focus is often narrower than men's.

Next, we measure tenure as the number of years elapsed from the first film to the last film that the individual directed. According to this measure of experience, women tend to have shorter careers than men, partially due to the proportion of women growing in recent years. Unsurprisingly, women tend to work with a smaller budget and complete projects for minor studios, and their films tend to generate lower box office revenue. However, women-directed films are generally reviewed more favorably than men-directed films. Hence, women appear to produce high-quality films relative to the budget they are given. In contrary, the audience rates the women-directed films and men-directed films similarly on average. Based on this difference in the reviews by critics and the audience may imply that women direct films that are more aligned with critics' preferences. Alternatively, critical reviews may be biased toward women if the underlying distribution of quality is similar for women and men.

Another important data point is the average time gap between two films. According to the data, women tend to take a longer time between two films. This could be an indication that women are given fewer opportunities. Alternatively, it could imply that women tend to work with a smaller team, which makes it more difficult to be a prolific director. It could also be that women prefer to take longer breaks in between films. We explore this further later in this article. Lastly, women directors are more likely to work with women producers and to cast women in the leading role.

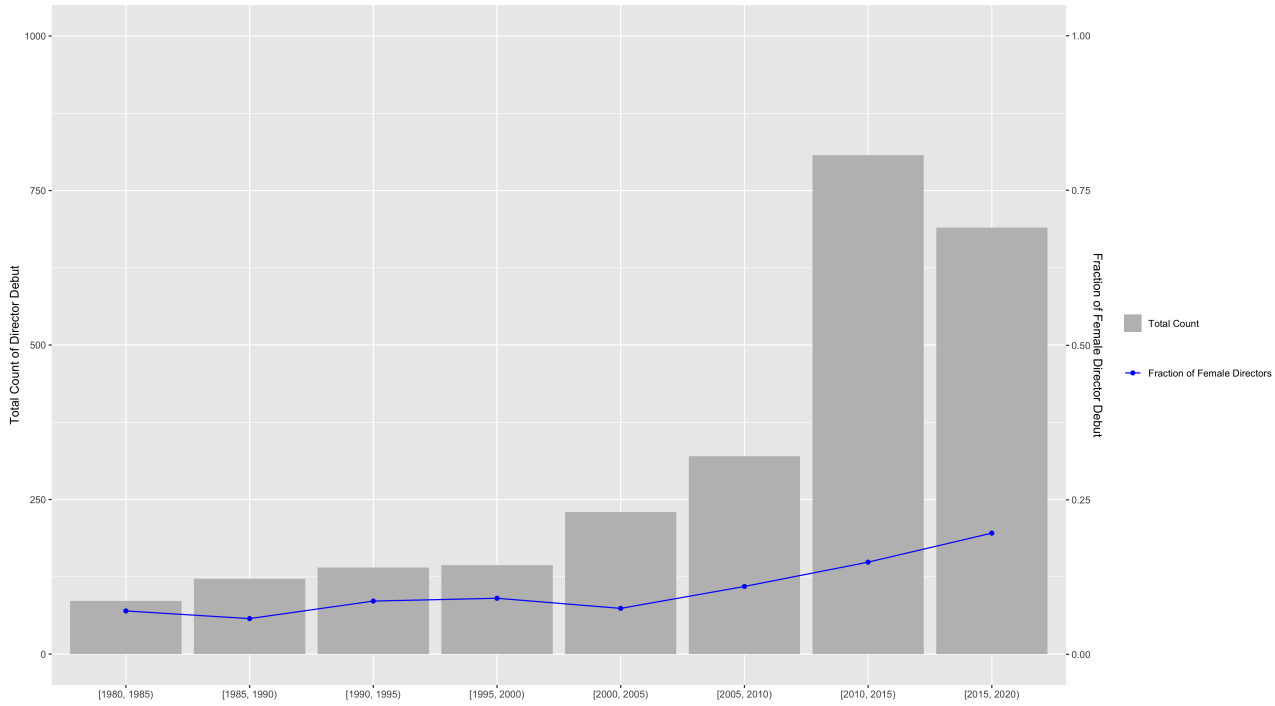
Table 3.1: Summary Statistics

	Female (371)	Male (2,596)	Diff	p-value
Total number of films[a]	538	5,967		
Avg. number of films	1.45	2.30		
Genre				
Documentary	0.22	0.07	0.15	0.00
Drama	0.33	0.25	0.08	0.00
Comedy	0.28	0.24	0.03	0.08
Other	0.01	0.02	-0.01	0.48
Western	0.01	0.01	-0.01	0.32
Horror	0.03	0.08	-0.05	0.00
Adventure	0.04	0.09	-0.05	0.00
Thriller/Suspense	0.05	0.11	-0.06	0.00
Action	0.03	0.11	-0.08	0.00
Genre flexibility	0.04	0.07	-0.03	0.00
Tenure				
0-10	0.94	0.88	0.06	0.00
11-20	0.04	0.07	-0.03	0.02
21-30	0.02	0.03	-0.01	0.18
Above 30	0.00	0.02	-0.02	0.01
Year gap	4.57	3.42	1.15	0.00
Film Scale				
Budget (in millions)	27.36	43.90	-16.54	0.00
Big Budget	0.02	0.08	-0.06	0.00
Gross Box Office (in millions)	37.60	82.71	-45.11	0.00
Major studios (top 40)	0.53	0.73	-0.19	0.00
Major studios (top 10)	0.32	0.53	-0.22	0.00
Reviews				
Tomatometer rating	64.09	56.47	7.63	0.00
Audience rating	61.11	60.26	0.84	0.32
Crew and cast composition				
Multiple roles	0.73	0.67	0.07	0.00
Producer female composition[b]	0.39	0.24	0.16	0.00
At least one female producer[c]	0.52	0.31	0.21	0.00
Actor female composition	0.48	0.32	0.16	0.00
At least one female actor	0.81	0.57	0.24	0.00

Notes: ***, **, and * indicate significance at 1%, 5%, and 10% levels.

[a]The sample is based on films with both gross box office and Tomatometer rating records.
[b]The statistic is calculated excluding films with the same person being the director and producer.
[c]The statistic is calculated excluding films with the same person being the director and producer.

Figure 3.1: Director Debut and Gender Composition



3.4 Regression Results

The main regression model compares the probability of directing another film between women and men with similar quality in the previous film. In Table 3.2, the estimates measure how the director's gender and the previous film's performance affect the probability that the individual directs another film. The regression is a linear probability model. So, the outcome variable is an indicator that takes the value of one if the individual directs another film. Each observation constitutes an event in which the director with any experience of releasing a film decides to either direct another film or choose not to direct another film. The table is divided into four panels, and each panel is based on different sample restrictions. There are three columns in each panel. The columns represent different measures of productivity: the Tomatometer (critics) rating, the average

audience rating, and the box office revenue from left to right.

In the top left panel, we include all observations with the restriction that the previous film must be directed by 2013 or earlier to give the director enough time to direct another film. The first row lists the estimated effect of director's gender on the probability of directing another film, holding the productivity measure constant. The negative estimates indicate that women are less likely to direct another film compared to men, but it is significant only with the box office revenue as the measure of productivity. The second row lists the estimated effect of the director's productivity on the outcome variable. Unsurprisingly, the estimates are positive and significant. Lastly, the third row interacts the director's gender and the productivity measure. It can best be understood as the measure of the gender difference in returns to productivity. Positive estimates in this row would imply that the gender gap in the probability of directing another film grows small for more productive directors. Overall, the estimates suggest that women are treated worse than similarly qualified men, but the effects are small and mostly insignificant. Also, we do not find any significance in the gender gap in returns to productivity.

Table 3.2: Chance and Productivity

Dependent Variable:	Chance of a Next Film			Chance of a Second Film		
Model:	(1)	(2)	(3)	(1)	(2)	(3)
	Tomatometer	Audience	Box Office (in billion)	Tomatometer	Audience	Box Office (in billion)
Female Director	-0.1084	-0.0092	-0.0631*	0.0601	0.1067	0.0047
	(0.0830)	(0.1141)	(0.0353)	(0.1218)	(0.1806)	(0.0519)
Productivity	0.3782***	0.4437***	0.3868***	0.5300***	0.5947***	0.6158***
	(0.0247)	(0.0321)	(0.0579)	(0.0518)	(0.0897)	(0.1502)
Female Director × Productivity	0.0529	-0.1008	-0.1348	-0.1294	-0.1760	-0.3453
	(0.1326)	(0.1697)	(0.3705)	(0.1753)	(0.2695)	(0.5686)
Adjusted R^2	0.18616	0.17184	0.15723	0.16363	0.13921	0.11487
Observations	4,042	4,039	4,042	1,605	1,603	1,605
Dependent Variable:	Chance of a Third Film			Fourth or Beyond		
Model:	(1)	(2)	(3)	(1)	(2)	(3)
	Tomatometer	Audience	Box Office (in billion)	Tomatometer	Audience	Box Office (in billion)
Female Director	-0.4448***	-0.6448**	-0.0055	-0.0946	-0.0721	-0.0665
	(0.1129)	(0.2527)	(0.0882)	(0.1427)	(0.2781)	(0.0612)
Productivity	0.3833***	0.5038***	0.7615***	0.1129***	0.1663***	0.1741***
	(0.0489)	(0.0941)	(0.1348)	(0.0389)	(0.0392)	(0.0541)
Female Director × Productivity	0.7795***	1.075***	0.4492	0.0143	-0.0279	-0.2461
	(0.1150)	(0.3296)	(0.4785)	(0.1849)	(0.4213)	(0.4402)
Adjusted R^2	0.10886	0.09885	0.07952	0.06848	0.06987	0.04739
Observations	650	649	650	1,627	1,627	1,787

Notes: We control for Distributor, Release Year, Genre, and MPAA rating fixed effects in all four panels. Clustered (Distributor) standard-errors in parentheses. *Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

There are reasons to suspect that aggregating the results mask some of the heterogeneity in the sample. Most notably, the studio executives may change their hiring practice depending on how many films the director has previously released. With new directors with only one film under their belt, the studio executives may rely primarily on the previous film's performance to weed out the low-productivity directors. Alternatively, the studio executives may rely more on gender stereotypes as they do not have enough sample size to make judgement regarding the director's ability. To address this issue, we divide the sample into three groups: observations with only one previous film in record, observations with two previous films, and observations with three or more previous films. The next three panels present the results from estimating these subsamples.

The top right panel restricts the sample to the observations in which the outcome is the probability of directing the second film given that the director has directed one film previously. The estimates suggest that performance in the previous film is important, but the director's gender is not, in the decision to hire the director for their second film. The gender gap estimates are insignificant and mostly positive, indicating that the gender gap, if it exists, is biased toward women when the sample is restricted to these early-career directors.

