

Incomplete early draft!

How Quantitative Political Scientists Compare Immigration Law *Tools, Research Designs, and Interdisciplinary Futures*

Samuel D. Schmid, University of Lucerne

Introduction	Currently	800 words
Understanding Immigration Law: Conceptual Tools	Currently	1'800
Measuring Immigration Law: Methodological Tools	Currently	1'500
Quantifying Immigration Law: Indexing Tools	Currently	2'500
Categorizing Immigration Law: Typological Tools	Projected	1'500
Mapping Variation in Immigration Law	Projected	1'000
Explaining Variation in Immigration Law	Projected	1'000
Estimating Effects of Immigration Law	Projected	1'000
Conclusion	Projected	1'000
	TOTAL	12'100 words

Abstract

The time is ripe for the interdisciplinary and quantitative study of comparative immigration law. It can build on a rich literature in comparative political science as well as an increasing number of large datasets and comprehensive databases. In this chapter, I present a toolbox of quantitative solutions to the methodological challenges that this enterprise faces. To do so, I first lay out a synthetic conceptual map of this legal field. In the next two sections, the chapter explains the basics for of translating laws into quantitative measures and discusses key examples from the field of comparative immigration policy analysis in political science. The aim is to show how this field has evolved and how valid and reliable indices of various dimensions of immigration law can be created across various units of analysis. The next section then discusses some key typologies that are, in contrast to many earlier ones, rooted in sophisticated quantitative datasets. These three sections on measurement, indexing applications, and typological approaches is the most extensive because, in my view, it is mostly here where more traditional comparative lawyers have most to contribute. I then proceed to briefly flesh out three comparative research designs in the next three sections: descriptive comparisons that map variations in immigration law across space and time, comparisons that aim to test causal explanations of variations in immigration law, and comparisons that estimate causal effects of variations in immigration law on various social and individual-level outcomes. I do so by unpacking some pertinent examples from the literature. I conclude by elaborating the potentials for interdisciplinary conversations and cross-fertilizations and delineate specifically what more traditional comparative lawyers can bring to the table of the future of quantitative immigration law. Instead of arguing that comparative legal scholarship should practice quantitative methods on its own, I propose some alternative ways of how comparative lawyers and comparative political scientists can usefully divide labor and collaborate productively while avoiding some dangerous pitfalls.

1. INTRODUCTION

When we think of scientific endeavors, many of us will picture numbers and people thinking and discussing in front of whiteboards with formulas and calculations. Few, if any, will think of paragraphs and legal scholars analyzing and comparing their function in various social settings (as the predominant functionalist approach in comparative law would have it; see Zweigert and Kötz 1998; Kischel 2019). Against this background, it may seem outlandish to suggest that comparative legal scholarship can and should embrace scientific methods that aim for using some sort of quantification. Yet, this is what is happening in the intense methodological discussion on how to advance comparative law. While some remain skeptical (e.g. Kischel 2019: 135-140), there are many voices calling for greater use of the many quantitative comparative tools developed by the social sciences, not only to measure but also to explain and estimate effects of immigration law. However, thus far comparative case study methods – which are qualitative in nature – have received most attention in comparative law, particularly in comparative constitutional law (e.g. Hirschl 2005, 2014). Overall, quantitative methods have remained at the margins and their value for comparative law has been recognized only more recently without fully fleshing out a workable methodology (e.g. Siems 2014, 2019).

I want to mention two notable exceptions. One is the quantitative cross-national study by Whytock (2008), and the other the book *How Constitutional Rights Matter* by Chilton and Versteeg (2020), who combine statistical analyses with other methods. It is no coincidence that these studies have emerged in the interdisciplinary field of comparative constitutional studies, where scholars of comparative politics had already produced statistical analyses of the causes and effects of constitutional features and provisions using extensive global and longitudinal datasets and measures such as those from the *Comparative Constitutions Project* (the current version at the time of writing is Elkins and Ginsburg 2021). To learn more about an object of study, both interdisciplinary exchange and the generation of large datasets are thus key. Regarding the former, we must acknowledge that especially comparative political science and comparative law have ignored each other for too long when it is so obvious how much they have to say to each other (Kischel 2019: 23-4, 26). Regarding the latter, as many legal domains under scrutiny have shown, it is large quantitative datasets that go beyond the confines of single studies that can unleash the full potential of comparative analysis of these domains. By establishing benchmark measures to comparatively map variation in key dimensions and classify cases, such datasets can fertilize both quantitative and qualitative investigations. This, in turn, is useful for political science and comparative law as well as for other disciplines, and for the conversation between them.

