

McKenzie, David; Yang, Dean

**Working Paper**

## Field and Natural Experiments in Migration

CREAM Discussion Paper Series, No. 26/22

**Provided in Cooperation with:**

Rockwool Foundation Berlin (RF Berlin)

*Suggested Citation:* McKenzie, David; Yang, Dean (2022) : Field and Natural Experiments in Migration, CREAM Discussion Paper Series, No. 26/22, Centre for Research & Analysis of Migration (CREAM), Department of Economics, University College London, London

This Version is available at:

<https://hdl.handle.net/10419/295695>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# Field and Natural Experiments in Migration

*David McKenzie & Dean Yang*

November 2022

# Field and Natural Experiments in Migration\*

David McKenzie<sup>1</sup> and Dean Yang<sup>2</sup>

<sup>1</sup>Development Economics Research Group, World Bank

<sup>2</sup>Department of Economics, University of Michigan

Version: November 28, 2022

## Abstract

Many research and policy questions surrounding migration are causal questions. We want to know what causes people to migrate, and what the consequences of migration are for the migrants, their families, and their communities. However, answering these questions requires dealing with the self-selection inherent in migration choices. Field and natural experiments offer methodological approaches that enable answering these causal questions. We discuss the key conceptual and logistical issues that face applied researchers when applying these methods to the study of migration, as well as providing guidance for practitioners and policymakers in assessing the credibility of causal claims. For randomized experiments, this includes providing a framework for thinking through what can be randomized; discussing key measurement and design issues that arise from issues such as migration being a rare event, and in measuring welfare changes when people change locations; as well as discussing ethical issues that can arise. We then outline what makes for a good natural experiment in the context of migration, and discuss the implications of recent econometric work for the use of difference-in-differences, instrumental variables (and especially shift-share instruments), and regression discontinuity methods in migration research. A key lesson from this recent work is that it is not meaningful to talk about “the” impact of migration, but rather impacts are likely to be heterogeneous, affecting both the validity and interpretation of causal estimates.

*JEL Classification:* F22, J61, O15, C93, C23, C26

*Keywords:* Experimental Methods, Difference-in-Differences, Instrumental Variables, Regression Discontinuity, Natural Experiment, Migration

---

\*We thank the editors, Kirill Borusyak, Peter Hull and Xavier Jaravel for helpful comments. Ryan McWay and Laston Manja provided excellent research assistance.

# 1 Introduction

Deciding where to live is one of the most important decisions most individuals will ever make. Milanovic (2015) illustrates this starkly, showing that two-thirds of the variability of individual percentiles of income across the world is driven by the country in which you live. However, this decision of whether or not to migrate to somewhere else is a choice made by individuals and households, with this choice depending on abilities, wealth, risk preferences, information, family ties, ambition, and a host of other observable and unobservable characteristics. This self-selection lies at the heart of classic migration theories such as Borjas (1987), and poses a severe challenge for researchers and policymakers attempting to estimate the impacts of migration and migration policies. For example, suppose we observe that households with a migrant are more likely to be running a business than households without a migrant. This could reflect a causal effect of migration, where the remittances sent back by migrants enable households to invest in bigger businesses and knowledge flows provide new sources of ideas and trade linkages. But conversely, it could just as easily reflect selection effects, where households with more wealth and high risk-taking propensities are both more likely to engage in migration, and to start up and run new businesses. As a result, even if we condition on many observable characteristics, comparisons of migrants and non-migrants are unlikely to give convincing estimates of the impacts of migration.

Many of the key research questions of interest in the study of migration do involve trying to answer causal questions about either the impacts of migration on some outcome, or the impacts of some policy or contextual feature on migration. For example, what is the effect of migration on the well-being of other household members? How has immigration affected the employment prospects of natives? How have increasing climate risks changed the likelihood of migrating? How effective are government policy actions to encourage high-skilled migration? Answering these questions requires overcoming the selection issues at the heart of migration decisions.

Causal inference methods provide tools that can be used by researchers to help answer these types of questions. Experimental methods aim to overcome these difficulties by exploiting situations where the reason one household engages in migration or remits and another does not is truly the result of random chance. Randomized experiments do this by explicitly aiming to randomize migration opportunities or different variables entering into migration decisions. Natural experiments aim to mimic this process by finding situations where “nature” provides an exogenous source of variation. In the first edition of this handbook, we provided an introduction to this approach (McKenzie and Yang, 2012). The subsequent decade has seen rapid growth in the use of these methods, and enabled a number of methodological lessons about how to best apply these approaches to be learned.

Our goal in this chapter is to set out the key methodological approaches and a framework for thinking through the design and use of experimental and non-experimental causal

inference methods in studying migration. We do not attempt to provide a comprehensive literature review, or to summarize the key findings of this literature, but rather to draw out more general lessons about how to find or induce random variation, and of key considerations that need to be kept in mind when using these approaches.

We begin with a short discussion to illustrate the identification problem that these methods aim to solve. The first half of the paper then considers the use of randomized experiments in migration. We discuss what can be randomized, including both changes in migration policies, as well as changing different elements of the migration decision-problem such as the costs of migrating, the ability to pay these costs, the information and beliefs individuals have in making their migration decisions, the risks that they face, the wages or amenities that are available at origin or destination, and different behavioral factors. We outline five logistical and methodological issues to consider when implementing migration experiments: migration is a rare event, measuring outcomes when people move is difficult, it can be crucial to measure beliefs, measuring welfare gains becomes more complicated when people change locations, and the importance of being clear about what treatment effect is being estimated and for whom it applies. We conclude our discussion of experiments by discussing two ethical issues that particularly apply when doing experiments in a migration context: the stakes can be much higher than for many other experiments commonly done by researchers, and some migrant populations, such as refugees, can be particularly vulnerable.

The second half of the paper then turns to the use of natural experiments. The terms “natural experiment” and “quasi-experiment” are sometimes used interchangeably (e.g. Meyer (1995)), but others (e.g. Remler and Ryzin (2014)) make a distinction between natural experiments as situations where there is a transparent exogenous source of variation, as with policy changes or shocks coming from nature that were not explicitly intended to influence the outcome of interest; and quasi-experiments as cases where a treatment was implemented, but assignment was not random, and different statistical methods and functional form assumptions are used to try to remove any potential selection effects. We discuss what makes for a good natural experiment, as well as common approaches used in the migration literature that we think are less convincing as natural experiments. We begin by outlining the issues that arise in attempting to measure causal impacts in migration research through the use of quasi-experimental methods like regression and matching that require selection on observables, and discuss the conditions needed for these methods to be at least somewhat convincing. One key consideration here is whether there is an external policymaker with limited information that differs from the person making migration choices. We then provide more in-depth discussion of two non-experimental methods that have been the focus of substantial recent econometric research. The first is difference-in-differences (DiD) estimation.<sup>1</sup> We outline the types of DiD commonly used in migration research, and how new

---

<sup>1</sup>Lozano and Steinberger (2013) provide an introduction to some of the early work on difference-in-differences in immigration studies in the first edition of this handbook.

econometric research affects their interpretation and validity. The second is instrumental variables estimation. We note that there are many bad or unconvincing instruments used in migration research, and that Bartik or shift-share instruments have been thought of as more plausible, but have also been the subject of several recent econometric advances. We outline in detail issues to consider in using shift-share instruments for migration. Finally, we also consider the possibilities for using regression-discontinuity methods for migration research, discussing how language tests, points systems, and discontinuities in time can and cannot be used.

An underlying theme in much of the recent econometric research is that treatment effects are likely to be heterogeneous, and this heterogeneity has implications both for the validity and interpretation of many estimates. We conclude by noting this means we should aim to move away from talking about “the” impact of migration, and that likely areas for future work will involve approaches that work to better identify who is most likely to be affected by migration policies, use machine-learning methods to better estimate treatment effect heterogeneity, and consider external validity issues more fully.

## 2 What is the problem experimental methods help to solve?

Many research and policy questions surrounding migration are causal questions.<sup>2</sup> In the immigration literature this typically involves estimating the impact of immigration on wages or employment rates of natives. A large part of the development literature in migration attempts to answer questions like what is the impact of migration on the incomes, health, and life opportunities of migrants; and what is the impact of having a household member migrate or of receiving remittances on outcomes of household members remaining in the sending community, such as the education and health of children, their levels of entrepreneurship and labor supply, and on community poverty and inequality levels. A common approach to answering such questions is to use data on migrants and non-migrants, and attempt to control for differences between them in a linear regression framework. For example, researchers may attempt to estimate an equation of the form:

$$\mathbf{Y}_i = \alpha + \beta \text{Migrant}_i + \delta' \mathbf{X}_i + \varepsilon_i \quad (1)$$

Where  $\text{Migrant}_i$  is a dummy variable which takes the value 1 if the individual is a migrant, and zero if he or she is not, and  $\mathbf{X}_i$  observed characteristics of the individual which are presumed not to have changed with migration (e.g. age, sex and education level,

---

<sup>2</sup>This section draws heavily on our chapter in the first edition of this handbook (McKenzie and Yang, 2012).

location of birth, ethnicity, religion, parental education, etc.).<sup>3</sup> Then in order for the linear regression estimate of  $\beta$  to give the causal impact of migration on the outcome of interest, we require that:

$$E(Migrant_i \varepsilon_i | \mathbf{X}_i) = 0 \tag{2}$$

That is, we require the unobserved determinants of the outcome of interest (income in our example) to be uncorrelated with whether an individual migrates once we have conditioned on the observable characteristics of these individuals. But in the absence of experimental variation in migration, this assumption is unlikely to hold. Indeed, the seminal migration selectivity model of Borjas (1987) has migrants deciding whether or not to migrate in part on the basis of the  $\varepsilon_i$  they would expect to have at home versus abroad. It is easy to think of a whole range of typically unmeasured variables, such as entrepreneurial prowess, ambition, language proficiency, and health status which would both affect whether someone migrates, and also directly affect their income or other outcome of interest. Likewise, liquidity constraints will likely determine both the pattern of self-selection of migrants (McKenzie and Rapoport, 2010), as well as the range of job opportunities and consumption-smoothing opportunities that individuals will have at home. As a result, equation 2 will almost always be violated in practice, so that linear regression on equation 1 will result in biased estimates of the impact of migration.

Equation 2 therefore says that we can only estimate the causal impact of migration if the only reason one person migrates and another does not is random (conditional on observed characteristics). Natural experiments and randomized experiments both aim to use a source of random (or “as good as random”) variation to identify these causal impacts.

### 3 Using randomized experiments to study migration

In a randomized experiment, the units of interest (individuals, households, communities, etc.) are randomly assigned into different groups using a randomization device such as lottery, computer random number generator, or physical drawing. The simplest example involves just two groups: a treatment group who are offered some intervention (e.g. the opportunity to migrate) and a control group who are not. More complicated experiments can involve additional groups in order to compare the effectiveness of different types of treatments (e.g. Beam et al. (2016) randomly assign individuals in the Philippines to a control group or to one of four different treatment groups to compare different approaches

---

<sup>3</sup>For simplicity we consider only the case of an individual-level outcome here, assuming that all migrants and non-migrants are observed. When the comparison involves households with and without migrants, a second form of selectivity is involved, since households can also choose whether all members migrate, or only some. Return migration also introduces a third form of selectivity. See Gibson et al. (2013) for discussion of this more complicated case.

to facilitating international migration), or to explore potential complementarities between treatments (e.g. Doi et al. (2014) randomize households in a 2x2 design, where financial education is provided either to the migrant, to the remaining household member, to both, or to neither in order to measure whether there are complementarities from having both the migrant and family member trained).<sup>4</sup> Since which group a unit is in is randomly assigned, in expectation it is not correlated with any other determinants of the outcome of interest, and so equation (2) holds. Comparing treatment and control households then enables estimation of the causal effect of being offered treatment.

Duflo et al. (2007) and Glennerster and Takavarasha (2013) provide introductions to the use of randomized experiments in estimating causal impacts, and plenty of practical advice for designing and implementing experiments. Our focus will then be on issues that are specific to applying this methodology in a migration context.

### 3.1 What can be randomized?

In McKenzie and Yang (2015), we survey the evidence on the effectiveness of different types of policies to enhance the development impacts of international migration. We note there that there are a wide range of migration policies that occur throughout the different stages of the migration process. These include policies that occur prior to migration in order to offer more opportunities to migrate, and to ensure people are well-informed when making these decisions; policies that occur while migrants are abroad such as efforts to help migrants integrate successfully into their new communities, and to ensure they have access to appropriate financial products including ways to remit money back to remaining family members; and policies designed to help those who are planning on returning to their origin countries, such as reintegration programs, and return incentives.