However, the probability of directing the third film in the bottom left panel tells an entirely different story. The estimated gender gaps are negative in this specification and significant in two of three productivity measures. The estimates are surprisingly large. The first two columns indicate that women with the score of 0 in the productivity ratings (by critics and audience, respectively) are 44.5 pp and 64.5 pp less likely to direct another film compared to similarly qualified men. This gap decreases as the director's productivity improves, however. The gap disappears for women who received 60 out of 100 in the productivity ratings, and the gap reverses to favor women at ratings above 60. In other words, women with low quality films fare much worse than men, but top

women directors fare better than similarly qualified men. Hence, the results indicate that productivity is a better indicator of career continuity in women than it is for men.

Lastly, the bottom right panel indicates that women's experience beyond the first three films is largely similar to their experience in the third film. However, the gender gap estimates grow smaller and insignificant. Specifically, the director's fourth film demonstrates similar patterns observed in the third film, and these effects become smaller with each subsequent film as we show in the Appendix. Although we do not have enough data to determine exactly why this heterogeneity exists, but there are some hypotheses. For example, it could be that the studio executives put more weight on the director's gender once the low-quality directors are pushed out of the industry after their debut work. Among those who get another chance, men fare better because the connections and the gender stereotypes become more important while the productivity becomes less important. Eventually, only the best, most persistent women remain in the sample competing with men, and they have long careers in directing. Alternatively, the productivity become less relevant for men's probability of getting rehired, but for women, the productivity remains important. In other words, women need to continuously prove themselves to the studio executives that they can direct quality films to stay around but the pressure is lower for men possibly due to the work culture and gender stereotypes in the film industry that are more favorable to men.

Next, we examine other outcome variables to further discuss the validity of this hypothesis. First, we consider production budget as an outcome variable. The studios often have the power to determine the budget of the film, and directors sometimes may themselves use their connection to add to this budget. Hence, it is partly a measure of the studio's investment on the film and partly it is a reflection of the connections that directors have in the business. The estimates and the accompanying figure show that better critical and audience reception in the previous film generally leads to higher budget

in the subsequent film. However, women-directed films are often low-budget compared to men-directed films from directors of similar track record. Hence, the monetary returns, in the form of the budget allocation, is greater for men than women. This gap persists for both the women with good and bad track record as the interaction term is largely insignificant even if they are very noisy, and we cannot clearly discern whether the gap increases among high-achieving women as the negative coefficients would imply. The figure similarly shows that this gap is persistent regardless of the quality of the director's previous film. There are significant heterogeneities, however. For the director's second film, the gender disparity in budget allocation appears to be nonexistent, it appears to grow in subsequent films (Table 3.3).

These disparities may discourage women from continuing their career, and we may expect higher attrition rate among women than men. To investigate, we first construct an indicator variable that takes a value of one if the director does not have another film in their portfolio for at least five years since the last directing credit. This measure would include both retirements and hiatuses. We see that the estimates are consistent with our hypothesis that women are more likely to retire or take more than five years to direct another film. The gap persists regardless of the performance of the director's previous film among critics. The figure on the right shows that the gap widens if the director's previous film did well in the box office, but the difference is statistically insignificant throughout the distribution.

Table 3.3: Production Budget and Productivity

Dependent Variable:	Budget (Overall)			Budget (Two Films)		
Model:	(1) Tomatometer	(2) Audience	(3) Box Office (in billion)	(1) Tomatometer	(2) Audience	(3) Box Office (in billion)
Variables						
Female Director	-6.642*	-10.93	-8.259**	3.699	-18.06	-2.453
	(3.506)	(8.749)	(3.252)	(11.54)	(24.48)	(3.168)
Productivity	23.45***	30.63***	68.31***	20.93***	30.16***	101.9***
	(3.871)	(4.337)	(7.345)	(5.960)	(5.143)	(24.09)
Female Director × Productivity	-10.17	-1.995	-17.10	-21.03	14.24	-24.68
	(8.001)	(13.88)	(23.16)	(17.32)	(35.84)	(29.18)
Fit statistics						
Adjusted R^2	0.47735	0.47641	0.52613	0.41237	0.41312	0.54723
Observations	2,140	2,140	2,292	549	549	599
Dependent Variable:	Budget (Three Films)			Budget (Four or More Films)		
Model:	(1) Tomatometer	(2) Audience	(3) Box Office (in billion)	(1) Tomatometer	(2) Audience	(3) Box Office (in billion)
Variables						
Female Director	13.42	36.84***	4.816	-12.54	-19.40	-7.068
	(18.92)	(13.16)	(4.628)	(10.30)	(14.37)	(7.982)
Productivity	32.24***	42.23***	93.00***	26.13***	35.31***	50.97***
	(7.813)	(12.11)	(11.06)	(4.388)	(6.605)	(6.674)
Female Director × Productivity	-23.78	-56.70***	-69.41	11.27	19.19	-21.54
	(33.14)	(20.28)	(46.64)	(20.34)	(24.44)	(73.79)
Fit statistics						
Adjusted R^2	0.49869	0.49639	0.54864	0.47771	0.47808	0.50048
Observations	378	378	400	1,213	1,213	1,293

Notes:

Figure 3.2 shows the probability of a director exiting the labor market and the productivity of the previous film. Market exit of a director is defined to be no observation of films within five years of the last film's release. Panel A shows that women are more likely to exit the market than the similarly performed men using critics' rating as the measure for productivity. It also suggests that a director is less likely to exit the market when her film gets a higher critic's score, and there is no discernible gender disparity in the return to such productivity for market exiting. Panel B shows that previous film's box office performance is a spurious predictor for women's exit. One hypothesis is that other reasons may factor into the women's exiting the market than the commercial reception alone, not like men's exit shows significant link with their previous films' commercial performance. Another factor that could confound is that limited observations in women work on films that have potential in getting high commercial returns, and so the estimation itself bears higher standard errors.

In Figure 3.3, we plot the estimation results with identical analyses as in Table 3.3 for the director's next film's production budget. The production budget is used to measure the scale of the next film, and the analyses is to understand the gender disparity in progressing into a larger scale film controlling for the previous productivity. Both panels of Figure 3.3 show that women direct significantly lower budget films than similarly qualified men, whereas the returns to productivity, the slopes of the two fitted lines, are close for women and men. Specifically, Panel A shows that women get to direct a lower budget film than men given receiving the same critics' review for the previous film. Panel B shows a similar result that women direct a lower budget film than men, however, it also shows that the difference is insignificant for directors who have outstanding box office outcome in the previous film. This could suggest that there is no discernible disadvantage in returns to in commercial success for high productivity women. However, it could also be a result of limited observations of women directors in directing potentially high box

office films as is suggested in Table 3.1.

Figure 3.2: Exit and Productivity

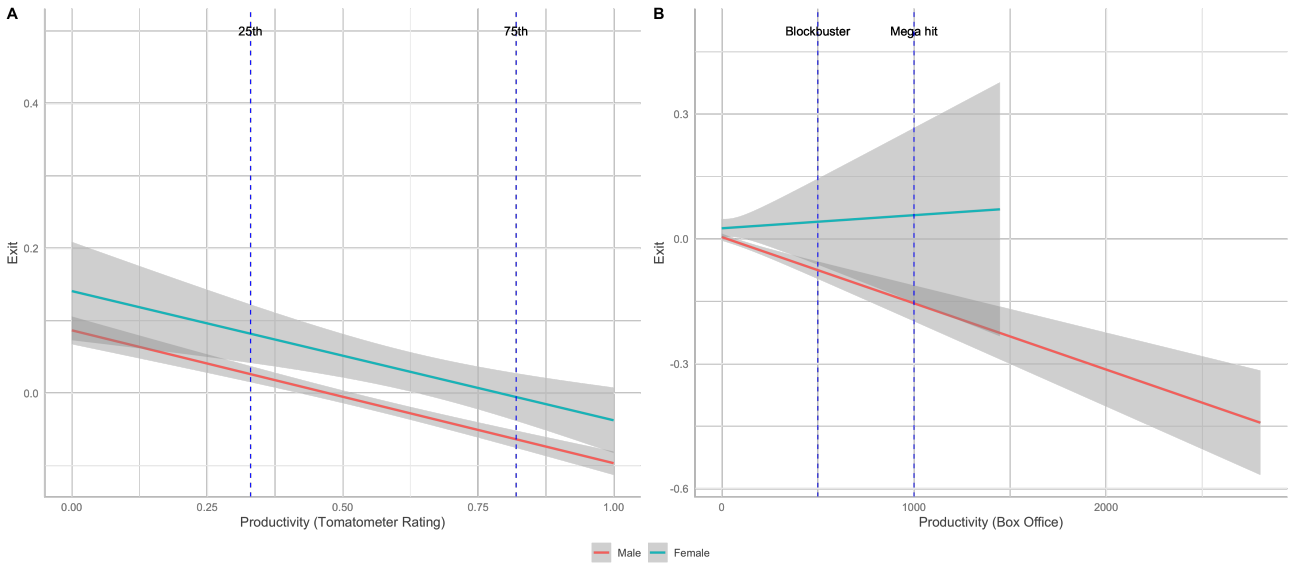
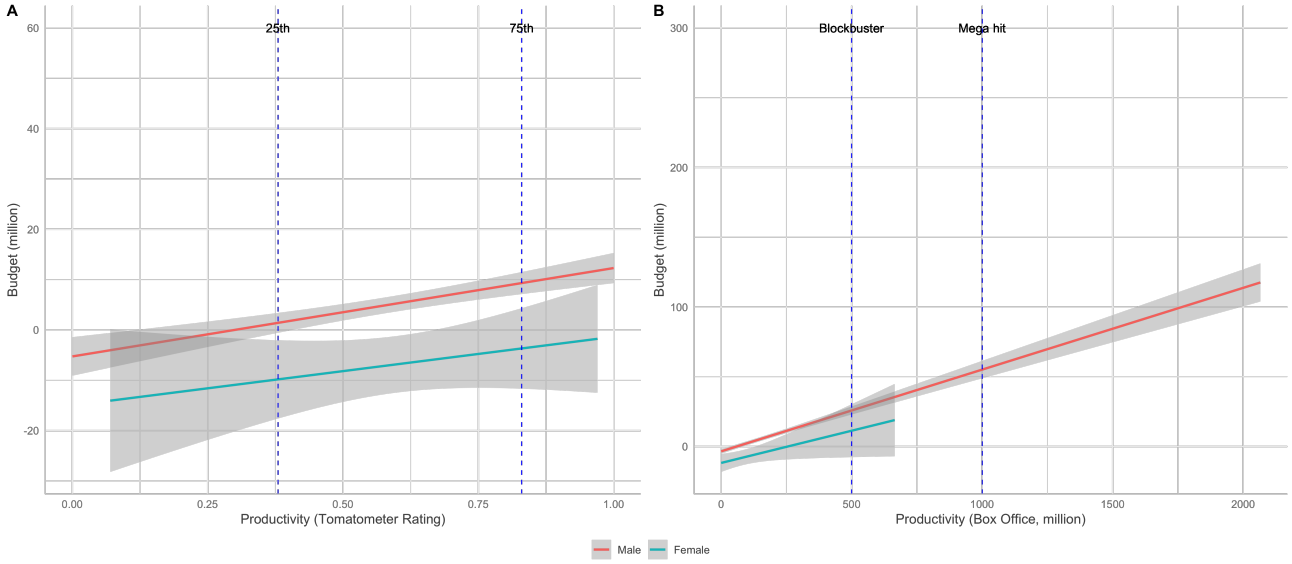


Figure 3.3: Budget and Productivity



3.5 Conclusion

This paper has studied the gender gap in employment outcomes in film directing. We use audience and critical responses and box office to measure a director's previous film's productivity. We find that there are no discernible gender gaps among high-productivity directors, but low-productivity woman directors are much less likely to direct another film than comparable men controlling for the productivity level. Our paper contributes to the economic literature on gender gaps and has important implications for the labor markets that involve frequent job switching or freelancers.

