Two-Sided Matching Model

by

Anh Le

Department of Department of Political Science
Duke University

Date: __________________________
Approved:

____
Michael Ward, Co-Supervisor

____
Eddy Malesky, Co-Supervisor

____
Daniel Stegmueller

____
Peter Hoff

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Department of Political Science
in the Graduate School of Duke University
2018

ABSTRACT

Two-Sided Matching Model

by

Anh Le

Department of Department of Political Science
Duke University

Date: ____________________________
Approved:

____
Michael Ward, Co-Supervisor

____
Eddy Malesky, Co-Supervisor

____
Daniel Stegmueller

____
Peter Hoff

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Department of Political Science
in the Graduate School of Duke University
2018

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Abbreviations

EM	Expectation Maximization.
FDI	Foreign Direct Investment.
IPA	Investment Promotion Agency.
IPE	International Political Economy.
MCMC	Markov Chain Monte Carlo.
MH	Metropolis-Hastings.
MLE	Maximum Likelihood Estimation.
MNC	Multinational Corporation.
MVN	Multivariate Normal.

1

Introduction

In recent decades, the global flow of foreign direct investment (FDI) has steadily increased, rising from almost nothing in the 1970s to over $2.3 trillion dollars in 2016 and becoming the an important source of global capital. For developing countries, FDI flow is remarkably robust to global downturns, leading to enthusiastic endorsement by major international organizations as a key factor to economic development (Figure 1.1) (Mallampally and Sauvant, 1999; World Economic Forum, 2013). This claim is also shared widely within the International Political Economy (IPE) literature, where much of the literature starts with the assumption that countries seek FDI for its many benefits. These works focus on *how* countries can attract FDI, and do not question *why* they want to do so (Jensen, 2003; Li and Resnick, 2003; Li, 2006; Ahlquist, 2006).[1]

In addition to bringing capital to and creating jobs in the host economy, FDI holds an important promise that is the spillover of productivity from foreign to domestic firms. As well-known from neoclassical growth theory, diminishing returns

[1] Two recent exceptions are Pinto (2013); Pandya (2016), which are the first to examine countries' demand for FDI.

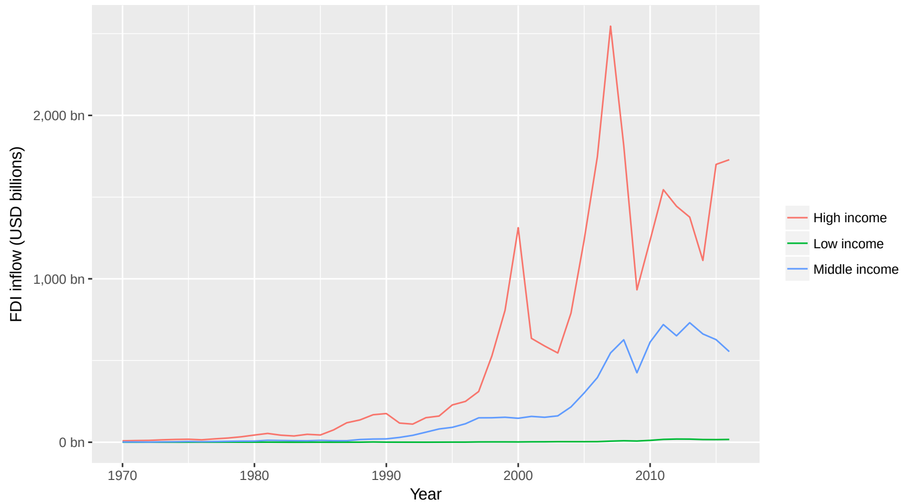

FIGURE 1.1: FDI global inflow, 1970-2006. The last four decades witness the explosive growth of FDI into the most important source of global capital. Source: World Bank's World Development Indicators.

to capital will at one point stop capital from accumulating further, preventing long-run economic growth from being driven by capital accumulation alone (Solow, 1956). Findlay (1978)'s groundbreaking model of FDI and growth shows how FDI can counteract this dynamic. In this model, technology spillover from foreign firms shift the domestic factor-price frontier to the right, allowing more output from the same input, ultimately resulting in a continually increasing capital stock for the domestic sector. In this view, FDI is welfare-enhancing, providing spillover benefits to local firms in ways that foreign firms do not take into account in their private calculations. This claim about the positive effect of FDI provides a justification for countries' using investment policies to rectify the undersupply of FDI.

However, this prevailing view highlighting the various benefits of FDI overlooks the fact that not all FDI are the same, and its effects are highly conditional. For example, there is no conclusive evidence of FDI having a positive effect on growth (Nair-Reichert and Weinhold, 2001; Carkovic and Levine, 2002) or poverty reduction (Guerra et al., 2009), prompting a substantial literature on how the growthenhancing and spillover effect of FDI is conditional on the absorptive capacity of the host economies, i.e. its level of human capital, technological sophistication, and financial market development (Durham, 2004; Nunnenkamp and Spatz, 2004; Fu, 2008; Willem, 2004). In addition, while the capital brought and jobs created by FDI are unconditionally good for the overall economy, its distributional effects different constituencies in the host economy are also conditional, varying across both sectors and geographies (Chintrakarn et al., 2012; Goldberg and Pavcnik, 2007; Nunnenkamp et al., 2007)

If the effect of FDI is so conditional, the earlier assumption in the IPE literature that countries all want FDI in the same way. This dissertation's goal is I want to bring the politics back in. The political science literature on FDI has focused largely on how politics shapes the flow of FDI across countries, while I want to study what are different countries' preference for different types of FDI. In order to do so, I have to solve three long standing issues in the literature.

This dominant approach in the literature has three long-standing issues that my paper will address. First, the majority of the literature relies on FDI stock and flow data as the outcome of interest even though they are often not an appropriate measure for the scale of MNCs' activities (Kerner, 2014). While it would be ideal to use firm-level data instead, both the lack of cross-national firm-level data and a suitable statistical model have posed a challenge.

Second, while there has been much focus on MNCs choosing host countries, the literature has largely neglected the other side of the investment decision: what are countries' preferences regarding MNCs? Consider the established finding that democracies receive more FDI. Without controlling for countries' preferences, it is difficult to interpret this fact as democracies actively pursuing MNCs or as MNCs finding democracies attractive. Not only are countries' preferences central to the modeling of investment decision, arguably it is also more steeped with politics and deserves more attention. Pinto (2013) and Pandya (2016) are two pioneering works in this area of research, proposing partisan politics and regime types as factors shaping countries' preferences for FDI. However, while their theories are ground-breaking, the empirical estimation of countries' preferences remains difficult.

Third, in addition to empirical issues raised above, I propose that we need to theorize about countries' preferences for FDI quality. While the political science literature has largely focused on the quantity of FDI, national policies and discourses pay much attention to the quality of FDI, using various incentives and restrictions to target certain types of FDI. Indeed, MNCs come with varying capital, demand for labor, and technology, all of which have different effects on the host country's economy. For example, policy makers and scholars have highlighted high-tech MNCs as a source of technological transfer for developing host countries, allowing them to upgrade their technical capacity and improve their productivity (Findlay, 1978; Nunnenkamp and Spatz, 2004). While such high-quality FDI has been enthusiastically endorsed by the development community, I argue that only governments with a long time horizon want to attract high-tech FDI because technological transfer takes time to pay off.

In sum, the current literature would benefit from an analysis that is capable of using firm-level data to estimate both firms' and countries' preferences for each other's characteristics. To accomplish this goal, I adapt the two-sided matching model originally designed for the labor market and the marriage market. In this model, both firms and countries evaluate their available options and choose the best according to their utility functions. As in many social science contexts, we only observe the final firm-country matches and not the full set of available options (also known as the opportunity set). I solve this problem by using the Metropolis Hastings algorithm, a Markov chain Monte Carlo (MCMC) approach that repeatedly samples new opportunity sets and rejects them at an appropriate rate to approximate their true distribution. Since the two-sided matching model is derived explicitly from actors' utility functions, their parameters also enjoy a straightforward interpretation in the utility space instead of some aggregate outcomes.

The paper proceeds as follows. Section 1.1 discusses the three long-standing issues with the literature and how they can be improved. Section **??** lays out the utility structure in the two-sided matching model and describes the matching process. Section 6.2 shows an application of the model on a census of Japanese firms overseas. Section presents the result. Section 6.6 concludes.

1.1 Three Issues in the Literature of FDI's Political Determinants

Using flow: (Jensen and McGillivray, 2005): democracy and federalism -¿ more FDI flow (panel data), (**?**): tax rate does NOT matter for FDI flow (OECD), (Busse and Hefeker, 2007)

(I use manufacturing FDI, so most capital in manufacturing is fixed) (because of the the "obsolescing bargaining" argument.)

1.1.1 Measuring MNCs' Activities

As **?** argues, the political economy literature on FDI is a bit of a misnomer. Political scientists are rarely interested in FDI *per se*—rather, they are interested in the activities of MNCs, which in turn, affect other important issues such as nation-state autonomy (Mosley, 2005), economic development (Moran, 1998), labor standards (Mosley and Uno, 2007), and environmental policies (Prakash, 2007). However, while the theories involve MNCs as the central actor in the causal mechanism, the empirics often use FDI flow as the independent variable of interest. These two concepts—the level of MNCs' affiliate activities in a country and FDI inflow into a country—are not the same.

Consider the definition of FDI from UNCTAD, the main producer of FDI data widely used by researchers:

> FDI has three components: equity capital, reinvested earnings and intra-company loans.
>
> - Equity capital, i.e. the foreign investors purchase of shares of an enterprise [in the host country].
> - Reinvested earnings, i.e. the foreign investors share ... of earnings not distributed as dividends by affiliates, or earnings not remitted to the foreign investor.
> - Intra-company loans between direct investors and affiliate enterprises.
>
> (UNCTAD, 2007, 245)

In essence, FDI data captures the amount of capital that crosses border. It is a poor proxy for the scale of MNCs' activities in the host countries because it overlooks important components of MNCs' activities while including components that are only relevant for balance of payment statistics (Beugelsdijk et al., 2010).

Consider the argument that FDI is the driver for the diffusion of labor standards across countries. Mosley and Uno (2007) theorizes that FDI can have this effect through three channels. First, MNCs may pressure the host governments for better rule of law and social programs. For MNCs to be able to effectively pressure the host governments, they must prove themselves valuable to the government by providing jobs or tax revenues. Both of these factors are only tenuously related to the amount of foreign capital inside the host country. Indeed, a MNC can employ thousands of employees, pay millions in tax, but show up as a net 0 on FDI flow data because the profit is repatriated to the foreign investor or through intra-company loans.[2] The size of the MNCs' operation is further understated because FDI statistics does not take into account capital raised locally. It also does not take into account the productivity of MNCs, which acts as an important multiplier when translating the amount of capital to the amount of output.

Second, scholars argue that MNCs may bring along best practices for workers' rights and spread it to local firms. If this channel operates via competition as MNCs provide better working condition forcing local firms to compete, then MNCs must employ a lot of labor for this effect to be noticeable. If this channel operates via demonstration, then it must form a lot of linkages with local firms, as suppliers and buyers, for the diffusion of norms to happen. Both the size of the labor force and the type of linkages with the local economy are not captured by FDI flow statistics.

Third, scholars argue MNCs may care more about labor quality than its cost, and thus may invest in higher wages, better benefits, or more training. Once again, for this effect to be noticeable, the size of the MNCs' labor force matters, its industry, and its investment in productivity, matters a lot more than how much capital it brings in and out of the country. In addition, non-equity transactions between the foreign parent company and the local subsidiary are not counted in FDI flow statistics, such as transfer of knowledge, technology, and management practices, thus excluding a component that is arguably much more important to labor quality than the amount of capital.[3]

This mismatch may also be a reason behind the still unsettled debate on the

[2] The issue of intra-company loans is particularly fraught with issues because companies very frequently use intra-company loans to get out of paying tax in a country. These loans will be recorded on the book as a massive outflow, even though the MNC still has a large presence on the ground.

[3] These issues are not isolated to studies of FDI and labor standards, but are common to the whole IPE literature of the effect of FDI on policy convergence, such as environmental policies (Prakash, 2007).

effect of FDI on poverty reduction. Scholars have theorized that FDI can lead to economic development and through three channels. First, MNCs can simply provide cheaper and better goods by being more productive. Second, MNCs may improve the productivity of local economy through technology transfer. Finally, MNCs can bring tax revenue to the host government, which can then spend on the poor via investment into social programs. Once again, these causal mechanisms only work depending the scale and the type of MNCs activities in the host country, not on the amount of equity capital that crosses the border. For example, productivity spillover is highly conditional on how thick the linkages between the MNCs and the local suppliers are as well as how technologically advanced the MNCs' activities are on the ground. The effect of FDI via tax revenue is particularly fraught with issues, as MNCs frequently engage in transfer pricing to get out of paying tax, especially via intra-company transactions of goods and services, such as charging for internal IP, whose price can be set arbitrarily by the firm (Malesky, 2015), which are not recorded in FDI flow statistics, [4]

What about studies that use FDI as the dependent variable, and are thus perhaps interested in flow of capital in and of itself?[5] The vast majority of theories on the political determinants of FDI flow relies on the "obsolescing bargain" model. Originally developed by Vernon (1971), the model is so named because the bargaining dynamics between the MNC and the host government changes over time, initially favoring the MNC and gradually tips towards the host government as the MNC commits more fixed capital on the ground. Indeed, knowing that it is costly for the MNC to uproot its increasingly large and immobile operation, the host government

[4] A similar argument is about the relationship of FDI on economic development, especially on the technology spillover and tax revenue.

[5] Arguably, political scientists are not interested in the flow of capital in and of itself, but also because of its implications for development, state autonomy, and other effects on policy. The discussion above has shown how problematic it is to study these effect of FDI using FDI flow data.

can unilaterally alter the original bargain, most egregiously by expropriating the MNC's asset and profit, but more often via "creeping expropriation," e.g. increased tax or tougher regulation (Li, 2009). Political economists argue that MNCs are acutely aware of the "obsolescing bargain," and thus prefer to invest in countries whose governments can make a credible commitment that they will not alter the original bargain. This means MNCs prefer countries with democratic accountability (Jensen, 2003), a federal system (Jensen and McGillivray, 2005), membership in international trade agreements (Büthe and Milner, 2008), less political risk (Beazer and Blake, 2011; Graham, 2010), or more veto points (Choi and Samy, 2008).[6].

The linchpin of this argument is the assumption that FDI capital is illiquid and cannot be quickly removed from the host country at will. This assumption is not fully warranted. According to the US Bureau of Economic Analysis (BEA)'s 2004 survey, 43% of US MNCs' balance sheet comprises of liquid assets that can be liquidated within one year under normal operating situations. Among the 57% of the balance sheet that are illiquid, 24% are "other non-current assets," which include non-tangible assets like brand names, trademarks, and patents—some of which are not expected to be liquidated but can be removed from the host countries. Only another 24% of the balance sheet is made up of physical capital, i.e. Plant, Property, and Equipment (PPE), which cannot be easily moved and match most closely to what we have in mind as the "illiquid capital" in the obsolescing bargain model (Kerner and Lawrence, 2014, 113).

Besides the conceptual mismatch between FDI flow and MNCs' activities, from a statistical standpoint, this measurement error may also be a contributing factor to the still unsettled debate on the effect of FDI. Even when the measurement error is random, it will inflates the standard error of our estimate when FDI is the dependent

[6] The fact that FDI is understood as illiquid capital subject to the obsolescing bargain is the central theoretical difference between FDI and footloose equity capital (Ahlquist, 2006; Mosley and Singer, 2008)

variable, and bias our estimate towards 0 when FDI is the independent variable. These effects may explain Jensen (2012)'s surprising finding that lower corporate tax rate does not lead to more FDI flow, or the mixed empirical evidence for the relationship between FDI and development (Mold, 2004, 108).

Even more worryingly, the measurement error is unlikely to be random. For example, the amount of locally raised capital, something we care about but FDI statistics does not capture, is likely to correlate with how developed the local capital market or the fluctuation in the exchange rate. On the other hand, repatriated earnings, something that does not necessarily indicate reduced MNCs' activities but is recorded as an outflow in FDI statistics, is likely to correlate with the tax rate of not only the host country but also the tax rate of tax havens that the MNC may have an affiliate in.[7] [8]

To deal with the measurement error problem, scholars have tried to use measurements that are closer to the theory than FDI flow. Given that political scientists are interested in MNCs' activities, recent work emphasizes using MNCs' operational data directly. These firm-level data allow researchers to measure directly the quantities of interest. For example, re-visiting Li (2009)'s hypothesis that democracies are more attractive to MNCs, Kerner (2014) uses data on US MNCs' fixed capital expenditures to more precisely test the relationship between democratic institutions and *illiquid* capital, not just FDI in general. The author finds that there is no relationship between democratic institutions and FDI flow and stock, but there is a positive relationship between democracy and MNCs' fixed capital expenditures, confirming

[7] See Gallop and Weschle (2017) for a recent and more comprehensive discussion of measurement error in political science research.

[8] FDI stock calculated at market value fluctuates based on market price, unrelated to firms' behavior. FDI stock calculated at historical value, which records asset value at the time it was acquired, is more stable and appropriate to measure the scale of MNCs' activities. Unfortunately, due to onerous data requirements, most countries measures FDI stock by simply adding up FDI flow across years. See Kerner (2014, 809) for a more in-depth discussion of FDI stock and flow data.

the theoretical argument [9]

In another example, Arel-Bundock (2017) uses ORBIS data to study the location decision of firms. However it only does one sided (i.e. only looking at the characteristics of the host countries to predict incidence of investment). This is a bit of a missed opportunities because using his Random Forest / non-parametric approach, it would have been possible to incorporate characteristics from firms. Then the random forest would be able to take into account the interactions between the firms' characteristics and country characteristics (in the form of sequential tree split). Even then, since random forests do not produce interpretable coefficients, this black-box approach does not allow us to understand the preference of actors, how these preference are correlated with other characteristics, and how they may evolve over time. The only claim he can make is whether some factors add predictive power over other factors.

In sum, while the need to use better data is clear, and while firm-level data has become more abundant in recent years,[10] political scientists have not developed a model to estimate this data appropriately. Given the data structure of a set of firms interacting with a set of countries, one may consider a dyadic-based analysis, frequently used in the International Relations literature. In such analysis, the unit of observation is a firm-country dyad, and the model used is typically OLS regression. Each dyad is assumed to be independent of each other, and any bias caused by interdependency is fixed via post-estimation procedures, such as clustered standard errors (Dorff and Ward, 2013).

Unfortunately, this dyadic approach is inappropriate to analyze MNCs' invest-

[9] Another alternative is to use other variable, for example when **?** re-examines whether MNCs favor democratic regimes because they pose less political risk, the author avoids using FDI flow and use price data of political risk insurance agencies instead. In other areas of IPE, scholars are also paying more attention to using the data that maps more closely to the theoretical argument, e.g. (Karcher and Steinberg, 2013).

[10] Examples of firm-level data include the US Bureau of Economic Analysis (BEA)'s survey of all US firms abroad, Tokyo Keizai's Overseas Japanese companies database (*Kaigai Sinshutsu Kigyou Souran*), World Bank's Enterprise Survey, and Orbis database of companies worldwide.

ment location. Once a firm chooses to invest in a country, it is by definition not investing in another. Therefore, the values of firm-country dyads deterministically constrain one another and cannot be modeled as independent draws from a common distribution.

The two sided matching model solves this problem by considering one firm-country match as the unit of observation. The intuition is as follows. If we observe that a firm is welcome to invest in countries $j_1, j_2, \ldots, j_n$ but ends up investing in country j^*, it must mean country j^* offers the highest utility to firms. Continuing the previous example, if country j^* has more veto players than average, we can infer that MNCs indeed prefer countries with more veto players.

1.1.2 Estimating Countries' Demand for FDI

Recognizing that our model of investment location has not taken into account countries' demand for FDI, Pinto (2013) and Pandya (2016) recently broke ground in this area. Similar to the rich IPE literature in trade and exchange rate, these studies argue that countries' demand for FDI varies according to FDI's distributive effect on their domestic constituencies (Broz and Frieden, 2001; Milner and Kubota, 2005). In this theoretical framework, labor supports FDI because foreign firms bring capital that increases the demand for labor and raises productivity, both of which lead to higher wage. On the other hand, domestic firms oppose FDI because foreign firms compete for local labor, inputs, and markets. Both Pinto (2013) and Pandya (2016) formulate their theories as a variant of this labor-vs-business tension, which surfaces in the former work as left-vs-right governments, and in the latter as democratic-vs-authoritarian regimes.