The time is ripe for the interdisciplinary and quantitative study of comparative immigration law.¹ It can build on a rich literature in comparative political science as well as an increasing number of large datasets and comprehensive databases. In this chapter, I present a toolbox of quantitative solutions to the methodological challenges that this enterprise faces. To do so, I first lay out a synthetic conceptual map of this legal field. In the next two sections, the chapter explains the basics for of translating laws into quantitative measures and discusses key examples from the field of comparative immigration policy analysis in political science. The aim is to show how this field has evolved and how valid and reliable indices of various dimensions of immigration law can be created across various units of analysis. The next section then discusses some key typologies that are, in contrast to many earlier ones, rooted in sophisticated quantitative datasets. These three sections on measurement, indexing applications, and typological approaches is the most extensive because, in my view, it is mostly here where more traditional comparative lawyers have most to contribute. I then proceed to briefly flesh out three comparative research designs in the next three sections: descriptive comparisons that map variations in immigration law across space and time, comparisons that aim to test causal explanations of variations in immigration law, and comparisons that estimate causal effects of variations in immigration law on various social and individual-level outcomes. I do so by unpacking some pertinent examples from the literature.

The chapter closes by concluding that not only “comparative law seems to be too important to be left to comparative lawyers” (Siems 2014: 312). I argue that the same applies to comparative politics and political scientists. I elaborate the potentials for interdisciplinary conversations and cross-fertilizations and delineate specifically what more traditional comparative lawyers can bring to the table of the future of quantitative immigration law. Instead of arguing that comparative legal scholarship should practice quantitative methods on its own, I propose some alternative ways of how comparative lawyers and comparative political scientists can usefully divide labor and collaborate productively while avoiding some dangerous pitfalls.

¹ See Cope (2021) for a similar yet broader argument to incorporate social scientific methods in the study of comparative migration law.

2. UNDERSTANDING IMMIGRATION LAW: CONCEPTUAL TOOLS

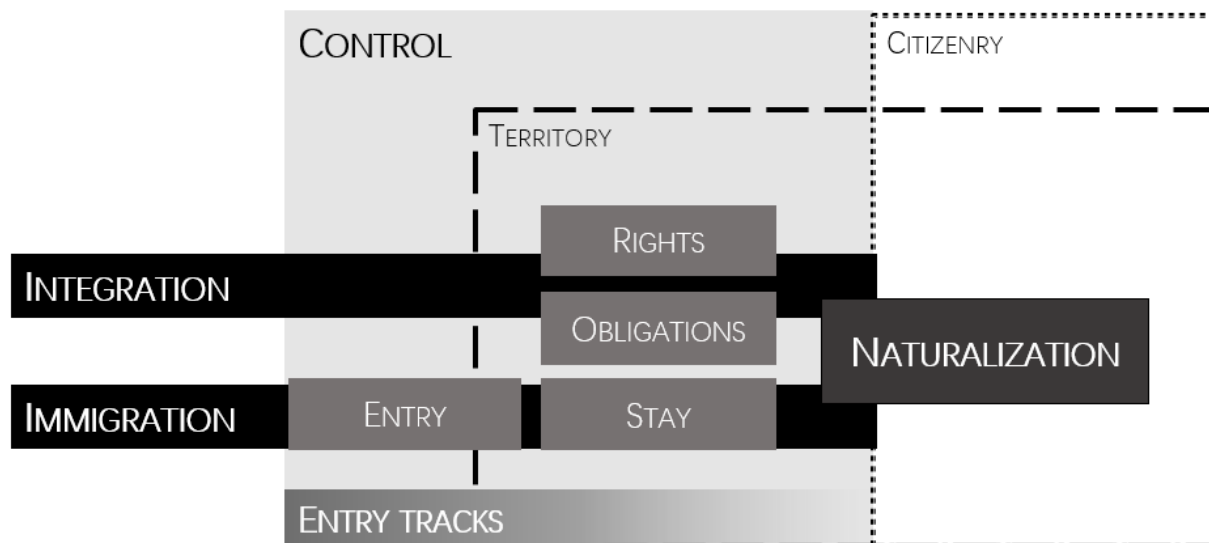
The basis for any comparative exercise is to define the object under scrutiny. In the language of comparative law, this object constitutes the *tertium comparationis*. It can be defined as the “the equivalent attribute” (Cope 2021: 3) of various legal systems that is to be compared across these systems. Most often this attribute relates to the substance of law and can be expressed as an abstract concept (Kischel 2019: 5). In the social sciences, leading methodologists have long recognized that valid and reliable measurement of any abstract concept requires rigorous conceptualization (Adcock and Collier 2001; Goertz 2006, 2020). However, many quantitative studies in comparative politics do not devote enough attention to proper concept specification, and sometimes this is also the case in the comparative study of immigration law. This is why this chapter does immediately jump to technicalities of measurement. It goes back even a step further. I first provide conceptual synthesis that encompasses the relevant components that immigration-related laws may exhibit. In describing this map, I hope to elucidate the architecture of immigration law and its potential variations across space and time. As will become clear, it is crucial to understand this architecture before elaborating abstract concepts based on which we can devise adequate and useful measures of immigration law. Measurement is the topic of the next two sections.