Appendix A

Appendix

A.1 Deriving the Estimation Equation From a Simple Evolutionary Model of Education, Tacit and Book Knowledge, and Comparative Advantage

Consider a standard 2-country, P-product Ricardian trade model, with labor, L, as the only factor of production. Value added in product p given by $Q_p = A_p L_p$ in Home and $Q_p^* = A_p^* L_p^*$ in the rest of the world. As usual, equilibrium in period t requires a ranking of products by Home's relative productivity, $\Omega_t(p) = A_{p,t}/A_{p,t}^*$, and that Home allocates its labor to meet global demand for the products at the top of this ranking. Thus, the likelihood that Home has a comparative advantage in p increases monotonically in $\Omega_t(p)$.

While relatedness can arise for many reasons (see footnotes 3 and 7), we adopt here the most widely discussed assumption in the literature—that it arises due to the importance of tacit knowledge for productivity. Let productivity in product p be log-separable in the tacit and book knowledge relevant to producing it: $A_{p,t} = B_{p,t}^{\theta_B} T_{p,t}^{\theta_T}$. We represent current tacit knowledge relevant to p as translog in prior familiarity with similar products,

F_{p,t_0}, and education, E, so that, ignoring quadratic terms: $\ln(T_{p,t_1}) = \tau_0 + \tau_1 \ln(F_{p,t_0}) + \tau_2 \ln(E) + \tau_3 \ln(F_{p,t_0}) \ln(E)$. Naturally, education and familiarity with p could increase the tacit knowledge relevant to p, so that $\tau_1, \tau_2 \geq 0$. If education permits tacit knowledge to be acquired in, or transferred from, less related production activities, then $\tau_3 < 0$. We can also represent book knowledge relevant to p as translog in the complexity of product p, C_p, and in education. Again, ignoring quadratic terms: $\ln(B_{p,t}) = \beta_0 + \beta_1 \ln(C_p) + \beta_2 \ln(E) + \beta_3 \ln(C_p) \ln(E)$. Education could make book knowledge easier to acquire $\beta_2 > 0$, and we hypothesize that this is especially true of the book knowledge required to produce the most complex products: $\beta_3 > 0$.[1]

Then, assuming that these knowledge production functions are the same in Home and the rest of the world, and using lower case to represent logarithms of variables described earlier, we have:

$$\ln \Omega_{t_1}(p) = \overbrace{[\theta_B \beta_2 + \theta_T \tau_2]}^{(+)}(e - e^*) + \overbrace{\theta_B \beta_3}^{(+)} c_p(e - e^*) + \underbrace{\theta_T \tau_1}_{(+)}(f_{p,t_0} - f^*_{p,t_0}) + \underbrace{\theta_T \tau_3}_{(-)}(e f_{p,t_0} - e^* f^*_{p,t_0}) \tag{A.1}$$

Notice that f_{p,t_0} and f^*_{p,t_0} are functions of the full vector of $\Omega_{t_0}(p)$, so that the P instances of (A.1) together determine how comparative advantage evolves between t_0 and t_1. As the probability that Home has a comparative advantage in p is monotonic in $\Omega_t(p)$, this equation becomes the right-hand side for a binary model of comparative advantage. In a multicountry setting, small-country assumptions make e^* and f^*_{p,t_0} constant across countries, and the first term in (A.1) is absorbed into the country fixed effects. Assume

[1] The restrictions on functional form just enumerated are either without loss, or are testable. Three of the four quadratic terms dropped from the tacit and book knowledge functions are absorbed into country and product fixed effects; and we have tested for robustness to correcting for familiarity-squared (available on request).

that education is exponential in measured schooling, and normalize education so that $e^* \approx 0$. What remains becomes the first line on the right-hand side of specification (1.1), with the following expectations for the parameters in that specification:

(i) Industrial development is path dependent, in the sense that $\beta_F = \theta_T \tau_1 > 0$.

(ii) Education can reduce path dependence by facilitating tacit knowledge acquisition in unfamiliar products: $\beta_{EF} = \theta_T \tau_3 < 0$.

(iii) Countries with more educated workforces are likely to export more complex products: $\gamma_{EC} = \theta_B \beta_3 > 0$.

Figure A.1: Q-Q Plot for $\ln(RCA)$

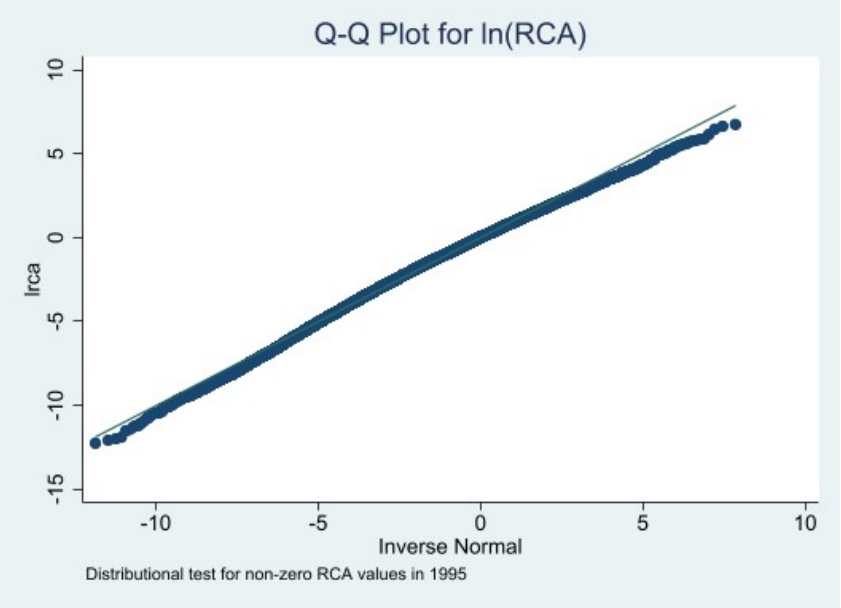
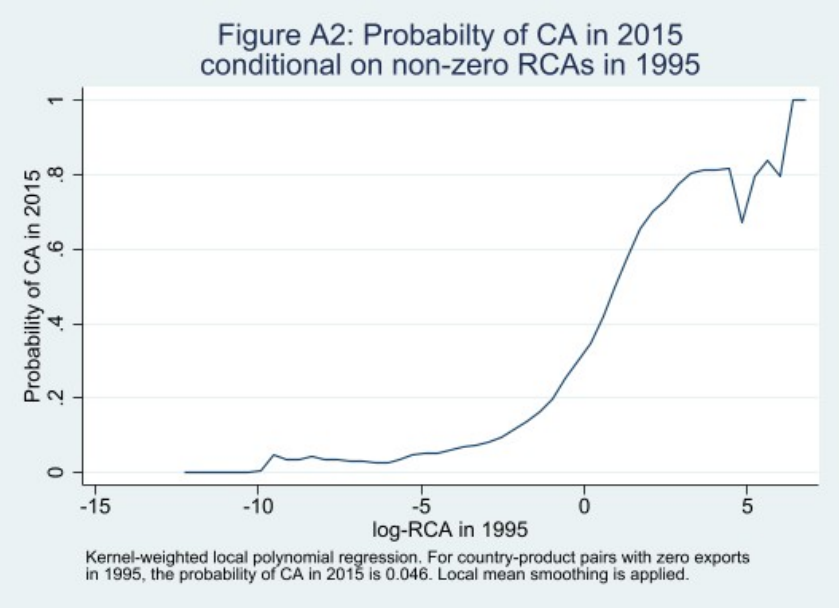


Figure A.2: Probability of Comparative Advantage in 2015 Conditional on Nonzero Revealed Comparative Advantages in 1995



A.2 Film Clustering with Plot Summary

In this section, I use machine learning tools to cluster films based on their plot summaries. A film plot summary is a document containing about 300-700 words that summarize the film. A typical plot summary contains explanation of the film's story line, major plot points, and key characters. Here is an example of a film plot summary:

American millionaire Frederick Harmon (played by Lionel Barrymore) is in Paris, France, for business and pleasure. While enjoying the Parisian night life, he meets and falls in love with Ginette (played by Hope Hampton), a fashion model who moonlights as an apache dancer in a nightclub.

They marry and he returns to New York with her. When Harmon meets the urbane divorcee Nina Olmstead (played by Louise Glaum) he becomes involved in an affair. Ginette discovers her husband's infidelity and decides to win him back by going out with an old boyfriend, Jean (played by Jean Del Val), a member of the Paris underworld.

Nina schemes to end the marriage of the Harmons using the seeming romance between Ginette and Jean. Harmon learns of Nina's treachery and her attempt to estrange the couple fails. He realizes that Ginette was merely trying to make him jealous and that he completely trusts her loyalty to him. They are happily reconciled.[2]

To find similar plot summaries across the documents, I first use Bidirectional Encoder Representations from Transformers (BERT) language models to encode the plot summaries into vectors. BERT is one of the most cutting edge language models developed by Google (Devlin et al., 2018). One key advantage of BERT is that the model not only takes into consideration of the semantic meaning of the words, it also considers the relative position of the words and thus to distinguishes the contextual meanings of the same word based. BERT outperforms other language models in interpreting contextual meanings of words and paragraphs, and is proved to be efficient in many language processing tasks without substantial task-specific architecture modifications. I use pre-

[2]The plot summary is for the film *Fifty-Fifty* directed by Henri Diamant-Berger released in 1925.

trained BERT model to convert each plot summary into a vector with the size of 768, and then use cosine similarities to measure the closeness of the plots based on the vectorization. To avoid BERT to capture the mechanical relationships across the plot summaries based on the common character names, I use named entity recognition with NLTK to exclude the character names from the plot summaries in the pre-processing stage.[3]

I then use cosine similarity to measure the distance between documents. Cosine similarity is a metric that measures similarity of the documents irrespective of the vector size. It is applied pairwise to the vectorized plot summaries and calculates the angle between the two vectors. Mathematically, it takes the following formula:

$$\text{cosine similarity} \equiv cos(\theta_{i,k}) = \frac{V_i \cdot V_k}{\|V_i\|\|V_k\|},$$

where V_i and V_k are vector representation of the documents, and $\theta_{i,k}$ is the angle between the two vectors. The measure takes the value between -1 and 1, and gets closer to 1 when two documents are similar.