While these pioneering works have enriched our understanding of the relationship between politics and FDI, their empirical approaches do not satisfactorily measure countries' demand for FDI, leaving their theoretical arguments untested.

Consider Pinto (2013)'s approach, which controls for economic and institutional factors that affect FDI flow into a country. The author then claims that the country's openness towards FDI is what's left in the residual.[11] For this approach to be valid, every economic, institutional, and endowment factors that affect FDI flow have to be controlled for, leaving only the country's demand in the error term. This claim is much stronger than the regular assumption of exogenous and normally distributed error, which is valid as long as the omitted factors are uncorrelated with the independent variable of interest. Framed substantively, since the residual is likely to contain more than just the country's demand for FDI, if we observe an abnormally high level of FDI, we do not know whether it is because the country welcomes FDI or because MNCs find something attractive in the country.[12]

In contrast to Pinto (2013)'s statistical approach, Pandya (2014, 2016) substantively measures countries' demand for FDI, using the annual US Investment Climate Reports to code the number of industries that have foreign ownership restrictions or face investment screening. The advantages of this measurement are its ease of interpretation and its availability for many countries. However, two problems remain. First, adding up the raw count of restricted industries is not appropriate because industries are not all the same. For example, given the reach of the banking sector into all corners of the economy, a country's opening up its financial industry indicates much more FDI-friendliness than, say, allowing foreign furniture makers to set up shops. Since the theoretical argument is driven by FDI's distributive effect, it is

[11] Specifically, the estimation of FDI openness involves two steps. First, the author runs a gravity model explaining bilateral FDI flows, estimating the intercept as the host country-year fixed effect. Second, this fixed effect is then regressed on several economic and endowment factors of that country-year (i.e. GDP, GDP per capita, average school years, arable land). The residual in the second stage is considered the country's "FDI openness" in that year.

[12] In addition, the data requirement of bilateral FDI flows, ideally disaggregated by sectors, is very demanding. Therefore, this approach is limited to OECD countries only (Pinto and Pinto, 2008). During the period the authors study, 1980-2000, OECD countries accounted for 95% of global FDI outflow and 90% of inflow. However, since then the role of the developed world in global FDI has declined sharply, reduced to 60.8% of outflow and 40.6% of inflow in 2014 (UNCTAD, 2015).

suspect to ignore the varying impact of FDI across sectoral constituencies.

Second, according to the coding rules, an industry is coded as free if there is no mention of restriction. If an industry receives little FDI, it may not be worth mentioning as being restrictive and yet still coded as open. Therefore, "zero restriction" in the dataset can either mean that a country is very closed or very open to FDI. This concern is not hypothetical. Figure 1.2 shows that, following the coding of the US Investment Climate Reports, China seemed 100% open to FDI up until 1986 when it started imposing restrictions. The reality is the opposite. Prior to 1986, only limited FDI was allowed as joint-venture in Special Economic Zones (SEZ). The year of 1986 was, in fact, the first time China allowed any wholly owned FDI outside of SEZs.

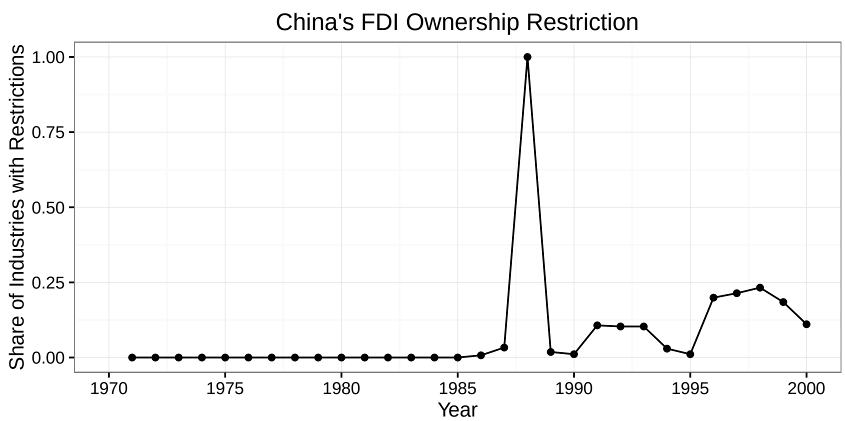

FIGURE 1.2: China's FDI Ownership Restriction, as coded in Pandya (2010). Prior to 1986, FDI in China was limited to few experimental Special Economic Zones, and thus not mentioned in US Investment Reports. The sharp spike in 1988 also does not seem to correspond to any actual change in policy, and likely another artifact of reporting. (See Zebregs and Tseng (2002) for a historical overview of China's FDI policy.)

The two-sided matching model circumvents these thorny measurement issues by incorporating countries' utility function directly into the model. If we observe that country j welcomes firms $i_1, i_2, \ldots, i_n$ to invest but not others, we can compare the characteristics of firms $i_1, i_2, \ldots i_n$ with the others to infer country j's preference.

1.1.3 Estimating Countries' Preferences for FDI's Technological Intensity

Laura Alfaro: is all FDI equal? - Use human capital from German firms as a proxy for all sectors (this works for the OECD sample of that paper, but not more generally) - Use IPA policy, but this could just be image building by the country (everyone says that they want advanced manufacturing (cite the picture))

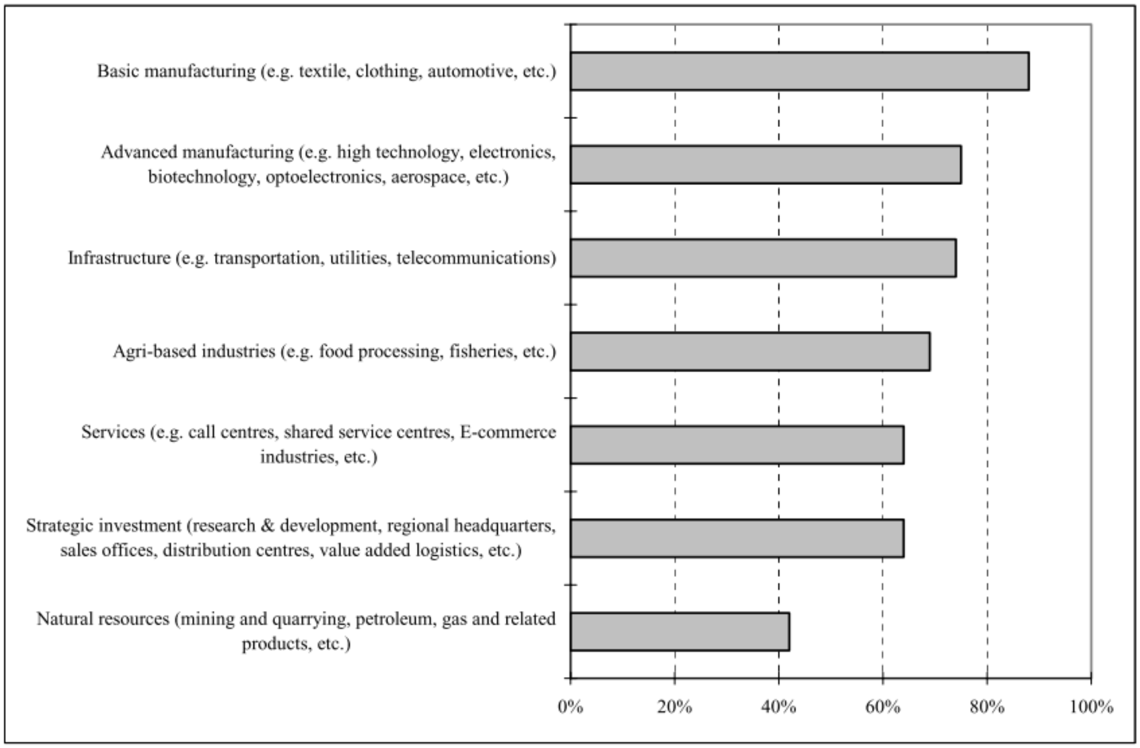

FIGURE 1.3: Target industries by IPA around the world. Because of the image building aspect of investment promotion, almost all IPAs say that they want to attract "advanced manufacturing." Therefore, using what is listed as investment priorities may not be a reliable way to measure countries' preference for FDI. Source: UNCTAD (2001)

While the political science literature has focused almost exclusively on the quantity of FDI, treating all FDI as one homogeneous flow of capital, policy makers seem to pay much more attention to distinguishing types of FDI. Commenting on the role of International Investment Agreements (IIAs), UNCTAD (2015) says, "Today, increasing the quantity of investment is not enough. What matters is its quality, i.e.

the extent to which investment delivers concrete sustainable development benefits." Governments in developing countries, from Ghana to China, all offer various forms of tax incentives and fee waivers to attract FDI that invests in a remote region, brings new technology, or focuses on exporting (Ricupero, 2000). Since 2006, China's official FDI policy has been "quality over quantity," promoting FDI with intense R&D in high-productivity sectors (Guangzhou, 2011). Indeed, for developing countries, the hope is that MNCs will transfer their technologies to the domestic economy by training workers or partnering with local suppliers.

Despite the importance of disaggregating FDI by its quality, data unavailability remains the bottleneck. The few existing attempts use detailed data from only one country or limit the sample to OECD countries (Alfaro, 2003; Alfaro and Charlton, 2007; Javorcik, 2004). With cross-country firm level data now available, we often have information on the firms' industry or even research and development (R&D) expenditure. With the two-sided matching model, I will be able to estimate countries' preferences for firms' technological intensity. I hypothesize that, since MNCs' technologies takes time to diffuse to local businesses, a country' preference of high-tech FDI is shaped by its time horizon.

Indeed, we see examples of various kind targeting and FDI policies.

For example, Ireland provides foreign investors with lower tax rate, lower land price, and cash grants for R&D that do not need to be repaid. China also offers a tax holiday (two years of no tax and three years of half the normal tax rate) in special economic zones to attract more foreign firms (Telford and Ures, 2001). We see the same widespread use of investment incentives in Southeast Asia (Fletcher, 2002). In Vietnam, provincial governments even defy the central government's directive and offer extra-legal incentives to FDI firms (Anh et al., 2007). Not only do these measures not work in attracting more FDI, they also deprive countries of revenues that could be spent on improving the local labor quality and investment climate—

factors that are much more conducive to spillover effect and growth.

2

Two-sided matching

Much of our social, economic, and political life is governed by two-sided matching markets. In these matching markets, actors from two disjoint sets evaluate the characteristics of someone on the other side and voluntarily form a match if both deem each other satisfactory.[1] Marriage is a prominent example of such matching process. Others include the matching between firms and workers, federal judges and law clerks, the *formateur* of a coalition government and other minority parties, or countries and multinational corporations (MNCs) that are looking for a location to invest.

Two-sided matching market is substantively consequential because it often involves scarce, indivisible goods, such as life commitment to a marital partner or political allegiance in a coalition government. It is also intellectually interesting because the market outcome depends on the actions of both sides, demanding a different analytical approach from what's used for one-sided markets.

This chapter will proceed as follows. First, I discuss the game theory literature

[1] Throughout the dissertation, I use "two-sided matching market" and "matching market" interchangeably. On the other hand, note that a two-sided market is not necessarily a matching market (Rysman, 2009).

of two-sided matching models, where much of the terminology and insight originate. I will highlight key results that are relevant to our goal of estimating actors' preference in matching markets. Second, I examine existing empirical studies of matching markets. Where existing studies have not taken into account the market's two-sided nature, I discuss how doing so can improve our understanding of the subject area. Where existing studies do model the two-sided dynamics, I discuss how they may or may not be used to study subjects that political scientists are interested in.

2.1 Game theory models of matching markets

Gale and Shapley (1962) was the first to study the matching market, using marriage as an example. In this market, there are two finite and disjoint sets of actors: men and women. Each man has preferences over the women, and vice versa. Each man's preference can be represented as an ordered list, ranking each woman based on how much he likes her.

The outcome of this market is a set of marriages, with none of some of people prefer to remain single. We call such a set of marriages a *matching* μ, which is a one-to-one function that matches a man with a woman. We refer to $\mu(x)$ as the *mate* of x. For convenience, we say that if an individual decides to remain single, they are matched with themselves.

We define a matching μ as *stable* if it cannot be improved by any individual or any pair of agents. A matching can be improved in two ways. First, an individual may prefer to remain single than to be matched with his or her mate $\mu(x)$ under the current matching μ. Second, a man and a woman may prefer to be with one another rather than whom they are currently matched with. Therefore, if a matching is stable, no one has a better option than their current situation.

The first key result from the game theory literature is that for any set of preference, there always exists a stable matching (Gale and Shapley, 1962). The proof is constructive, describing the "deferred acceptance" procedure that is guaranteed to produce a stable matching.[2] This result provides some justifications for us to assume that the matching we observe in real matching market is stable, and that the agents' utility cannot be further improved. Our empirical model of two-sided matching markets thus needs to describe a process that produces a stable matching.

While a central coordinator employing the *deferred acceptance* algorithm is guaranteed to come up with a stable matching, it is unclear whether decentralized markets, such as the labor market or the FDI market, would be able to reach this outcome by themselves.[3] The second key result from the game theory literature is that stable matching in decentralized matching market is indeed possible, even likely. For example, Roth and Vate (1990) show that, starting from an arbitrary matching, the market can converge to a stable matching with probability 1 if we allow random blocking pairs, i.e. two individuals that are not matched but prefer each other to their current match, to break off and form their own match. In addition, Adachi (2003) shows that a random search process, in which pairs of man and woman randomly meet and decide whether each other is better than their current mates, will

[2] The "deferred acceptance" procedure works as follows. In the first stage, every man proposes to his preferred mate. Every woman rejects all of her suitors except the one that she most prefers. However, she does not yet accept her (so far) favorite suitor, but keeps him along. In the second stage, every man that was rejected in the previous round proposes to his second choice. Every woman then picks her favorite from the set of new proposers and the man she keeps along from the previous round. The procedure continues until there is no longer any woman that is unmatched, at which point women finally accept their current favorite choices. (This procedure is called *deferred acceptance* to capture the fact that women defer accepting her favorite choice until the last round in case better options become available.) The resulting match is stable because, throughout the procedure, every woman has received all the offers that would have been made to her, and she has chosen her favorite among all of those offers. If there were any other man that she would prefer to her current match, that man would not have been available to her. Therefore, the final match cannot be further improved by any man or woman.

[3] The deferred acceptance procedure was used in the market for US medical residency with enthusiastic participation from medical students and hospitals. The high participation rate indicates that the matching produced is stable enough to entice students and hospitals away from arranging their own matches outside of the centralized market.

converge towards a stable matching if the search cost is negligible.[4] These results further suggest that the matching we observe in decentralized markets is likely stable. Therefore, our empirical model of matching markets to describe a process that produces a stable matching.

The third key result is that all conclusions regarding the one-to-one matching market (e.g. marriage) generalize to the many-to-one matching market (e.g. college admission, labor market), albeit requiring additional assumptions Roth and Sotomayor (1992). One important assumption is that firms treat workers as substitutes, not complements. In other words, firms never regret hiring a worker even if another worker is no longer available. Therefore, when we conduct empirical analysis of many-to-one markets, we should focus on markets where agents have such "substitutable preference." Otherwise, a stable matching is not guaranteed, agents' utility functions are interdependent, and it becomes unclear what kind of matching process our empirical model should approximate.

2.2 Empirical models of matching markets

The game theory literature takes the agents' preference as given and proves the existence of a stable matching. In contrast, empirical models of matching markets takes the observed matching as given and attempt to estimate the agents' preference.

Unfortunately, most extant empirical models fail to adequately account for the structure of a two-sided matching market. Often, researchers simply analyze the market from one side, e.g. estimating a firm's preference by looking at the type of workers it hires. This approach does not take into account the fact that a match depends not only on the agent's preference but also his opportunity. For example, a farm may prefer to hire highly-educated workers but cannot do so because highly-

[4] In this model, searching has a time cost. Thus, negligible search cost is modeled as agents having a time discount close to 1.

educated workers do not want to work on farms. Modeling this interaction between preference and opportunity is the key contribution of this dissertation.

Alternatively, some researchers measure agents' preferences by surveying them directly (Posner, 2001; Sprecher et al., 1994). While this approach circumvents the need to disentangle preference and opportunity, it can only measure agents' *stated* preference. In addition, such surveys require a high data collection effort while data on final matching (e.g. married couples, workers' current job, country location of MNCs) are widely available. This dissertation aims to make use of such available data to estimate agents' *revealed* preference.

Below I discuss existing empirical models of matching markets. First, I discuss two markets of interest to political scientists: the US federal clerkship market and the "market" for forming a coalition government. Researchers in both subject areas have not approached the problem with an empirical model that adequately captures its two-sided dynamics.

Second, I examine models from other disciplines that do take into account the two-sided dynamics of matching markets. I start with machine learning models applied to online marketplaces and dating sites. Then, I discuss the statistical models of the labor market (Logan, 1996) and the marriage market (Logan et al., 2008), which are most relevant to our goal of estimating agents' preference based on observed match data. These statistical models serve as the foundation of my empirical approach.

2.2.1 US federal clerkship market

In the US, graduates at top law schools vie for the best federal clerkships every year. These temporary, one-to-two-year positions are the launching pad for Supreme Court clerkships, prestigious teaching jobs, or employment at top law firms. On the other side, federal judges also compete for the best law graduates, who help reduce the judges' workload from copy-editing to drafting opinions (Gulati and Posner, 2016; Posner, 2001). Because the first clerkship tends to have an outsized ideological influence on law graduates, this matching market has important implications for the polarization of the judicial branch (Ditslear and Baum, 2001; Liptak, 2007).

The market for US federal clerkship has been noted as a classic case of a two-sided market. Clerks look for positions that provide not only prestige and connection but also comfortable quality of life (Posner, 2001). Judges select law graduates based on not only academic credentials but also, some argue, ideology, gender, and race (Slotnick, 1984). This market also suffers from strategic behavior emblematic of a matching market, such as offers being made aggressively early and with a short time to accept (Posner, 2001; Posner et al., 2007).

One approach to estimating the preference of agents in this market is to survey clerks and judges directly (Peppers et al., 2008). However, as discussed, this approach only measures stated preference, which is likely to suffer from social desirability bias when it comes to dimensions that we care about most such as matching based on ideology, gender, or race.

Other approaches estimate revealed preference by using observed hiring outcome. However, no existing study has properly taken into account the two-sided nature of the market, thus confusing the effects of preference and opportunity. For example, Bonica et al. (2017) use political contribution data (DIME dataset) to measure political ideology, then correlate the ideology of the hiring judge and the ideology of his clerks. This approach does not take into account the pool of applicants, leading to conclusions such as conservative judges hire more liberal clerks than conservative clerks (Bonica et al., 2017, 31). This curious finding has a potentially simple explanation: the pool of top law graduates tend to be overwhelmingly liberal, leaving conservative judges with no choice. Despite this issue, the authors proceed to measure judges' ideology by taking the average of their clerks' ideology. Without taking the pool of applicants into account, they may wrongly conclude that conservative judges are more liberal than they actually are.

In another approach, Rozema and Peng (2016) model the process as a discrete choice problem, in which clerks are differentiated products that Supreme Court justices select to maximize their utilities. Their model does not consider what clerks think about the offer because of their focus on Supreme Court clerkships, whose unparalleled prestige ensures that any offer made will be accepted. However, if we want to extend the model to the broader market of federal clerkship, such assumption is untenable.

2.2.2 The market for forming a coalition government

Besides election, government formation is the most consequential political process in determining which government people are subject to. Most extant studies of government formation are either game theoretic models or thick, “inside-the-Beltway” narratives. Potential advances can be made if we consider government formation as a many-to-one matching market, with the *formateur* party on one side and other minority parties on the other.[5]

A two-sided matching model of government formation would complement the game theory literature that models politicians as policy-seeking (as opposed to office-seeking) (Laver, 1998). When politicians are policy-seeking, parties have policy positions that can be modeled as their characteristics. Then, parties choose one another to form a coalition based on their policy positions, akin to men and women choosing one another to form a marriage based on their height or income.[6] As the game theory literature suggests, ideologically compact coalitions are more valuable because they

[5] The *formateur* party could be the one with the procedural power to set up the coalition, e.g. the incumbent party, or the largest party in established coalitions.