A first avenue for using randomized experiments in studying migration is to randomize who gets to benefit from these policies, or to randomly test alternative versions of these policies. Often this is justified in terms of there being limited resources or opportunities that prevent the program from being offered to everyone at once, or it can be an explicit attempt by policymakers to understand which version of a program works best. This typically requires researchers to work with governments, NGOs or private sector organizations that are administering these programs, and enables estimation of the program effect. Several examples of this approach from our own work include: evaluating the effect of migrating through visa lotteries from Tonga and Samoa to New Zealand (McKenzie et al., 2010; Gibson et al., 2013), working with the government of Indonesia to test alternative ways of teaching

---

<sup>4</sup>A newer area is to use adaptive experiments, in which individuals are randomized to different treatments in batches or cohorts, and the results from the first batches are then used to inform the treatment assignment for subsequent individuals. This works best when the outcome of interest is one that can be measured relatively quickly. Caria et al. (2020) provide an example, in adaptively testing different approaches to helping refugees in Jordan find immediate employment.



financial literacy to migrants (Doi et al., 2014), working with a private bank in Italy to test the effect of a new product that allowed migrants to directly pay for school fees in the Philippines (de Arcangelis et al., 2015), working with an NGO to give U.S.-based migrants a new mechanism to fund students’ education in El Salvador (Ambler et al., 2015), working with a private bank to test the savings-accumulation impact of giving U.S.-based migrants more control over bank accounts in El Salvador (Ashraf et al., 2015), and working with the government of the Philippines to test different versions of a pre-departure orientation program (Barsbai et al., 2020). In practice this can involve prospective studies that actively work with a government or other policy actor to randomize a new policy, as well as ex post utilization of randomization conducted by governments, as was the case with the visa lottery work of Gibson et al. (2017) and Mobarak et al. (2020).

An alternative approach is to take the existing policy environment as given, and conduct randomized experiments that change other parameters of the decision problem. For example, consider a potential migrant deciding between staying in their home country where they receive home wages  $W_H$  and amenities  $A_H$ , or paying a cost  $C$  to migrate abroad, where they would face uncertain wages  $W_M$  and amenities  $A_M$ . Then the individual will choose to migrate if the expected gain in utility from migrating exceeds the costs of migrating, and they can afford the costs of migrating. That is, their decision is:

$$\begin{aligned} \max_M E[U(W_M, A_M) - U(W_H, A_H)] - C \\ \text{s.t. } C < \lambda B \end{aligned} \tag{3}$$

where  $B$  is wealth, and  $\lambda$  reflects how tightly liquidity constraints bind.

This framework then highlights several areas where researchers can conduct experiments around migration decisions without needing to change or randomize migration policies. A first area is to experiment with *changing the costs*  $C$ . For example, Hainmueller et al. (2018) randomly lower the cost of naturalizing by giving immigrants vouchers to cover the costs of the naturalization fee; and Beam et al. (2016) randomly lower the cost of migrating by fully subsidizing the cost of a passport. The cost may not just be a monetary cost, but also the logistical and bureaucratic costs of filling out paperwork, waiting in lines, etc., and these studies also provided treated subjects with assistance to lower these costs. Lowering costs can also be a useful way of understanding other parts of migrant decision-making, such as remitting decisions and communication. For example, Ambler et al. (2014) randomly assign migrants different discounts on remittance transaction fees, while Batista and Narciso (2018) randomly vary the costs of calling home by giving migrants international calling credit.

A second area, rather than changing the costs, is to *change an individual’s ability to pay these costs*, either by increasing their wealth  $B$ , or their access to credit, and thus how tightly liquidity constraints bind,  $\lambda$ . Bryan et al. (2014) provide an example of both approaches in

attempting to spur internal migration in Bangladesh. They give some households cash that they can use to pay for the costs of a bus into the city, and offer others a loan to pay for these costs. This also offers the scope for researchers to examine how non-migration-specific programs that increase household wealth affect migration decisions. For example, Gazeaud et al. (2022) examine how a cash transfer randomly offered to households in Comoros through a public works program affects their migration to Mayotte. Angelucci (2015) studies how the randomized Mexican Progresa cash transfer program affected migration to the U.S.

Thirdly, hidden in the expectation term  $E$  in equation 3 is that migration decisions depend on the *information and beliefs* that potential migrants have about the wages and amenities available abroad. In many cases this information may be incomplete, and beliefs may be wrong. This raises the possibility of information interventions designed to change these beliefs. For example, Shrestha (2020) randomly gives Nepalese potential migrants information about both the wages and mortality incidences at destination, and Baseler (2021) provides rural Kenyans with information about earnings in urban areas. Seshan and Yang (2014) offer migrants a motivational workshop encouraging them to improve their financial habits (e.g., raising their savings levels). Shrestha and Yang (2019) provide immigrants in Singapore with information intended to facilitate job-to-job mobility. Barsbai et al. (2020) give new U.S. immigrants information to promote assimilation and settlement. Information campaigns are particularly popular among policymakers seeking to deter dangerous and irregular migration, such as migration across the Mediterranean to Europe. Tjaden (2020) provides several examples of IOM experiments designed to test these awareness campaigns, and Bah et al. (2022) test the impact of an information intervention on irregular migration from The Gambia.

Fourth, the expected utility term,  $E(U)$  also recognizes that migration is a risky and uncertain endeavor, and so decisions will also depend on risk preferences and the ability to insure against risk. This suggests the scope for *experiments which provide insurance*. This is a relatively underexplored area for experimentation. Bryan et al. (2014) provide one exception, conducting experiments in which loans for migrating were bundled with insurance, so that migrants would not need to repay if there was excessive rainfall at destination that would reduce job opportunities.

Fifth, experiments could aim to *change wages  $W$  or amenities  $A$  at either destination or origin*. One way of doing this would be to connect migrants to labor brokers, to help them match to jobs with either a higher likelihood of employment, or that offer better wages or better conditions. Beam et al. (2016) try this by setting up a website to directly link potential migrants to reputable recruitment agencies and to allow these agencies to post job opportunities. Bazzi et al. (2021) provide potential migrants with quality ratings information about labor intermediaries in the Philippines, with the aim of helping them choose intermediaries who can link them to better job conditions abroad. Other examples

could include experimenting with the housing as in Qiu et al. (2021), social opportunities, and other amenities migrants have at destination, and experimenting with improving job opportunities in the origin region through training programs or other job programs.

Finally, there are other aspects of the migration decision that this simple model in equation 3 does not explicitly incorporate, but where experimentation may also be possible. A notable example is behavioral factors. Some of these (such as attention costs) could be included in  $C$ , others (such as loss aversion and impatience) could be included in more general formulation with decision-making over multiple periods and with an alternative to expected utility maximization. This can lead to experiments designed to overcome some of these behavioral constraints. For example, Hainmueller et al. (2018) experiment with behavioral nudges such as reminders and text messages to encourage immigrants who are eligible for a federal fee waiver to take advantage of this.

In addition to the above experiments which change actual migration policies or inputs into the migration decision, there are other types of experiments which just vary perceived characteristics. One example is audit studies, which randomly vary characteristics of individuals on resumes and then measure the impact of this characteristic on interview callbacks. In the migration domain, this has been used to examine whether employers value migration experience of returnees (e.g. Abarcar (2015) and Wah (2018)), and whether they discriminate against immigrants from different countries (e.g. Duguet et al. (2010)).

### **3.2 Some key logistical and methodological issues that arise when conducting migration experiments**

Running any randomized experiment involves many key design issues, data collection issues, and implementation issues. We focus here on some specific issues that are particularly important to consider when conducting migration experiments.

*Migration is usually a rare event.* This is particularly the case when considering international migration. For example, even in a country known for international migration like Mexico, where 870,000 Mexicans moved from Mexico to the U.S. in the five-year period between 2009 and 2014 (Gonzalez-Barrera, 2015), this five-year flow represents less than one percent of Mexico’s population. This means that if researchers start with a random sample of the population in the origin country and implement an experiment designed to change migration decisions, they will need incredibly large samples, or lack statistical power for detecting impacts. This was the case in work we did in the Philippines, in Beam et al. (2016), where we found that few people migrated even after we offered substantial migration assistance. This is still useful for understanding whether certain barriers to movement are binding, but is a challenge for researchers intending to examine migration outcomes. An alternative is to select a sample at much higher risk of migration. For example, Shrestha (2020) recruited subjects who had just finished applying for passports in Nepal. An in-

between option can be to implement some general screening for characteristics associated with higher migration likelihoods, such as focusing on young men in high migration areas.

One complementary suggestion is for researchers to make sure they collect detailed information on the whole set of steps needed for migration to occur. For example, an intervention to increase migration may first raise an individual's desire or intention to migrate, then lead them to take steps to look for job opportunities in different places, then lead them to get their appropriate documentation in order and to apply for jobs, and finally to get them to move. Even if studies lack sufficient statistical power to detect impacts on this final stage of the migration process, they may find impacts on these migration steps, and in doing so, learn about additional constraints that bind once individuals have decided they want to migrate.

The rareness of migration is also an issue when attempting to conduct experiments with immigrants, since it can be difficult and expensive to collect a representative sample of immigrants. McKenzie and Mistiaen (2009) show how intercept-point sampling can be used to achieve reasonably representative samples at a cost much lower than census-based sampling. Intercept-point sampling involves sampling individuals during set time periods at a pre-specified set of locations where households in the target group are likely to congregate, and then re-weighting the sample so that individuals who go to more of the locations (and are hence more likely to be sampled) receive less weight. An alternative has been to use convenience samples of customers of remittance organizations, clients of an NGO working with immigrants, or other such approaches.

*Measurement of outcomes can be challenging when people are moving.* It is costly and challenging to track individuals who move. In many randomized experiments, migration therefore manifests itself as attrition, and researchers try and assure the reader that it is limited and balanced across treatment groups. But when the focus is on migration as an outcome, or mobile individuals as the subjects of interest, much more care and effort needs to be devoted to tracking individuals. Of particular concern is the possibility that all of the treatment effect occurs among attriters. For example, consider an experiment designed to deter irregular migration, perhaps through an information campaign. If researchers return a year later and re-interview 80% of individuals, they might find very low rates of irregular migration in both the treatment and control groups. But the 20% of individuals who cannot be found might be precisely the ones who are most mobile and most likely to be affected by the treatment. There are seldom easy approaches to dealing with this issue, but rather the need for a range of methods to be used. These can include collecting lots of contact information at the beginning for friends and family members who may be able to report on individuals who have moved and provide assistance relocating them, trying to use technologies such as Whatsapp or other online communication methods that migrants may retain when they move, attempting to track individuals and interview them in their new

locations, and then employing different bounding approaches and reweighting approaches to examine how sensitive the results might be to different assumptions about individuals who are not found. McKenzie (2015) provides more details on data collection for migration experiments.

*Measurement of beliefs can be crucial.* Equation 3 makes clear that migration decisions depend not just on the actual wages and amenities in different locations, but also on what individuals believe these to be. Haaland et al. (2022) offer a nice recent guide to conducting information experiments, which offers several lessons for things to measure in migration experiments. First, the expected directional response to an information treatment depends on what people believe in the first place, and we should expect heterogeneity in the responses. For example, individuals who under-estimate wages at the destination may increase their migration likelihood when given the correct information, whereas individuals who over-estimate wages may decrease this likelihood. Looking only at the average effect without measuring baseline beliefs would miss this important heterogeneity. Second, there are a lot of design issues in how to provide information, and it is important to think carefully about whether the information provided is credible, and whether potential migrants will trust the information. Third, you want to then measure how beliefs are updated, so that we can distinguish between a treatment not working because the information did not change beliefs, versus it not working because other constraints were more important.

*Measuring welfare improvements can be difficult when people change locations.* Migration can deliver large gains in incomes. For example, Gibson et al. (2017) use a visa lottery to show that migrating from Tonga to New Zealand leads to a 263 percent gain in income in the first year, with these benefits persisting over time for a lifetime benefit of US\$237,000. However, the change in location also means migrants face different costs of living, different amenities (e.g. different quality schools, differences in opportunities for their children, differences in access to nature, differences in access to cultural opportunities, etc.). The logic of equation 3 and revealed preference would suggest that migrating (at least in expectation) should make people better off. But, as we have discussed, these expectations may be based on incorrect information in some cases, or migrants may have correct expectations but still be unlucky and not earn as much as expected.