Lastly, I use K-means to cluster the films given the similarity metric. Specifically, given the number of clusters (K), K-means solves the following question:

$$\underset{C \in \mathbf{C}}{\text{argmin}} \sum_{C=1}^{K} \frac{1}{2|C|} \sum_{i,k \in C} \|1 - cos(\theta_{i,k})\|^2.$$

Table **??** shows an example of film cluster with $K = 280$. Four films are clustered in the same group. Table A18 shows an example of the film cluster with $K = 28$.[4]

[3]NLTK is a language model platform to build Python programs for language data.

[4]Given that there are 2,823 film plot summaries matched into the main dataframe for analysis, the numbers of clusters are selected to have ten and a hundred films on average in each cluster respectively. However, as K-means optimizes based on the distance, each cluster does not necessarily have balanced size of film observations.

A.3 Tables

Table A1: List of Countries

Countries	ISO Codes	Countries	ISO Codes
Albania	ALB	Mauritius	MUS
Algeria	DZA	Mexico	MEX
Argentina	ARG	Moldova	MDA
Armenia	ARM	Mongolia	MON
Brazil	BRA	Morocco	MAR
Bulgaria	BGR	Nigeria	NGA
Chile	CHL	Panama	PAN
Colombia	COL	People's Republic of China	PRC
Costa Rica	CRI	Peru	PER
Croatia	HRV	Philippines	PHI
Egypt	EGY	Poland	POL
El Salvador	SLV	Republic of Korea	KOR
Estonia	EST	Romania	ROU
Ghana	GHA	Russian Federation	RUS
Honduras	HND	Slovak Republic	SVK
Hungary	HUN	South Africa	ZAF
India	IND	Thailand	THA
Indonesia	INO	Trinidad and Tobago	TTO
Iran	IRN	Tunisia	TUN
Jordan	JOR	Turkey	TUR
Kazakhstan	KAZ	Ukraine	UKR
Kyrgyz Republic	KGZ	Uruguay	URY
Latvia	LVA	Viet Nam	VIE
Lithuania	LTU	Zimbabwe	ZWE
Malaysia	MAL		

Note: ISO=International Organization for Standardization.

Table A2: Best Director 2010-2019

Year	Director	Film	Origin
2010	Thomas Hooper	The King's Speech	British-Australian
2011	Michel Hazanavicius	The Artist	French
2012	Ang Lee	Life of Pi	Taiwanese/Chinese
2013	Alfonso Cuarón	Gravity	Mexican
2014	Alejandro González Iñárritu	Birdman	Mexican
2015	Alejandro González Iñárritu	The Revenant	Mexican
2016	Damien Sayre Chazelle	La La Land	French-American
2017	Guillermo del Toro	The Shape of Water	Mexican
2018	Alfonso Cuarón	Roma	Mexican
2019	Bong Joon-ho	Parasite	South Korean

Notes: The table shows the directors who have received Best Director Oscar awards in the 2010s. All eight directors are foreign-born, and seven of the films are US-produced (co-produced) films.

Table A3: Highest Grossing Films

Rank	Title	Director	Country of Origin	Distributor	Year	Domestic (million, $)	Worldwide (million, $)
1	Avengers: Endgame	Russo Brothers	United States	Walt Disney	2019	858.37	2797.80
2	Avatar	James Cameron	Canada	20th Century Fox	2009	760.51	2788.70
3	Titanic	James Cameron	Canada	Paramount Pictures	1997	659.36	2208.21
4	Star Wars: The Force Awakens	J.J. Abrams	United States	Walt Disney	2015	936.66	2065.48
5	Avengers: Infinity War	Russo Brothers	United States	Walt Disney	2018	678.82	2048.36
6	Jurassic World	Colin Trevorrow	United States	Universal	2015	652.27	1670.40
7	The Lion King	Jon Favreau	United States	Walt Disney	2019	543.64	1656.81
8	Furious 7	James Wan	Malaysia	Universal	2015	353.01	1518.72
9	The Avengers	Joss Whedon	United States	Walt Disney	2012	623.36	1515.10
10	Frozen II	Chris Buck	United States	Walt Disney	2019	477.37	1447.55
11	Avengers: Age Of Ultron	Joss Whedon	United States	Walt Disney	2015	459.01	1396.10
12	Black Panther	Ryan Coogler	United States	Walt Disney	2018	700.06	1341.91
13	Harry Potter and the Deathly Hallows: Part 2	David Yates	United Kingdom	Walt Disney	2011	381.19	1341.65
14	Star Wars: The Last Jedi	Rian Johnson	United States	Walt Disney	2017	620.18	1332.54
15	Jurassic World: Fallen Kingdom	J. A. Bayona	Spain	Universal	2018	417.72	1308.47
16	Frozen	Chris Buck	United States	Walt Disney	2013	400.74	1268.23
17	Beauty and the Beast	Bill Condon	United States	Walt Disney	2017	504.01	1258.21
18	Incredibles 2	Brad Bird	United States	Walt Disney	2018	608.58	1242.81
19	The Fate of the Furious	F. Gary Gray	United States	Universal	2017	225.76	1236.70
20	Iron Man 3	Shane Black	United States	Walt Disney	2013	408.99	1215.39
21	Minions	Pierre Coffin	France	Universal	2015	336.05	1159.63
22	Captain America: Civil War	Russo Brothers	United States	Walt Disney	2016	408.08	1153.28
23	Aquaman	James Wan	Malaysia	Warner Bros.	2018	335.06	1143.97
24	Spider-Man: Far From Home	Jon Watts	United States	Sony Pictures	2019	390.53	1129.95
25	Captain Marvel	Anna Boden	United States	Walt Disney	2019	426.83	1129.73
26	Transformers: Dark of the Moon	Michael Bay	United States	Paramount Pictures	2011	352.39	1123.79
27	The Lord of the Rings: The Return of the King	Peter Jackson	New Zealand	New Line	2003	377.85	1120.22
28	Skyfall	Sam Mendes	United Kingdom	Sony Pictures	2012	304.36	1110.53
29	Transformers: Age of Extinction	Michael Bay	United States	Paramount Pictures	2014	245.44	1104.05
30	The Dark Knight Rises	Christopher Nolan	United Kingdom	Warner Bros.	2012	448.14	1082.23

Notes: The table shows the top 30 grossing US-produced films of all time based on the worldwide box office (nominal). One-third of the directors were born outside of the U.S. The table is adapted from the listings on Wikipedia: Highest-grossing films by adding the country of origin information for the directors.

Table A4: Directors Who Received Two or More Oscar Best Director Awards 1927/1928-2019

Generation	Director	Place of Birth	Year	Film
First Generation				
	Lewis Milestone	Russia	1927/1928	Two Arabian Knights
			1929/1930	All Quiet on the Western Front
	Frank Lloyd	Scotland	1928/1929	The Divine Lady
			1932/1933	Cavalcade
	Frank Capra	Italy	1934	It Happened One Night
			1936	Mr. Deeds Goes to Town
			1938	You Can't Take It With You
	William Wyler	Germany	1942	Mrs. Miniver
			1946	The Best Years of Our Lives
			1959	Ben-Hur
	Billy Wilder	Austria	1945	The Lost Weekend
			1960	The Apartment
	Elia Kazan	Constantinople	1947	Gentleman's Agreement
			1954	On the Waterfront
	Fred Zinnemann	Austria	1953	From Here to Eternity
			1966	A Man for All Seasons
	David Lean	England	1957	The Bridge on the River Kwai
			1962	Lawrence of Arabia
	Milos Forman	Czechoslovakia	1975	One Flew Over the Cuckoo's Nest
			1984	Amadeus
	Ang Lee	Taiwan/China	2005	Brokeback Mountain
			2012	Life of Pi
	Alfonso Cuarón	Mexico	2013	Gravity
			2018	Roma
	Alejandro González Iñárritu	Mexico	2014	Birdman
			2015	The Revenant
Second Generation				
	Frank Borzage		1927/1928	7th Heaven
			1931/1932	Bad Girl
	John Ford		1935	The Informer
			1940	The Grapes of Wrath
			1941	How Green Was My Valley
			1952	The Quiet Man
	Leo McCarey		1937	The Awful Truth
			1944	Going My Way
	Joseph L. Mankiewicz		1949	A Letter to Three Wives
			1950	All About Eve
Third and Higher Generations				
	George Stevens		1951	A Place in the Sun
			1956	Giant
	Robert Wise		1961	West Side Story (Co-directed with Jerome Robbins)
			1965	The Sound of Music
	Oliver Stone		1986	Platoon
			1989	Born on the Fourth of July
	Clint Eastwood		1992	Unforgiven
			2004	Million Dollar Baby
	Steven Spielberg		1993	Schindler's List
			1998	Saving Private Ryan

Notes: The table is adapted from Table 4 in Hirschman (2005). It shows the list of directors who have one the Best Director Oscar award twice or more times since the establishment of the Oscars until 2019. More than one half of the directors are foreign-born, and the other half being second- or higher-generation immigrants. The origins of the directors show brief coincidence with the immigration waves in the U.S.; the early directors are from the European countries and the recent ones from Latin American and Asian countries.