[6] In contrast, when politicians are office-seeking, the only coin of the realm is the number of legislative seats that a party controls. It determines both the inclusion of the party in the government and its portfolio allocation. In this framework, concepts like power indices and dominant parties are all about how parties can bring its controlled seats to a coalition to turn it into a winning coalition.

entail a smaller cost in terms of policy compromises (De Swaan, 1973). With the empirical matching model, we can test if parties do indeed prefer others that are ideologically close to themselves.

In addition, an advantage of the two-sided matching approach is its ability to consider multidimensional policy spaces. By considering a party's positions on various policies as their covariates, we would be able to estimate parties' relative preference for ideological proximity across policy dimensions.

2.2.3 The FDI market

To be introduced here, or kept until its own empirical chapter?

2.2.4 Recommender system for online two-sided markets

In recent years, the Internet underwent a proliferation of two-sided matching markets such as online marketplaces (e.g. AirBnB), dating sites (e.g. eHarmony), or job board (e.g. Elance). To help their users discover a match quicker, these sites often build a recommender system that suggests potential matches.[7] To maximize user engagement and profitability, these sites are incentivized to make recommendations that resemble a stable matching so that their users get the best match possible. And fo find the stable matching, they have to first estimate the preferences of their users.

While most of these algorithms are proprietary, some academic publications have addressed this problem. An interesting approach is the paper by Tu et al. (2014), which uses the Latent Dirichlet allocation (LDA) model to uncover the latent types of users based on their activities on an online dating platform.[8] In the original

[7] To clarify, the term "recommender system" typically refers to systems that recommend items to users based on the reviews of users like them. That is not what we are discussing here. Instead, we focus on matching markets where the recommender system recommends users to one another.

[8] Besides Tu et al. (2014), Hitsch et al. (2010); Goswami et al. (2014) are two other attempts to estimate users' preference in online matching markets. However, these papers take a simple one-sided approach, ignoring the interplay between preference and opportunity. Therefore, I don't discuss them further here.

application of LDA model in topic modeling, each document is a mixture of latent topics, and each topic is a distribution over words. In this application, each user is a mixture of latent "types," and each type is a distribution signifying relative preference over various mates' features. For example, the "outdoor type" may have higher preference for athleticism or dog ownership over other traits.

While the LDA model works well for the online dating market, it is not applicable to most social science problems for two reasons. First, this model requires data of users reaching out to multiple partners rather than just the final match. Second, while the LDA model uncovers users' latent types, most social scientists want to estimate the preference of specific, known types (e.g. how different regime types may prefer different characteristics of an MNC).

2.2.5 Two-sided models for the labor and marriage markets

To be introduced here, or in their own chapters?

2.3 Conclusion

Roadmap for the rest of the dissertation

3

Two-Sided Matching Model

Here I present a behavioral model of the two-sided matching market, focusing on the case of many-to-one matching, proposed by Logan (1996). For easier exposition, throughout the chapter I will use the example of the labor market, where many workers can be matched to one firm.

We assume that the matching process in the labor market happens in two stages. In the first stage, each firm evaluates each worker in the sample, deciding whether to hire that worker or not. At the end of this stage, each worker will have received a set of offers from firms, which we call his *opportunity set*. In the second stage, each worker evaluates the firms in his opportunity set and chooses the firm that he likes best. This constitutes the final, observed match between a worker and a firm. This is a many-to-one matching problem because a firm can make offers to multiple workers, none, some, or all of which can be accepted by workers.

Our model only needs data on 1) the covariates of firms and workers, and 2) the job that workers accept. Such data is widely available in many social science surveys of the job market. Importantly, we do not need to observe the opportunity set. Therefore, our model obviates the need to follow the matching process and record who makes offer to whom, which is rarely possible for researchers.

If we assume that firms and workers are utility-maximizing agents, at the end of the matching process, no firm or worker would voluntarily change their final matches. As discussed in Section 2.1, this property is called *stability* in the game theoretic two-sided matching literature. We want our model to have this property because matching markets tend to produce stable matching. Indeed, Roth and Sotomayor (1992) show that for any given set of preferences, a stable match always exist. Furthermore, Roth and Vate (1990) and Adachi (2003) show that a decentralized market with agents making independent, utility-maximizing decisions can also reach a stable match by itself.

This stability property does not imply that the matches will never change. Indeed, if actors' preference shifts, their characteristics change, or new actors enter the market, the matches will also change as a result of actors' recalculating their utility and adjusting their decisions. Therefore, since we are estimating actors' preference using only a snapshot of matching market, we are making the assumption that on a systemic level, the average characteristics of the actors and their preference remain sufficiently static for our estimates to be meaningful.

This chapter will proceed as follows. First, I discuss the utility model for how firms make offers to workers. Second, I discuss the utility model for how workers choose the best offer among those extended by firms. Third, I show how we can use a Bayesian MCMC approach to estimate the model. Fourth, I analyze US labor data and demonstrate how to interpret the model's result.

3.1 Modeling firms' decision making

A firm j's decision on whether to hire worker i rests on two utility functions. First, firm j's utility for hiring worker i is:

$$U_j(i) = \boldsymbol{\beta}_j' X_i + \epsilon_{1ij} \tag{3.1}$$

where β_j is a vector of firm j's preference for worker characteristics, x_i is a vector of worker i's measured values on those characteristics, and ϵ_{1ij} is the unobserved component that influences firm j's utility.

On the other hand, the utility of not hiring worker i is:

$$U_j(\neg i) = b_j + \epsilon_{0ij} \tag{3.2}$$

where b_j is the baseline utility of firm j, and ϵ_{0ij} is the unobserved component that influences firm j's utility.

Firm j will make an offer to hire worker i if $U_j(i) > U_j(\neg i)$. Relevant worker characteristics (i.e. X_i) that a firm may consider are age, education, or experience. The corresponding β's represent the firm's preference for these characteristics.

This model makes two important assumptions about firms' hiring process. First, whether a firm decides to hire worker A depends on the characteristics of worker A alone, and it will continue to hire worker A even if worker B is no longer available. In other words, firms regard workers as substitutes rather than complements.[1] This assumption is not universally true. A Hollywood producer may want to hire two specific lead actors for their chemistry, and if one is unavailable, the other also has to be replaced. However, for large firms where workers are closer to swappable cogs than unique superstars, this assumption is reasonable.

Second, the model assumes that the utility of hiring a worker does not depend on how many other workers accept the offer. In other words, the firm is large enough

[1] In the terminology of Roth and Sotomayor (1992), firms are assumed to have "substitutable preference," or firms' preference is assumed to have the property of substitutability. As discussed in Section 2.1, this assumption is necessary to prove the existence of stable matching in the case of many-to-one matching.

to employ all the workers to whom it extends offer without feeling the effect of diminishing marginal productivity of labor. This assumption is less restrictive than it may seem. Indeed, we can model the fact that the workers under consideration are less productive than the previous batch of workers by allowing firm j to have a high baseline utility b_j. Therefore, we are not assuming that there is never any diminishing marginal productivity of labor, only that there is negligible diminishing effect between the first and the last of the workers under consideration. This assumption is a reasonable approximation if the firm's labor force is large compared to the number of workers being considered.[2]

In addition to the two above assumptions about the process of firm's decision making, we make three parametric assumptions that are standard in the discrete choice literature. First, we assume a linear utility function. Second, we assume that the error terms $\epsilon_{1ij}, \epsilon_{0ij}$ are uncorrelated with one another and across firms. Third, we assume that the as error terms $\epsilon_{1ij}, \epsilon_{0ij}$ follow the Gumbel distribution.[3] The choice of the Gumbel distribution is largely motivated by convenience since it allows us to derive the probability of firm j making an offer to worker i as the familiar binomial logit form:

$$\begin{aligned} Pr(o_{ij} = 1) &= Pr(U_j(i) > U_j(\neg i)) && (3.3) \\ &= Pr(\epsilon_{0ij} - \epsilon_{1ij} < \boldsymbol{\beta}_j' X_i - b_j) && (3.4) \\ &= \frac{\exp(\boldsymbol{\beta}_j' X_i)}{1 + \exp(\boldsymbol{\beta}_j' X_i)} && (3.5) \end{aligned}$$

[2] While not concerned with diminishing marginal productivity, Roth and Sotomayor (1992) also assume that firms' quota, i.e. the number of workers they can accept, is sufficiently large to hire everyone in the set of workers under consideration. This assumption simplifies the proof that a stable match always exists in the case of many-to-one matching.

[3] The Gumbel distribution is very similar to the normal, only with a slightly fatter tail that allows for slightly more extreme variation in the unobserved utility. Its density function is $\exp^{-(x+\exp^{-x})}$, with mode 0, mean 0.5772, and fixed variance $\frac{\pi^2}{6}$. In practice, the difference between using Gumbel and independent normal error terms is small (Train, 2009).

The term b_j is absorbed into $\boldsymbol{\beta}$ when we add an intercept term to the covariate matrix X.

Once firms have made their offers, each worker i will have a set of offers from which to pick her favorite. We call this set of offers the *opportunity set* of worker i, denoted O_i. Since unemployment is always an available option, every opportunity set includes unemployment as an "offer."[4]

The probability of worker i obtaining the opportunity set O_i is:

$$p(O_i|\boldsymbol{\beta}) = \prod_{j \in O_i} p(o_{ij} = 1|\boldsymbol{\beta}) \prod_{j \notin O_i} p(o_{ij} = 0|\boldsymbol{\beta}) \tag{3.6}$$

$$= \prod_{j \in O_i} \frac{\exp(\boldsymbol{\beta_j'} X_i)}{1 + \exp(\boldsymbol{\beta_j'} X_i)} \prod_{j \notin O_i} \frac{1}{1 + \exp(\boldsymbol{\beta_j'} X_i)} \tag{3.7}$$

3.2 Modeling workers' decision making

Worker i's utility for the accepting an offer from firm j is:

$$V_i(j) = \alpha' W_j + v_{ij} \tag{3.8}$$

where α is a vector of workers' preference for relevant characteristics of firms, W_j is a vector of firm j's measured values on those characteristics, and v_{ij} is the unobserved component that influences worker i's utility.

Worker i evaluates all the firms in her opportunity set and selects the offer that brings the highest utility. This decision of worker i concludes the matching process, resulting in the observed final match between a worker and her chosen firm in our data.

[4] In our model setup, firms and workers decide sequentially, with firms making offers first in order for workers to have opportunity sets to choose from. While firms and workers in real life certainly do not act in this sequential manner, the idea of the opportunity set is still applicable. Workers in the real labor market may not know their exact set of offers, but they can certainly guess which firms are within their reach based on their characteristics and on guesses about firms' preference.

We make two assumptions in modeling the worker's decision making. First, for simplicity, we assume that all workers share the same set of preferences–hence α does not have a subscript i. The model can be extended so that there is heterogeneous preference among workers, either by estimating a separate model for each worker type (i.e. no pooling) or by building a hierarchical model for worker preference (i.e. partial pooling).

Second, we assume that the error term v_{ij} are uncorrelated across j. In other words, the unobserved factors in the utility of one job offer is uncorrelated to the unobserved factors in the utility of another job offer.[5] This assumption is most likely not true: if worker i values some unobserved factors of an offer, she is likely to consider those same factors in another offer as well. The hope is that we have modeled the observed portion sufficiently well that the remaining unobserved factors are close to white noise. In any case, this issue afflicts any application of discrete choice models and is not unique to our setup.[6]

Similar to our model of firm's utility, our model of worker's utility has three additional parametric assumptions that are standard in the literature. First, we assume that utility is linear. Second, the error term v_{ij} are uncorrelated across i. Third, we model v_{ij} having a Gumbel distribution so that the probability that worker i will accept the offer of firm j out of all the offers in its opportunity set O_i takes the conditional logit form (Cameron and Trivedi, 2005):

[5] This assumption also gives rise to the Independence of the Irrelevant Alternatives (IIA) property. IIA implies that the relative odds of choosing between two alternatives depend only on the two alternatives under consideration. It does not depend on whether other alternatives are available or what their characteristics may be. Hence, other alternatives are considered "irrelevant."

[6] The discrete choice literature has developed solutions for such correlated error structure, such as nested logit, probit, and mixed logit, that can be applied here if we suspect that the unobserved portion is strongly correlated.

$$p(A_i = a_i | O_i, \alpha_i) = \frac{\exp(\alpha' W_{a_i})}{\sum\limits_{j:j \in O_i} \exp(\alpha' W_j)} \tag{3.9}$$

where a_i is the index of the firm that i accepts to to work for. Unemployment is indexed as 0.

3.3 Model estimation

Our goal is to estimate the preference of firms and workers, i.e. β_j and α. The key insight is that, conditional on the opportunity set being observed, the model of firms' and workers' decision making is a straightforward application of the binary logit and conditional logit model. Both models can be estimated with familiar tools like Maximum Likelihood Estimation (MLE).

However, in most social science research problems, the researcher only observes the final match A and not the opportunity set O. For example, labor market data typically does not include the set of offers a worker received (or would have received if she had applied), while data on her current job is widely available. Similarly in the marriage market or the FDI market, researchers often do not have the data on the offers being made, and only observe the final matching between men and women (i.e. who is married to whom) and between MNCs and countries (i.e. which factory is located where).

Logan (1998)'s solution to this problem is to use the Expectation-Maximization (EM) algorithm, an iterative method capable of finding the maximum likelihood estimates when the model depends on unobserved latent variables (i.e. the unobserved opportunity set in this case) (Dempster et al., 1977). Our innovation is to estimate the model using a Bayesian MCMC approach, which offers several advantages. First, our MCMC approach produces the full posterior distribution, making inference easy.

In contrast, EM only produces point estimates out of the box.[7] Second, our MCMC approach can be faster than EM when the latent variable, i.e. the opportunity set, is high dimensional (Rydén, 2008).[8] Third, within the Bayesian framework, we can naturally put a hierarchical structure on firms' preference. This allows us to borrow information across firms, producing more precise estimates even when there is not a lot of data for a specific firm.

The rest of this section describes how we conduct model estimation.

3.3.1 Estimating the model using Metropolis-Hastings

We are interested in the posterior distribution of workers' and firms' preference given the observed final match, i.e. $p(\alpha, \beta|A)$. Unconditioned on the opportunity set, this posterior is difficult to derive or sample from. Therefore, we instead sample from the augmented posterior $p(\alpha, \beta, O|A)$, whose density is much simpler to derive.[9] Specifically,

$$p(\alpha, \beta, O|A) = \frac{p(A|\alpha, \beta, O)p(\alpha, \beta, O)}{p(A)} \tag{3.10}$$

$$\propto p(A|O, \alpha)p(O|\beta)p(\alpha)p(\beta) \tag{3.11}$$

where $p(A|O, \alpha)$ is derived in (3.9), $p(O|\beta)$ is derived in (3.7), $p(\alpha)$ and $p(\beta)$ are prior distributions for α and β. A key insight of this equation is that the acceptance of offers, i.e. $p(A|O, \alpha)$, depends only on the opportunity set and on the

[7] Jamshidian and Jennrich (2000) propose a method for estimating the standard error of EM estimates. However, for hypothesis testing, we need further assumptions about the distribution of the EM estimates.

[8] Indeed, our opportunity set O is a $(I \times J)$ matrix of 0s and 1s, where I is the number of workers and J is the number of firms. Thus, there are 2^{IJ} potential values for the opportunity set, which quickly becomes untenable even for a small number of I and J. The high dimension of O forces Logan (1998) to reduce the data dimension by aggregating 17 employers in the data into 5 employer types, e.g. professional or blue collar jobs.

[9] See Tanner and Wong (1987) for a discussion of such data augmentation techniques.

workers' preference. Similarly, the opportunity sets, i.e. $p(O|\beta)$, depend only on firms' preference.

Because the opportunity set O is a discrete matrix of 0's and 1's, there is not any convenient conjugate model for (3.11), making Gibbs sampling impossible. Therefore, we use Metropolis-Hastings instead, a technique to sample from an arbitrary distribution $p(\theta)$ using the following steps:

1. Start from an arbitrary value of θ
2. Generate a proposal value θ^* from the proposal distribution $q(\theta^*|\theta)$
3. Calculate the acceptance ratio $MH_\theta = \frac{p(\theta^*)q(\theta|\theta*)}{p(\theta)q(\theta^*|\theta)}$
4. Accept the proposed value θ^* with probability $\max(1, MH_\theta)$
5. Repeat step 2-4 until convergence

In our case, we will use symmetric proposal distributions, i.e. $p(\theta^*|\theta) = p(\theta|\theta^*) \forall \theta, \theta^*$, so that the MH acceptance ratio simplifies to $MH_\theta = \frac{p(\theta^*)}{p(\theta)}$. In addition, because preference parameters tend to be correlated, we use an adaptive proposal distribution so that our MCMC samples have a faster convergence rate (Haario et al., 1999, 2001).[10]

Below we describe how to sample from the posterior of each parameter in the model using the Metropolis-Hastings (MH) algorithm. More detailed derivation of the Metropolis acceptance ratio is included in Appendix A. We ensure that our derivation and implementation of the acceptance ratio is correct using the unit-testing approach suggested by Grosse and Duvenaud (2014).[11]

[10] Description of the Adaptive Metropolis procedure?

[11] Describe the unit-testing framework to ensure the correctness of MCMC code?

3.3.2 Posterior of the opportunity set $p(O|A, \alpha, \beta)$

For each worker i, we propose a new value O_i^* by flipping random cells in the current value O_i from 0 to 1 and 1 to 0. Substantively, this is equivalent to perturbing the opportunity set by randomly making new offers or withdrawing existing offers. Note that this proposal distribution is indeed symmetric because proposing O_i^* from O_i and proposing O_i from O_i^* both involve flipping the same cells in the opportunity set. Hence, $p(O_i^*|O_i) = p(O_i|O_i^*) =$ the probability of selecting these particular cells out of the opportunity set.

The Metropolis acceptance ratio for the proposed opportunity set O_i^* is

$$MH_O = \frac{p(O_i^*|A_i, \alpha, \boldsymbol{\beta})}{p(O_i|A_i, \alpha, \boldsymbol{\beta})} \tag{3.12}$$

$$= \frac{\sum\limits_{j:j\in O_i} \exp(\alpha' W_j)}{\sum\limits_{j:j\in O_i} \exp(\alpha' W_j) \pm \exp(\alpha' W_{j*})} \times \exp(\pm \boldsymbol{\beta}'_{j*} X_i) \tag{3.13}$$

where $\pm$ evaluates to $+$ if j^* is a new offer being added to the current opportunity set, and evaluates to $-$ if j^* is an existing offer being withdrawn from the current opportunity set.

To understand the intuition behind this formula for MH_O, consider the scenario in which we propose a new opportunity set for worker i by adding an offer from firm j. Since worker i now has one more choice to choose from, it becomes less likely that worker i's accepted job is the best choice. This makes the proposed opportunity set less consistent with the observed data than the current opportunity set, and MH_O should decrease accordingly. This is reflected in the formula for MH_O by the $\exp(\alpha' W_{j*})$ term in the denominator.

On the other hand, whether we should add the offer to the opportunity set also depends on firm j's preference for worker i. If hiring worker i brings firm j net positive utility (i.e. $\boldsymbol{\beta}'_{j*} X_i > 0$), we should add the offer. This is reflected in the formula for MH_O by the multiplier $\exp(\boldsymbol{\beta}'_{j*} X_i)$, which is larger than 1 when $\boldsymbol{\beta}'_{j*} X_i > 0$.

3.3.3 Posterior of firms' preference $p(\alpha|A, O, \beta)$

At the beginning of the MCMC chain, we propose a new α^* using a Normal proposal distribution centered on the current value α with a hand-tuned diagonal covariance matrix. Later in the MCMC chain, the covariance matrix of the proposal distribution is adapted based on past samples to take into account the correlations across preference parameters (Haario et al., 2001).