A partial solution to this problem is to attempt to measure a broader range of outcomes, including direct measurement of amenities and survey measures that may better capture well-being than income or expenditure. Examples include subjective happiness, or mental health. But these subjective well-being measures also face difficulties when people change locations, since their reference groups may change. Stillman et al. (2015) find that impacts on subjective well-being differ depending on whether migrants use their previous life as a reference or compare themselves to their contemporaneous surroundings. Likewise, tastes and preferences may change over time as migrants assimilate into a new culture and form

new friendships and networks at destination. Sarvimaki et al. (2022), for example, provide a model of habit formation for residential location, where people grow more attached to a place the longer they live there. McKenzie (2022) highlights a number of psychological factors such as a bias toward the status quo and the inability to picture what one is giving up by not migrating, that can result in people not moving, even when they would benefit from movement. Researchers should therefore take care when measuring welfare to be clear what they are measuring, be careful about relying on revealed preference, and potentially use vignettes or more detailed descriptions to try to ensure respondents are using the same reference points and scales.

*What is the treatment effect being estimated, and for whom does it apply?* In most experiments, what is randomized is the offer of an intervention or program, and then comparing the treatment and control groups will give the intention-to-treat (ITT) effect for the sample of individuals exposed to the experiment. Non-compliance with this treatment can arise because of all the other constraints that affect migration (so that many of those offered a migration treatment may still not migrate), and/or because the sample selected is one that is so eager to migrate that they will find alternative ways of doing so if not selected for treatment (some of those in the control group may still migrate). If the treatment only affects outcomes of interest (e.g. income, well-being) through migration, then the local-average-treatment-effect (LATE) can be estimated using assignment to treatment as an instrument for migration, giving the treatment effect for compliers. If it is further the case that none of the control group can migrate, this can also give the treatment effect on the treated. McKenzie et al. (2010) provides an example in the context of a visa lottery. However, this relates back to the point of being clear as to what the sample of interest for the experiment is, and raises the standard issue of external validity. Under a Borjas (1987) model of self-selection, individuals will self-select into migration based in part on unobserved components of their wages at home and abroad, so that the gains from migration to those most interested in migrating may exceed those in the general population. However, if the main barrier to migration lies in the costs of migrating and ability to pay these costs, or in incorrect information, it could conversely be that the individuals who are least interested in migrating to begin with are the ones who have the most to gain from actually migrating. Indeed, Sarvimaki et al. (2022) provide an example and links to several other papers showing that forced migration appears to have positive long-term benefits on agricultural workers and youth who were otherwise unlikely to have moved.

### **3.3 Ethical considerations in conducting migration experiments**

There has been a vigorous debate about a number of ethical issues involved in conducting randomized experiments. Glennerster and Powers (2014) provide a discussion of many of the general issues and advice for researchers in balancing the risks and benefits of this approach.

We note here two specific issues that arise when conducting experiments around migration.

*Migration can involve much bigger stakes than many other experimental decisions.* Migrating to a new place is one of the most major decisions people can make in their lives. The potential benefits can be enormous, but so can the potential costs in some cases (e.g. family separations, dangerous journeys). This is often coupled with immigration being a topic that attracts substantial political debate and public attention. The result is likely to be a much stronger emotional response from the public to experimentation than might be the case when randomizing whether someone gets sent a few text messages or gets given a small grant. For example, in an experiment testing different alternative approaches to reducing irregular migration from the Gambia to Europe (Bah et al., 2022), one of the authors has seen people raise concerns about (i) what right do researchers have to try to reduce such migration, when it offers the potential to dramatically increase incomes far more than any other option individuals have available at home; but also (ii) whether the participants may actually overstate the risks of this dangerous migration, and so giving them the correct information could spur more of them to undertake these risky journeys. These conflicting concerns are indeed a prime reason for experimentation and careful measurement before widespread policy adoption of such policies, but often ethical concerns can bias towards inaction.

*Some migrant populations are especially vulnerable and require particular care.* A classic example has been refugee populations, where many of the organizations working with these groups have a desire to help everyone (even if their resources do not allow it), and feel particular unease about control groups that do not receive assistance for some time (even if they would not be able to serve everyone anyway). Testing different approaches to find out how best to help refugees as in Caria et al. (2020) may raise fewer concerns. Another example is that in many contexts, migration may occur without documentation. Asking specifically about legal migration status may raise concerns about this information getting used for law enforcement purposes. Taking additional care to separate personal identifying information from the main data is important here, and in contexts where this is a particular concern, methods of indirectly measuring sensitive status could be employed. For example, McKenzie and Siegel (2013) illustrate how list randomization can be used to measure the proportion of migrants without documents. A downside is that these approaches for eliciting sensitive information can reduce statistical power with a given sample size compared to direct elicitation.

## 4 Using natural experiments to study migration

When a randomized experiment is not feasible, researchers can instead use “natural experiments” to understand causal relationships. When analyzing a natural experiment, a

researcher exploits variation that happens “naturally” in the real world, rather than in the context of a randomized experiment, to reveal the causal impact of a right-hand-side (independent) variable on a left-hand-side (dependent) variable. Dinardo (2008) calls natural experiments “serendipitous randomized trials”. Examples in migration research include using sudden policy changes (with difference-in-differences), shocks like exchange rate shocks, weather shocks, or shocks to earnings opportunities (in panel data or instrumental variables settings), and policy thresholds (in regression discontinuities). This requires using non-experimental causal inference methods. Good introductions to these methods are covered in Angrist and Pischke (2009) and Cunningham (2021). We start by outlining some key considerations of what makes for a good and not-so-good natural experiment, and explain why non-experimental methods that do not rely on a natural experiment such as regression, matching, and simple panel data fixed effects models can be subject to concerns in migration work. We then highlight key issues involved in employing different methods of causal inference with natural experiments in migration research.

#### **What is a “good” natural experiment?**

A “good” natural experiment makes a credible, explicit case that the right-hand-side (independent variable) variation being exploited is exogenous. A case for the credibility of a natural experiment typically has two components. First, researchers need to make a “rhetorical” or *a priori* case that the variation they are exploiting is plausibly exogenous. A rhetorical case for exogeneity is often more convincing when the variation being exploited cannot be controlled or influenced by the units of observation in the study (e.g., individuals, households, or other groupings such as localities). For example, changes in government policy (e.g., immigration policy in destination countries), or overseas economic shocks can often be credibly argued to be exogenous from the standpoint of individual prospective migrants or their origin households. To make a plausible rhetorical case, researchers should be able to argue that the variation they are exploiting could not have been influenced by the units of observation, was unanticipated, and is not correlated with their pre-existing or underlying characteristics.

Second, researchers need to show “statistical” plausibility: formal statistical tests that help establish that the variation they are exploiting is plausibly exogenous. The most appropriate such tests depend on the exact natural experimental empirical approach, and we go into details below. Such tests often include placebo or “false” experiments in pre-periods (periods prior to the occurrence of the natural experiment). For this reason, statistical plausibility is easier to establish when one’s empirical analyses involve panel data. Panel data allows inclusion of unit-of-observation fixed effects, which help control for time-invariant differences across units, shifting focus to explaining *changes* over time. It is also very helpful when panel data include more than one time period of data for each unit prior to the occurrence of the natural experiment, to allow statistical tests of pre-trends (e.g., tests of



the parallel-trend assumption in difference-in-difference analyses).

### Signs of a “bad” natural experiment (“bad IVs”)

With these guidelines in mind, we highlight some danger signs of a “bad” natural experiment, one that does not make a credible case for the exogeneity of its key right-hand-side variation.

It is often difficult to make an *a priori* case for rhetorical plausibility when seeking to achieve causal identification using only cross-sectional data (data at just a single point in time for each observation). With cross-sectional data, the researcher cannot control for fixed effects for each unit (e.g., individuals or households). Inclusion of unit fixed effects in a regression helps control for time-invariant differences across units, but requires panel data (more than one observation per unit). Only in unusual cases can one make a case for rhetorical plausibility. An example is Shrestha (2017), who uses cross-sectional Nepalese data to analyze the impact of a change in an educational requirement for a migration opportunity varies plausibly exogenously across cohorts (birth-year groups) and localities.

Another danger sign is simply using one existing variable in a dataset as an instrumental variable (IV) for the key endogenous (right-hand-side) variable of interest, without a clear argument for the rhetorical plausibility of the instrument. Examples of variables that have been used to instrument for the migration decision include network instruments, such as leave-one-out instruments (instrumenting own migration with that of others)<sup>5</sup>, or village lagged migration rates; being the oldest child in a household; the gender of oldest child; parental education; proximity to transport; and proximity to remittance services. These variables are often found to be correlated with migration (or related outcomes like remittances), and thus are argued to have a strong “first stage” in IV estimation. But each one is quite likely to violate the IV exclusion restriction: the assumption that the instrument affects the second-stage outcome of interest only via the instrumented endogenous variable. The exclusion restriction cannot be tested directly, so it is crucial to make a strong rhetorical case for its plausibility.

## 4.1 Selection on Observables by Regression and Matching approaches

Before discussing how natural experiments can be used, we note that the traditional way of estimating impacts has typically been based on an attempt to control for observable factors in regression or matching. Consider estimating the impact of migration on income in an equation like equation 1. In the absence of a randomized experiment, we might be concerned that migrants and non-migrants differ in ways that also determine income. For example, migrants may tend to be younger, more educated, and male. The hope with linear regression and matching is that controlling for these observable differences will be sufficient to allow the regression error term to be uncorrelated with migration, so that equation 2 holds. For

---

<sup>5</sup>McKenzie (2021a) discusses the issues involved in more detail

example, Clemens et al. (2019) compare the wages of immigrants from 42 countries in the United States with those of individuals of the same gender, age, and education level in the origin countries to get an estimate of the wage gain from migrating. Methods such as propensity score matching attempt to match migrants to similar non-migrants on the basis of much larger sets of covariates, whilst ensuring that these comparisons lie in the common support and do not involve the linear extrapolation that regression can entail. For example, Gibson and McKenzie (2014) match Tongan and ni-Vanuatu households that participate in a seasonal worker program to similar households matched on household demographics, characteristics of the males most likely to migrate such as their health status and work experience, previous migration histories and networks, household assets and infrastructure, geography, and past wage and salary history. Another example is Dustmann and Fabbri (2003), who use a matching approach to estimate the impact of English language proficiency among UK immigrants.

The challenge for these methods is that migration theory is full of examples of how individuals may self-select into migrating on the basis of variables that may be difficult to observe or measure, such as unobserved ability (Borjas, 1987), risk preferences, language skills, exposure to shocks, entrepreneurial aptitude, etc. This raises two issues for practical use. The first is that since many surveys will not measure all of these features, it will typically be difficult to convince a skeptical reader that one has controlled for all of the important characteristics of individuals that both influence their migration decisions and that may also be correlated with the outcomes of interest. The second is that because migration is typically a rare event, most individuals in the population will not be good matches for those who have migrated, dramatically reducing the effective sample size. For example, McKenzie et al. (2010) match migrants from Tonga who applied for a migration lottery program in New Zealand to Tongans in a nationally representative labor force survey. They find that their sample drops from over 4,000 observations to just 354 when restricting to individuals with propensity scores between 0.05 and 0.95 – reflecting that most individuals in the labor force survey are too dissimilar to be used as controls for the migrants.

The result is that in many cases we do not believe matching (or regression) will yield reliable and precise point estimates of the impacts of migration. How much this matters depends a bit on the question of interest. In some cases, the impacts of migration may be so large that the bias from selection is second-order (McKenzie, 2020). For example, McKenzie et al. (2010) use a randomized lottery to estimate the income gain from migrating to be NZ\$274 per week, a 263% increase on the control mean. Using regression or matching overstates this gain, with estimates of around NZ\$330-350 per week, which are 20-28% higher than the experimental estimate. But in both cases one would conclude that migrating resulted in a massive gain in income, and this bias from matching is an order of magnitude smaller than the treatment effect. Bounding methods can also be used to ask how sensitive

the results are to different degrees of selection, as is done by Clemens et al. (2019).