Table A5: Worldwide Box Office

Dependent Variable:	Log Worldwide					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Foreign-born	-0.0418	0.0518	0.0041	0.0068	0.0204	0.0404
	(0.0658)	(0.0502)	(0.0506)	(0.0512)	(0.0510)	(0.0500)
Log Budget	1.121***	0.7330***	0.7194***	0.7486***	0.7110***	0.6974***
	(0.0380)	(0.0272)	(0.0301)	(0.0335)	(0.0342)	(0.0349)
Fixed-effects						
Distributor		Yes	Yes	Yes	Yes	Yes
Genre			Yes	Yes	Yes	Yes
Year				Yes	Yes	Yes
Month					Yes	Yes
MPAA Rating						Yes
Fit statistics						
Adjusted R^2	0.41365	0.64079	0.65123	0.66036	0.66571	0.68048
Observations	3,595	3,517	3,517	3,517	3,517	3,443

Notes: This table analyzes the effect of a foreign-born director on the film outcome measured by the worldwide box office (gross revenue). Column (1) displays a simple OLS estimation of the effect of foreign-born director on the film outcome controlling for the production budget only as a proxy for the expected film quality. A higher quality film is associated with a higher film gross revenue. However, given the same level of film quality, the director's nationality does not affect the gross revenue. In columns (2)-(6), I add the Distributor, Genre, (Release) Year, (Release) month, and MPAA rating fixed effects consecutively. The changes in the specifications do not alter the effect of a foreign-born director on the film's gross revenue overall, according to the coefficients in the first row. The production budget is positively correlated with the gross revenue across the specifications; 1% increase in the budget results in 0.69% increase in the revenue as column (6) shows, and the marginal effect of the production budget is diminishing. The comparison of adjusted R^2 across the columns shows that the variation in film outcomes is the most significant according to the individual distributor effect; the adjusted R^2 increases by about 55% after adding the Distributor fixed effect. All the standard errors are clustered at the individual film level in brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Table A6: Domestic Box Office

Dependent Variable:	Log Domestic					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Foreign-born	-0.2848***	-0.1777***	-0.1925***	-0.1895***	-0.1822***	-0.1753***
	(0.0584)	(0.0480)	(0.0481)	(0.0487)	(0.0483)	(0.0478)
Log Budget	0.8609***	0.5322***	0.5299***	0.6001***	0.5693***	0.5552***
	(0.0249)	(0.0238)	(0.0264)	(0.0317)	(0.0326)	(0.0330)
Fixed-effects						
Distributor		Yes	Yes	Yes	Yes	Yes
Genre			Yes	Yes	Yes	Yes
Year				Yes	Yes	Yes
Month					Yes	Yes
MPAA Rating						Yes
Fit statistics						
Adjusted R^2	0.39697	0.62107	0.62903	0.63621	0.64294	0.65487
Observations	3,526	3,477	3,477	3,477	3,477	3,405

Notes: This table analyzes the effect of a foreign-born director on the film outcome measured by the domestic box office. The domestic box office includes the ticket sales in the U.S. and Canada. Column (1) displays a mean comparison of the film outcome between the foreign- and native-born directors after controlling for the production budget as a proxy for the expected film quality. The production budget is positively associated with the domestic revenue. Given the same level of film production budget, a foreign-born director is negatively associated with the domestic revenue. In columns (2)-(6), I add the Distributor, Genre, (Release) Year, (Release) month, and MPAA rating fixed effects consecutively. The changes in the specifications do not alter the effect of a foreign-born director on the film's domestic revenue according to the coefficients in the first row. In the main specification shown in column (6), a foreign-born director results in a 17.5% lower domestic box office given the same level of the production budget. The production budget is positively correlated with the gross revenue across the specifications; 1% increase in the budget results in 0.56% increase in the revenue as column (6) shows. The marginal effect of the production budget is diminishing and the magnitude is smaller than the coefficient for worldwide box office, which suggests that the domestic consumers do not have the same preference for the big-budget films as in the rest of the world. The comparison of adjusted R^2 across the columns shows that the variation in film outcomes is again the most significant based on the individual distributor effect; the adjusted R^2 increases by about 56% after adding the Distributor fixed effect. All the standard errors are clustered at the individual film level in brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Table A7: International Box Office

Dependent Variable:	Log International					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Foreign-born	0.4228***	0.4430***	0.3176***	0.3298***	0.3572***	0.3591***
	(0.0718)	(0.0722)	(0.0724)	(0.0700)	(0.0697)	(0.0685)
Log Budget	1.233***	1.148***	1.109***	1.005***	0.9439***	0.9339***
	(0.0361)	(0.0468)	(0.0517)	(0.0482)	(0.0489)	(0.0493)
Fixed-effects						
Distributor		Yes	Yes	Yes	Yes	Yes
Genre			Yes	Yes	Yes	Yes
Year				Yes	Yes	Yes
Month					Yes	Yes
MPAA Rating						Yes
Fit statistics						
Adjusted R^2	0.41458	0.45159	0.47301	0.51553	0.52570	0.53084
Observations	3,048	3,020	3,020	3,020	3,020	3,001

Notes: This table analyzes the effect of a foreign-born director on the film outcome measured by the international box office. The international box office includes all the ticket sales outside of the U.S. and Canada. Column (1) displays a mean comparison of the film outcome between the foreign- and native-born directors after controlling for the production budget as a proxy for the expected film quality. The production budget is positively associated with the international revenue. Given the same level of film production budget, a foreign-born director is positively associated with the international revenue. In columns (2)-(6), I add the Distributor, Genre, (Release) Year, (Release) month, and MPAA rating fixed effects consecutively. The changes in the specifications do not alter the effect of a foreign-born director on the film's international revenue according to the coefficients in the first row. In the main specification shown in column (6), a foreign-born director results in a 35.9% higher international box office given the same level of the production budget. The production budget is positively correlated with the gross revenue across the specifications; 1% increase in the budget results in 0.93% increase in the revenue as column (6) shows. The marginal effect of the production budget is diminishing and the magnitude is greater than the coefficient for worldwide box office and also for the domestic box office. It suggests that the international consumers have a greater preference for the big-budget American films than the domestic consumers. The comparison of adjusted R^2 across the columns shows that the variation in film outcomes is again the most significant based on the individual distributor effect; the adjusted R^2 increases by about 56% after adding the Distributor fixed effect. All the standard errors are clustered at the individual film level in brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Table A8: International Box Office (IV)

Dependent Variables:		Foreign-born	Log International
IV stages		First	Second
Model:	Baseline	(2)	(3)
Variables			
Foreign-born	0.2994**		0.5387***
	(0.0853)		(0.1185)
Log Budget	1.020***	0.0275	1.024***
	(0.1234)	(0.0153)	(0.1333)
Number of prod. countries		0.1850***	
		(0.0230)	
Number of languages		-0.0586	
		(0.0358)	
Fixed-effects			
Distributor-Year	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes
Month	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes
Fit statistics			
Adjusted R^2	0.62053	0.04373	0.61578
Observations	3,001	2,870	2,870
F-test (IV only)		65.143	3.3565
Wald (IV only), p-value		3.22×10^{-23}	5.68×10^{-6}

Notes: This table analyzes the effect of foreign-born director using instrumental variable specification. A film is more likely to be directed by a foreign-born director when it is produced by multiple countries. Column (2) shows the first-stage estimation using the production budget, an indicator for the film produced by multiple countries, and an indicator for a film having lines in multiple languages as instrumental variables. In the second stage, the coefficient has an implication that for the films produced by multiple countries a foreign-born director yields 39.4% higher international box office than a native-born director. The magnitude is greater than the coefficient in the baseline specification (column (1)), whereas the coefficient for the production budget is consistent. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAAs Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Table A9: Domestic Box Office (IV)

Dependent Variables:		Foreign-born	Log Domestic
IV stages		First	Second
Model:	Baseline	(2)	(3)
Variables			
Foreign-born	-0.1639*		-1.216***
	(0.0763)		(0.0117)
Log Budget	0.6014***	0.0297	0.6372***
	(0.0438)	(0.0162)	(0.0520)
MultiCountry		0.3035***	
		(0.0324)	
MultiLanguage		-0.0571	
		(0.0510)	
Fixed-effects			
Distributor-Year	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes
Month	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes
Fit statistics			
Adjusted R^2	0.68115	0.05522	0.62022
Observations	3,405	3,001	3,001
F-test (IV only)		87.038	2.2989
Wald (IV only), p-value		5.25×10^{-42}	5.35×10^{-19}

Notes: This table analyzes the effect of foreign-born director on the domestic box office using instrumental variable specification. A film is more likely to be directed by a foreign-born director when it is produced by multiple countries. Column (2) shows the first-stage estimation using the production budget, an indicator for the film produced by multiple countries, and an indicator for a film having lines in multiple languages as instrumental variables. In the second stage, the coefficient has an implication that for the films produced by multiple countries a foreign-born director yields 121.6% higher domestic box office than a native-born director. The magnitude is larger than the coefficient in the baseline specification (column (1)), whereas the coefficient for the production budget is consistent. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAAs Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Table A10: Domestic Box Office (IV)

Dependent Variables:		Foreign-born	Log Domestic
IV stages		First	Second
Model:	Baseline	(2)	(3)
Variables			
Foreign-born	-0.1639*		-1.182***
	(0.0763)		(0.0943)
ln(*Budget*)	0.6014***	0.0305*	0.6369***
	(0.0438)	(0.0134)	(0.0358)
Number of prod. countries		0.1800***	
		(0.0200)	
Number of languages		-0.0568	
		(0.0320)	
Fixed-effects			
Distributor-Year	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes
Month	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes
Fit statistics			
Adjusted R^2	0.68115	0.04373	0.61578
Observations	3,405	2,870	2,870
F-test (IV only)		65.143	3.3565
Wald (IV only), p-value		3.22×10^{-23}	5.68×10^{-6}

Notes: This table analyzes the effect of foreign-born director on the domestic box office using instrumental variable specification. A film is more likely to be directed by a foreign-born director when it is produced by multiple countries. Column (2) shows the first-stage estimation using the production budget, an indicator for the film produced by multiple countries, and an indicator for a film having lines in multiple languages as instrumental variables. In the second stage, the coefficient has an implication that for the films produced by multiple countries a foreign-born director yields 121.6% higher domestic box office than a native-born director. The magnitude is larger than the coefficient in the baseline specification (column (1)), whereas the coefficient for the production budget is consistent. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAAs Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Table A11: Market Power and Director Origin

Dependent Variables:	Baseline	-China	-HK TW	-CN and Chn dir	-China and *ROW*
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Foreign-born	0.3746*	0.3764*	0.3571*	0.4013*	0.3799*
	(0.1580)	(0.1609)	(0.1647)	(0.1686)	(0.1657)
Log Budget	0.9467***	0.9159***	0.9173***	0.9150***	0.9202***
	(0.2013)	(0.1920)	(0.1912)	(0.1925)	(0.1934)
Fixed-effects					
Distributor-Year	Yes	Yes	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Adjusted R^2	0.52411	0.51009	0.50775	0.50959	0.50972
Observations	3,442	3,403	3,384	3,397	3,402

Notes: This table analyzes whether the effect of a foreign-born director on the international box office is driven by certain importing countries' market power. The table shows with the market power of China as an example. Column (1) shows the baseline result with the total international box office as dependent variable. In column (2), the dependent variable is the international box office subtracting the box office generated in mainland China for the associated film. The positive effect of a foreign-born director persists with a similar magnitude as in the baseline specification. In column (3), I further subtract the box office from Hong Kong and Taiwan, China, and the coefficient shows qualitatively similar effect of a foreign-born director. In column (4), I subtract the box office generated in China and exclude the outcome of the films by the Chinese-born directors. It shows that a foreign-born director from the countries other than China generate positive effect on film box office revenue in the rest of the international market excluding China. In column (5), I exclude the box office recorded with spurious categorization as "Rest of World" for the international box office, which may bias the result if the revenue implicitly contains the revenue from China. The coefficient is consistent with the previous columns. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAA Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Table A12: Films Exported to English-speaking Countries

Samples:	Sub-sample		Overall (baseline)	
Dependent Variables:	Log Domestic	Log International	Log Domestic	Log International
Model:	(1)	(2)	(3)	(4)
Variables				
Foreign-born	0.0507	0.5668**	-0.1639*	0.2994**
	(0.1211)	(0.1652)	(0.0763)	(0.0853)
Log Budget	0.4620***	0.6220**	0.6014***	1.020***
	(0.0566)	(0.1919)	(0.0438)	(0.1234)
Fixed-effects				
Distributor-Year	Yes	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
Fit statistics				
Adjusted R^2	0.35055	0.40770	0.68115	0.62053
Observations	874	882	3,405	3,001