The Metropolis acceptance ratio for the proposed α^* is[12]

$$MH_\alpha = \frac{\alpha^*|A, O, \beta)}{p(\alpha|A, O, \beta)} \tag{3.14}$$

$$\begin{aligned}\log MH_\alpha = \sum_i \Big[& (\alpha^* - \alpha)' W_{a_i} + \\ & \log \left(\sum_{j: j \in O_i} \exp(\alpha' W_j) \right) - \log \left(\sum_{j: j \in O_i} \exp(\alpha^{*\prime} W_j) \right) \Big] + \\ & \log p(\alpha^*) - \log p(\alpha)\end{aligned} \tag{3.15}$$

3.3.4 Posterior of workers' preference $p(\beta|A, O, \alpha)$

We propose a new $\boldsymbol{\beta}^*$ using a Normal, adaptive proposal distribution similar to α. Because $\boldsymbol{\beta}$ is high dimensional, with one set of β for each employer, in each MCMC iteration we randomly choose and update only a part of $\boldsymbol{\beta}$.

The Metropolis acceptance ratio for the proposed β is

[12] We log-transform the Metropolis acceptance ratio for better numerics.

$$MH_\beta = \frac{p(\beta * |A, O, \alpha)}{p(\beta|A, O, \alpha)} \tag{3.16}$$

$$\log MH_\beta = \sum_i \left[\sum_{j \in O_i} \left(\beta_j^{*\prime} X_i - \beta_j^{\prime} X_i \right) + \sum_j \left(\log(1 + \exp(\beta_j^{\prime} X_i)) - \log(1 + \exp(\beta_j^{*\prime} X_i)) \right) \right]$$

$$+ \log p(\boldsymbol{\beta}^*) - \log p(\boldsymbol{\beta}) \tag{3.17}$$

3.3.5 Posterior of β's hyperparameters μ_β, τ_β

As discussed above, the Bayesian approach to estimating our two-sided model allows us to put a hierarchical structure on the preference parameter. Here, we model firms' preference $\boldsymbol{\beta}$ as being drawn from the multivariate normal distribution $MVN(\mu_\beta, \tau_\beta)$, where μ_β is the mean and τ_β is the precision.

When the prior $p(\beta)$ is also normal, we have a conjugate multivariate normal model, where μ_β and τ_β are the parameters while β is considered the "data".

Since the model is conjugate, we can sample from the posterior of μ_β and τ_β with Gibbs sampling. Their full conditional distribution of μ_β is:

$$p(\mu_\beta) \sim MVN(\mu_0, \Sigma_0) \tag{3.18}$$

$$p(\mu_\beta|\beta, \tau_\beta) \sim MVN(m, V) \text{ where} \tag{3.19}$$

$$V = (\Sigma_0^{-1} + n\tau_\beta)^{-1} \tag{3.20}$$

$$m = (\Sigma_0^{-1} + n\tau_\beta)^{-1}(\Sigma_0^{-1}\mu_0 + n\tau_\beta\bar{\beta}) \tag{3.21}$$

The full conditional distribution of τ_β is:

$$p(\tau_\beta) \sim \text{Wishart}(\nu_0, S_0^{-1}) \tag{3.22}$$

$$p(\tau_\beta|\beta, \mu_\beta) \sim \text{Wishart}(\nu, S^{-1}) \text{ where} \tag{3.23}$$

$$\nu = \nu_0 + n \tag{3.24}$$

$$S^{-1} = \left(S_0 + \sum(\beta - \mu_\beta)(\beta - \mu_\beta)'\right)^{-1} \tag{3.25}$$

3.4 Parameter interpretation

The coefficients can be interpreted as the relative influence of a factor on the utility of the decision maker.

3.5 Results for US labor data

To be written ...

4

Simulation results

In this chapter, I simulate the matching process of a labor market, using real data on workers' and firms' characteristics and assigned model parameters. I then apply the two-sided logit model to show that the model is able to recover the underlying parameters and to diagnose the properties of the MCMC sampling. I compare the results of the two-sided logit model with the one-sided conditional logit model, showing that the one-sided approach produces biased estimates of workers' and firms' preference. This result demonstrates how the one-sided approach, despite being the default method for analyzing two-sided markets, can be misleading and unable to disentangle the effect of one side's preference from the other side's.

4.1 Labor market data

To ensure that my simulation result generalizes to real situations, I use real data on workers' and firms' characteristics from the US General Social Survey (GSS), 1982-1990.[1] On one side of the matching market is 2149 workers, a representative sample of US male workers between 25 and 44 years old. Table 4.1 shows the summary

[1] I thank Professor Michael Newton and Professor John Allen Logan for sharing the dataset.

statistics for workers. On average, a worker is 33 (± 5.7) years old and has 13 years (± 3.1) of education. 11% of workers in our sample are non-white.

Table 4.1: Summary statistics of workers' education, age, and race. The data come from the GSS, 1982-1990, for male workers in the US between 25 and 44 years old.

Statistic	N	Mean	St. Dev.	Min	Max
Years of education	2,149	13.103	3.111	2	20
Age	2,149	33.524	5.716	25	44
Non-white	2,149	0.113	0.316	0	1

On the other side of the matching market are five firms, representing five job categories: professional, managerial, sales / clerical / services, manufacturing blue collar, and other blue collar. Table 4.2 shows their characteristics. *Prestige* is the Hodge-Seigel-Rossi score, used in the GSS to measure the prestige of a job (Hodge et al., 1964; NORC, 2014). *Autonomy* is calculated as the odds of having a supervisor, multiplied by -1 so that a higher score is associated with a higher level of autonomy.[2] The prestige and the autonomy scores of a firm in our dataset is the average scores reported by workers who currently work in that job categories. Unemployment by itself does not have a prestige and autonomy score. Following Logan (1996)'s study on the labor matching market, I calculate them as 50% of the prestige of the last job held and as the average autonomy scores of all workers (Logan, 1996).[3]

[2] In other words, autonomy $= -\frac{P(\text{having a supervisor})}{P(\text{not having a supervisor})}$

[3] Our measurement of firms' prestige may be biased if workers that have a high regard for a certain type of job tend to work in that kind of job. Here we are making the assumption that workers are largely homogeneous so that workers in different jobs still have the same opinion regarding a job's prestige. While it is a strong assumption, it is not an extra burden since we have already assumed that workers have the same set of preferences in the modeling step. In addition, in other applications, such as the FDI market where MNCs are matched with countries, we can obtain objective data on countries' GDP, growth, and human capital, etc.

Table 4.2: Firms' characteristics

Firm category	Prestige	Pr(Supervisor)	Autonomy
Unemployment	18	0.204	-0.256
Professional	59.670	0.163	-0.483
Managerial	48.141	0.442	-3.237
Sales, Clerical, Services	34.545	0.100	-0.112
Manufacturing Blue Collar	34.330	0.071	-0.077
Other Blue Collar	34.035	0.175	-0.214

4.2 Simulated matching process

I assign values to workers' and firms' preference parameters, choosing values to achieve some level of realism and to have a sample of workers in each job. Table 4.3 describes the utility functions. I normalize firms' utility of not hiring to 0 so that firm j will extend and offer to worker i if the utility of hiring is positive (i.e. $U_j(i) > 0$. The magnitude of the intercept can thus be interpreted as how selective a firm is in making an offer. For example, professional and managerial firms are highly selective with intercepts of -24 and -22, while the other firms are less so with intercepts of $-9, -8$ and -6. The coefficients represent how much a firm values a worker's trait. For example, managerial and professional firms have a similar preference for a worker's education, with coefficients of 1 and 1.3. On the other hand, managerial firm values a worker's age twice as much as professional firm does, with coefficients of 0.2 and 0.1.

While the two-sided logit model can be extended to accommodate different worker types as well, for this simulation I assume that workers have homogeneous preference, sharing one utility function.

Each utility function has a random component, represented by the Gumbel-distributed error term ϵ. While normally distributed error is more common in simulations, justified by the claim that the error term is the sum of many independent unobserved variables, here I use the Gumbel distribution so that the coefficient estimates from the two-sided logit model can be directly compared with the true preference parameter values.[4] Practically the Gumbel distribution is very similar to the normal distribution, and I discussed the implication of using the Gumbel distribution in further depth in Section 3.1.

Table 4.3: Utility functions of firms and workers used in labor market simulation. x_{i1}, x_{i2}, x_{i3} are worker i's education, age, and race (nonwhite is coded as 1). w_{j1}, w_{j2} are firm j's prestige and autonomy, with $j \in \{1, \ldots, 5\}$. The ϵ's are Gumbel-distributed error terms.

Firms' utility functions	
Professional	$U_j(i) = -24 + 1.3x_{i1} + 0.1x_{i2} + 1x_{i3} + \epsilon$
Managerial	$U_j(i) = -22 + 1x_{i1} + 0.2x_{i2} + 1x_{i3} + \epsilon$
Sales, Clerical, Services	$U_j(i) = -9 + 0.75x_{i1} + -0.05x_{i2} + 0x_{i3} + \epsilon$
Manufacturing blue collar	$U_j(i) = -8 + 0.5x_{i1} + 0.02x_{i2} + 0x_{i3} + \epsilon$
Other blue collar	$U_j(i) = -6 + 0.5x_{i1} - 0.01x_{i2} + 1x_{i3} + \epsilon$
Workers' utility function	$V_i(j) = 0.01w_{j1} + 0.1w_{j2} + \epsilon$

With these utility functions, I simulate the matching process as follows:

- First stage: Each firm j evaluates each worker i, calculating the utility of hiring. If the utility of hiring is positive, it will extend an offer. Unemployment is always an option for workers, and is thought of as a "firm" that extends an offer to every worker. After this stage, we generate an opportunity set that is a 2149×6 matrix, in which cell (i, j) is 1 if worker i receives an offer from firm j and 0 if not. We typically do not observe this opportunity set matrix in real datasets.

- Second stage: Worker i evaluates each firm that extends him an offer in the first stage, calculating the utility of working for that firm. Worker i then chooses a firm (or unemployment) if it offers the highest utility. After this stage, we

[4] If I use normally distributed error terms, the coefficient estimates have to be divided by $\frac{\pi^2}{3}$ to be comparable with the true values.

generate a choice vector that is a 2149×1 vector, in which cell i equals j if worker i decides to work at firm j. This choice vector is what we observe in real datasets such as the GSS.

4.3 Simulation results

I estimate the two-sided logit model using the MCMC approach described in Chapter 3. I use $N(0, \mathrm{sd} = 10)$ as a weak and proper prior for the preference parameters. The results below are from a MCMC chain of 50,000 iterations with a thinning interval of 5, resulting in 10,000 saved iterations. To propose new samples in the Metropolis-Hastings algorithm, I use normal distributions with a scale hand-picked by examining the trace plots of discarded runs.

Figure 4.1 shows the trace plots of the MCMC samples for workers' preference. We see that the MCMC chain mixes well and quickly converge to the true value indicated by the red line. This result gives us confidence that the MCMC algorithm is implemented correctly and can achieve convergence within a reasonable time frame.

Figure 4.2 shows the trace plots of the MCMC samples for professional firm's preference. We see that the MCMC chain is also able to converge to the true parameter values, albeit slower and with more autocorrelation between iterations.[5] There are several reasons for this poorer mixing.

First, while we can use the entire sample to estimate the preference of workers, only a subset of the sample works at a particular firm, resulting in a smaller sample that we can use to estimate each firm's preference. This problem is clearest in the trace plots for the managerial employer, which only has a sample of 40 workers, or 1.9% of the total sample. To partially combat this issue, I use a hierarchical model

[5] To improve the mixing of the MCMC chain, I standardize workers' characteristics so that they have mean 0. Therefore, the intercept term has to be changed accordingly. The true intercept values displayed in the plots are the standardized intercepts, which is different from those reported in Table 4.3.

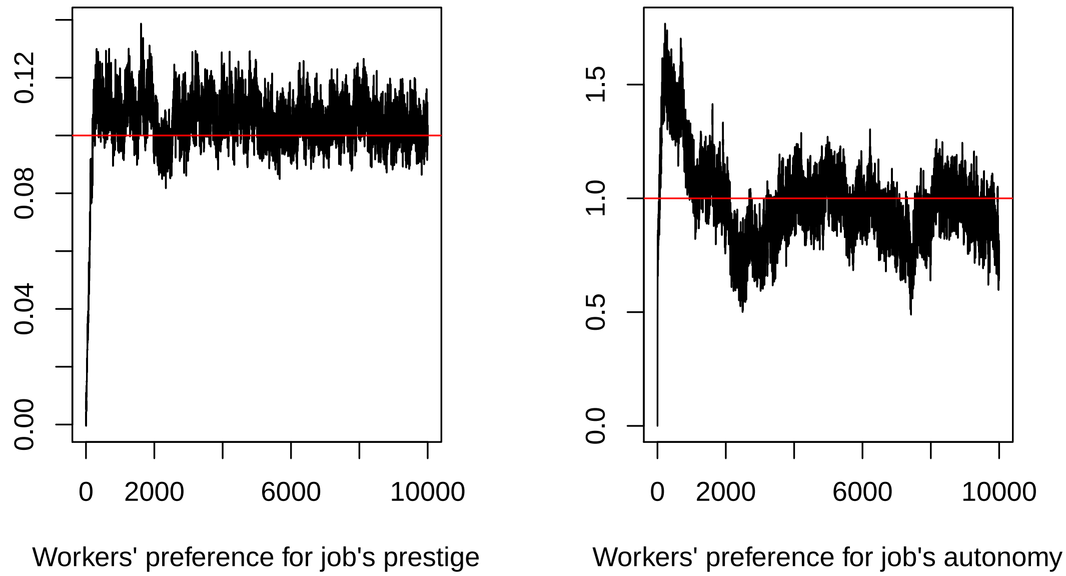

FIGURE 4.1: Two-sided logit estimates for workers' preference. The black line plots are the trace plots of the MCMC samplers, and the red line indicates the true parameter values. The trace plots show that the MCMC chain is able to converge to the true value after 10,000 iterations (2,000 saved iterations $\times 5$ thinning interval).

in which firms' preference parameters are drawn from a common distribution. By doing so, I "partially pool" the sample across firms, pulling the estimate for firms with small sample sizes towards the common mean, and thus producing estimates that have more predictive power (Gelman and Hill, 2006). On a related note, the MCMC chain of the preference parameter for *non-white* has a particularly poor mixing, likely because *non-white* is a binary variable, having less variation and thus information that our model can use.

Second, while workers only have two preference parameters (for firm's prestige and autonomy), each firm has four preference parameters (for worker's education, age, race, and an intercept term), resulting in a total of 24 parameters. Updating the MCMC chain in such high dimension is inherently difficult—to update one parameter

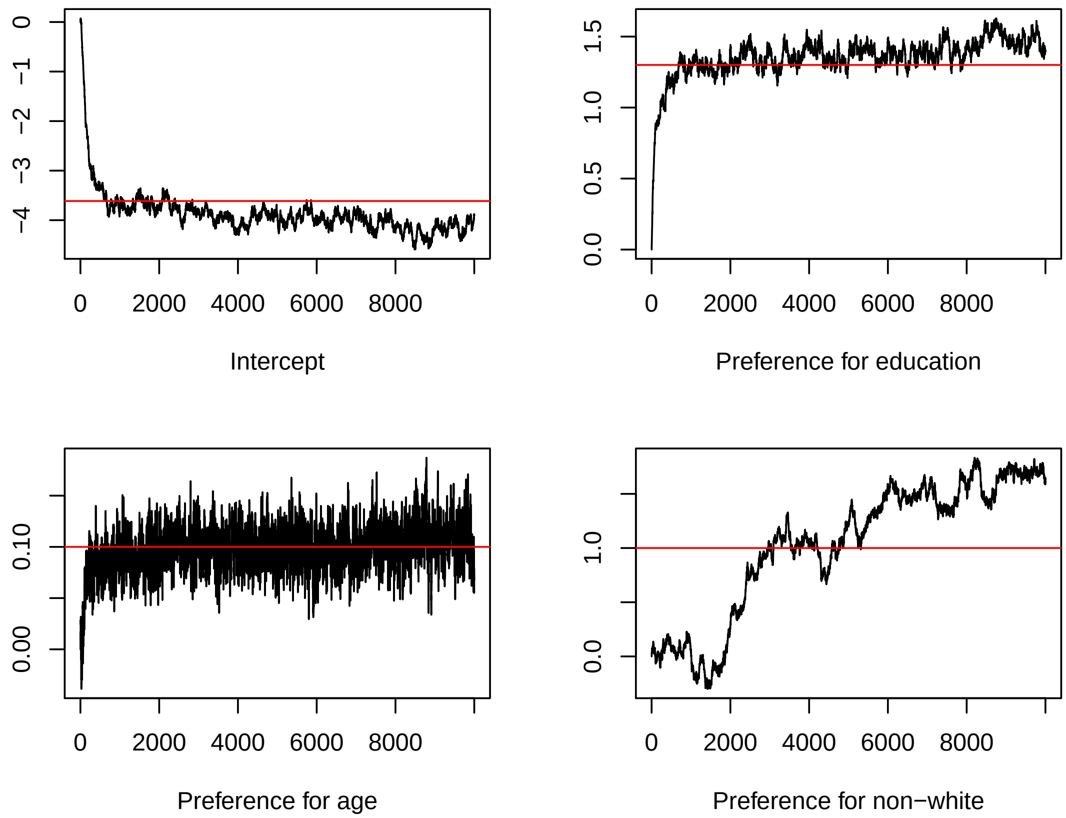

FIGURE 4.2: Two-sided logit estimates for professional firm's preference. The MCMC chain is able to converge to the true parameter value, indicated by the red line, albeit with more autocorrelation than the MCMC chain for worker's preference in Figure 4.1.

we only need to come up with one good proposal, but to update 24 parameters we need to come up with good proposals for each of them.

Third, while firms' preference and the opportunity set are highly correlated, our proposals for these parameters are independent, not take into their correlation, and thus causing the MCMC to get stuck at local mode. Section 4.5 discusses this issue and potential remedies in more details.

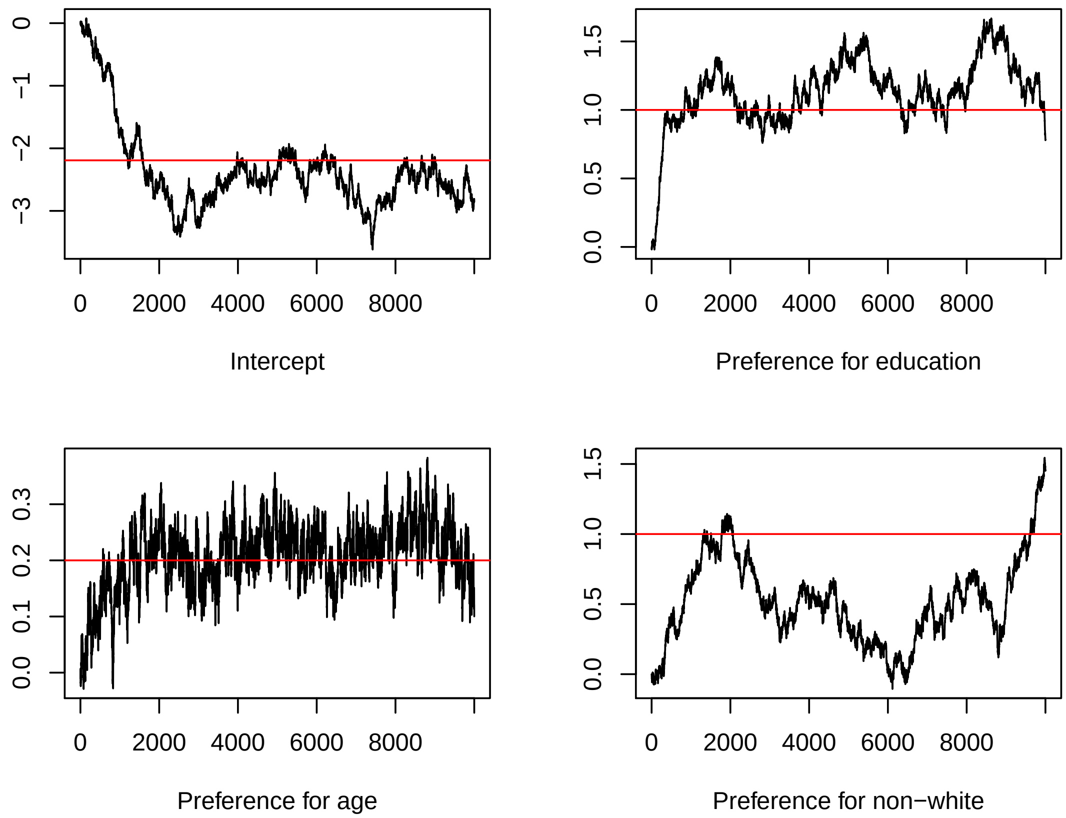

FIGURE 4.3: Two-sided logit estimates for managerial firm's preference. Because the managerial firm only has a small sample size of 40 workers, or 1.9% of the total sample, its preference is estimated more poorly than other's.