Immigration-related laws are primarily constituted by two parallel legal domains – *immigration* and *integration* (cf. Hammar 1985). These domains no longer or only partially apply once a third crucial legal domain – that of *naturalization*. It transforms the legal subjects of non-naturalized immigrants into resident citizens (cf. Hammar 1990). Mechanisms of *control* and *enforcement* constitute the fourth and final domain. It defines how borders and regulations of immigration, integration, and naturalization are policed and how these measures materialize in various government activities as well as physical and non-physical infrastructures. Each of these attributes of immigration law may be *differentiated according to the category of immigrants* it applies to. These categories are primarily defined by the specific entry channel – such as labor immigration or asylum – or by the nationality or origin of immigrants, or by a combination of both. These differentiations can vary across both space and time, as different receiving countries possess different immigration systems with different legal architectures at different points in time.

The domain of *immigration* defines the rules that determine who can enter and stay, and who must leave. I reduce these aspects into *entry* and *stay*, conceiving of the latter as the negation of who must leave – these are those that cannot stay. Entry regulations are *external* in nature – they apply extra-territorially. They consist of various requirements and conditions that immigrants must fulfill to qualify

for the establishment of legal residence in a receiving country. Laws that regulate stay operate *internally* – they apply within the territory of a nation-state. They confer residence rights that define how long a certain immigrant can stay in the receiving country based on the permit under which s*he as entered. These regulations may not allow all category of migrants to renew their visas or transition toward permanent residence in case the entry permit is temporary. In other words, obtaining a permanent permit may depend on additional conditions. In some cases, for instance based on the Green Card in the US, immigrants can also receive permanent residence rights upon territorial admission.

Figure 1 The National Architecture of Immigration Law



This potential differentiation across immigrant categories and entry permits brings us to the notion of *entry tracks*. Constituting distinct ways of entering a receiving country, these tracks are distinguished based on “the purpose of migration as well as on various characteristics of the applicant” (Beine et al. 2016) and are “identified for any situation where possible applicants receive distinct treatment in the law, based on any given characteristic of that applicant’s profile evoked in the wording of the law or regulation” (REF). The nature and number of entry tracks depends on the receiving countries immigration regime and its architecture at a given point in time. These architectures can vary widely across space and time. Different countries may recognize and facilitate different purposes and types of migration, and they may regard to different characteristics as relevant for their laws. This cross-national variation is compounded by potential intra-national variation over time. For instance,

not all countries target specific categories of labor immigrants in the same way; some countries at some points in time may have specific policies on low-skilled labor immigrants while not having specific ones on high-skilled immigrants, whereas others may have the opposite configuration and still others only have non-skill-specific entry tracks for labor immigration, entry tracks that are differentiated along criteria other than skill, or any combination of the above. In other words, the entry tracks that constitute the architecture of different immigration systems are likely never to be congruent across countries. They are more likely to be congruent within countries over time as adding and removing entry tracks is a rarer occurrence, but also within-country variation is possible.