Notes: I display the effect of a foreign-born director for the films exported to at least one of the English-speaking countries in comparison to the overall sample. The sub-sample reflects a selection of the films that are exported to the countries that have a closer linguistic distance to the US than the other countries, and it also reflects a selection of the films from the right tail of the box office distribution, as those films are more likely to have a complete records of international box office breakdowns at the country level. Comparison of column (1) shows that foreign-born directors' effect on the domestic market for the films in the sub-sample is not negative as compared to the baseline result in column (3). The dependent variable in column 2 is the sum of the international box office generated within those English-speaking countries. In comparison to the baseline result (column (4)), column (2) shows that the magnitude of the foreign-born directors' effect on the international box office within the sub-sample is larger. This implies that for the top grossing films, the effect of foreign-born directors is larger in magnitude. Columns (2) and (3) show that even comparing with the other English-speaking countries, the US has a distance in preference for films directed by foreign-born directors. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAA Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Table A13: Star Cast

Variable	Native-born	Foreign-born	Overall
Lead 1 Star			
Count (%)	5278 (71.02%)	2154 (28.98%)	7432
0	2043 (38.71%)	708 (32.87%)	2751 (37.02%)
1	3235 (61.29%)	1446 (67.13%)	4681 (62.98%)
Lead 2 Star			
Count (%)	5278 (71.02%)	2154 (28.98%)	7432
0	2155 (40.83%)	774 (35.93%)	2929 (39.41%)
1	3123 (59.17%)	1380 (64.07%)	4503 (60.59%)
Number of Star Leads			
Count	5278	2154	7432
Mean (SD)	1.20 (0.87)	1.31 (0.84)	1.24 (0.86)

Notes: The table shows the summary statistics on the star power of the films. *Lead 1 Star* and *Lead 2 Star* are the indicators for the leading roles of the film to be a *Star*. The *Star* is defined based on the cast's previous film performance in terms of worldwide box office; if the leading cast's previous film made into the top 25 percentile of the worldwide box office distribution in the corresponding year, then the cast is defined to be a star and the indicator takes the value of 1. The statistics show that overall 63.0% of the films have the first leading cast to be a star, and 60.6% of the films have the second leading cast to be a star; on average, a film has 1.24 star casts leading the roles. Columns (1) and (2) show the difference in the star power in films directed by native- vs foreign-born directors. A foreign-born director's film has a higher probability of having either one of the leading casts to be a star, and as well as having a higher number of the star casts, than a native-born director's film.

Table A14: Star effect

Dependent Variable:	Log International		
Model:	Baseline	(2)	(3)
Variables			
Foreign-born	0.2994**	0.2988***	0.2989**
	(0.0853)	(0.0814)	(0.0801)
Log Budget	1.020***	0.7733***	0.7741***
	(0.1234)	(0.1201)	(0.1208)
Lead 1 Star		1.016***	
		(0.0915)	
Lead 2 Star		0.9065***	
		(0.0289)	
Number of Stars			0.9597***
			(0.0256)
Fixed-effects			
Distributor-Year	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes
Month	Yes	Yes	Yes
Fit statistics			
Adjusted R^2	0.62053	0.66872	0.66883
Observations	3,001	3,001	3,001

Note: The table checks the robustness of the foreign-born director effect on the international box office when a film has star leading cast. *Lead 1 Star* and *Lead 2 Star* are the indicators for the leading role of the film to be a *Star*. The *Star* is defined based on the actors/actresses previous film performance being made into the top 25 percentile of the worldwide box office distribution in the corresponding year. Adding the *Star* effect either with an indicator (column (2)) or the number of star (column (3)) does not alter the effect of a foreign-born director on the film. Instead, it results in the attenuation of the effect of the production budget on the film outcome. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAA Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

Table A16: Films by Director Type

Variable	Foreign-born	Native-born	Overall
Type			
Count (%)	1903 (36.29%)	3341 (63.71%)	5244
(Col %)			
DirectorOnly-Type	907 (47.66%)	2095 (62.71%)	3002 (57.25%)
Actor-Type	67 (3.52%)	101 (3.02%)	168 (3.20%)
Producer-Type	126 (6.62%)	176 (5.27%)	302 (5.76%)
Writer-Type	358 (18.81%)	307 (9.19%)	665 (12.68%)
Other-Type	445 (23.38%)	662 (19.81%)	1107 (21.11%)

Variable	Foreign-born	Native-born	Overall
Type			
Count (%)	3396 (28.21%)	8641 (71.79%)	12037
(Col %)			
DirectorOnly-Type	1374 (40.46%)	4471 (51.74%)	5845 (48.56%)
Actor-Type	174 (5.12%)	440 (5.09%)	614 (5.10%)
Producer-Type	522 (15.37%)	1073 (12.42%)	1595 (13.25%)
Writer-Type	763 (22.47%)	1344 (15.55%)	2107 (17.50%)
Other-Type	563 (16.58%)	1313 (15.20%)	1876 (15.59%)

Notes: The table summarizes the number of directors in each director type (upper panel), and the number of films directed by each director type (lower panel). Directors are categorized into five types based on their proclaimed job titles. For example, a director is producer-type if the director has one of her first two job titles being "Producer" besides being a "Director." Foreign- and native-born directors show comparable distribution across the types. The upper panel of the table shows that *DirectorOnly-Type* of directors are most common for both foreign- and native-born directors, and *Actor-Type* directors are least common. There are a higher fraction of *DirectorOnly-Type* among the native-born directors than the foreign-born directors, this suggests that the native-born directors are subject to negative selection due to a relatively lower cost of entry.

Table A17: Director Type

Dependent Variables:	Log Worldwide	Log Domestic	Log International
Model:	(1)	(2)	(3)
Variables			
Log Budget	0.8065***	0.6316***	1.074***
	(0.0596)	(0.0417)	(0.1117)
DirectorOnly-Type	-0.3579**	-0.3626***	-0.4718***
	(0.0965)	(0.0820)	(0.1494)
Producer-Type	0.0703	0.0201	0.0686
	(0.0859)	(0.0764)	(0.0747)
Writer-Type	-0.0111	-0.0588	-0.1202
	(0.0962)	(0.1031)	(0.1092)
Other-Type	-0.0178	0.0472	-0.3160
	(0.1655)	(0.1200)	(0.2082)
Foreign-born	0.3455*	0.0047	0.6640***
	(0.1698)	(0.2183)	(0.1502)
Foreign-born × DirectorOnly-Type	-0.1907	-0.0298	-0.2928***
	(0.1275)	(0.1926)	(0.0406)
Foreign-born × Producer-Type	-0.5424**	-0.3350	-0.8063***
	(0.1765)	(0.2206)	(0.1081)
Foreign-born × Writer-Type	-0.3338**	-0.1071	-0.4206**
	(0.1115)	(0.1668)	(0.1764)
Foreign-born × Other-Type	-0.2815	-0.3770*	-0.1573
	(0.2028)	(0.1593)	(0.2491)
Fixed-effects			
Distributor-Year	Yes	Yes	Yes
Genre-Year	Yes	Yes	Yes
MPAA Rating	Yes	Yes	Yes
Month	Yes	Yes	Yes
Fit statistics			
Adjusted R^2	0.72780	0.69218	0.63579
Observations	3,261	3,227	2,845

Notes: The table analyzes the effect of a foreign-born director on the box office outcomes categorized by their proclaimed job titles. A director is a producer-type director if one of her first two proclaimed job titles is "Producer." DirectorOnly-Type of directors are the ones with "Director" as their job title only. I assume that, among the foreign-born directors, actor-type directors carry a higher level of linguistic and communication skills, and among the native-born directors, the linguistic skills are unanimous. The above regression analysis omitted the actor-type as the baseline category from the table. Column (1) shows that, compared to the baseline category, DirectorOnly-type makes 35.8% lower gross revenues, and other types of native-born directors make about the same. Foreign-born actor-type of directors make 34.6% higher gross revenue than the native-born actor-type director, foreign-born producer-type makes 19.7% (0.3455-0.5424) lower than the native-born actor-type directors, and foreign-born writer-type makes 1.2% (0.3455-0.3338) lower than the native-born actor-type directors; in other words, a foreign-born actor-type director makes the most among the foreign-born directors in gross revenue. Similarly, for the international revenues, foreign-born actor-type of directors makes more than the other foreign-born director types, but an actor-type of native-born director does not show an advantage over the other types of native-born directors. This implies that the linguistic skill is important for film directing and the skill could also be a factor which is highly linked to cultural understanding affects the translation capability of the foreign-born directors. Domestically, there is little evidence showing the effect of linguistic skill of a foreign-born director. All the standard errors are clustered at the Distributor-Year, Genre-Year, MPAAs Rating, and (Release) Month in the brackets. ***, **, and * indicate significance at 1%, 5%, and 10% levels.

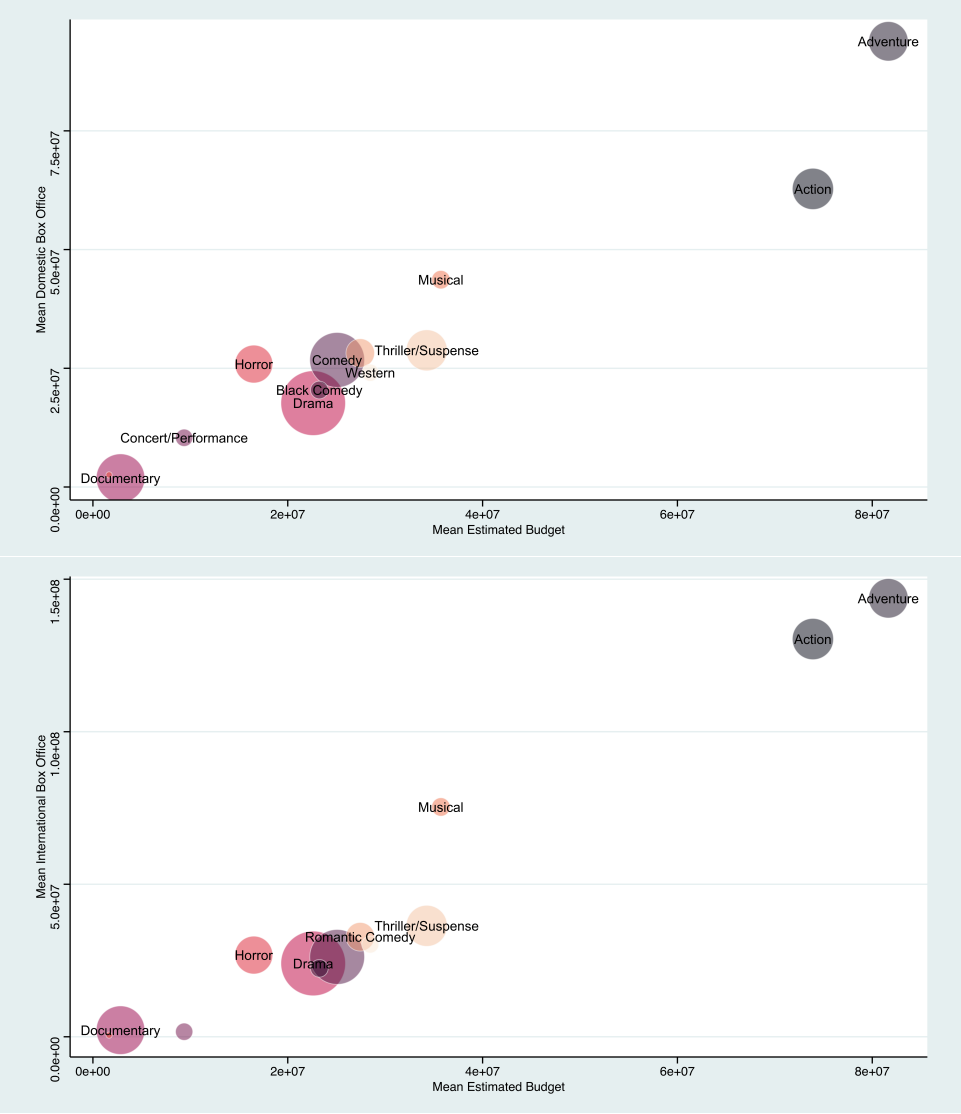
Table A18: Example of Film Cluster ($K = 28$)