4.4 Comparing two-sided logit model and one-sided models

In this section, I demonstrate that, without taking into account the two-sided nature of the matching market, one-sided models produce biased estimates of the actors' preference. While it may be unsurprising that one-sided models fail when the data generating process is so different from their assumptions, this is a worthwhile exercise given that many empirical researches rely on these models. For example, using discrete choice models (multinomial logit, conditional logit), Cheng and Stough (2006) models Japanese MNCs' location choice across Chinese provinces and Aw and Lee (2008) models Taiwanese firms' decision to stay home or to open a factory in China and the US.[6] Using count models (Poisson, negative binomial), Wu (1999) models MNCs' location choice in Guangzhou, China. Arauzo-Carod et al. (2010) provides a literature review of how these methods are used in studying the location choice of firms.

I estimate a conditional logit model in which workers choose the best firm to work for as if all firms were available in their opportunity set. This assumption is not satisfied by our two-sided data generating process. Figure 4.4 shows that the one-sided conditional logit model is not robust when this assumption is violated, producing biased estimates of workers' preference. Worse yet, its estimate has little uncertainty and can cause researchers to be overly confident in the wrong result.[7]

Examining the big difference between the two-sided and one-sided estimates for *prestige* demonstrates a situation in which the one-sided approach confounds one side's preference with the other's. Figure 4.5 (left) shows the binary heat map for the true opportunity set—a dark blue cell indicates that an offer is made by firm in column j to worker in row i. The columns for professional and managerial firms are quite similar, reflecting the fact that they make offers to the same kind of workers. In contrast, in the observed choice (Figure 4.5, right), the columns for professional and managerial firms are very different, reflecting the fact that the professional firm is slightly more desirable, causing workers that receive offers from both firms to overwhelmingly choose to work for the professional firm over the managerial firm. Therefore, there are very few workers at the managerial firm. To the one-sided conditional

[6] In the empirical literature, researchers often use the term "multinomial logit" and "conditional logit" interchangeably to refer to a discrete choice model of unordered choices. In this discussion, I follow the terminology in McFadden (1974)'s seminal paper on discrete choice models, distinguishing "multinomial logit" as the model whose independent variables are the choosers' characteristics, and "conditional logit" as the model whose independent variables are the choices' characteristics.

[7] This conditional logit model is equivalent to a Poisson model in which the dependent variable is the count of workers at each firm, as shown in Guimaraes et al. (2003). Both models, estimated with MLE, would produce exactly the same estimates for the coefficients and their covariance matrix. Therefore, the argument against one-sided logit applies fully to Poisson.

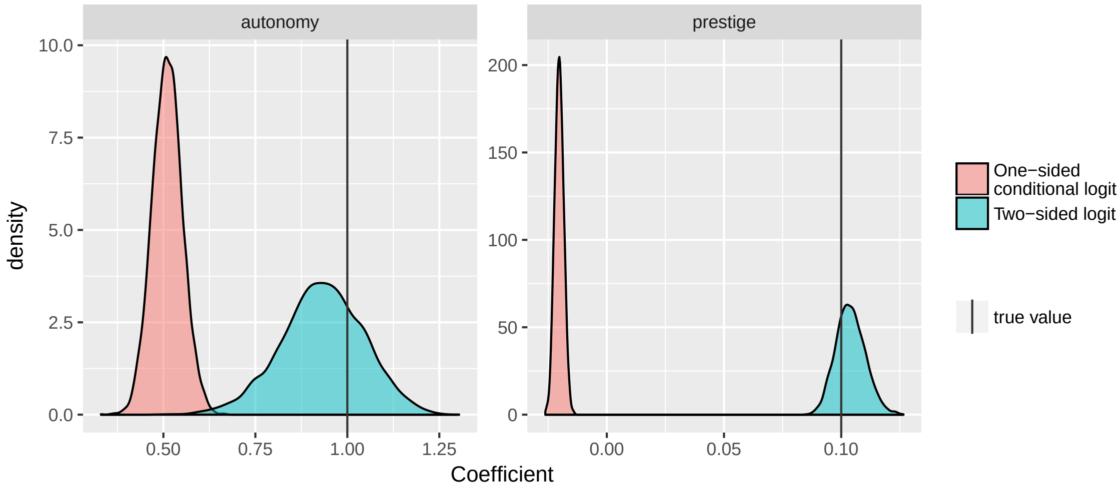

FIGURE 4.4: Estimates of workers' preference, produced by two-sided logit and conditional logit. The density plots show that the two-sided logit's 95% credible interval includes the true value, indicated by the black line, while the conditional logit's 95% confidence interval is far from it.

logit model, it looks as if the managerial firm—a highly prestigious job—were less desirable than even the services and blue collar firms. Therefore, it severely underestimate workers' preference for *prestige* to such an extent that *prestige* is considered a negative trait. This example shows how misleading it can be to estimate workers' preference by assuming that all the choices are available. Indeed, the managerial firm is rarely chosen not because it is undesirable, but because it has to compete with the professional firm for the same pool of highly educated and experienced workers.

4.5 Issues with MCMC convergence

The MCMC chain for firms' preference parameters β is poorly mixed because it is highly correlated with the opportunity set. Intuitively, at any point during the MCMC, we cannot propose a new opportunity set that is very different from the current one because it would be too unlikely given the current value of β. Likewise, we cannot propose too different a value for β because it would be rejected given the current opportunity set.

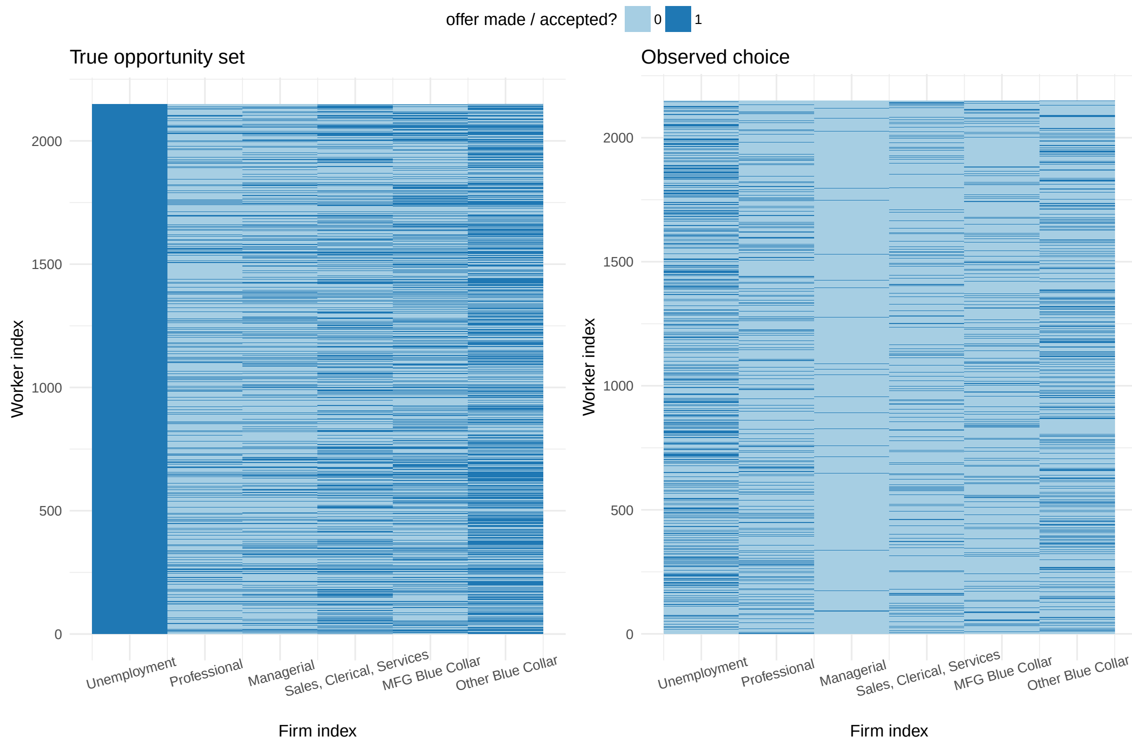

FIGURE 4.5: Binary heat map for the true opportunity set (left) and observed choice (right). A dark blue cell indicates that an offer was made (or accepted) between the firm in the corresponding column and the worker in the corresponding row.

This problem is especially severe for firms with only a few workers. If we propose a new opportunity set in which those firms now extends the offer to a new worker, then this new worker will heavily affect the estimate for β especially if it is different from the current workers. In contrast, for firms with a large sample size, there is already a lot of information to precisely estimate their preference. Making one new offer in these cases will not substantially change the estimate.

Currently, I make random-walk proposals for β and the opportunity set, which insufficiently takes into account this correlation, causing poor mixing. A potential solution to this problem is to make a correlated proposal for β and for the opportunity set: if we propose a new β that puts a high emphasis on workers' education, then we should also perturb the opportunity set to make more offers to highly-educated worker. While conceptually simple, this approach is not straightforward to implement, and is left for future research.[8]

[8] Alternatively, we may reparameterize the model entirely and eliminate the opportunity set, whose binary nature makes it impossible to use more modern MCMC approach such as Hamiltonian Monte Carlo. A potential alternative parameterization is Logan et al. (2008)'s, which samples directly from the utility space.

5

US labor market

In this chapter I examine the US labor market, where the two sided logit model is originally developed for. Check the results of our model against real data and look at the fit.

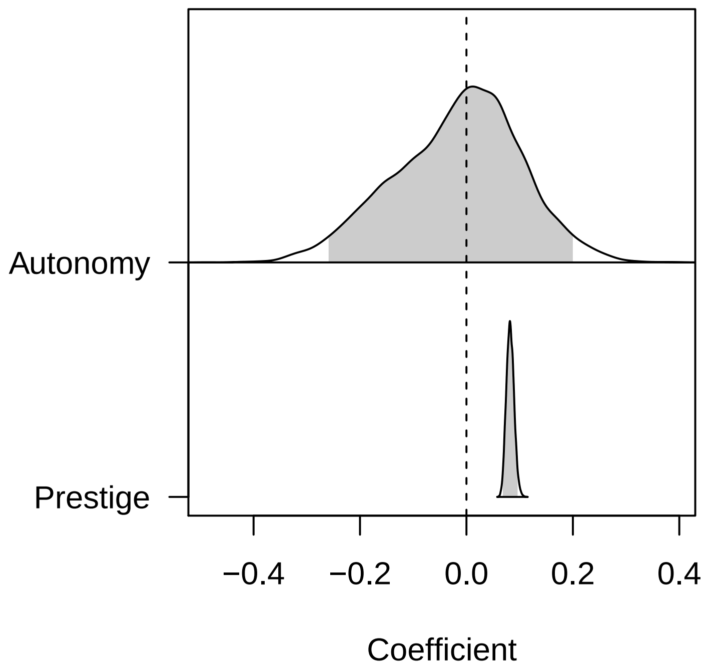

Figure 5.1: Preference of workers for firms' characteristics. The density plot and the shaded region show the posterior distribution and the 95% credible interval after burn-in. While prestige is highly valued by workers, autonomy seems to be less importance.

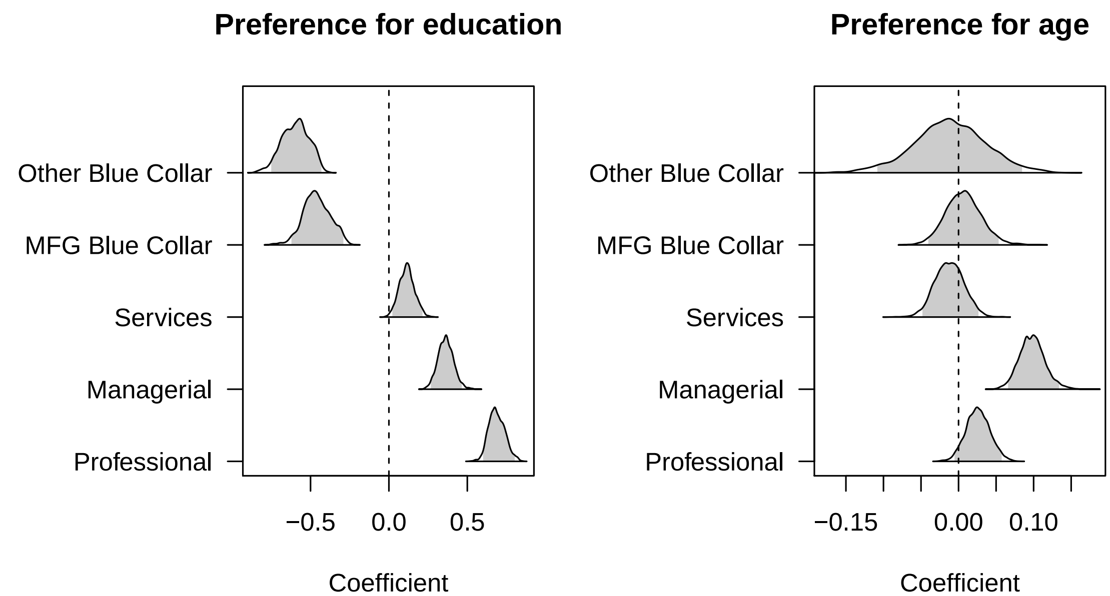

FIGURE 5.2: Preference of firms for workers' education and age. Professional and managerial firms have a strong and positive preference for highly educated workers. While most firms do not highly value older workers, except managerial firm stands out in their preference for age (likely as a proxy for experience).

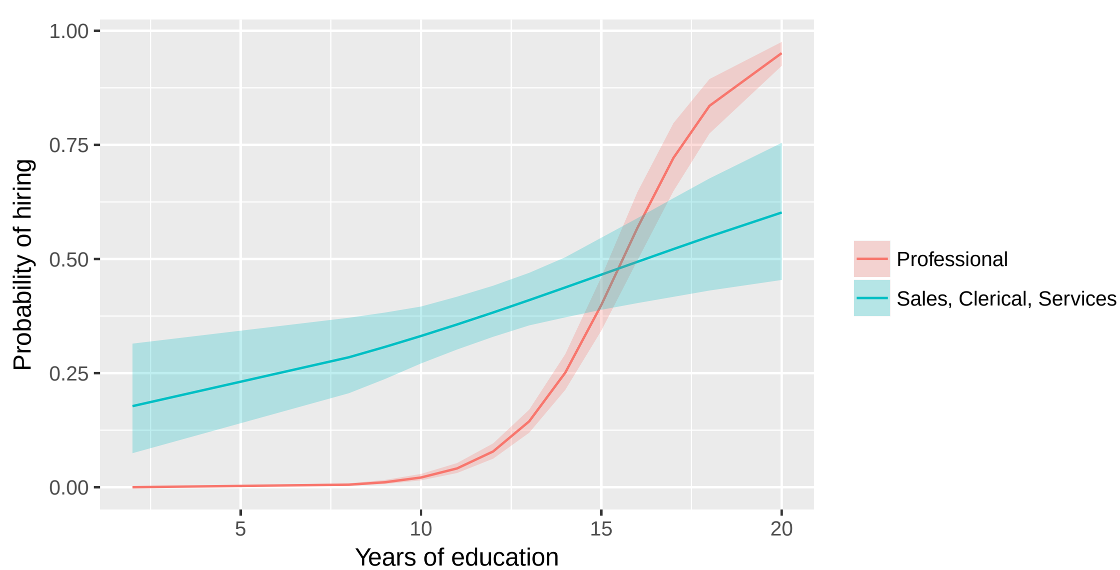

FIGURE 5.3: The effect of education on the probability of a worker being hired into a professional and a services job.

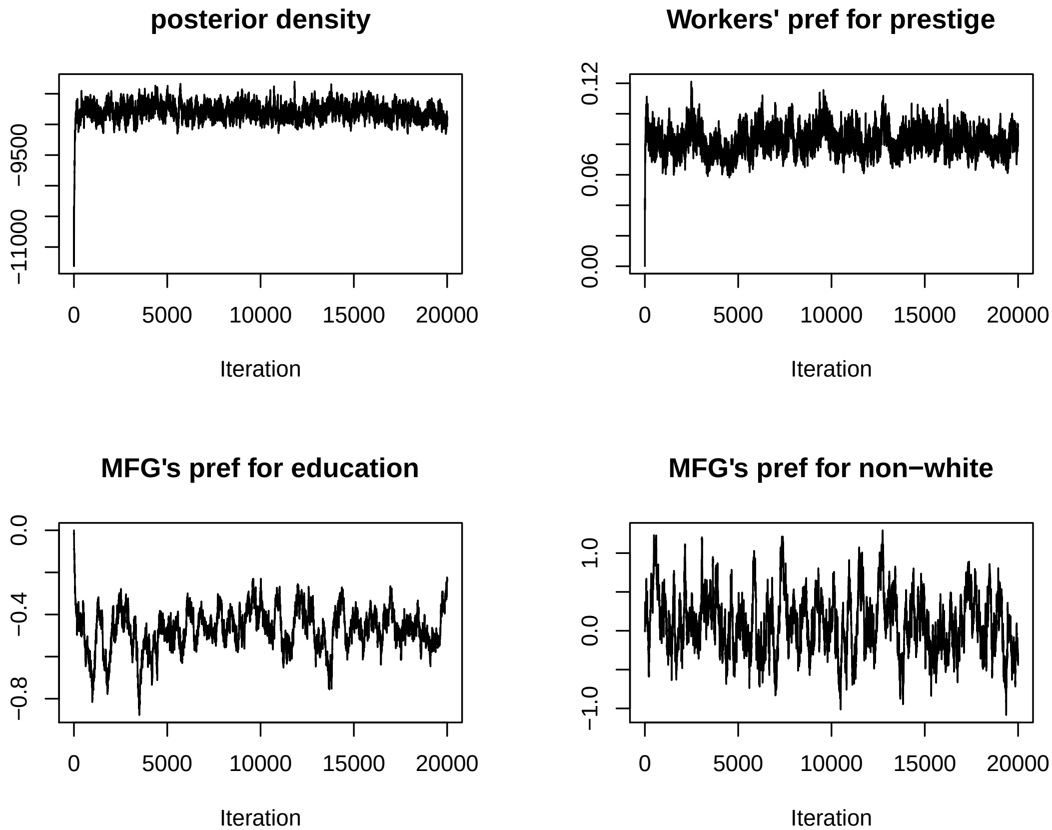

FIGURE 5.4: Trace plots of the posterior density and parameter samples, showing quick convergence.

6

FDI

6.1 The literature on Japanese MNCs investment

Make a graph about the size of Japanese FDI relative to Asia FDI here

Calculate the size of Japanese FDI into several countries (nominator from JETRO, denominator from IMF)

Relative exchange rate (Takagi2011). Assume that the capital market is imperfect, that external borrowers face a premium. So if a host country currency depreciates, inflow FDI will increase because the same amount of source currency can buy more input in the host country. The relationship is especially strong between exchange rate and inflow of FDI into industries with a lot of firm-specific assets. (e.g. Japanese firms buy a US firm with innovation for cheap, then use that innovation to improve its production back home in yen)

(Originally people don't think exchange rate matter because if you can buy an asset for cheap in the host country, when you repatriate the profit back to the home country it's a wash)

FDI is all about the relative factors (because a firm thinks about a country in terms of how that country is relative to its host). So the two sided matching framework makes sense

General FDI: Why don't firms export or license, but open their own plant overseas? Argument is that they have firm-specific asset that cannot be fully exploited otherwise. A licensee won't be able to exploit the entire asset, both sides can't agree on a price beforehand (This is the OLI framework, ownership-location-internalization, also the internationalization hypothesis). Empirically, firm specific asset is unobservable, so people use R&D intensity and advertising intensity instead. R&D does correlate strongly with multinationality.