However, in many other cases the effects we are trying to measure are not so large or obvious, and then the bias from matching or regression is much more of a concern. Then McKenzie (2021b) argues that there needs to be a rhetorical argument made for why two similar units have a different treatment status, with several examples offered of when this might occur. In the migration context, two such cases may be most useful to consider. The first is when there is a separate decision-maker with limited information who is deciding on treatment. For example, individuals may self-select into applying for a migration program, and then an employer or recruiting agency may just decide which individuals to recruit based on a few variables on an application form (e.g. gender, height, language, work experience). If the researcher is able to observe the same information as this decision-maker, then they might be able to plausibly argue that, conditional on having the same application information, which individuals end up getting selected by the recruiter or employer is as good as random. A second case is when there are capacity limits so that not everyone who wants to migrate is able to, and some small frictions or noise determining who gets to take advantage of this opportunity. Gibson and McKenzie (2014) provides an example combining both ideas. They consider a new seasonal migration opportunity in Tonga and Vanuatu, in which workers decide to apply for a pool, that employers then choose from. Since the program is new, not everyone finds out about the program right away, and employers in New Zealand recruit batches of workers at a time, and only observe limited information about each worker when deciding who to hire. The result is that similar individuals with the same desire to migrate end up in different migration statuses through no choice of their own.

The above examples are cases where migration is essentially driven by some randomness, such as in recruiter or employer idiosyncratic choices or through the introduction of a new policy and some small frictions. They also highlight when matching or regression are less likely to be applicable. For example, if there are no limits on migration (as with internal migration, or irregular migration), and individuals or households are self-selecting into whether or not to migrate, then those who do migrate are likely to differ in unmeasured ways from those who do not. Likewise, matching is unlikely to be helpful in estimating the impacts of remittances, since in almost all cases, the choice of whether or not to remit will be done by the migrant and household, not some external decision-maker, and there are unlikely to be capacity constraints that prevent some people who want to remit from doing so while allowing others to send money.

A final note for researchers using these approaches is it is crucial to match only on the basis of variables that are unaffected by the migration decision of interest. This raises a further challenge in cross-sectional studies, since many of the key determinants of migration decisions that we would like to match on (household structure, household wealth, networks,

language skills, poverty, etc.) are themselves likely to have been affected by the migration decision. As such, anyone designing migration surveys that are intended to estimate causal impacts using these methods should endeavor to ask retrospective questions about household characteristics at some point in time prior to the migration events.

## 4.2 Panel Data, Fixed Effects, and Difference-in-Differences

### 4.2.1 Fixed Effects

Panel data provides multiple observations over time of the unit of observation (e.g., individual, household, or locality).<sup>6</sup> This allows controlling for not just the observable characteristics of these units, but also non-observable time-invariant characteristics that might be correlated with both migration and the outcome of interest. For example, consider estimating the impact of migration of a household member on child schooling in that household. The concern with using regression or matching is that migrant households may differ in unobservable ways from migrant households due to factors that also influence schooling. For example, they may live in communities with worse schools (and migrate to escape poor infrastructure), they may care a lot more about schooling (and migrate to give their children schooling opportunities) or they might speak a different language from non-migrants (that is rewarded more at destination). With panel data, equation 1 is supplemented by adding the individual fixed effect term  $\alpha_i$ , and time effect  $\lambda_t$  to give:

$$Y_{i,t} = \alpha_i + \lambda_t + \beta \text{Migrant}_{i,t} + \delta' \mathbf{X}_{i,t} + \varepsilon_{i,t} \quad (4)$$

This individual fixed effect  $\alpha_i$  will capture the influence of any time-invariant characteristic like how much the household cares about education, the starting infrastructure in their neighborhood, and the language they taught their children, even when we cannot directly observe these characteristics. Estimation of the impact of migration,  $\beta$ , then comes from units that *change* migration status; households that happen to have a migrant member in all time periods in the data, or no migrant members in any time period, do not contribute directly to the estimates.<sup>7</sup>

The fact that identification is now coming from units that change migration status raises three important questions. The first is how many units actually change status? As we have discussed, migration is typically a rare event, and so over a short panel, most households may not change their migration status. For example, Chen et al. (2009) use a panel of 1,358 Chinese households between 2002 and 2006, and find that only 220 (16%) had

<sup>6</sup>Funkhouser (2013) provides an introduction to the use of longitudinal data in studying migration and remittances in the first edition of this Handbook.

<sup>7</sup>Households that never have changes in whether they have a migrant member may still affect coefficient estimates *indirectly* by contributing to other coefficient estimates in a regression equation, such as time-period fixed effects.

engaged in new internal migration over this 4-year period. With international migration, even fewer households are likely to change status, reducing statistical power and requiring large samples. The second question is who are these households that change migration status, and in particular, are the impacts of migrating different for them than for those who do not change migration status? In equation 4, we had written the impact of migration,  $\beta$ , without any subscript. But in practice, the impacts may be heterogeneous, and so the causal impact identified by fixed effects may differ from the average treatment effect. This point has been part of an important recent debate in the literature about the extent to which there is mis-allocation of labor through too little rural-to-urban migration. Incomes and productivity are much higher in urban areas than rural areas in many developing countries. However, using panel data from Kenya and Indonesia, Hamory et al. (2021) find the wage gains from moving to be relatively small. One potential explanation is that the individuals who would gain most from moving are constrained by lack of information, networks, or wealth, and so, since they are never observed to migrate, do not contribute to the estimate of the migration impact. The large gains in income seen by Baseler (2021) in a randomized experiment which provides information is consistent with this viewpoint.

The third, and most critical, question for the use of fixed effects is what caused some units to suddenly change their status? That is, our concern is that

$$E[\Delta \text{Migrant}_{i,t} \Delta \varepsilon_{i,t}] \neq 0 \tag{5}$$

For example, in the Chen et al. (2009) study, we might be concerned that households experienced a shock that both affected their desire to migrate and the educational performance of their children. For example, a child may experience a negative health shock, which might reduce their performance in school, and entail large health expenditures in the household and less money to spend on education. A household member may then migrate to earn additional income to pay these expenses. Such an effect would bias the effect of migration downwards, since it is occurring due to shocks that also reduce educational outcomes. Conversely, there could be other omitted variables causing the change in migration that result in a positive bias. For example, the unobserved change could be that a relative outside the household gets an opportunity to migrate overseas. This relative could provide information or financial assistance to a household member to migrate. But that outside relative might also send remittances to the household, which could be spent on schooling and tutoring. If these remittances are unobserved, they would be an omitted variable leading the estimated effect of migration on child schooling to be biased in a positive direction.<sup>8</sup> The migration

---

<sup>8</sup>Even if remittances are observed in the data, simply controlling for remittances would be an imperfect solution, for a few reasons. First, remittances are typically not recorded in datasets in sufficient detail to allow separately controlling for remittances from non-household members and household members. One would want to control only for remittances from non-household members in this case. Second, if there is measurement error in remittances, they would only imperfectly control for this omitted variable, allowing

literature contains many theoretical and applied papers arguing how households can use migration to respond to shocks, and so relying on fixed effects requires having a compelling story as to why one household suddenly changed migration status at the time that it did, and why this is not otherwise related to the outcome of interest.

Fixed effects may potentially raise fewer concerns when migration or migrant outcomes like employment and wages are the dependent variable, and the fixed effects are being used to help ensure that impacts are being identified from shocks in an independent variable. A first example is McKenzie et al. (2014). Their analysis involves panel data on migration from Philippines to various countries. The unit of observation is the destination country of migrants, and outcomes of interest are counts and wages of migrants going to those countries. Fixed effects for destination country account for time-invariant omitted variables at the destination level. The focus is on the impact of changes in destination log GDP on migrant flows and migrant wages. Reverse causation is plausibly ruled out because Philippine migrants are a small share of destination labor forces. Omitted variable concerns that remain are about whether changes in GDP *per se* are the causing factor, as opposed to changes that happen to coincide with GDP changes (such as changes in public policies). Changes in public policies (such as immigration policies) are not necessarily identification concerns, however, if they are mechanisms through which GDP shocks lead to changes in migration. A second example of a fixed effects approach is Mahajan and Yang (2020), who examine the effect of hurricanes on international migration to the U.S. The unit of observation is combinations of origin-country and year from 1980 to 2002, and the outcome of interest is immigration to the U.S. (expressed as a share of origin country population). The analysis regresses immigration on an origin country's hurricane index (a measure of the origin country's hurricane exposure) in a given year, including origin-country fixed effects, year fixed effects, and differential origin-country-specific linear time trends. Fixed effects for origin country in a panel data setting are crucial here, to account for an important time-invariant omitted variable: a country's *propensity* to be struck by hurricanes. If this analysis was done with cross-sectional data, or if country fixed effects were not included in the regression, interpretation of the coefficient on hurricanes would be confounded by the effect of the country's general exposure to hurricanes – a geographic characteristic of a country that could itself predict the outcome of interest. For example, hurricane exposure is highly correlated with location in the tropics, which itself could have effects on many outcomes of interest, including migration.<sup>9</sup>

---

omitted variable bias to remain. Third, these remittances could themselves be endogenous, responding to shocks to education.

<sup>9</sup>Another example is Abarcar (2017), who examines how return migration by Australian immigrants is affected by exchange rate shocks in immigrants' home countries, taking advantage of exchange rate variation over time while controlling for household fixed effects. The analysis actually looks at return migration in first-differences, which is equivalent to inclusion of household fixed effects when there are only two time periods.

The general point is that fixed-effects methods do not solve all omitted variable problems. Concerns remain about time-varying omitted variables that may be correlated with the dependent and independent variables of interest. This brings to the fore difference-in-difference methods that seek to exploit policy or other exogenous natural-experimental variation.

#### 4.2.2 Difference-in-differences

The fixed effects panel data model in equation 4 is equivalent to a difference-in-differences regression when there are only two time periods: it then compares changes in outcomes for units that did and did not change their migration status. However, our focus is on the use of the difference-in-differences method (DiD) as a natural experimental approach, where there is a sharp and relatively well-defined change in policy or other aspect of the economic environment (the “treatment”) for some units, for which the researcher can make a plausible case for exogeneity, from both a rhetorical and statistical standpoint. Analysis then compares differences in outcomes for these treated units to differences in outcomes for units unaffected by this policy change. A classic early example in the migration literature is Card (1990), who estimates the impact of the Mariel boatlift (that brought thousands of Cubans to Miami suddenly) on changes in native-born wages, using changes in wages in other cities as the counterfactual. Lozano and Steinberger (2013) provide an introduction to the early use of these methods in studies of immigration. A key underlying assumption is the “parallel trend” or “common trend” assumption: if it were not for this policy change, treated and untreated units would have followed the same time trend in outcomes.

Although difference-in-differences is one of the most widely used tools in causal inference, the last few years have seen a number of econometric theory papers that highlight issues in how these methods have traditionally been employed. The frontier for applying these techniques is changing rapidly, and researchers using DiD for migration will likely find that any new DiD study will face questions about how they have dealt with some of this new literature. Reviews of some of the key points of this new literature can be found in de Chaisemartin and D’Haultfoeuille (2022) and Roth et al. (2022). We attempt to draw out a few key lessons for migration work, while recognizing that this literature is evolving rapidly.

The simplest applications of difference-in-differences involve a single policy change that occurs at one point in time and affects one set of units and not others, as in the Mariel boatlift case. There are multiple other examples of this type of application in the migration literature. For example, McKenzie et al. (2014) estimate the impact of a policy change in which the Philippine government mandated an increase in the minimum wage (to US\$400 per month) for Filipino migrants departing to work as household service workers. They compare changes in migration flows to destinations affected by the policy change (where the

minimum wage was initially below US\$400), to changes in other destinations that already had minimum wages exceeding US\$400. Two more examples come from the brain gain literature, and illustrate how difference-in-differences can be used with cross-sectional data by considering how a policy change affected people of different ethnic groups and birth-year cohorts. Shrestha (2017) considers a policy change in Nepal in 1993, where the British Army (that had historically employed men from the Gurkha ethnic group) imposed a requirement that applicants must have at least an 8th grade education. Taking the differences in schooling for old cohorts (whose schooling was unaffected) and newer cohorts (who were making schooling decisions after the reform), and then comparing these differences for Gurkhas (treated individuals) vs non-Gurkhas (control individuals) gives the impact of a change in the incentives to acquire schooling for migration. Chand and Clemens (2019) use a similar approach in Fiji, for a policy change that affected the returns to migrating for Indian-Fijians versus indigenous-Fijians.