	Film Title	Director Name	Nationality	Release Year	Domestic Box Office	International Box Office
1	2001: A Space Odyssey	Stanley Kubrick	Native-born	1968	59936321	11927191
2	Beneath the Planet of the Apes	Ted Post	Native-born	1970	17489009	NA
3	The Andromeda Strain	Robert Wise	Native-born	1971	12376563	NA
4	Star Trek: The Motion Picture	Robert Wise	Native-born	1979	82258456	56741544
5	Firefox	Clint Eastwood	Native-born	1982	45785720	NA
6	Star Trek III: The Search for Spock	Leonard Nimoy	Native-born	1984	76471046	10528954
7	Spaceballs	Mel Brooks	Native-born	1987	38119483	NA
8	The Hunt for Red October	John McTiernan	Native-born	1990	120709866	79790134
9	Die Hard 2	Renny Harlin	Foreign-born	1990	117323878	122490010
10	Spaced Invaders	Patrick Read Johnson	Native-born	1990	15000000	NA
11	Jurassic Park	Steven Spielberg	Native-born	1993	402523348	643049687
12	Waterworld	Kevin Reynolds	Native-born	1995	88246220	176000000
13	Star Trek: First Contact	Jonathan Frakes	Native-born	1996	92027888	57972112
14	Broken Arrow	John Woo	Foreign-born	1996	70645997	77700000
15	Mars Attacks!	Tim Burton	Native-born	1996	37771017	63600000
16	Men in Black	Barry Sonnenfeld	Native-born	1997	250690539	337100000
17	The Lost World: Jurassic Park	Steven Spielberg	Native-born	1997	229086679	389552320
18	The Fifth Element	Luc Besson	Foreign-born	1997	63570862	200327899
19	Event Horizon	Paul W.S. Anderson	Foreign-born	1997	26673242	711
20	Volcano	Mick Jackson	Foreign-born	1997	47546796	72553204
21	Lost In Space	Stephen Hopkins	Foreign-born	1998	69117629	66929688
22	Armageddon	Michael Bay	Native-born	1998	201578182	353021818
23	Space Cowboys	Clint Eastwood	Native-born	2000	90454043	38420000
24	Jurassic Park III	Joe Johnston	Native-born	2001	181166115	184733885
25	Planet of the Apes	Tim Burton	Native-born	2001	180011740	182200000
26	Black Hawk Down	Ridley Scott	Foreign-born	2001	108638745	51052340
27	Evolution	Ivan Reitman	Foreign-born	2001	38311134	60030798
28	The Sum of All Fears	Phil Alden Robinson	Native-born	2002	118471320	75028680
29	Impostor	Gary Fleder	Native-born	2002	6114237	1860370
30	The Core	Jon Amiel	Foreign-born	2003	31111260	43009532
31	I, Robot	Alex Proyas	Foreign-born	2004	144801023	203828562
32	United 93	Paul Greengrass	Foreign-born	2006	31567134	46067901
33	The Day the Earth Stood Still	Scott Derrickson	Native-born	2008	79366978	158003372
34	Monsters vs. Aliens	Rob Letterman	Native-born	2009	198351526	183335854
35	G.I. Joe: The Rise of Cobra	Stephen Sommers	Native-born	2009	150201498	152267519
36	District 9	Neill Blomkamp	Foreign-born	2009	115646235	96807196
37	Skyline	Brothers Strause	Native-born	2010	21416355	46848618
38	Battle: Los Angeles	Jonathan Liebesman	Native-born	2011	83552429	129911547
39	The Darkest Hour	Chris Gorak	Native-born	2011	21443494	41388221
40	Apollo 18	Gonzalo Lopez-Gallego	Foreign-born	2011	17686929	8830890
41	Prometheus	Ridley Scott	Foreign-born	2012	126477084	275971181
42	Star Trek Into Darkness	J.J. Abrams	Native-born	2013	228778661	238602923
43	Oblivion	Joseph Kosinski	Native-born	2013	89107235	198809398
44	Interstellar	Christopher Nolan	Foreign-born	2014	188017894	464172502
45	Captain America: The Winter Soldier	Anthony Russo	Native-born	2014	259746958	454654931
46	Dawn of the Planet of the Apes	Matt Reeves	Native-born	2014	208545589	502098977
47	Godzilla	Gareth Edwards	Foreign-born	2014	200676069	328400000
48	The Signal	William Eubank	Native-born	2014	600896	2064426
49	Jurassic World	Colin Trevorrow	Native-born	2015	652306625	1017673342
50	Avengers: Age of Ultron	Joss Whedon	Native-born	2015	459005868	936311111
51	Star Trek Beyond	Justin Lin	Foreign-born	2016	158848340	176825368
52	Alien: Covenant	Ridley Scott	Foreign-born	2017	74262031	164259216

Notes: The table presents an example of film cluster with 28 pre-specified film clusters. The films are sorted based on the release year.

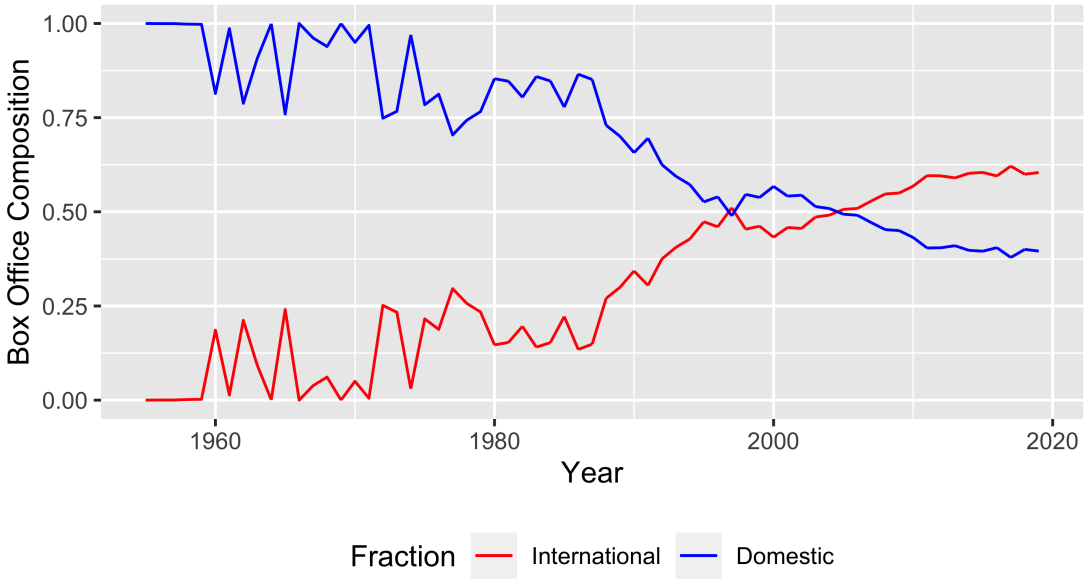
A.4 Figures

Figure A.1: Production Budget and Box Office



Notes: The figures show the correlation between the production budget and box office. They also show the mean production budget and box office level for each genre, where the size of the bubble indicates the number of films belonging to each genre. The production budget is positively correlated with both the domestic and international box office; Adventure and Action films make the top in distributions for both the production budget and box office.

Figure A.2: Box Office Composition



Notes: The figure shows the composition of the gross box office revenue for the U.S. film industry in each film release year since the 1960s. The domestic box office revenues dominate the total revenue sales before the 1980s. The importance of the international market increased since the late 1980s, and the international market has surpassed the domestic market in making the larger part of the total revenue since the early 2010s.

Bibliography

Alberto Abadie, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research, 2017.

Manuel R Agosin, Roberto Alvarez, and Claudio Bravo-Ortega. Determinants of export diversification around the world: 1962–2000. *The World Economy*, 35(3):295–315, 2012.

Nadir Altinok, Claude Diebolt, and Jean-Luc Demeulemeester. A new international database on education quality: 1965–2010. *Applied Economics*, 46(11):1212–1247, 2014.

Joseph G Altonji, Prashant Bharadwaj, and Fabian Lange. Changes in the characteristics of american youth: Implications for adult outcomes. *Journal of Labor Economics*, 30 (4):783–828, 2012.

Alice Hoffenberg Amsden et al. *The rise of "the rest": challenges to the west from late-industrializing economies*. Oxford University Press, USA, 2001.

Joshua D Angrist and Jörn-Steffen Pischke. Mostly harmless econometrics. In *Mostly Harmless Econometrics*. Princeton university press, 2008.

The Growth Lab at Harvard University. International Trade Data (SITC, Rev. 2), 2019a. URL `https://doi.org/10.7910/DVN/H8SFD2`.

The Growth Lab at Harvard University. International Trade Data (HS, 92), 2019b. URL `https://doi.org/10.7910/DVN/T4CHWJ`.

Dany Bahar, Ricardo Hausmann, and Cesar A Hidalgo. Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics*, 92(1):111–123, 2014.

Dany Bahar, Samuel Rosenow, Ernesto Stein, and Rodrigo Wagner. Export take-offs and acceleration: Unpacking cross-sector linkages in the evolution of comparative advantage. *World Development*, 117:48–60, 2019.

Asian Development Bank. *The Asian Development Outlook 2007: Growth amid change*. Manila, 2007.

World Bank. The east asian miracle. 1993.

Robert J Barro and Jong Wha Lee. A new data set of educational attainment in the world, 1950–2010. *Journal of development economics*, 104:184–198, 2013.