6.2 Applying the Two-Sided Matching Model to Japanese MNCs

In this section, I apply the two-sided matching model to study the investment location of Japanese firms overseas. The data comes from a dataset compiled by Andrew Delios from *Kaigai Shinshutsu Kigyou Souran* (Japanese Overseas Investments-by Country), between 1986-1999 editions, a publication that contains information about the foreign affiliates of listed Japanese firms.[1] Tokyo Keizai, Inc. collect these data via annual surveys of the overseas operations of both listed and non-listed firms. This database is reputed to include all Japanese firms overseas (Yamawaki, 1991). (Delios and Keeley, 2001) compares the coverage of the Japanese Overseas Investment with other sources of publicly listed firms and found that 98.5% of public firms are included, which has 99.5% of the foreign subsidiaries. Since there is no public information on non-public firms, they cannot check the coverage of the Japanese Overseas Investment for these firms. However, the result for the public firms makes us confident that our data is close to the population of Japanese FDI overseas. This dataset has been used to study X, Y, and Z (Delios and Henisz, 2000). [2]

[1] I thank Professor Andrew Delios for generously sharing the data.

[2] There are similar datasets if scholars want to replicate this study in other context. On a global scale, the ORBIS dataset claims to have data of FDI firms across countries. (cite the paper that I reviewed). However, there are concerns about its data quality, given that the data is collected via

Sample choices:

- I only use the data of subsidiaries that are founded in year 1996 (which is different from the list of companies who are in existence in year 1996). Reasons: + the utility function is only modeled as a linear combinations of the country / firms covariates. It does not take into account the cost of uprooting a firm to move to another country once they are already there. This is important for FDI because, unlike equity investor, FDI are less foot-loose. Indeed, the fact that it is not footloose is an important quality of FDI that's appealing to countries (cite). The political economy literature on FDI has also derived its insights largely from this "obsolescing bargain" problem, so it's important that our model takes this into account. In past applications of the two sided matching approach, researchers use a random sample in time (i.e. a sample of all couples who's married in a certain year), which is fine, because the cost of leaving a job or leaving a partner may not be too onerous. However, for FDI, this fixed cost is more central. By limiting the sample to the firms who are founded in 1996, we examine their decisions as they are all looking for potential locations.[3]

+ the data is largest for this year. (some summary statistics here). There may be some concerns about this year being a special year, in the year leading up to the 1997 Asian Financial Crisis. But 1) we're only looking at manufacturing firms, not equity investors or land developers, 2) FDI during the crisis is largely the same as before the crisis (UNCTAD, 1998). Indeed, this is because FDI firms are largely

public governmental or municipal sources. Due to the differences of reporting across jurisdiction, the data quality is much less consistent than the Japanese Overseas Survey. For the US, there is a census of US firms overseas, which should be similarly high quality. However, this dataset requires citizenship. Tokyo Keizai, Inc. also continue to publish this data series. However, the cost is prohibitive and require understanding Japanese to work with.

[3] Of course, a subsidiary's foundation year in 1996 does not preclude the decision process to happen outside of this year. However, I consider this a reasonable approximation, given that many country characteristics, especially the political and institutional ones, do not change drastically within a window of several years.

looking at countries' fundamentals, such as labor cost, market potential, and thus not affected by the fluctuations in the financial markets. Essentially, the types of firms that invest before, during, and after crisis are still the same types of firms.[4]

+ I only consider Japanese FDI into East Asian and Southeast Asian economies to make sure that it's realistic to say that all these companies have the same preference parameters. (Pak and Park, 2005) finds that Japanese FDI in the West seeks to augment their global competitiveness, while Japanese FDI in the East focuses on exploiting their core competencies. Japanese FDI in the West are ones with oligopolistic power in their domestic market (so they are in a strong position to compete) and require R&D and marketing capabilities. They have different level of equities. This suggests that the two types of firms are fundamentally different.[5]

The final sample includes 6474 Japanese foreign affiliates in 2003, spreading across 37 countries, with China and the US leading as the two top destinations for Japanese MNCs (Table **??**).

For firms' characteristics that countries consider, I include:

- Capital size (in US$): A main argument for the benefit of FDI is that it brings capital to the country, improving labor productivity. MNCs' capital is especially important for developing countries, which cannot muster much domestic capital from their poor population. The capital size of a firm is included in the Japanese Overseas Business dataset.

[4] One potential concern is that our data does not capture the firms who thought about making an investment but decided not to. This could be a problem if the MNCs in our dataset is the most risk-seeking firms, then essentially we've only estimated the preference of the very risk-seeking or the very risk-averse firms. AFC too high short term interest, too high local currency that is fixed. However, the exit rate of Japanese firms in Thailand, the epicenter of the financial crisis, is the same, indicating that the types of firms who invest are not so affected by the financial crisis. Looking at the exit is the reverse way of looking at who would have invested but didn't. (Delios and Keeley, 2001)

[5] Even though the difference in theory could be due to the preference of the countries. However, it doesn't make sense why the level of Japanese ownership is also different (because countries would not care about this).

- Labor size: Similarly, a reputed benefit of FDI is that it creates jobs, generating not just economic growth but also increasing the government's popularity among the populace. The total number of employees of a firm is included in the Japanese Overseas Business dataset.

- Technology intensity: I proxy for a firm's technology intensity by the industry to which it belongs. OECD (2009) categorizes ISIC industries into four levels of technology intensity—low, medium low, medium high, and high—according to the level of R&D expenditure divided by sales. I convert the industry classification of firms in my data from SIC 3 to ISIC and categorize their technology intensity from 1 to 4, with 1 being low and 4 being high. On several occasions, one industry in SIC 3 matches to multiple ISIC (rev 3) industries or none at all. In the former case, I take the average across matched ISIC industries. In the latter case, the data is missing and later removed from the analysis.[6][7]

- Export intensity (ratio of export to sale):

Summary statistics table for these covariates.

For countries' characteristics that firms consider, I include:

- Market size: MNCs are expected to prefer countries with a large market size, which present MNCs with many potential customers. Indeed, this has been often cited as the allure of China to MNCs (Luo et al., 2010), as well as confirmed in larger studies. This is also a major variable in the gravity model, which has become a standard model for analyzing FDI flows. Bergstrand and Egger (2007) provides the theoretical framework for the use of gravity model.

[6] Bergstrand and Egger (2007) discusses the difference between R&D intensity and advertising intensity, and find that R&D intensity is higher for manufacturing firms compared with consumer product firms. Plus R&D intensity is much more important for firms' performance than advertising intensity.

[7] Definition of R&D intensity: the amount spent on R&D as a percentage of sale

I follow the standards in the literature and include log GDP (constant 2005 US$), taken from the Penn World Table.[8]

- Level of development: MNCs are expected to prefer countries with a high level of development. A developed economy has consumers with high purchasing power and better infrastructure. It can also measure capital abundance, in which case a higher GDP per capita imply less flow because the simple model of FDI frames FDI as the movement of capital from the capital rich countries to the capital poor countries. To measure development, I use log GDP per capita (constant 2005 US$) from World Development Indicators.

- GDP growth may be a proxy of potential returns,

- Labor quality: As one primary factor of production, labor matters greatly to firms' productivity and profit. To measure labor quality, I use the average years of schooling of adult, taken from the UNDP's Human Development Report.[9]

- Democracy: Democracy has been a mainstay in the political science literature on FDI. Scholars have argued that MNCs want to invest in democratic regimes for various reasons, including stable policy, credible commitment, and strong property rights (Ahlquist, 2006; Li and Resnick, 2003; Jensen, 2003). On the other hand, recent works have also argued that democratic regimes want FDI more than autocratic regimes (Pandya, 2016). Thus, it is unclear whether the observed high level of FDI in democracies is due to the push or the pull factors. By controlling for countries' preference in the two-sided matching model, I can better estimate the effect of democracies on firms' utility. I measure democracy

[8] An advantage of the Penn World Table is that it compiles data for Taiwan, an important destination that the World Bank Development Indicators does not include.

[9] Since Taiwan is not included in UNDP's and World Bank's data, I collected its statistics from the Taiwanese Statistical Website.

using the binary Demoracy & Dictatorship, developed by Cheibub et al. (2009).

6.3 Result

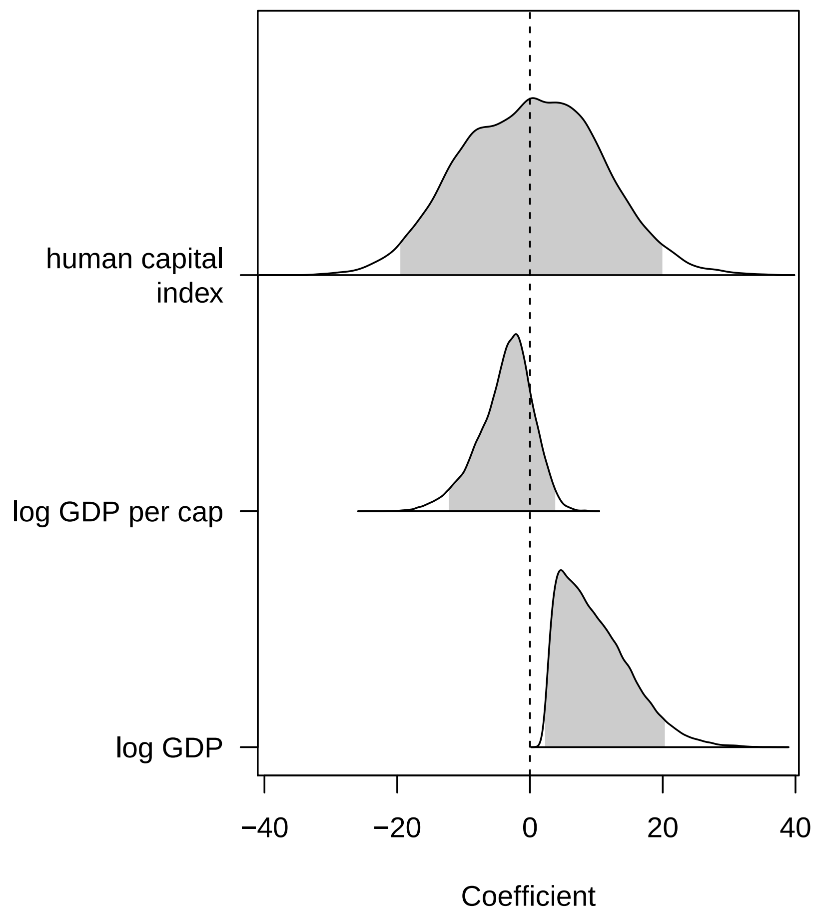

FIGURE 6.1: Preference of MNCs for countries' characteristics. The density plot and the shaded region show the posterior distribution and the 95% credible interval.

6.4 Model fit

Profiles of firms

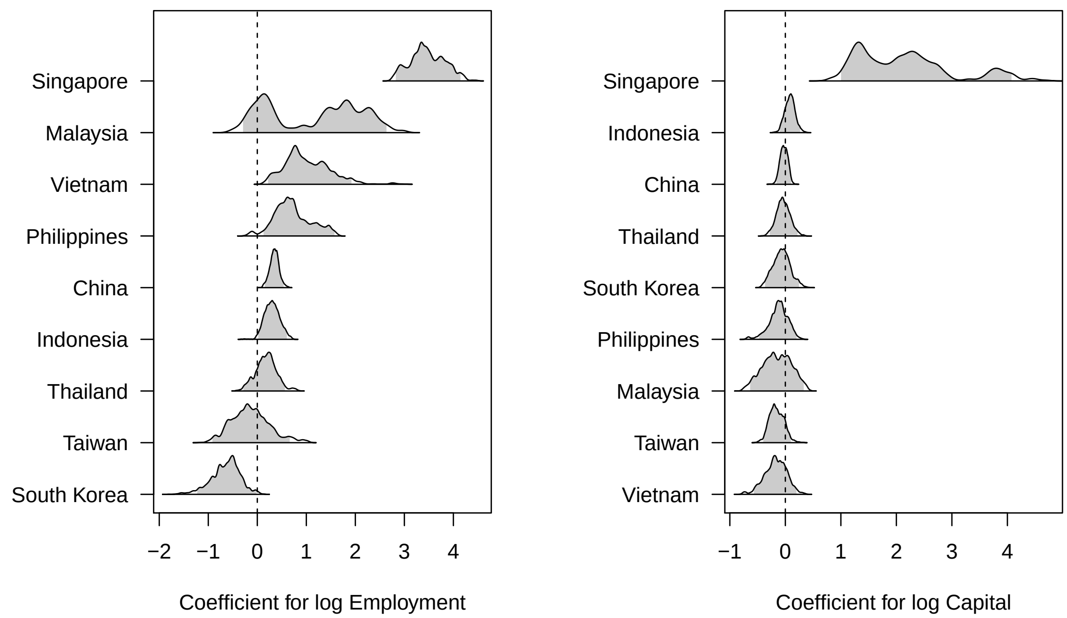

FIGURE 6.2: Preference of countries for firms' size, measured by their labor force (left) and capital (right). The density plot and the shaded region show the posterior distribution and the 95% credible interval.

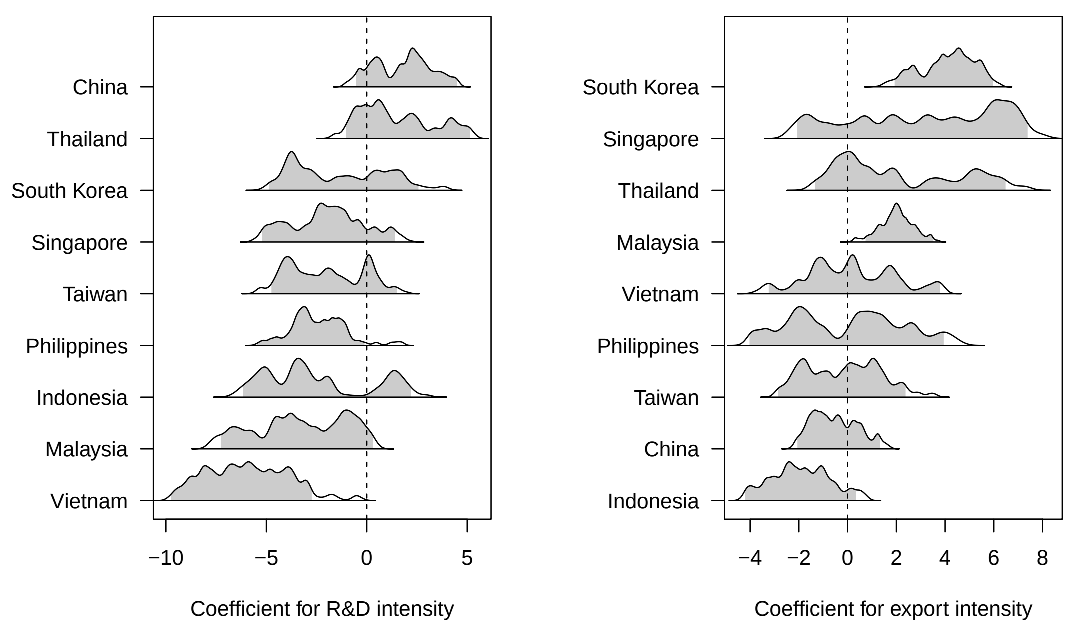

FIGURE 6.3: Preference of countries for firms' intangible assets, i.e R&D intensity (left) and export intensity (right). The density plot and the shaded region show the posterior distribution and the 95% credible interval.

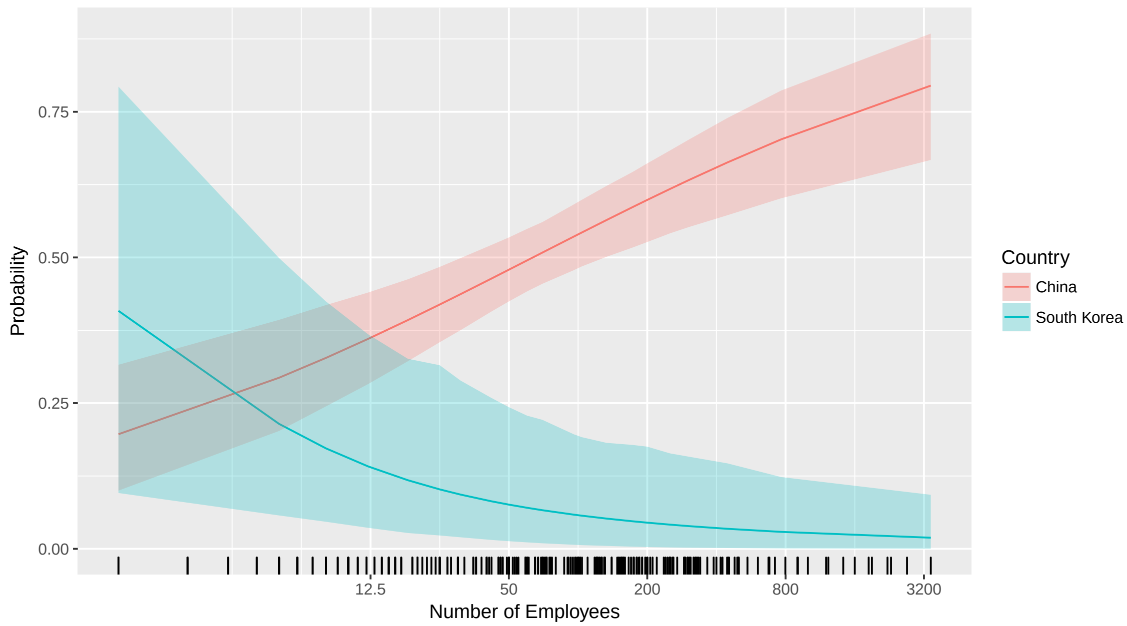

FIGURE 6.4: The effect of employee size on the probability to be offered by China and South Korea

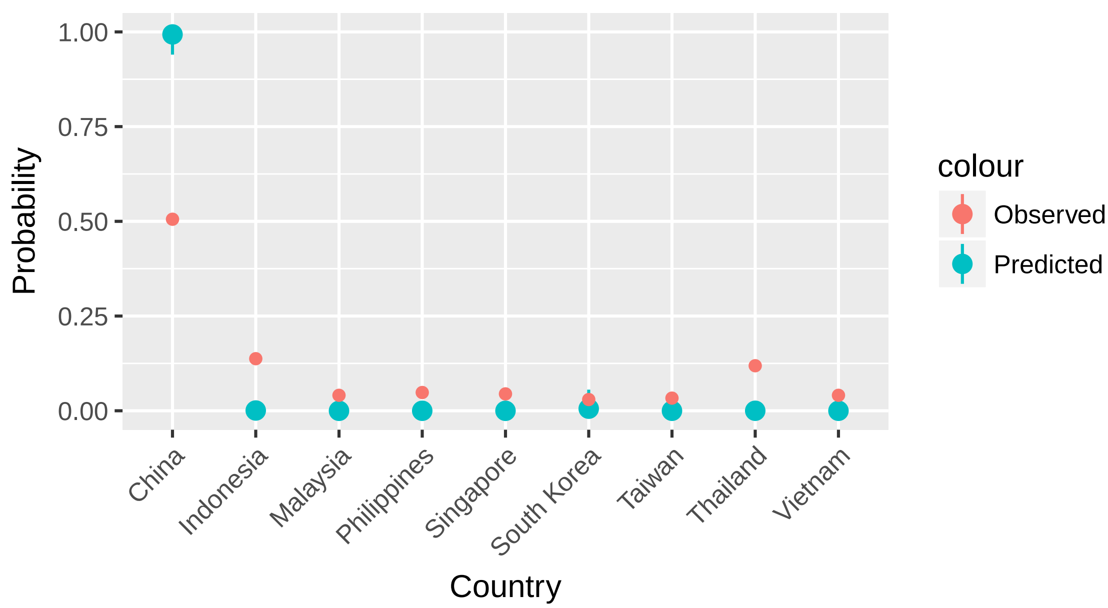

FIGURE 6.5: Predicted and observed probabilities that an MNC chooses to locate in a country, unconditional on the preference of countries. The point and the error bar show the posterior mean and the 95% credible interval.