One of the key points of the newer DiD literature is thinking through whether difference-in-differences is being considered as a form of identification from functional form, relying on the parallel trends assumption, versus a design-based approach where it is intended to mimic a quasi-experiment in which the policy change randomly affects some units and not others (Rambachan and Roth, 2020). For example, with a randomized experiment, standard empirical practice is to cluster the standard errors at the level of randomization. With difference-in-differences estimation of a policy implemented at the state-level, this viewpoint would also suggest clustering standard errors at the state level. But this raises questions for how one should analyze cases such as the Mariel Boatlift, or changes in a policy that affect one ethnic group and not another - since effectively there is only one treated city or one treated ethnic group. One solution has been to instead employ synthetic control methods to these cases, as was done by Peri and Yasenov (2019) for the Mariel Boatlift case. Alternatively, more care needs to be taken in making clear that identification is coming from a parallel trends functional form assumption, and justifying why this assumption seems reasonable. For example, are we certain whether parallel trends should hold for the outcome expressed in levels, logs, or in some other transform? See Roth and Sant'Anna (2022) for discussion of this point. It is common to use multiple periods of data before the policy change as a placebo test of this assumption, although some of the recent literature has pointed out that these tests often have low statistical power, and that it may be better to examine robustness to allowing for some violation of common trends. In addition to statistical tests, it is also useful to provide a description of why the policy change was made to apply to some units and not others, and why this is not likely to be correlated with anticipated trends in the outcome. For example, a DiD analysis of the impact of giving refugees work permits in some cities and not others will not be valid if governments decide only to introduce work permits in cities where unemployment is low and falling, and not in



cities where unemployment is expected to be continue to be high or rise further. Our view is the common trends assumption is likely to be more plausible when the units being compared are more similar to one another to begin with, which is where combining matching with DiD can be of use (e.g. Gibson and McKenzie (2014)). Newer, “doubly-robust” approaches to DiD build on this idea to yield valid estimates when either the selection on observables needed for propensity scores or the parallel trends modelling assumption used in DiD holds.

A second common form of difference-in-differences estimation in migration studies is to examine the impact of staggered policies introduced into different states or different countries at different points in time. For example, Otsu (2021) uses panel data from 1999 to 2020 to examine the impact of sanctuary cities (cities that do not share data on immigrants with the Federal government) on crime, with cities adopting this policy over time. At the country level, Kahanec et al. (2016) examine the impact of different countries joining the European Union’s free movement of workers on bilateral migration flows. In cases like these, where treatments are staggered over time, recent literature has shown that in the presence of treatment effect heterogeneity or dynamic treatment effects, the standard two-way fixed effects difference-in-differences estimator is a weighted average of many different comparisons, and it is possible that some of these weights may even be negative, making it hard to interpret the estimate (Goodman-Bacon, 2021). Different estimators have been developed to deal with this problem, which involve making sure that treated units are only ever compared to ‘clean controls’ and not to earlier treated units (see de Chaisemartin and D’Haultfoeuille (2022) and Roth et al. (2022) for a survey). A key point to note here is that with treatment heterogeneity, researchers need to clearly specify precisely which weighted average they are interested in estimating.

A third type of difference-in-differences commonly used in the migration literature involves situations in which there is different exposure to some policy change across different regions, sometimes due to some pre-existing migration networks. In these cases the treatment is no longer binary, but can be continuous or have multiple discrete values. For example, Abarcar and Theoharides (2022) examine the impact of changes in U.S. working visas (H-1B visas) for medical professionals, which affected some parts of the Philippines more than others based on their prior participation in medical worker migration to the U.S. Theoharides (2020) considers a natural experiment that occurred in the Philippines in 2006, in the form of a nearly-complete shutdown of a large specific migration channel – that of Filipino female “overseas performance artists” (OPAs) going to Japan. The identification strategy exploits the fact that OPA migration was unevenly distributed across provinces prior to 2006, so pre-2006 province shares of OPA migration are associated with greater declines in migration from pre- to post-2006. Dinkelman and Mariotti (2016) provide a third example, examining the impact of migrant work opportunities (male outmigration to the South African mines) on education in Malawian origin areas. She compares cohorts of

different ages, and localities that historically had different numbers of recruiting stations.

Recent work by Callaway et al. (2021) has highlighted that difference-in-differences with continuous treatments or different doses of a treatment results in additional complications to consider relative to the binary treatment case. In particular, a stronger parallel trends assumption needs to be employed, and if treatment effects are heterogeneous across units receiving different doses of the treatment, two-way fixed effects specifications can be hard to interpret, even when treatment is not staggered over time. They discuss several possible solutions, and migration work that relies on this differential exposure across areas will need to discuss these issues more in the future.

A final set of issues that can arise occurs when policies can get reversed or “turned off”. While this can raise further complications in the presence of dynamic effects and treatment heterogeneity, it may also increase the credibility of causal claims. For example, observing an effect in one direction when a policy is introduced, and then that the estimated effects go (as would be predicted) in the opposite direction after a policy is reversed, may serve as further proof that the policy is responsible for the changes in outcomes observed. Studies that take advantage of such reversals include Dinkelman and Mariotti (2016) and Abarcar and Theoharides (2022).

### 4.3 Instrumental Variables

The instrumental variables (IV) approach exploits variation in the real world to establish causal relationships. When implemented credibly, IV deals with concerns about omitted variable bias and reverse causation.

Consider the following regression equation estimated using a cross-section of data on households, indexed by  $i$ . The outcome of interest is  $Y_i$ , for example the level of household per capita consumption. We are interested in estimating the impact of an independent variable  $X_i$ , such as whether the household has a member earning income overseas as a migrant worker.

$$Y_i = \alpha + \beta X_i + \varepsilon_i \tag{6}$$

In general, we worry that the coefficient  $\beta$  on  $X_i$  is biased due to omitted variables and/or reverse causation. When using the IV approach, a researcher uses an *instrument*  $Z_i$  that helps isolate exogenous variation in the independent variable  $X_i$  in determining the impact of  $X_i$  on outcome  $Y_i$ . The following “first stage” regression captures the impact of the instrument  $Z_i$  on  $X_i$ :

$$X_i = \gamma + \delta Z_i + \varepsilon_i \tag{7}$$

There are key assumptions underlying IV. First, the instrument must be *relevant*:  $Z_i$

must strongly predict  $X_i$ . In practice, the F-statistic representing the significance of the empirical relationship between  $Z_i$  and  $X_i$  must exceed a certain threshold (see Stock and Yogo (2005), and more recently Lee et al. (2022)).

The other key assumption behind IV is the *exclusion restriction*: the instrument  $Z_i$  only affects the “second stage” outcome of interest  $Y_i$  through its effect on  $X_i$ . In other words, aside from  $X_i$ , there can be no other channel or mechanism through which  $Z_i$  affects  $Y_i$ .

IV has a long history in economics, going back roughly a century (see Angrist and Pischke (2009) for a brief historical overview), and it is straightforward to implement. But credible instruments are rare, in practice. Most instruments that fail to be convincing fail due to a violation of the exclusion restriction.

The most convincing instruments derive from randomization. In a sample of Tongan households with a member who applied to work in New Zealand, Gibson et al. (2011) instrument for whether a household has a member working overseas in New Zealand with a dummy variable for whether someone in the household won a lottery for a migrant worker visa. Similarly, in a sample of Bangladeshi households, all of which had a household member apply for a work visa to Malaysia, Mobarak et al. (2020) instrument for a household member migrating to Malaysia with the result of a lottery allocating the visas.

Another convincing recent example of the IV approach is Parsons and Vezina (2018), who study how Vietnamese refugee settlement in 1975 affects U.S. exports to Vietnam from particular U.S. states after the lifting of the U.S. trade embargo in 1994. The authors argue use the number of Vietnamese refugees resettled in a U.S. state in 1975 (after the fall of Saigon) as an instrument for the size of a state’s Vietnamese population in 1995 (immediately after the lifting of the trade embargo). The instrument has a strong first stage (1975 Vietnamese refugees strongly predict 1995 Vietnamese population), although one challenge with using state-level variation is that identification comes from a relatively small number of units (only 50 states plus D.C.). The authors argue that the exclusion restriction is satisfied, because refugees were intentionally dispersed across the U.S. in a relatively haphazard fashion after the immediate influx of 140,000 refugees in 1975. The resulting distribution of refugees is thus quasi-random. The authors find that more Vietnamese immigrants leads states to see larger new export flows to Vietnam after the trade embargo was lifted. Overall, this seems like a credible use of the IV strategy. One might raise questions about the validity of the exclusion restriction: are there other channels through which states with high 1975 Vietnamese refugee flows might have higher exports to Vietnam in 1995? One channel might be *return* migrants – some of the refugees could have returned prior to 1995, and be responsible for some trade facilitation. This would be a related but slightly different mechanism than (non-returned) 1995 Vietnamese immigrants facilitating trade. Given the political and economic situation of Vietnam in the two decades after 1975, it is perhaps plausible that return migration flows would not have been large.

## 4.4 Shift-share approaches

A very popular research design in trade and migration studies involves using “shift-share” instruments (also known as “Bartik” instruments, after the first such study, Bartik (1991).) In shift-share designs, the researcher analyzes the impact of multiple shocks, each of which affects some units of observation more than others. The “share” measures the degree to which a unit is exposed to a particular shock, and the “shift” refers to the size of a particular shock. This approach has been widely used in economic research on migration.

For example, one might be interested in the impact of immigration flows into a particular location (say, U.S. cities) from different countries around the world. The shift-share instrument  $B_{l,t}$  would be as follows:

$$B_{l,t} = \sum_{j=1}^J z_{l,j,t-1} m_{j,t} \quad (8)$$

where  $z_{l,j,t-1}$  (the “share”) is the initial share of immigrants from migration-source country  $j$  in location  $l$  (as a share of all immigrants to the U.S. from country  $j$  in the initial year  $t - 1$ ), and  $m_{j,t}$  (the “shift”) is the normalized change in immigration from country  $j$  into the U.S. as a whole.  $B_{l,t}$  is then the predicted inflow of migrants to destination  $l$  in time  $t$ : the weighted average of the national inflow from each country (the shift), where the weights (the shares) are the initial fractions of migrants across destinations  $l$  from each country of origin  $j$ .

There have been a number of advances in the econometrics of shift-share research designs in recent years. The recent work has made clear how to clearly justify the exogeneity of the shift-share instrument. In addition, there has been important work clarifying how to calculate correct standard errors in shift-share designs. Researchers will need to make a case why either the shifts and/or the shares can be considered exogenous (uncorrelated with the error term), and then the effective sample size that determines standard errors will depend on how many independent shifts or shares there are. We next discuss some of these issues, while acknowledging that this is a still fluid and developing literature, with consensus best practice still evolving.

### 4.4.1 Justifying exogeneity of the shift-share instrument

In many studies, the shift-share variable is calculated analogously to equation 8 above, in which the shifter is an aggregate flow (in this case, aggregate immigration inflows from origin country  $j$  to the entire U.S.). The canonical Bartik (1991) study is an example. Other examples include Cortes (2008), who studies the impact of low-skilled immigration to the U.S. on the prices of goods and services provided by such immigrants; Gonzalez and Ortega (2013) who study the impact of immigration to Spain on housing prices; and Glitz

(2012), who examines the impact of new inflows of ethnic Germans to Germany from Eastern Bloc countries in the late 1990s on labor market outcomes of native workers. More recently, Theoharides (2017) examines the impact of international outmigration from the Philippines, using a shift-share instrument for outmigration in which the shift is aggregate Philippine outmigration to a given overseas destination, and the share is a Philippine location’s initial share of total Philippine outmigration to the overseas destination.<sup>10</sup>

Goldsmith-Pinkham et al. (2020) show that this shift-share specification is numerically equivalent to using the initial shares (interacted with time fixed effects when there are multiple periods) as instruments in a weighted GMM regression. In this formulation, the shifts do not provide exogenous variation, and simply serve as weights in the regression. Causal identification under this approach then requires one to make an argument that the shares ( $z_{l,j,t-1}$  in equation 8) are exogenous. One implication is that no other shifters should affect outcomes through the same (or correlated) exposure shares. For example, a common concern in migration research is that shares that capture connections to migration origins may be correlated with shares that capture connections to trade or FDI. Goldsmith-Pinkham et al. (2020) provide suggestions for empirical tests that can help authors justify the claim that the shares can be taken as plausibly exogenous, including overidentification tests that can reveal whether any of the multiple instruments are endogenous. They also note that this method puts different weights on different shares, and frequently a small number of the shares account for much of the identifying variation in the data. This offers a guide for researchers needing to argue why it is plausible that the shares are exogenous, since they can focus on providing some institutional background and context about the shares that matter most for identification.

Alternatively, the claim of exogeneity may be more credible if one can use explicit shocks as the shifts in the shift-share instrument. One might replace the aggregate migration inflows ( $m_{j,t}$ ) in equation 8 with shocks  $s_{j,t}$  representing an exogenous determinant of migration outflows from country  $j$  to the U.S. in time  $t$ . An example of such an exogenous determinant of outmigration might be a measure of country  $j$ ’s affectedness by natural disasters (which could stimulate outmigration, as in Hanson and McIntosh (2012)), or by U.S. visa restrictions affecting country  $j$  (as in Peri et al. (2015)).