Entry tracks can be conceptually grouped under various broader migrant categories with which immigration laws recognize different migrant groups and their (putative) purposes for migration (Helbling et al. 2017: 83).² In the literature as well as in databases, we typically find at least three of categories: labor immigration (migration for economic purposes), asylum and refugee admission (migration for humanitarian purposes), and family reunification or formation (migration for social purposes). In addition, while some conceptualizations add co-ethnics (migration for cultural reasons; Helbling et al. 2017), others also consider international student migration (for educational reasons; Beine et al. 2016). The most comprehensive conceptualization – which can be found in the DEMIG database (de Haas et al. 2015) – completes this catalogue with the consideration of irregular migrants as a separate category. Labor immigrants as well as asylum seekers and refugees have been the most prominent migrant groups in both academic and political debates. Economic and humanitarian entry tracks are sometimes also seen as the primary means of entry (Schultz et al. 2021). Migrants that reunite with family members or move to form a family are a secondary and cross-cutting group as they can be (future) members of families whose sponsor can belong to any kind of migrant group listed above, including irregular migrants. Family migration is cross-cutting also in another sense. It can be considered as an *immigration right* of (future) family members or, alternatively, as an *immigrant right* of the sponsor.

This brings us to the domain of *integration*. It defines the *rights* and *obligations* that are applied to (specific) immigrants after establishing residence in a receiving country. Hence, these regulations operate *internally*. It is useful to distinguish rights and obligations *tied to permits* from rights and obligations *tied to other criteria*. Depending on the number of permits and their variation in legal content, *permit-based discriminations* do not only exhibit cross-national but also intra-national variation across

² These are purposes as recognized by law and are not to be confused with actual migrant motivations, which are often much more complex.

specific permits for specific immigrants and entry tracks (and in federal contexts they might even vary across territorial sub-units). By contrast, *more general rights and obligations* mainly show cross-national rather than intra-national variation. For instance, the right to work is often tied to permits, while anti-discrimination rights are more general. Non-citizen voting rights, by contrast, often discriminate based on immigrants' nationalities. This shows how other criteria may add to the differentiated architecture of immigration regimes. It also shows that immigration and integration domains may interact or intersect. Not only can a specific right apply only to holders of a specific permit or nationality, but obligations such as language tests may be needed to qualify for longer or permanent stay. Conversely, a permanent residence permit may come with a host of additional rights and potentially also obligations. Language learning and testing may also form part of entry regulations that apply externally (Goodman 2014).

The domain of *naturalization* defines the conditions under which immigrants can transform into full and equal members of the receiving national community and thus no longer be subject to immigration law. Citizenship status fortifies residence and re-entry rights and all remaining differentials in rights are no longer based on immigration law but on other criteria deemed relevant by states, for instance by specific regulations of the welfare state.³ Still, the regulation immigrants' access to citizenship itself – abbreviated as *citizenship law* here – can be considered as a part of immigration law broadly conceived. Regarding immigration and citizenship acquisition (rather than emigration and the loss of citizenship), citizenship law comprises the regulations that define the conditions for immigrants to acquire the legal status of full membership in the sense of nationality or passport-holding. These regulations range from conditions based on residence (in terms of duration and immigration status) to conditions to renounce other nationalities to integration conditions such as citizenship and language tests as well as economic and criminal record requirements. An expansive definition may also consider the conditions under which citizenship is conferred to children of non-citizen parents based either on birth on the territory of the receiving country – *ius soli* birthright citizenship – or socialization in the receiving society. Even more expansive definitions may also consider otherwise facilitated naturalization, as is often the case for refugees or spouses. This again shows that differential treatment can be based on various criteria, with some of them based on entry tracks. Note that this linear conception of immigration, integration, and naturalization is meant to capture the constituent components of law rather than immigrant behavior, which may be non-linear or otherwise more

³ There can be exceptions for dual citizens or naturalized citizens, however. For instance, some Latin American states restrict electoral rights (voting and candidacy) for these categories of citizens (Schmid et al. 2019).

complex. Crucially, not every immigrant will end up being a citizen or permanent resident of the receiving country. At the same time, some who naturalize may do so from abroad without ever setting foot in the country whose passport they acquire, for instance by claiming descent-based birthright citizenship or by buying citizenship through investment.

The fourth and final domain is *control* and comprises both mechanisms to monitor enforcement as well as enforcement, policing, and border control measures themselves. Control can apply both outside as well as inside national territories as borders and their physical and non-physical manifestations are shifting both outwards and inwards to achieve the various ends of different national immigration laws (Shachar 2020). In principle, they can also apply to not only to immigration but also to integration and naturalization, yet most government efforts to control and enforce borders are concentrated in the domain of immigration. Finally, again, control may be differentiated along entry tracks, immigrant categories, their nationalities, or according to other criteria.