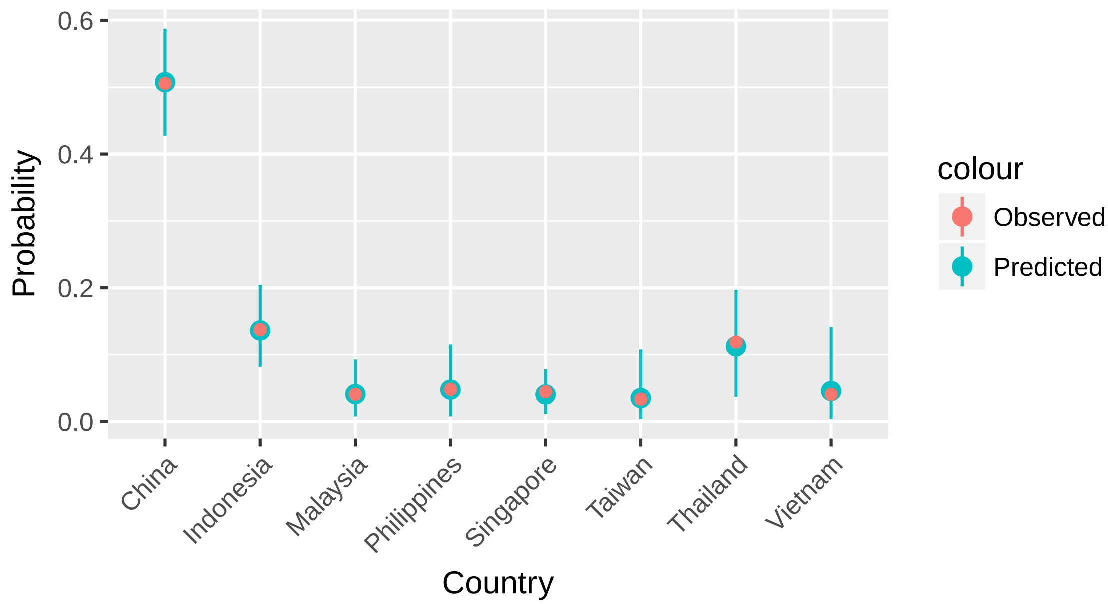

FIGURE 6.6: Predicted and observed probabilities that an MNC chooses to locate in a country, conditional on the preference of countries. The point and the error bar show the posterior mean and the 95% credible interval.

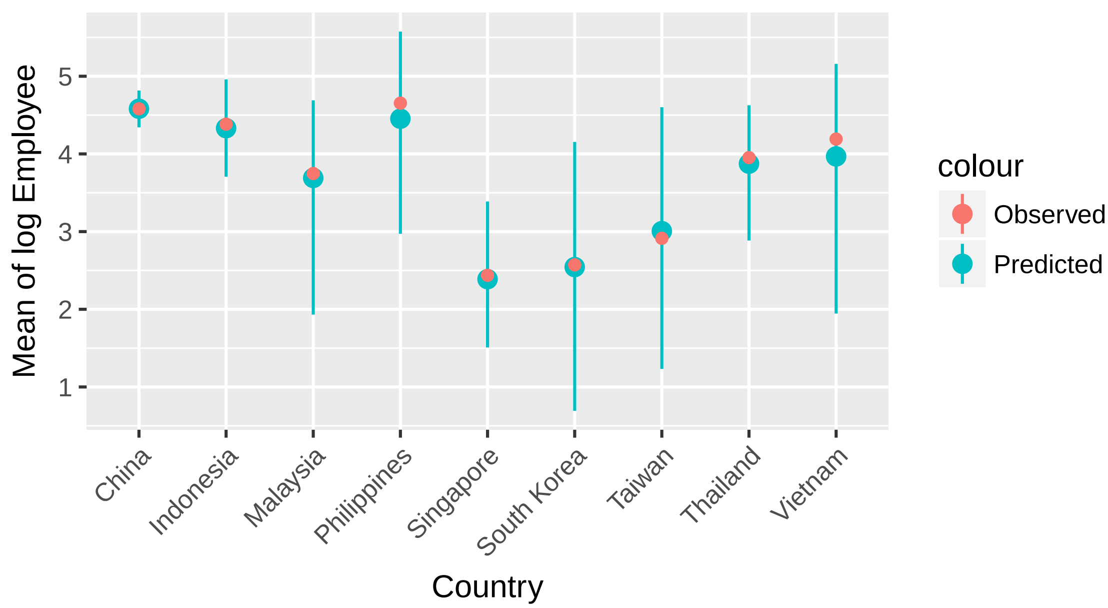

FIGURE 6.7: Average of MNCs' labor size across countries. The point and the error bar show the posterior mean and the 95% credible interval.

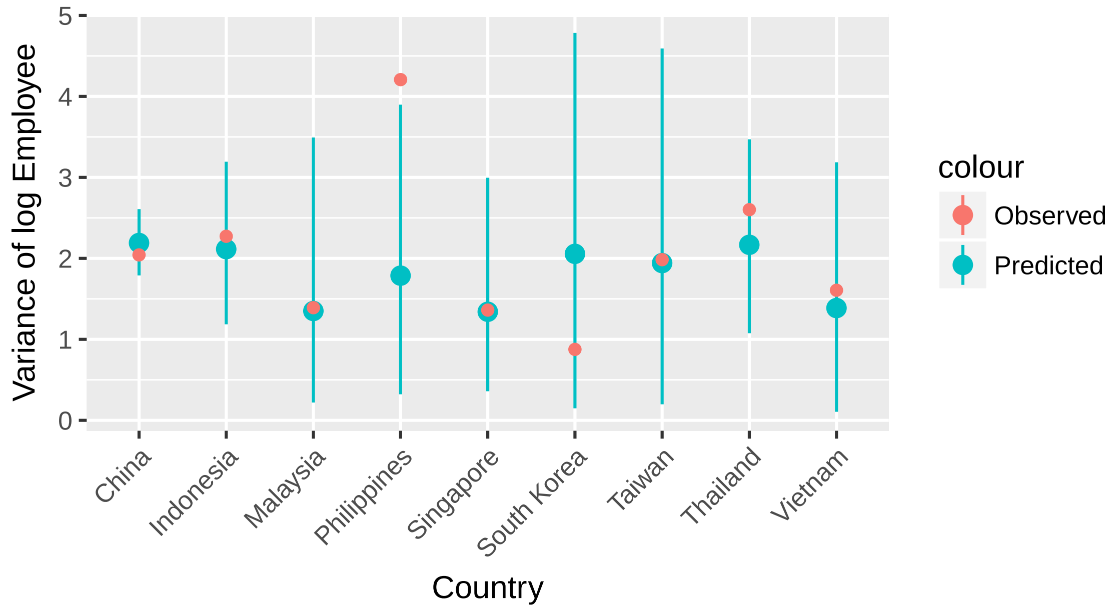

Figure 6.8: Variance of MNCs' labor size across countries. The point and the error bar show the posterior mean and the 95% credible interval.

6.5 Political determinants of FDI

(Nunnenkamp, 2002) finds that for developing countries market based factors are still the most important

6.6 Conclusion

In this paper, I propose the two-sided matching model to estimate firms' and countries' preferences, solving three persistent issues in the literature of FDI's political determinants. The results indicate that, for Japanese MNCs, only a country's level of development matters and not its market size, labor quality, or regime type. This finding suggests that we should take a closer look at the relationship between democracies and MNCs. Since previous works in the literature have not controlled for countries' preferences, they may have mistaken democracies' love for FDI as FDI's fondness for democracies.

On the other hand, the model's estimation of countries' preference remains lacking. Since each country has its own set of parameters, the parameter space seems too large for the current implementation of the Metropolis-Hastings algorithm to fully explore. Several solutions are possible. First, we can collapse countries into categories of interest, e.g. regime types, (categorical) time horizon length. Second, we can build a hierarchical model, modeling countries' preferences as draws from a common distribution. Such model will allow us to pool information across countries and reduce the parameter space.

7

Basic Document Class Features

This chapter is an example of how to format normal material in the dissertation style. Most of this information is standard to LaTeX.

7.1 Intra-chapter divisions: Sections

Section headlines are `\Large` and in the standard font. Compare them to subsections below.

7.1.1 Subsections: Wow! Italics!

Yes, italics. You may now dance. Isn't it funny that upright letters are called "roman" while slanted letters are "italic". That's like Italian, and Romans are Italians too. What gives?

Subsubsections: Smaller and smaller

Subsubsections are allowed, but are not numbered and don't appear in the table of contents. Likewise, you can use the next level of sectioning.

Paragraphs These divisions are unnumbered and do not appear in the Table of Contents.

Subparagraphs This is the finest division possible. It's also unnumbered and omitted from the Table of Contents.

7.2 Let's do some math

Let's look at an equation:

$$\partial f t = f(t) \quad \text{subject to} \quad f(0) = c. \tag{7.1}$$

We've used the `\newcommand` defined in the preamble of `dissertation.tex` to produce the derivative. You can get a second derivative like $\partial^2 f t^2$ by adding some sneaky superscripts. Fancy.

More advanced equation formatting is available in the AMS environments. See the guide amsmath user's guide. Here are some nice examples of cases people usually have trouble with.

An equation that's too long for one line — use `multline`:

$$\begin{multline} a + b + c + d + e + f + g + h + i + j + k + l + m + n + o \\ = p + q + r + s + t + u + v + w + x + y + z. \tag{7.2} \end{multline}$$

An equation with multiple parts and one number per line — use `align`:

$$a_1 = b_1 + c_1 \tag{7.3}$$

$$a_2 = b_2 + c_2. \tag{7.4}$$

The same equation, set inside the `subequations` environment:

$$a_1 = b_1 + c_1 \tag{7.5a}$$

$$a_2 = b_2 + c_2. \tag{7.5b}$$

Notice that by clever placement of labels, I can reference the pair via (7.5), the first (7.5a), or the second (7.5b). One number for multiple equations can be accomplished using the `split` environment:

$$\begin{split} a &= b + c - d \\ &\quad + e - f \\ &= g + h \\ &= i. \end{split} \tag{7.6}$$

People often struggle under the complicated and ugly 'eqnarray' environment. Don't do it! The AMS ones are easy. Other stumbling blocks are cases:

$$a = \begin{cases} b & \text{for } x > 0 \\ c & \text{otherwise,} \end{cases} \tag{7.7}$$

matrices:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \tag{7.8}$$

and evaluation bars:

$$a = \left. \frac{\partial u}{\partial x} \right|_{x=0}. \tag{7.9}$$

See the source file for details.

When we reference an equation with something like `\eqref` (7.1). If you click on the above references in the PDF, your viewer should scroll up to the above equation. It's handy. Labels and references may be attached to all sorts of objects. There is a `\label` attached to this chapter (it appears at the top of this file), and we may reference it by (`Chapter~\ref{chap:example}`), producing "Chapter 7". By default these ref's are hyperlinked as well. Later, we'll see labeled and referenced figures and tables. Particular pages may be labeled with standard `\label` commands in the text and referenced via `\pageref`.

You might also like the links from `\cite` commands to the corresponding bibliographic entry. Go look at this imaginary book by Stephen Colbert **?**. If you're not a bibtex expert, look in `mybib.bib` at the `@ARTICLE` that generated this entry. It shows an example of accents on author names and how to preserve upper-case for letters in the title. Other entries show the use of the `and` keyword between author names. You may order a particular author's name as either "first last" or as "last, first". The actual format of the bibliography is controlled by the `\bibliographystyle{}` command in `dissertation.tex`.

7.3 Table of Contents Behavior

Now is a good time to look back at the Table of Contents. Notice that you may click on entries here to warp to the corresponding document location. In Adobe Acrobat and many other viewers, you can open a 'bookmarks' pane. This should be populated with named and numbered sections and subsections identical to the Table of Contents.

7.4 Figures and footnotes

Figures are set with very little space between the caption and the bottom of the included graphic. This is because most graphics programs pad the edges of images. If you find the spacing unsatisfactory, you may always add a bit manually. The text of the caption is single-spaced, and the word 'Figure' is set in small caps. See Fig. 7.1. Notice the use of the nonbreakable space "`~`" between the "Fig." and the reference. Figures (and tables) are examples of 'floats — objects that LaTeXdecides where to place for you. You may give LaTeXsome hints. Change the `\begin{figure}[tbp]` to a `\begin{figure}[b!]` to restrict the placement. Inside the [ ], you can put the following

Figure 7.1: Longer caption for actual body of dissertation. Figure captions should be BELOW the figure.

- t Allow placement at the top of the page
- b Allow placement at the bottom of the page
- h Allow placement 'here', in the middle of the page close to the text that the figure environment appears next to.
- p Allow placement on a seperate 'floats page' that has no body text.
- ! Tighten the screws on the placement algorithm. This doesn't force things to happen as you say, but it makes it much more likely. Be careful: the bang option can cause figures to appear above the chapter title and in other bad locations.

Notice that each entry just changes what is *allowed*, but no preference among the entries can be registered. The default is [tbp], which is a very good default for a document like this, since floats in the middle of a page trap too much whitespace for double-spaced text. There is also a prohibition against having a page with more than 75% float. Instead, long floats will get kicked over onto float pages. Float pages are often a bad idea, as the creation of one will often cause a domino effect, with all subsequent figures appearing on float pages themselves, and all these float pages appearing together at the end of the chapter. (This is more like sinking than floating.) Avoid this by physically moving where the figure environment appears in your source file to an earlier location. Don't be afraid to put the environment before the first spot you reference it! Many float problems can be solved by a combination of relocating the figure environment and a little fiddling with the [ ] options.

Also notice the order of the graphic, caption, and label. If you deviate from this, strange things can happen. The caption of this figure shows the use of short captions (inside []). These caption appear in the List of Tables, while the captions appear in the body. If you omit the [ ] short caption, the long caption will be used in its place.

Another technical note: since this style sheet is designed for processing by pdflatex, `\includegraphics` looks for `PDF`s, `PNG`s, and `JPG`s instead of the usual `PS`, `EPS`, and `TIFF` formats. You can convert existing graphics with a vareity of tools. `PDF` graphics are preferred, as they scale nicely. The open-source software Inkscape runs on Mac OSX, Windows, Linux, and some UNIX variants. Versions 0.46 and beyond have great support for creating and editing `PDF`s. It can even be used to convert other docs.

7.4.1 List of Figures

If you've put even one measly figure in your document, grad school rules say you need a List of Figures. It's automatically generated for you if you do a `\listoffigures`

Table 7.1: Long table caption appears on in the body text. See the short caption in the List of Tables. Table captions need to be ABOVE the table.

Numbers	Letters	Symbols
1	a	†
2	b	$\smile$
3	c	$\times$
4	d	$\sharp$

in the master file (heck, it's there right now). Go look at the list of figures now. You should be able to click on the figure number to warp to the figure. You'll also see the result of the 'short caption' used above.

7.5 Table example

Just to make sure tables are formatted correctly, here's an example of a table float, see Table 7.1. You should note that [b] formatting (`\begin{table}[b]`) can cause floats to appear under the footnotes. Try changing it here and see the ugliness. Tables are identical to figures, except that the word 'Table' appears in the caption and its entry is in the List of Tables instead of the List of Figures.

7.5.1 Footnotes

Footnotes are allowed.[1] They are numbered with arabic numerals inside each chapter and appear at the bottom of the page.[2] The little footnote numbers are also hyperlinks. Try clicking them. You should place the footnote command immediately following the period of the sentence it is attached to. Any spaces or newlines will result in strange spacing between the number and the sentence.

[1] But, you should probably just work them into the text since it's annoying to jump around when reading.

[2] . . . rather than the end of the chapter or the thesis. Those would properly be endnotes, I guess.

7.6 Corner cases in formatting, such as very very very long section titles. Man, this goes on forever.

Common corner-cases involve very long titles (like above). In these cases, the long titles are set single-spaced both here and in the Table of Contents.

7.6.1 Figure and Table caption cases are neat, and this is an absurdly long subsection heading

Consider the shield logo again with an absurd caption, as in Fig. 7.2. Also examine the new table, Table 7.2. Both of these have been forced onto a floats page so you can see what that looks like.

FIGURE 7.2: The Duke logo again, but now with a really long rambling caption. This caption should be set single-spaced in the LoF and in the body text. What do you think about having graphics in the main directory of a project? I'd prefer them in a folder, then put 'foldername/picturename' as the argument to includegraphics.

Table 7.2: The same silly table again, but with a really long caption.

Numbers	Letters	Symbols
1	a	$\dagger$
2	b	$\ddot\smile$
3	c	$\times$
4	d	$\sharp$

Appendix A

Derivation of the Metropolis-Hastings Acceptance Ratio

A.0.1 Opportunity sets O

Target distribution for a firm i

$$p(O_i|A_i, \alpha, \boldsymbol{\beta}) = \frac{p(O_i, A_i, \alpha, \boldsymbol{\beta})}{p(A_i, \alpha, \boldsymbol{\beta})} \tag{A.1}$$

$$MH_O = \frac{p(O_i^*|A_i, \alpha, \boldsymbol{\beta})}{p(O_i|A_i, \alpha, \boldsymbol{\beta})} = \frac{p(O_i^*, A_i, \alpha, \boldsymbol{\beta})}{p(A_i, \alpha, \boldsymbol{\beta})} \times \frac{p(A_i, \alpha, \boldsymbol{\beta})}{p(O_i, A_i, \alpha, \boldsymbol{\beta})} \tag{A.2}$$

$$= \frac{p(O_i^*, A_i, \alpha, \boldsymbol{\beta})}{p(O_i, A_i, \alpha, \boldsymbol{\beta})} \tag{A.3}$$

$$= \frac{p(A_i|O_i^*, \alpha)p(O_i^*|\boldsymbol{\beta})}{p(A_i|O_i, \alpha)p(O_i|\boldsymbol{\beta})} \tag{A.4}$$

$$\tag{A.5}$$

where the factorization of the likelihood in (A.4) is due to the fact that the acceptance of firm i only depends on what is offered to it and what is its preference, $p(A_i|O_i^*, \alpha)$; what is offered to i depends on the preferences of all countries, $p(O_i^*|\boldsymbol{\beta})$.

If we plug in (3.9) and (3.7)

$$\frac{p(O_i^*|A_i,\alpha,\boldsymbol{\beta})}{p(O_i|A_i,\alpha,\boldsymbol{\beta})} = \frac{\sum\limits_{j:j\in O_i} \exp(\alpha' W_j)}{\sum\limits_{j:j\in O_i} \exp(\alpha' W_j) + \exp(\alpha' W_{j*})} \times \exp(\boldsymbol{\beta}'_{j*} X_i) \tag{A.6}$$

where j^* is the index of the newly sampled job. This is the case when the newly proposed job is not already offered, so it's added to the opportunity set.

When the newly proposed job is already offered, so it's removed from the opportunity set, we have

$$\frac{p(O_i^*|A_i,\alpha,\boldsymbol{\beta})}{p(O_i|A_i,\alpha,\boldsymbol{\beta})} = \frac{\sum\limits_{j:j\in O_i} \exp(\alpha' W_j)}{\sum\limits_{j:j\in O_i} \exp(\alpha' W_j) - \exp(\alpha' W_{j*})} \times \exp(-\boldsymbol{\beta}'_{j*} X_i) \tag{A.7}$$

A.0.2 Workers' parameters, α

Target distribution:

$$p(\alpha|A,O,\boldsymbol{\beta}) = \frac{p(O,A,\alpha,\boldsymbol{\beta})}{p(A,O,\boldsymbol{\beta})} \tag{A.8}$$

Metropolis-Hasting acceptance ratio:

$$MH_\alpha = \frac{p(\alpha^*|A,O,\boldsymbol{\beta})}{p(\alpha|A,O,\boldsymbol{\beta})} = \frac{p(A_i|O_i,\alpha^*)p(O_i|\boldsymbol{\beta})p(\alpha^*)}{p(A_i|O_i,\alpha)p(O_i|\boldsymbol{\beta})p(\alpha)} \tag{A.9}$$

$$= \frac{p(A_i|O_i,\alpha^*)p(\alpha^*)}{p(A_i|O_i,\alpha)p(\alpha)} \tag{A.10}$$

where (A.10) is due to the symmetric proposal distribution (so $\frac{p(\alpha^*|\alpha)}{p(\alpha|\alpha^*)} = 1$)

If we plug in (3.9),

$$MH_\alpha = \prod_i \left[\frac{\exp(\alpha^{*\prime} W_{a_i})}{\exp(\alpha' W_{a_i})} \times \frac{\sum\limits_{j:j\in O_i} \exp(\alpha' W_j)}{\sum\limits_{j:j\in O_i} \exp(\alpha^{*\prime} W_j)} \right] \times \frac{p(\alpha^*)}{p(\alpha)} \tag{A.11}$$

$$= \prod_i \left[\exp(\epsilon_\alpha' W_{a_i}) \times \frac{\sum\limits_{j:j\in O_i} \exp(\alpha' W_j)}{\sum\limits_{j:j\in O_i} \exp(\alpha^{*\prime} W_j)} \right] \times \frac{p(\alpha^*)}{p(\alpha)} \tag{A.12}$$

Finally, we log transform the MH acceptance ratio for numerical stability.

$$\log MH_\alpha = \sum_i \left[\epsilon_\alpha' W_{a_i} + \log\left(\sum_{j:j\in O_i} \exp(\alpha' W_j) \right) - \log\left(\sum_{j:j\in O_i} \exp(\alpha^{*\prime} W_j) \right) \right] + \log p(\alpha^*) - \log p(\alpha) \tag{A.13}$$

A.0.3 Firms' parameters, β

Target distribution:

$$p(\boldsymbol{\beta}|A, O, \alpha) = \frac{p(O, A, \alpha, \boldsymbol{\beta})}{p(A, O, \alpha)} \tag{A.14}$$

Metropolis-Hasting acceptance ratio:

$$MH_\beta = \frac{p(\beta^*|A, O, \alpha)}{p(\beta|A, O, \alpha)} = \frac{p(A_i|O_i, \alpha) p(O_i|\boldsymbol{\beta}^*) p(\boldsymbol{\beta}^*|\mu_\beta, \tau_\beta)}{p(A_i|O_i, \alpha) p(O_i|\boldsymbol{\beta}) p(\boldsymbol{\beta}|\mu_\beta, \tau_\beta)} \tag{A.15}$$

$$= \frac{p(O_i|\boldsymbol{\beta}^*) p(\boldsymbol{\beta}^*|\mu_\beta, \tau_\beta)}{p(O_i|\boldsymbol{\beta}) p(\boldsymbol{\beta}|\mu_\beta, \tau_\beta)} \tag{A.16}$$

where (A.15) is due to the symmetric proposal distribution.