Borusyak et al. (2022) argue that claims for the exogeneity of shift-share instruments are likely more credible when exploiting such plausibly-exogenous shocks as the shifts, rather than trying to make the case that the shares are exogenous. They develop a framework for shift-share estimation in which the shares can be endogenous, and in which causal identification derives from the exogeneity of the shifts. They provide an equivalence result, showing then that the shift-share IV regression coefficients are equivalent to a regression estimated at the level of the shocks, in which the outcome and treatment variables are first averaged,

---

<sup>10</sup>Other recent examples include Bazzi et al. (2016) and Bahar et al. (2022).

using shares as weights.

Studies that use a shift-share instrument with explicit shocks as shifts include Kinnan et al. (2019), who study how rural-urban migration in China affects development in migrants’ rural origin areas. Their shift-share instrument uses shares based on prior dyadic rural-urban connections created by a government program, while the shifts are urban labor demand shocks and relaxations of urban in-migration restrictions. Another example is Khanna, Murathanoglu, Theoharides, and Yang (Khanna et al.), who study the impact of international migration income in Philippine migrant-origin areas. The share is the magnitude of a Philippine province’s baseline migrant income, and the shift is exchange rate shocks due to the 1997 Asian Financial Crisis.<sup>11</sup>

In the Borusyak et al. (2022) framework, causal identification requires the assumption that the shifts are as good as randomly assigned. Conducting shift-share analyses in the framework of Borusyak et al. (2022) reveals ways to support this assumption. They suggest useful tests for the exogeneity of the shifts that are analogous to “balance tests” in analyses of randomized control trials. The framework also suggests that an important (and frequently omitted) type of control variable to include in the regression is appropriately-weighted averages of characteristics of the units generating the shocks (the units indexed by  $j$ ). They also highlight that in cases with “incomplete shares” (cases where the sum of shares varies across observations), this sum of shares is likely to induce bias if omitted from the regression, so it is important to include it as a control variable as well.

The framework of Borusyak et al. (2022) also offers ways to potentially achieve credible causal inference in shift-share designs when one does not have an explicitly exogenous shock to use as the shift (e.g., the original Bartik (1991) study). In their framework, the key concern about such settings is that there may be a “mechanical” correlation between the shift-share instrument and outcomes driven by shocks in the  $l$  units, that then get aggregated to the measure of  $m_{j,t}$ . In the case of equation 8 above, for example, the concern is that variation in aggregate U.S. immigration from country  $j$  could be endogenous to changes in U.S. locations  $l$  (e.g., demand for immigrant labor, and determinants thereof) that sum up to the aggregate change  $m_{j,t}$ .

Borusyak et al. (2022) describe conditions under which one may reasonably assume that the shift is exogenous.<sup>12</sup> First, if shocks in units  $l$  in the study can be taken as spatially uncorrelated, the measure of aggregate flows from unit  $j$  can use a “leave-one-out” estimator

---

<sup>11</sup>Household-level studies exploiting exchange rate shocks of migrant workers in different countries also involve shift-share approaches. Yang (2006) and Yang (2008) study how migrant return migration and origin-household outcomes are affected by overseas exchange rate shocks faced by migrants. The “shift” is the exchange rate shock in each particular destination, and the “share” is the fraction of a household’s migrants in a destination. Most households only have one migrant overseas (so that its share of migrants in any particular destination is 1 for one country and 0 for all other destinations), but a minority have multiple migrants, so that they nonzero shares in two or more destinations.

<sup>12</sup>See Borusyak et al. (2022) Section 4.1 and Appendix A.6 for a discussion of this in the context of the Bartik (1991) setting.

that excludes the own-unit  $l$  from calculation of the aggregate flow  $m_{j,t}$ , and this could provide more credible identification. Second, if the number of units  $l$  over which each  $m_{j,t}$  is calculated is large, the “mechanical” correlation between the instrument and the outcome may be ignorable.

Finally, researchers also sometimes use more complicated versions of shift-share instruments, that are non-linear. Examples include Basso and Peri (2015) who use a nonlinear shift-share instrument for immigration flows, and Allen et al. (2018) who use a nonlinear instrument to predict border walls. It can be harder to be clear where identification is coming from in these non-linear cases, and their use requires extra care. Appendix D.4. of Borusyak and Hull (2021) discusses these cases, and note that a type of “re-centering” is usually needed to make these instruments valid.

#### 4.4.2 Calculating correct standard errors

It has also been recently shown that conventional standard errors in shift-share designs may be invalid, due to the fact that units  $l$  that have similar exposure shares have similar residuals. For example, under the Borusyak et al. (2022) framework that relies on shifts being exogenous, the effective sample size depends on the number of shifters (number of independent exogenous shocks). Borusyak et al. (2022) and Adao et al. (2019) provide methods for constructing valid standard errors that account for such shared shock exposure across units, both of which are available as Stata packages. A key point here is that it is often the case that the effective sample size providing variation may be much lower than researchers have traditionally thought. For example, consider attempting to use a shift-share instrument to examine impacts of migration on different Mexican communities, where the shares are the share of migrants that historically when to different U.S. cities, and the shocks are immigration enforcement actions in different U.S. cities. Then there may only be 10 or 15 U.S. cities that most migrants are choosing between, and some of these cities may coordinate their policy changes, so that effectively we can think of policies being randomly varied over only 4 or 5 clusters of cities. Correctly accounting for this will typically result in much larger standard errors.

Researchers have thought less about the validity of standard errors under the Goldsmith-Pinkham et al. (2020) framework of identification based on exogenous shares. However, often it will not seem plausible to think that the shares are independently distributed across towns or regions, since there is often some spatial correlation. We currently recommend that researchers discuss explicitly what they believe is the cause of shares varying across units, and the use this to consider clustering standard errors at higher levels of aggregation (such as state), or otherwise explicitly modeling the assignment process that gave rise to shares differing.

### 4.4.3 Short- vs. Long-Run Dynamics

Goldsmith-Pinkham et al. (2020) note that two key assumptions in their work are that locations are independent (so there are no spatial spillovers), and that the data consist of a series of steady-states. Jaeger et al. (2018) provide a critique of this second assumption when studying the impacts of immigration, but their point also applies more generally to other settings where adjustment dynamics are important.

Consider a regression equation seeking to estimate the effect of immigration on wages:

$$\Delta W_{l,t} = \alpha + \beta I_{l,t} + \varepsilon_{l,t} \quad (9)$$

where  $\Delta W_{l,t}$  is the change in log wages,  $I_{l,t}$  is immigration in period  $t$ , and  $\varepsilon_{l,t}$  is the error term.

The standard concern is that there are contemporaneous factors (e.g. local demand shocks) that affect both local native wages and how many immigrants move in. A shift-share instrument, as in equation 8, is meant to be exogenous to these local demand shocks. Jaeger et al. (2018) demonstrate that, if it takes time for markets to adjust to shocks, then the error term  $\varepsilon_{l,t}$  can also include other terms, which reflect the ongoing general equilibrium adjustment effects of past immigrant supply shocks (e.g. capital adjustment). The result is that the shift-share instrument will conflate the short-term response (e.g. a fall in wages when new immigrants enter) and the long-term response (e.g. a positive move back as capital has time to adjust).

Their suggested solution is to control for these dynamics by adding lagged immigrant flows to this regression, and also instrumenting for this with the analogous shift-share instrument:

$$\Delta W_{l,t} = \alpha + \beta I_{l,t} + \theta I_{l,t-1} + \varepsilon_{l,t} \quad (10)$$

Two shift-share instruments are then used:

$$B_{l,t} = \sum_{j=1}^J z_{l,j,0} m_{j,t} \quad (11)$$

and

$$B_{l,t-1} = \sum_{j=1}^J z_{l,j,0} m_{j,t-1} \quad (12)$$

Then  $\beta$  will capture the short-run effect, and  $\theta$  captures the longer-term reaction to past supply shocks. However, for this to work, we need there to be independent variation in the two periods in where migrants are coming from. This can be a tall order, since the



country of origin mix of the flow of immigrants to the U.S. is so similar over time – since the 1980s the correlation across metro areas between the instrument and its lag is 0.96 to 0.99. The result is that shift-share instruments are usually going to be strongly serially correlated and not provide enough variation to separately identify dynamics. The authors suggest such a strategy worked better in the 1970s vs 1980s in the U.S., due to policy changes and other shocks dramatically changing the immigrant composition then, whereas since the 1980s, country of origin is too highly correlated. They also suggest it might work better in European countries where immigrant flows have been less stable over time.

This point about dynamics is also crucial for studies that attempt to use shocks in shift-share instruments for migration in studying the impacts of migration on development outcomes. For example, consider trying to estimate the impact of migration on business ownership, using the fact that historically people in different regions of the Philippines have migrated to different countries. Then an exchange rate shock that increases earnings in the U.A.E. relative to those in Korea or the U.S. may increase migration this year from communities that historically sent more migrants to the U.A.E. But then next year, new shocks will affect migration from other communities, and after multiple years, many of these shocks will average out, leaving a lot less variation and typically making these instruments weaker and weaker as more and more time periods are included. Such shocks may therefore be more useful for identifying impacts of migration on outcomes that happen at very time-sensitive points (e.g. on the progression of a child from secondary school to university), than on outcomes like business ownership, where the business could be started in any one of ten different years, depending on which year brings the positive shock.

#### **4.4.4 Local Average Treatment Effects (LATE)**

A key point relevant for IV in general, and shift-share IV approaches in particular, is that IV estimates identify a local average treatment effect (LATE). Therefore, if we find that IV gives a different result from OLS, it need not necessarily be because of an omitted variable bias that IV is solving. Rather, it could also be that there are heterogenous treatment effects and the local average identified by IV is different from the weighted average identified by OLS.

This is particularly important to also emphasize when thinking through shift-share designs as instruments for migration, since they will identify effects off people whose migration behavior changes in response to short-term shocks. These effects might be quite different from the impact of migration for those who are always takers, who may have planned more systematically to migrate for years, and who may have the most to gain from migrating (which is why they migrate regardless). Likewise, in the immigration-wage literature, the response of local labor markets to a sudden unexpected influx of migrants may be quite different from a steady and demand-driven growth in migration over time during which

migrants come in when firms have jobs they can't fill, firms have time to adjust capital, cities have time to adjust infrastructure, and other equilibrium adjustments. And then even further, within the shift-share, Goldsmith-Pinkham et al. (2020) point out that the results can be driven by a handful of the shares.

## 4.5 Regression discontinuity methods in migration impact evaluation

Regression discontinuity designs (RDDs) are often viewed as one of the most credible non-experimental methods for estimating causal effects. In this design, there is a variable called the running variable or score, and the probability of being treated or receiving some intervention changes discontinuously at some threshold value of the score. If units cannot perfectly sort around this threshold, then comparing units with scores just below the threshold to units with scores just above it can be used to obtain causal estimates. Cattaneo et al. (2019) provide an excellent practical introduction and code for the use of this design. A key point to note is that regression discontinuity will identify the causal effect only locally, for individuals with scores in the neighborhood of the threshold. We discuss here some examples of how RDD has been applied in the study of migration, and some key considerations in being able to recognize when this design may or may not be applicable.

A first case is when there literally is a scored test that is used to decide whether individuals can migrate or whether they get access to some program for migrants. These cases are rare, but are the most direct application of the RDD method. Clemens and Tiongson (2017) provide an example, using the case of Filipinos applying to a temporary work program in Korea. Applicants for this program had to take a 200-question Korean language test, and achieve a score of 120 or greater to secure a work permit. Large numbers of applicants take the test, and there are many applicants whose scores ended up being within 5 points on either side of this threshold. They tracked down and surveyed applicants who had scores in this range, and by comparing households with a member who scored just above this threshold to households with a member who scored just below the threshold, are able to estimate the causal impact of migrating on a variety of household outcomes. Lochmann et al. (2019) provide a second example, using a requirement that new immigrants to France from outside the European Union pass a basic French test, or be required to take language classes. Applicants had to score at least 50 out of 100, but this was a much less ideal case for RDD, since scores were in increments of 5, and there were not many individuals with scores close to the threshold of 50. As a result the authors have to rely more on estimation over a wider score range in order to measure the impacts of language training for immigrants.