Howard S Becker. *Art worlds: updated and expanded*. University of California Press, 2008.

Marianne Bertrand, Emir Kamenica, and Jessica Pan. Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614, 2015.

Denise D Bielby and William T Bielby. Women and men in film: Gender inequality among writers in a culture industry. *Gender & Society*, 10(3):248–270, 1996.

Denise D Bielby and William T Bielby. Hollywood dreams, harsh realities: writing for film and television. *Contexts*, 1(4):21–27, 2002.

William T Bielby and Denise D Bielby. Cumulative versus continuous disadvantage in an unstructured labor market: Gender differences in the careers of television writers. *Work and occupations*, 19(4):366–386, 1992.

Anne Booth. Education and economic development in southeast asia: myths and realities. *ASEAN Economic Bulletin*, pages 290–306, 1999.

George J Borjas. Self-selection and the earnings of immigrants. *The American economic review*, pages 531–553, 1987.

George J Borjas. The economic analysis of immigration. In *Handbook of labor economics*, volume 3, pages 1697–1760. Elsevier, 1999.

George J Borjas. Does immigration grease the wheels of the labor market? *Brookings papers on economic activity*, 2001(1):69–133, 2001.

George J Borjas and Kirk B Doran. The collapse of the soviet union and the productivity of american mathematicians. *The Quarterly Journal of Economics*, 127(3):1143–1203, 2012.

Konrad B Burchardi, Thomas Chaney, and Tarek A Hassan. Migrants, ancestors, and foreign investments. *The Review of Economic Studies*, 86(4):1448–1486, 2019.

Thomas Buser, Muriel Niederle, and Hessel Oosterbeek. Gender, competitiveness, and career choices. *The quarterly journal of economics*, 129(3):1409–1447, 2014.

A Colin Cameron and Pravin K Trivedi. *Microeconometrics: methods and applications*. Cambridge university press, 2005.

David Card. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64, 2001.

David Card. Immigration and inequality. *American Economic Review*, 99(2):1–21, 2009.

Ha-Joon Chang. *23 things they don't tell you about capitalism*. Bloomsbury Publishing USA, 2012.

Antonio Ciccone and Elias Papaioannou. Human capital, the structure of production, and growth. *The review of economics and statistics*, 91(1):66–82, 2009.

Nicola D Coniglio, Davide Vurchio, Nicola Cantore, and Michele Clara. On the evolution of comparative advantage: path-dependent versus path-defying changes. *Journal of International Economics*, 133:103522, 2021.

David A. Cook and Robert Sklar. *History of the motion picture*. Encyclopædia Britannica, inc., 2019.

Arthur De Vany. *Hollywood economics: How Extreme Uncertainty Shapes the Film Industry*. Routledge, 2003.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Thomas Dohmen and Armin Falk. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American economic review*, 101(2):556–90, 2011.

Barbara Freedman Doyle. *Make Your Movie: What You Need to Know about the Business and Politics of Filmmaking*. Routledge, 2012.

Robert C Feenstra, Robert Inklaar, and Marcel P Timmer. The next generation of the penn world table. *American economic review*, 105(10):3150–82, 2015.

Jesus Felipe and Cesar A Hidalgo. Economic diversification: implications for kazakhstan. In *Development and Modern Industrial Policy in Practice*. Edward Elgar Publishing, 2015.

Jesus Felipe, Aashish Mehta, and Changyong Rhee. Manufacturing matters... but it's the jobs that count. *Cambridge Journal of Economics*, 43(1):139–168, 2019.

Greg Felker, KS Jomo, and Rajah Rasiah. *Industrial technology development in Malaysia: Industry and firm studies*. Routledge, 2013.

Koen Frenken, Frank Van Oort, and Thijs Verburg. Related variety, unrelated variety and regional economic growth. *Regional studies*, 41(5):685–697, 2007.

David W Galenson. Innovators: Filmmakers. Technical report, National Bureau of Economic Research, 2010.

Ricard Gil and Ryan Lampe. The adoption of new technologies: Understanding hollywood's (slow and uneven) conversion to color. *The Journal of Economic History*, 74 (4):987–1014, 2014.

Ricard Gil and Pablo T Spiller. The organizational implications of creativity: the us film industry in mid-xxth century. Technical report, National Bureau of Economic Research, 2007.

Reuben Gronau et al. A useful interpretation of r2; in binary choice models (or, have we dismissed the good old r2; prematurely). Technical report, 1998.

Gordon H Hanson and Chong Xiang. International trade in motion picture services. In *International trade in services and intangibles in the era of globalization*, pages 203–222. University of Chicago Press, 2009.

Eric A Hanushek and Ludger Woessmann. The role of cognitive skills in economic development. *Journal of economic literature*, 46(3):607–68, 2008.

Dominik Hartmann, Miguel R Guevara, Cristian Jara-Figueroa, Manuel Aristarán, and César A Hidalgo. Linking economic complexity, institutions, and income inequality. *World development*, 93:75–93, 2017.

Ricardo Hausmann, Francisco R Rodriguez, and Rodrigo Andres Wagner. Growth collapses. 2006.

Ricardo Hausmann, Jason Hwang, and Dani Rodrik. What you export matters. *Journal of economic growth*, 12(1):1–25, 2007.

Ricardo Hausmann, César A Hidalgo, Sebastián Bustos, Michele Coscia, and Alexander Simoes. *The atlas of economic complexity: Mapping paths to prosperity*. Mit Press, 2014.

César A Hidalgo and Ricardo Hausmann. The building blocks of economic complexity. *Proceedings of the national academy of sciences*, 106(26):10570–10575, 2009.

César A Hidalgo, Bailey Klinger, A-L Barabási, and Ricardo Hausmann. The product space conditions the development of nations. *Science*, 317(5837):482–487, 2007.

César A Hidalgo, Pierre-Alexandre Balland, Ron Boschma, Mercedes Delgado, Maryann Feldman, Koen Frenken, Edward Glaeser, Canfei He, Dieter F Kogler, Andrea Morrison, et al. The principle of relatedness. In *International conference on complex systems*, pages 451–457. Springer, 2018.

Paul M Hirsch. Processing fads and fashions: An organization-set analysis of cultural industry systems. *American journal of sociology*, 77(4):639–659, 1972.

Charles Hirschman. Immigration and the american century. *Demography*, 42(4):595–620, 2005.

Michael Hobday et al. Innovation in east asia. *Books*, 1995.

Isaac R Holloway. Foreign entry, quality, and cultural distance: Product-level evidence from us movie exports. *Review of World Economics*, 150(2):371–392, 2014.

Kai-Lung Hui and Ivan PL Png. On the supply of creative work: Evidence from the movies. *American Economic Review*, 92(2):217–220, 2002.

Jennifer Hunt and Marjolaine Gauthier-Loiselle. How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, 2(2):31–56, 2010.

Michael Jetter and Andrés Ramírez Hassan. Want export diversification? educate the kids first. *Economic Inquiry*, 53(4):1765–1782, 2015.

Kose John, S Abraham Ravid, and Jayanthi Sunder. Performance and managerial turnover: Evidence from career paths of film directors. 2003.

Kose John, S Abraham Ravid, and Jayanthi Sunder. Managerial ability and success: Evidence from the career paths of film directors. *Journal of Corporate Finance*, 44: 425–439, 2017.

Chalmers Johnson. *MITI and the Japanese miracle: the growth of industrial policy, 1925-1975*. Stanford university press, 1986.

Kwame Sundaram Jomo. *Southeast Asian paper tigers*. Routledge, 2013.

Kwame Sundaran Jomo. *Globalization, liberalization and equitable development: lessons from East Asia*. United Nations Research Institute for Social Development Geneva, 2003.

Alan B Krueger and Mikael Lindahl. Education for growth: Why and for whom? *Journal of economic literature*, 39(4):1101–1136, 2001.

Martha M. Lauzen. The celluloid ceiling in a pandemic year: Employment of women on the top u.s. films of 2021, 2022.

Edward E Leamer. Sources of comparative advantage: theory and evidence, 1984.

Hong Luo and Laurina Zhang. Gender Orientation and Segregation of Ideas: #MeToo's Impact in Hollywood. page 53, 2021.

Hong Luo and Laurina Zhang. Scandal, social movement, and change: Evidence from# metoo in hollywood. *Management Science*, 68(2):1278–1296, 2022.

Andrew Maynard. *Films from the Future: The Technology and Morality of Sci-Fi Movies*. Mango Media, 2018.

Aashish Mehta and Jesus Felipe. Education and the journey to the core: path dependence or leapfrogging? *Available at SSRN 2467895*, 2014.

Peter Miskell. Hollywood films and foreign markets in the studio era: A fresh look at the evidence. 2014.

Petra Moser, Alessandra Voena, and Fabian Waldinger. German jewish émigrés and us invention. *American Economic Review*, 104(10):3222–55, 2014.

Richard R Nelson. *An evolutionary theory of economic change*. harvard university press, 1985.

Gianmarco IP Ottaviano and Giovanni Peri. The economic value of cultural diversity: evidence from us cities. *Journal of Economic geography*, 6(1):9–44, 2006.

Giovanni Peri and Vasil Yasenov. The labor market effects of a refugee wave: Applying the synthetic control method to the mariel boatlift. Technical report, National Bureau of Economic Research, 2015.

Joy Press. Promising young women. *Vanity Fair: the 27th Annual Hollywood Issue*, 2021.

S Abraham Ravid. Information, blockbusters, and stars: A study of the film industry. *The Journal of Business*, 72(4):463–492, 1999.

Stephen Redding. Dynamic comparative advantage and the welfare effects of trade. *Oxford economic papers*, 51(1):15–39, 1999.

Ronny Regev. *Working in Hollywood: How the Studio System Turned Creativity Into Labor*. UNC Press Books, 2018.

Violaine Roussel. *Representing talent: Hollywood agents and the making of movies*. University of Chicago Press, 2017.

Heather Sarsons, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. Gender differences in recognition for group work. *Journal of Political Economy*, 129(1):101–147, 2021.

Pier Paolo Saviotti and Koen Frenken. Export variety and the economic performance of countries. *Journal of Evolutionary Economics*, 18(2):201–218, 2008.

Theodore W Schultz. The value of the ability to deal with disequilibria. *Journal of economic literature*, 13(3):827–846, 1975.

Allen John Scott. *On Hollywood: The place, the industry*. Princeton University Press, 2005.

Joseph E Stiglitz. Some lessons from the east asian miracle. *The world Bank research observer*, 11(2):151–177, 1996.

Joseph E Stiglitz and Bruce Greenwald. Creating a learning society. In *Creating a Learning Society*. Columbia University Press, 2014.

Joe Studwell. *How Asia works: Success and failure in the world's most dynamic region*. Open Road+ Grove/Atlantic, 2013.

Jeffrey M Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.

Srilata Zaheer. Overcoming the liability of foreignness. *Academy of Management journal*, 38(2):341–363, 1995.