We plug in (3.7),

$$MH_\beta = \prod_i \left[\prod_{j \in O_i} \frac{\exp(\beta_j^{*\prime} X_i)}{\exp(\beta_j' X_i)} \times \prod_j \frac{1 + \exp(\beta_j^{*\prime} X_i)}{1 + \exp(\beta_j' X_i)} \right] \times \frac{MVN(\boldsymbol{\beta}^* | \mu_\beta, \tau_\beta)}{MVN(\boldsymbol{\beta} | \mu_\beta, \tau_\beta)} \quad \text{(A.17)}$$

$$\log MH_\beta = \sum_i \left[\sum_{j \in O_i} \beta_j^{*\prime} X_i - \beta_j' X_i + \sum_j \log(1 + \exp(\beta_j^{*\prime} X_i)) - \log(1 + \exp(\beta_j' X_i)) \right] \quad \text{(A.18)}$$

$$+ \log MVN(\boldsymbol{\beta}^* | \mu_\beta, \tau_\beta) - \log MVN(\boldsymbol{\beta} | \mu_\beta, \tau_\beta)$$

Bibliography

Adachi, H. (2003), "A search model of two-sided matching under nontransferable utility," *Journal of Economic Theory*, 113, 182–198.

Ahlquist, J. (2006), "Economic policy, institutions, and capital flows: portfolio and direct investment flows in developing countries," *International Studies Quarterly*, 50, 681–704.

Alfaro, L. (2003), "Foreign Direct Investment and Growth : Does the Sector Matter?" *Harvard Business School*, pp. 1–32.

Alfaro, L. and Charlton, A. (2007), "Growth and the quality of foreign direct investment: is all FDI equal," .

Anh, V. T. T., Thai, L. V., and Thang, V. T. (2007), "Provincial Extralegal Investment Incentives in the Context of Decentralisation in Viet Nam : Mutually Beneficial or a Race to the Bottom ?" *Forum American Bar Association*.

Arauzo-Carod, J. M., Liviano-Solis, D., and Manjón-Antolín, M. (2010), "Empirical studies in industrial location: An assessment of their methods and results," *Journal of Regional Science*, 50, 685–711.

Arel-Bundock, V. (2017), "The Political Determinants of Foreign Direct Investment: A Firm-Level Analysis," *International Interactions*, 43, 424–452.

Aw, B. Y. and Lee, Y. (2008), "Firm heterogeneity and location choice of Taiwanese multinationals," *Journal of International Economics*, 76, 403–415.

Beazer, Q. and Blake, D. (2011), "It's All Relative: Home Country Risk and FDI Flows," .

Bergstrand, J. H. and Egger, P. (2007), "A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises," *Journal of International Economics*, 73, 278–308.

Beugelsdijk, S., Hennart, J. F., Slangen, A., and Smeets, R. (2010), "Why and how FDI stocks are a biased measure of MNE affiliate activity," *Journal of International Business Studies*, 41, 1444–1459.

Bonica, A., Chilton, A. S., Goldin, J., Rozema, K., and Sen, M. (2017), "The political ideologies of law clerks," *American Law and Economics Review*, 19, 96–128.

Broz, J. L. and Frieden, J. (2001), "The political economy of international monetary relations," *Annual Review of Political Science*, pp. 317–343.

Busse, M. and Hefeker, C. (2007), "Political risk, institutions and foreign direct investment," *European Journal of Political Economy*, 23, 397–415.

Büthe, T. and Milner, H. (2008), "The Politics of Foreign Direct Investment into Developing Countries: Increasing FDI through International Trade Agreements?" *American Journal of Political Science*, 52, 741 – 762.

Cameron, A. and Trivedi, P. (2005), *Microeconometrics: methods and applications*, Cambridge University Press, Cambridge.

Carkovic, M. V. and Levine, R. (2002), "Does foreign direct investment accelerate economic growth?" *U of Minnesota Department of Finance Working Paper*.

Cheibub, J. A., Gandhi, J., and Vreeland, J. R. (2009), "Democracy and dictatorship revisited," *Public Choice*, 143, 67–101.

Cheng, S. and Stough, R. R. (2006), "Location decisions of Japanese new manufacturing plants in China: A discrete-choice analysis," *Annals of Regional Science*, 40, 369–387.

Chintrakarn, P., Herzer, D., and Nunnenkamp, P. (2012), "Fdi and income inequality: Evidence from a panel of U.S. states," *Economic Inquiry*, 50, 788–801.

Choi, S.-W. and Samy, Y. (2008), "Reexamining the Effect of Democratic Institutions on Inflows of Foreign Direct Investment in Developing Countries," *Foreign Policy Analysis*, 4, 83–103.

De Swaan, A. (1973), *Coalition Theories and Cabinet Formations: A Study of Formal Theories of Coalition Formation Applied to Nine European Parliaments after 1948*, Jossey-Bass.

Delios, A. and Henisz, W. J. (2000), "Japanese firms' investment strategies in emerging economies," *Acadnmy of Management Journal*, 43, 305–323.

Delios, A. and Keeley, T. D. (2001), "Japanese Foreign Direct Investment In Thailand: An Empirical And Qualitative Post-Crisis Analysis," *Journal of International Business and Economy*, 1, 91–118.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 39, 1–38.

Ditslear, C. and Baum, L. (2001), "Selection of Law Clerks and Polarization in the U.S. Supreme Court," *The Journal of Politics*, 63, 869–885.

Dorff, C. and Ward, M. (2013), "Networks, Dyads, and the Social Relations Model," *Political Science Research and Methods*.

Durham, J. B. (2004), "Absorptive capacity and the effects of foreign direct investment and equity foreign portfolio investment on economic growth," *European Economic Review*, 48, 285–306.

Findlay, R. (1978), "Relative Backwardness, Direct Foreign Investment, and the Transfer of Technology: A Simple Dynamic Model," *Quarterly Journal of Economics*, 92, 1–16.

Fletcher, K. (2002), "Tax Incentives in Cambodia, Lao PDR, and Vietnam," Tech. rep.

Fu, X. (2008), "Foreign Direct Investment, Absorptive Capacity and Regional Innovation Capabilities: Evidence from China," *Oxford Development Studies*, 36, 89–110.

Gale, D. and Shapley, L. S. (1962), "College Admissions and the Stability of Marrige," *The American Mathematical Monthly*, 69, 9–15.

Gallop, M. and Weschle, S. (2017), "Assessing the Impact of Non-Random Measurement Error on Inference: A Sensitivity Analysis Approach," *Political Science Research and Methods*, pp. 1–18.

Gelman, A. and Hill, J. (2006), *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press.

Goldberg, P. K. and Pavcnik, N. (2007), "Distributional Effects of Globalization in Developing Countries," *Journal of Economic Literature*, 45, 39–82.

Goswami, A., Hedayati, F., and Mohapatra, P. (2014), "Recommendation systems for markets with two sided preferences," *Proceedings - 2014 13th International Conference on Machine Learning and Applications, ICMLA 2014*, pp. 282–287.

Graham, B. (2010), "Political Risk and Diaspora Direct Investment," *APSA 2010 Annual Meeting Paper*, 42, 622–696.

Grosse, R. B. and Duvenaud, D. K. (2014), "Testing MCMC code," *CoRR*, abs/1412.5, 1–8.

Guangzhou (2011), "Implementation Opinions on Further Promoting the Work of Utilizing Foreign Investment," .

Guerra, E., de Lara, J., Malizia, A., and Díaz, P. (2009), "Supporting user-oriented analysis for multi-view domain-specific visual languages," .

Guimaraes, P., Figueirdo, O., and Woodward, D. (2003), "A Tractable Approach to the Firm Location Decision Problem," *The Review of Economics and Statistics*, 85, 201–204.

Gulati, M. and Posner, R. A. (2016), "The management of staff by federal court of appeals judges," *Vanderbilt Law Review*, 69, 479–497.

Haario, H., Saksman, E., and Tamminen, J. (1999), "Adaptive proposal distribution for random walk Metropolis algorithm," *Computational Statistics*, 14, 375.

Haario, H., Saksman, E., and Tamminen, J. (2001), "An Adaptive Metropolis Algorithm," *Bernoulli*, 7, 223.

Hitsch, G. J., Hortaçsu, A., and Ariely, D. (2010), "What makes you click?-Mate preferences in online dating," *Quantitative Marketing and Economics*, pp. 1–35.

Hodge, R. W., Siegel, P. M., and Rossi, P. H. (1964), "Occupational Prestige in the United States, 1925-63," *American Journal of Sociology*, 70, 286–302.

Jamshidian, M. and Jennrich, R. I. (2000), "Standard Errors for EM Estimation," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62, 257–270.

Javorcik, B. S. (2004), "Does Foreign Direct Investment Increase the Productivity of Domestic Firms ? In Search of Spillovers through Backward Linkages Does Foreign Direct lnvestment Increase the Productivity of Domestic Firms ? In Search of Spillovers Through Backward Linkages," *American Economic Review*, 94, 605–627.

Jensen, N. and McGillivray, F. (2005), "Federal institutions and multinational investors: Federalism, government credibility, and foreign direct investment," *International Interactions*, 31, 303–325.

Jensen, N. M. (2003), "Democratic Governance and Multinational Corporations: Political Regimes and Inflows of Foreign Direct Investment," *International Organization*, 57.

Jensen, N. M. (2012), "Fiscal Policy and the Firm: Do Low Corporate Tax Rates Attract Multinational Corporations?" *Comparative Political Studies*, 45, 1004–1026.

Jensen, N. M., Malesky, E., and Weymouth, S. (2014), "Unbundling the Relationship between Authoritarian Legislatures and Political Risk," *British Journal of Political Science*, 44, 655–684.

Karcher, S. and Steinberg, D. A. (2013), “Assessing the Causes of Capital Account Liberalization: How Measurement Matters,” *International Studies Quarterly*, 57, 128–137.

Kerner, A. (2014), “What We Talk About When We Talk About Foreign Direct Investment,” *International Studies Quarterly*, pp. 804–815.

Kerner, A. and Lawrence, J. (2014), “What’s the risk? Bilateral investment treaties, political risk and fixed capital accumulation,” *British Journal of Political Science*, 44, 107–121.

Laver, M. J. (1998), “Models of Government Formation,” *Annual Review of Political Science*, 1, 1–25.

Li, Q. (2006), “Democracy, autocracy, and tax incentives to foreign direct investors: A cross-national analysis,” *Journal of Politics*, 68, 62–74.

Li, Q. (2009), “Democracy, Autocracy, and Expropriation of Foreign Direct Investment,” *Comparative Political Studies*, 42, 1098–1127.

Li, Q. and Resnick, A. (2003), “Reversal of fortunes: Democratic institutions and foreign direct investment inflows to developing countries,” *International organization*.

Liptak, A. (2007), “A Sign of Court’s Polarization - Choice of Clerks,” .

Logan, J. A. (1996), “Opportunity and Choice in Socially Structured Labor Markets,” *American Journal of Sociology*, 102, 114.

Logan, J. A. (1998), “Estimating Two-Sided Logit Models,” *Sociological methodology*, 28, 139–173.

Logan, J. A., Hoff, P. D., and Newton, M. a. (2008), “Two-Sided Estimation of Mate Preferences for Similarities in Age, Education, and Religion,” *Journal of the American Statistical Association*, 103, 559–569.

Luo, Y., Xue, Q., and Han, B. (2010), “How emerging market governments promote outward FDI: Experience from China,” *Journal of World Business*.

Malesky, E. J. (2015), “Transfer pricing and global poverty,” *International Studies Review*, 17, 669–677.

Mallampally, P. and Sauvant, K. P. (1999), “Foreign Direct Investment in Developing Countries,” *Finance and Development*, 36.

McFadden, D. (1974), “Conditional logit analysis of qualitative choice behavior,” in *Frontiers in Econometrics*, ed. P. Zarembka, pp. 105–142, Academic Press, New York.

Milner, H. V. and Kubota, K. (2005), *Why the Move to Free Trade? Democracy and Trade Policy in the Developing Countries*, vol. 59.

Mold, A. (2004), "Fdi and Poverty Reduction: a Critical Reappraisal of the Arguments," *Weltbank, New York*.

Moran, T. H. (1998), *foreign direct investment and development: the new policy Agenda for developing countries and Economies in Transition*.

Mosley, L. (2005), "Globalisation and the state: Still room to move?" *New Political Economy*, 10, 355–362.

Mosley, L. and Singer, D. (2008), "Taking Stock Seriously: EquityMarket Performance, Government Policy, and Financial Globalization," *International Studies Quarterly*, 3, 405–425.

Mosley, L. and Uno, S. (2007), "Racing to the Bottom or Climbing to the Top ?" *Comparative Political Studies*, 40, 923–948.

Nair-Reichert, U. and Weinhold, D. (2001), "Causality tests for cross-country panels: a new look at FDI and economic growth in developing countries," 2, 153–171.

NORC (2014), "Appendix G: Prestige Scores and Socioeconomic Index Distributions," .

Nunnenkamp, P. (2002), "Determinants of FDI in Developing Countries: Has Globalization Changed the Rules of the Game?" *Kiel Institute for World Economics Working Paper*, 1122, 1–34.

Nunnenkamp, P. and Spatz, J. (2004), "FDI and economic growth in developing economies: how relevant are host-economy and industry characteristics," *Transnational Corporations*, 13.

Nunnenkamp, P., Schweickert, R., and Wiebelt, M. (2007), "Distributional effects of FDI: How the interaction of FDI and economic policy affects poor households in Bolivia," *Development Policy Review*, 25, 429–450.

OECD (2009), "Business R&D by technology intensity," in *OECD Science, Technology and Industry Scoreboard 2009*, OECD Publishing.

Pak, Y. S. and Park, Y. R. (2005), "Characteristics of Japanese FDI in the East and the West: Understanding the strategic motives of Japanese investment," *Journal of World Business*, 40, 254–266.

Pandya, S. (2010), "Labor markets and the demand for foreign direct investment," *International Organization*.

Pandya, S. S. (2014), "Democratization and Foreign Direct Investment Liberalization, 1970-2000," *International Studies Quarterly*, 58, 475–488.

Pandya, S. S. (2016), *Trading Spaces Foreign Direct Investment Regulation, 19702000*, Cambridge University Press, New York.

Peppers, T. C., Giles, M. W., and Tainer-Parkins, B. (2008), "Inside Judicial Chambers: How Federal District Court Judges Select and Use Their Law Clerks," *Albany Law Review*, 71, 8–23.

Pinto, P. (2013), "Partisan Investment in the Global Economy," .

Pinto, P. M. and Pinto, S. M. (2008), "The politics of investment partisanship: And the sectoral allocation of foreign direct investment," *Economics and Politics*, 20, 216–254.

Posner, R. A. (2001), "The Market for Federal Judicial Law Clerks," *The University of Chicago Law Review*, 68, 793–902.

Posner, R. A., Avery, C., Jolls, C., and Roth, A. (2007), "The New Market for Federal Judicial Law Clerks," *The University of Chicago Law Review*, 74, 447.

Prakash, A. (2007), "Investing Up: FDI and the Cross-Country Difusion of ISO 14001 Management Systems," *International Studies*, 51, 723–744.

Ricupero, R. (2000), "Tax incentives for foreign direct investment: A Global Survey," Tech. Rep. 16, UNCTAD, Geneva.

Roth, A. E. and Sotomayor, M. (1992), "Two-sided matching," in *Handbook of Game Theory*, eds. S. Hart and R. Aumann, vol. 1, chap. 16, pp. 486–541, Elsevier Science Publishers, 1st edn.

Roth, A. E. and Vate, J. H. V. (1990), "Random Paths to Stability in Two-Sided Matching," *Econometrica*, 58, 1475.

Rozema, K. and Peng, S. (2016), "The Role of External Referrals in Hiring: Evidence from Judicial Law Clerks," *Available at SSRN 2822952*.

Rydén, T. (2008), "EM versus Markov chain Monte Carlo for Estimation of Hidden Markov Models: A Computational Perspective," *Bayesian Analysis*, 3, 659–688.

Rysman, M. (2009), "The Economics of Two-Sided Markets," *Journal of Economic Perspectives*, 23, 125–143.

Sawant, R. J. (2010), "The economics of large-scale infrastructure FDI: The case of project finance," *Journal of International Business Studies*, 41, pp. 1036–1055.

Slotnick, E. E. (1984), "The paths to the federal bench: gender, race and judicial recruitment variation," *Judicature*, 67, 371–388.

Solow, R. M. (1956), "A Contribution to the Theory of Economic Growth," *The Quaterly Journal of Economics*, 70, 65–94.

Sprecher, S., Sullivan, Q., and Hatfield, E. (1994), "Mate selection preferences: gender differences examined in a national sample." *Journal of personality and social psychology*, 66, 1074–1080.

Tanner, M. A. and Wong, W. H. (1987), "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Telford, T. G. and Ures, H. A. (2001), "The Role of Incentives in Foreign Direct Investment," *Loyola of Los Angeles International and Comparative Law Review*, 23.

Train, K. E. (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

Tu, K., Ribeiro, B., Jiang, H., Wang, X., Jensen, D., Liu, B., and Towsley, D. (2014), "Online Dating Recommendations: Matching Markets and Learning Preferences," *CoRR*, abs/1401.8.

UNCTAD (1998), "The Financial Crisis in Asia and Foreign Direct Investment : an Assessment," Tech. rep.

UNCTAD (2001), "World Investment Report 2001: Promoting Linkages," Tech. rep., United Nations, New York and Geneva.

UNCTAD (2007), "World Investment Report 2007: Transnational Corporations, Extractive Industries and Development," Tech. rep., UNCTAD, New York and Geneva.

UNCTAD (2015), "World Investment Report 2015: Reforming International Investment Governance," Tech. rep., UNCTAD, Geneva.

Vernon, R. (1971), *Sovereignty at Bay: The Multinational Spread of U.S. Enterprises*, Basic Books, New York.

Willem, D. (2004), "Foreign Direct Investment and Income Inequality in Latin America by and Income Inequality in Latin America," *ODI Research Papers*.

World Economic Forum (2013), "Foreign Direct Investment as key driver for Trade, Growth and Prosperity: The case for a multilateral agreement on investment," *World Economic Forum*, p. 36.

Wu, F. (1999), "Intrametropolitan FDI firm location in Guangzhou, China: A Poisson and negative binomial analysis," *The Annals of Regional Science*, 33, 535–555.

Yamawaki, H. (1991), "Exports, and Foreign Distributional Activities: Evidence on Japanese Firms in the United States," *The Review of Economics and Statistics*, 73, 294–300.

Zebregs, M. and Tseng, M. (2002), *Foreign direct investment in China: Some lessons for other countries.*