A key requirement for the use of RDD is that individuals cannot sort themselves around the threshold. In the case of language test scores, it is therefore important to only consider the score the first time an applicant takes the test, since otherwise more determined potential

migrants who marginally fail a test may study more and re-take the tests again and again until they pass. The result then would be that the set of individuals left just failing the test will be different in important ways from those who just pass. Individuals are also much less likely to be able to sort around the threshold if they do not know where the threshold lies. Several countries like Canada, Australia, and New Zealand use points systems for immigration, where potential migrants are assigned points for characteristics like their age, language skills, education, work experience, job offers, and presence of family links, and must reach some points threshold in order to be able to immigrate. This may sound like a great opportunity for an RDD, by comparing individuals who just exceed the points threshold to those who just miss out. However, the points thresholds are public information, and the immigration agencies have online calculators where potential applicants can get a good approximation of how many points they would be awarded. As such we might expect substantial sorting around the threshold, with many of those who would score just below the threshold waiting to apply until they could improve their points, making the few applicants who end up getting awarded points just below the threshold not being comparable to those with points just above the threshold. As such, we believe most points systems will not yield valid RDD causal estimates. An exception may occur when the thresholds are more opaque. For example, Carpio et al. (2016) use RDD to estimate the impact of a tax incentive program for high-skilled emigrants in Malaysia, which used education-specific work experience abroad thresholds that were not publicly known.

A second potential use of the RDD in studies of migration occurs with date or time as the running variable. Migration amnesty policies typically restrict eligibility to those who can prove presence in the country as of some earlier date prior to the announcement of the reform. Examples include the 1986 Immigration Reform and Control Act (IRCA) in the United States, which required applicants to have been in the U.S. as of January 1, 1982, and the September 9, 2002 Italian amnesty program, which required employers to have employed the immigrant for at least three months prior to the legalization, that is, at least since June 11, 2002. In principle, one could use a RDD that compares irregular migrants who entered the country just before the threshold date to those who entered just after the date. However, in practice, researchers either have data that does not have fine-tuned data on data of arrival (e.g. just year of entry) or that have few observations exactly around the threshold. As a result, these amnesties have instead been evaluated using difference-in-difference approaches, that compare a cohort on one side of the threshold to another on the other side of the threshold (e.g. Pan (2012) for the IRCA and Devillanova et al. (2018) for the Italian amnesty). Another practical concern here is that while migrants cannot retrospectively sort around the announced eligibility date, some migrants who just miss the cutoff may be able to use fake documentation to make themselves appear eligible.

Regression discontinuity in time approaches for migration will work better when re-

searchers have much more fine-tuned data on dates, and many observations near the threshold. This may occur when applications are processed on a first-come, first-served basis with many applications arriving close in time to one another. A striking example comes from Pinotti (2017), who examine an Italian amnesty system where employers must submit applications on certain “click days”, and are then processed on a first-come, first-served basis, until a quota is reached. He uses a large sample of over 110,000 applicants that had applications processed within one hour around the cutoff, and considers bandwidths ranging from 1 minute to 30 minutes. This fine and continuous data on application times enables RDD to be used to estimate the impact of legalization on crime. A second example is found in Hainmueller et al. (2017), who examine the protective effect of the Deferred Action for Childhood Arrivals (DACA) program in the U.S. on mental health of their young children. Eligibility for the program required a mother to under age 31 as of 15 June, 2012. Using the exact date of birth, they can compare outcomes in children whose mothers were born just before June 15, 1981 (and were hence ineligible) to those born just after (who were eligible).

While language test scores and migration policy eligibility date thresholds seem the most likely avenues for the use of RDD in the study of migration, there are also other opportunities where comparing people who just marginally qualify for some policy to those who just marginally miss out can allow for the use of this method. For example, Hainmueller et al. (2017) use an unusual feature of Swiss immigration policy, which subjected naturalization applications to municipal referendums, where local voters voted on whether or not to accept individual candidates as citizens. Comparing those applicants who received just above the threshold of 50% to those who received just below this threshold in votes enabled them to estimate the causal impacts of naturalization on immigrants’ social integration. Additional applications will be possible when considering the impacts on migration of social welfare programs that use various eligibility score thresholds.

A couple of final notes for researchers planning on using this method. The first is that it can be very data intensive: typically considerably larger sample sizes are needed for RDD than for randomized experiments to have the same statistical power for identifying a given effect size. Secondly, the method is usually much more credible when the analysis occurs in a narrow bandwidth of the threshold, and researchers should be particularly cautious about using global polynomials or relying on observations far from the threshold in identifying the causal effect.

## 5 Conclusion

Both experimental and non-experimental causal inference methods have increasingly been used to answer a wide variety of research questions surrounding migration. Careful and creative use of these methods have enabled researchers to overcome many of the self-selection

challenges inherent in the study of migration, and provided new understanding of the drivers and consequences of migration and of migration policies. Recent methodological developments have helped to clarify the conditions needed for these methods to be used, and we expect to see continued growth in their use over the coming decade.

We conclude by noting several emerging areas of research interest. The first is around targeting of policy interventions. Migration, especially international migration, is typically such a rare event that the sample size of people actually changing migration status is very small in many experiments (and in panel data approaches that rely on people changing migration status). Work on the design of experiments needs to develop better methods of targeting those individuals whose migration status is most likely to change with interventions, both for the purpose of having sufficient statistical power to measure effects, as well as for more effective use of policy tools.

Second, much recent methodological work has highlighted the importance of considering and properly accounting for treatment effect heterogeneity. This is clearest in the new literature on staggered difference-in-difference approaches, as well as in interpreting encouragement design experiments and parameters in IV estimation, but also is an issue in all estimation. The result is that it does not make sense to talk about “the” effect of migration, since effects will differ across units and potentially also over time. A key area for future work is to better understand and model this heterogeneity of treatment effects. This has been an area of active research in applying machine learning methods to estimate heterogeneity in randomized experiments, and such methods are likely to be incorporated more in non-experimental methods in the near future. A challenge is that these methods tend to be quite data-intensive, requiring large samples, and so migration research may need to work with larger scale projects or use big administrative datasets to better measure this heterogeneity.

Finally, another consequence of treatment heterogeneity is that the impacts measured for one sample, in one study, may not automatically translate to another context. This has two implications for future research. The first is to use research approaches that attempt to explicitly model external validity, meta-analysis approaches, and Bayesian approaches that will help better use the evidence from one setting to inform policy knowledge in other settings. However, these methods are only as good as the inputs available. Much of our existing research base comes from just a few countries with high migration intensities (e.g., Mexico and the Philippines). A second implication is thus the need for researchers to study migration in a much broader range of countries and settings.

## References

- Abarcar, P. (2015). Do Employers Value Return Migrants? An Experiment on the Returns to Foreign Work Experience. *Mathematica Policy Research* (48).
- Abarcar, P. (2017). The Return Motivations of Legal Permanent Migrants: Evidence from Exchange Rate Shocks and Immigrants in Australia. *Journal of Economic Behavior and Organization* 144, 62–77.
- Abarcar, P. and C. Theoharides (2022). Medical Worker Migration and Origin-Country Human Capital: Evidence from U.S. Visa Policy. *Review of Economics and Statistics*, forthcoming.
- Adao, R., M. Kolesar, and E. Morales (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics* 134(4), 1949–2010.
- Allen, T., C. Dobbin, and M. Morten (2018). Border Walls. *NBER Working Paper* (25267).
- Ambler, K., D. Aycinena, and D. Yang (2014). Remittance Responses to Temporary Discounts: A Field Experiment among Central American Migrants. *National Bureau of Economic Research* (20522).
- Ambler, K., D. Aycinena, and D. Yang (2015). Channeling Remittances to Education: A Field Experiment among Migrants from El Salvador. *American Economic Journal: Applied Economics* 7(2), 207–232.
- Angelucci, M. (2015). Migration and Financial Constraints: Evidence from Mexico. *The Review of Economics and Statistics* 97(1), 224–228.
- Angrist, J. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press: Princeton, NJ.
- Ashraf, N., D. Aycinena, C. Martinez, and D. Yang (2015). Savings in Transnational Households: A Field Experiment among Migrants from El Salvador. *The Review of Economics and Statistics* 97(2), 332–351.
- Bah, T., C. Batista, F. Gubert, and D. McKenzie (2022). Can Information and Alternatives to Irregular Migration Reduce "Backway" Migration from the Gambia? *World Bank Policy Research Working Paper* (10146).
- Bahar, D., C. Ozguzel, A. Hauptmann, and H. Rapoport (2022). Migration and Post-Conflict Reconstruction: The Effect of Returning Refugees on Export Performance in the Former Yugoslavia. *Review of Economics and Statistics*, forthcoming.
- Barsbai, T., V. Licuanan, A. Steinmayr, E. Tiongson, and D. Yang (2020). Information and the Acquisition of Social Network Connections. *National Bureau of Economic Research* (27346).
- Bartik, T. (1991). *Who Benefits from State and Local Economic Development Policies?* Upjohn Institute: Kalamazoo, MI.
- Baseler, T. (2021). Hidden Income and the Perceived Returns to Migration. *SSRN* (3534715).
- Basso, G. and G. Peri (2015). The Association between Immigration and Labor Market Outcomes in the United States. *IZA Working Paper no. 9436*.
- Batista, C. and G. Narciso (2018). Migrant Remittances and Information Flows: Evidence from a Field Experiment. *World Bank Economic Review* 32(1), 203–219.
- Bazzi, S., L. Cameron, S. Schaner, and F. Witoelar (2021). Information, Intermediaries, and International Migration. *UC San Diego*.
- Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2016). Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia. *The American Economic Review* 106(9), 2658–2698.
- Beam, E. A., D. McKenzie, and D. Yang (2016). Unilateral Facilitation Does Not Raise International Labor Migration from the Philippines. *Economic Development and Cultural Change* 64(2), 323–368.
- Borjas, G. J. (1987). Self-Selection and the Earnings of Immigrants. *The American Economic Review* 77(4), 531–553.
- Borusyak, K. and P. Hull (2021). Non-Random Exposure to Exogenous Shocks: Theory and Applications. *Working Paper*.
- Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-Experimental Shift-Share Research Designs. *Review of Economic Studies* 89, 181–213.

- Bryan, G., S. Chowdhury, and A. M. Mobarak (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica* 82, 1671–1748.
- Callaway, B., A. Goodman-Bacon, and P. Sant’Anna (2021). Difference-in-Differences with a Continuous Treatment. *ArXiv* (2107.02637).
- Card, D. (1990). The Impact of the Mariel Boatlift on the Miami Labor Market. *Industrial and Labor Relations Review* 43, 245–257.
- Caria, S., G. Gordon, M. Kasy, S. Quinn, S. Shami, and A. Teytelboym (2020). An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan. *CES* (8535).
- Carpio, X. V. D., C. Ozden, M. Testaverde, and M. Wagner (2016). Global Migration of Talent and Tax Incentives: Evidence from Malaysia’s Returning Expert Program. *The World Bank* (7875).
- Cattaneo, M., N. Idrobo, and R. Titiunik (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge Elements: Quantitative and Computational Methods for Social Science, Cambridge University Press, forthcoming.
- Chand, S. and M. Clemens (2019). Human Capital Investment under Exit Options: Evidence from a Natural Quasi-Experiment. *Center for Global Development* (152).
- Chen, X., Q. Huang, S. Rozelle, Y. Shi, and L. Zhang (2009). Effect of Migration on Children’s Educational Performance in Rural China. *Comparative Economic Studies* (51), 523–543.
- Clemens, M., C. Montenegro, and L. Pritchett (2019). The Place Premium: Bounding the Price Equivalent of Migration Barriers. *Review of Economics and Statistics* 101(2).
- Clemens, M. and E. Tiongson (2017). Split Decisions: Household Finance When a Policy Discontinuity Allocates Overseas Work. *The Review of Economics and Statistics* 99(3), 531–543.
- Cortes, P. (2008). The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data. *Journal of Political Economy* 116.
- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press: New Haven.
- de Arcangelis, G., M. Joxhe, D. McKenzie, E. Tiongson, and D. Yang (2015). Directing Remittances to Education with Soft and Hard Commitments: Evidence from a Lab-in-the-Field Experiment and New Product Take-up among Filipino Migrants in Rome. *Journal of Economic Behavior and Organization* 111, 197–208.
- de Chaisemartin, C. and X. D’Haultfoeuille (2022). Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey. *Econometrics Journal*, forthcoming.
- Devillanova, C., F. Fasani, and T. Frattini (2018). Employment of Undocumented Immigrants and the Prospect of Legal Status: Evidence from an Amnesty Program. *ILR Review* 71(4), 853–881.
- Dinardo, J. (2008). *Natural Experiments and Quasi-Natural Experiments*. Palgrave Macmillan, London.
- Dinkelman, T. and M. Mariotti (2016). The Long-Run Effects of Labor Migration on Human Capital Formation in Communities of Origin. *American Economic Journal: Applied Economics* 8(4), 1–35.
- Doi, Y., D. McKenzie, and B. Zia (2014). Who You Train Matters: Identifying Combined Effects of Financial Education on Migrant Households. *Journal of Development Economics* 109, 39–55.
- Duflo, E., R. Glennerster, and M. Kremer (2007). *Chapter 61 Using Randomization in Development Economics Research: A Toolkit*, Volume 4. Elsevier.
- Duguet, E., N. Leandri, Y. L’Horty, and P. Petit (2010). Are Young French Job Seekers of Ethnic Immigrant Origin Discriminated against? A Controlled Experiment in the Paris Area. *Annals of Economics and Statistics* (99/100), 187–215.
- Dustmann, C. and F. Fabbri (2003). Language Proficiency and the Labour Market Performance of Immigrants in the UK. *Economic Journal* 113.
- Funkhouser, E. (2013). *Using longitudinal data to study migration and remittances*. Edward Elgar.
- Gazeaud, J., E. Mvukiyehe, and O. Sterck (2022). Cash Transfers and Migration: Theory and Evidence from a Randomized Controlled Trial. *Review of Economics and Statistics*, forthcoming.
- Gibson, J. and D. McKenzie (2014). The Development Impact of a Best Practice Seasonal Worker Policy.

- The Review of Economics and Statistics* 96(2), 229–243.
- Gibson, J., D. McKenzie, H. Rohorua, and S. Stillman (2017). The Long-term Impacts of International Migration: Evidence from a Lottery. *The World Bank Economic Review* 32(1), 127–147.
- Gibson, J., D. McKenzie, and S. Stillman (2011). The Impacts of International Migration on Remaining Household Members: Omnibus Results from a Migration Lottery Program. *Review of Economics and Statistics* 93(4), 1297–1318.
- Gibson, J., D. McKenzie, and S. Stillman (2013). Accounting For Selectivity And Duration-Dependent Heterogeneity When Estimating The Impact Of Emigration On Incomes And Poverty In Sending Areas. *Economic Development and Cultural Change* 61(2), 247–280.
- Glennerster, R. and S. Powers (2014). *Balancing Risk and Benefit: Ethical Tradeoffs in Running Randomized Evaluations*. Oxford University Press.
- Glennerster, R. and K. Takavarasha (2013). *Running Randomized Evaluations: A Practical Guide*. Princeton University Press: Princeton, NJ.
- Glitz, A. (2012). The Labor Market Impact of Immigration: A Quasi-Experiment Exploiting Immigration Rules in Germany. *Journal of Labor Economics* 30.
- Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review* 110(8), 2586–2624.
- Gonzalez, L. and F. Ortega (2013). Immigration and Housing Booms: Evidence from Spain. *Journal of Regional Science* 53.
- Gonzalez-Barrera, A. (2015). More Mexicans Leaving than Coming to the U.S. *Pew Research Center*.
- Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics* 225(2), 254–277. Themed Issue: Treatment Effect 1.
- Haaland, I., C. Roth, and J. Wohlfar (2022). Designing Information Provision Experiments. *Journal of Economic Literature*, forthcoming.
- Hainmueller, J., D. Hangartner, and G. Pietrantuono (2017). Catalyst or Crown: Does Naturalization Promote the Long-Term Social Integration of Immigrants? *American Political Science Review* 111(2), 256–276.
- Hainmueller, J., D. Lawrence, J. Gest, M. Hotard, R. Koslowski, and D. D. Laitin (2018). A Randomized Controlled Design Reveals Barriers to Citizenship for Low-Income Immigrants. *Proceedings of the National Academy of Sciences* 115(5), 939–944.
- Hainmueller, J., D. Lawrence, L. Marten, B. Black, L. Figueroa, M. Hotard, T. R. Jimenez, F. Mendoza, M. I. Rodriguez, J. J. Swartz, and D. D. Laitin (2017). Protecting Unauthorized Immigrant Mothers Improves their Children’s Mental Health. *Science* 357(6355), 1041–44.
- Hamory, J., M. Kleemans, N. Li, and E. Miguel (2021). Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata. *Journal of the European Economic Association* 19(3), 1522–1555.
- Hanson, G. and C. McIntosh (2012). Birth Rates and Border Crossings: Latin American Migration to the US, Canada, Spain and the UK. *Economic Journal* 122, 707–26.
- Jaeger, D. A., J. Ruist, and J. Stuhler (2018). Shift-Share Instruments and the Impact of Immigration. *NBER* (24285).
- Kahanec, M., M. Pytlikova, and K. Zimmermann (2016). *The Free Movement of Workers in an Enlarged European Union: Institutional Underpinnings of Economic Adjustment*. Labor Migration, EU Enlargement, and the Great Recession. Springer: Berlin.
- Khanna, G., E. Murathanoglu, C. Theoharides, and D. Yang. Abundance from Abroad: Migrant Income and Long-Run Economic Development. *NBER Working Paper* (29862).
- Kinnan, C., S.-Y. Wang, and Y. Wang (2019). Access to Migration for Rural Households. *American Economic Journal: Applied Economics* 10(4), 79–119.
- Lee, D., J. McCrary, M. J. Moreira, and J. Porter (2022). Valid t-ratio Inference for IV. *American Economic Review* 112, 3260–90.



- Lochmann, A., H. Rapoport, and B. Speciale (2019). The Effect of Language Training on Immigrants' Economic Integration: Empirical Evidence from France. *European Economic Review* 113, 265–296.
- Lozano, F. and M. Steinberger (2013). *Empirical methods in the economics of international immigration*. Edward Elgar.
- Mahajan, P. and D. Yang (2020). Taken by Storm: Hurricanes, Migrant Networks, and US Immigration. *American Economic Journal: Applied Economics* 12(2), 250–77.
- McKenzie, D. (2015). *Chapter 4 Learning about Migration Through Experiments*. Oxford Scholarship Online.
- McKenzie, D. (2020). If it Needs a Power Calculation, Does it Matter for Poverty Reduction? *World Development* 127.
- McKenzie, D. (2021a). I'm not a fan of leave-one-out/spatial instruments. *Development Impact Blog*.
- McKenzie, D. (2021b). What Do You Need to Make a Matching Estimator Convincing? Rhetorical vs. Statistical Checks. *Development Impact Blog*.
- McKenzie, D. (2022). Fears and Tears: Should More People Be Moving within and from Developing Countries, and What Stops This Movement ? *World Bank Policy Research Working Paper* (10128).
- McKenzie, D. and J. Mistiaen (2009). Surveying Migrant Households: A Comparison of Census-Based, Snowball and Intercept Point Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172(2), 339–360.
- McKenzie, D. and H. Rapoport (2010). Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migrant Networks. *The Review of Economics and Statistics* 92(4), 811–821.
- McKenzie, D. and M. Siegel (2013). Eliciting Illegal Migration Rates through List Randomization. *Migration Studies* 1(3), 276–291.
- McKenzie, D., S. Stillman, and J. Gibson (2010). How Important is Selection? Experimental VS. Non-Experimental Measures of the Income Gains from Migration. *Journal of the European Economic Association* 8(4), 913–945.
- McKenzie, D., C. Theoharides, and D. Yang (2014). Distortions in the International Migrant Labor Market: Evidence from Filipino Migration and Wage Responses to Destination Country Economic Shocks. *American Economic Journal: Applied Economics* 6(2), 49–75.
- McKenzie, D. and D. Yang (2012). *Chapter 12: Experimental Approaches in Migration Studies*. Edward-Elgar: Northampton, MA.
- McKenzie, D. and D. Yang (2015). Evidence on Policies to Increase the Development Impacts of International Migration. *World Bank Research Observer* 30(2), 155–192.
- Meyer, B. (1995). Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics* 13, 151–161.
- Milanovic, B. (2015). Global Inequality of Opportunity: How Much of our Income is Determined by Where we Live? *The Review of Economics and Statistics* 97(2), 452–460.
- Mobarak, A. M., I. Sharif, and M. Shrestha (2020). Returns to Low-Skilled International Migration: Evidence from the Bangladesh-Malaysia Migration Lottery Program. *SSRN* (9165).
- Otsu, Y. (2021). Sanctuary Cities and Crime. *The Center for Growth and Opportunity* (2020.015).
- Pan, Y. (2012). The Impact of Legal Status on Immigrants' Earnings and Human Capital: Evidence from the IRCA 1986. *Journal of Labor Research* 33, 119–142.
- Parsons, C. and P.-L. Vezina (2018). Migrant Networks and Trade: The Vietnamese Boat People as a Natural Experiment. *The Economic Journal* 128(612), F210–F234.
- Peri, G., K. Shih, and C. Sparber (2015). STEM Workers, H-1B Visas, and Productivity in US Cities. *Journal of Labor Economics* 33(2), 225–255.
- Peri, G. and V. Yassenov (2019). The Labor Market Effects of a Refugee Wave: Synthetic Control Method Meets the Mariel Boatlift. *Journal of Human Resources* 54(2), 267–309.
- Pinotti, P. (2017). Clicking on Heaven's Door: The Effect of Immigrant Legalization on Crime. *The*

- American Economic Review* 107(1), 138–168.
- Qiu, H., J. Hong, X. Wang, and M. Filipksi (2021). Home Sweet Home: Impacts of Living Conditions on Rural-Urban Labor Migration Using Evidence from a Housing Lottery. *Working Paper*.
- Rambachan, A. and J. Roth (2020). Design-Based Uncertainty for Quasi-Experiments. *ArXiv* (2008.00602).
- Remler, D. K. and G. G. V. Ryzin (2014). *Research Methods in Practice: Strategies for Description and Causation*, Second Edition.
- Roth, J. and P. Sant’Anna (2022). When Is Parallel Trends Sensitive to Functional Form? *Econometrica*, *forthcoming*.
- Roth, J., P. Sant’Anna, A. Bilinski, and J. Poe (2022). What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *Working Paper*.
- Sarvimaki, M., R. Uusitalo, and M. Jantti (2022). Habit Formation and the Misallocation of Labor: Evidence from Forced Migrations. *Journal of the European Economic Association*, *forthcoming*.
- Seshan, G. and D. Yang (2014). Motivating Migrants: A Field Experiment on Financial Decision-Making in Transnational Households. *Journal of Development Economics* 108, 119–127.
- Shrestha, M. (2020). Get Rich or Die Tryin’: Perceived Earnings, Perceived Mortality Rates, and Migration Decisions of Potential Work Migrants from Nepal. *World Bank Economic Review* 34(1), 1–27.
- Shrestha, S. A. (2017). No Man Left Behind: Effects of Emigration Prospects on Educational and Labour Outcomes of Non-migrants. *The Economic Journal* 127(600), 495–521.
- Shrestha, S. A. and D. Yang (2019). Facilitating Worker Mobility: A Randomized Information Intervention among Migrant Workers in Singapore. *Economic Development and Cultural Change* 68(1), 63–91.
- Stillman, S., J. Gibson, D. McKenzie, and H. Rohorua (2015). Miserable Migrants? Natural Experiment Evidence on International Migration and Objective and Subjective Well-Being. *World Development* 65, 79–93.
- Stock, J. and M. Yogo (2005). *Testing for Weak Instruments in Linear IV Regression*. Cambridge University Press.
- Theoharides, C. (2017). Manila to Malaysia, Quezon to Qatar International Migration and Its Effects on Origin-Country Human Capital. *The Journal of Human Resources* 53(4), 1022–1049.
- Theoharides, C. (2020). The Unintended Consequences of Migration Policy on Origin-Country Labor Market Decisions. *Journal of Development Economics* 142, 1–22.
- Tjaden, J. (2020). *Chapter 37 Assessing the Impact of Awareness Raising Campaigns on Potential Migrants – What we have Learned So Far*. International Organization for Migration.
- Wah, S. H. (2018). Do Employers in Myanmar Prefer Workers who Accumulated Skills in More Advanced Countries? Evidence from a Field Experiment. *IZA Journal of Development and Migration* 8(1).
- Yang, D. (2006). Why Do Migrants Return to Poor Countries? Evidence from Philippine Migrants’ Responses to Exchange Rate Shocks. *Review of Economics and Statistics* 88(4), 715–735.
- Yang, D. (2008). International Migration, Remittances and Household Investment: Evidence from Philippine Migrants’ Exchange Rate Shocks. *The Economic Journal* 118(528), 591–630.