3. MEASURING IMMIGRATION LAW: METHODOLOGICAL TOOLS

The first attempts to measure immigration-related policies in the social sciences relied on a “national model approach” based on simple and often binary typologies (Koopmans 2013: 696). One of the earliest and most influential dichotomous distinctions goes back to Brubaker (1992). He proposed that there are two “idioms of nationhood” that are at the heart of national citizenship laws: an exclusive “ethnic” type based on descent-based birthright citizenship (*jus sanguinis*) and an inclusive “civic” type based on territorial birthright citizenship (*jus soli*). For Brubaker, a prime instance of the former is Germany whereas the latter manifests clearly in France. Arguing that these types of nationhood are deeply entrenched and persistent, Brubaker anticipated that birthright citizenship regimes in Germany and France would not converge over time.

In the following decades, this “national model approach” and its usefulness for empirical analysis has been questioned more and more. It has come to be considered too static, too simplistic, and too normative (Helbling and Vink 2013: 552; see also Finotelli and Michalowski 2012: 233-5; Bertossi and Duyvendak 2012). Instead, as the previous section has summarized, it has become clear that contemporary immigration-related laws – including citizenship law itself – constitute complex and dynamic regimes that rarely correspond to neat categorical distinctions. They reflect differences in degree much more than differences in kind (Bauböck 1996: 67; VINK REF), and while they continue to be constrained by past trajectories, they are only completely path-dependent but instead can change

significantly over time (see e.g. Howard 2009; Koopmans et al. 2012; Goodman 2014). After Germany had introduced a conditional form of *jus soli* in 2000, even Brubaker (REF) acknowledged that birthright citizenship regimes are not as persistent as he had theorized.

Against this background, a large strand of the empirical literature in comparative politics has moved beyond single and small-N comparative case studies and has instead developed increasingly composite indices and comprehensive databases that we can use to measure immigration-related laws comparatively in a quantitative way. This has led to significant progress in the field. The study of citizenship law again provides a prime example. Using fine-grained measures of numerous legal components of citizenship regulations, an innovative comparative study has demonstrated that European states' citizenship laws today in fact combine *jus soli* and *jus sanguinis* birthright citizenship principles in different ways and to different degrees (Vink and Bauböck 2013). Germany is a case in point. Thus, through the increasing sophistication in measuring citizenship law, we now know that across Europe *jus soli* and *jus sanguinis* provisions are part of two independent dimensions, thus defying what initially seemed to be an easy “ethnic-civic” distinction. Whether this holds beyond Europe remains an open question, but it surely will be tackled soon as global datasets in the field of citizenship are now available (Vink et al. 2021)

To understand how such potentially ground-breaking measures can be created, this section explains the basic approaches to and key challenges of index-building using legal indicators. The next section then discusses how two specific measurements that quantify certain aspects of immigration law have approached and addressed these challenges of index building. Based on my critical appraisal of each application, I hope to show how valid and reliable legal indicators and composite indices in the field of immigration law are constructed.

Two basic philosophies to measure social science concepts can be distinguished. The first is based on *theoretical deduction* and an *ontological and semantic approach* to concept formation. This approach finds its most sophisticated elaboration in Goertz (2006, 2020). For him, social science concepts are best theorized using a three-level framework. To conceptualize is to define what a concept means – in an ontological and semantic sense – at its basic level, what dimensions constitute the concept at its second level, and what indicators can be used to measure these dimensions at the third level. This concept structure is then translated into mathematical aggregation rules by determining how third-level indicators relate to each other in constituting second-level dimensions, and how these dimensions relate to each other to yield composite measures of the basic-level concept. In a nutshell, Goertz argues

that when we build aggregate measures of any social science concept, we should do so based on a sound and elaborate three-level specification of the concept structure that we aim to capture with the aggregate measure.

The second strategy is based on *empirical induction* and a *latent variable approach* to concept formation (the classic account is Bollen 1989). Within this approach, what belongs to which dimension of a concept is derived mainly from empirical rather than theoretical considerations. Only if the statistical correlation between all indicators is very high, they are assumed to reflect the same dimension, latent variable, or concept. Indeed, the latter terms can be used interchangeably within this two-level approach – it does not operate with a more complex three-level and most often also not with multi-dimensional concept structures. Instead, it devotes most attention to the selection of indicators that are sufficiently correlated so that they can be reduced to a single consistent, reliable, and one-dimensional statistical measure. Most often researchers simply use the resulting latent variable itself also as the aggregate measure. This means that potentially complex relationships between indicators that might better capture the ontological nature of concepts are typically not considered. Latent variable models also determine empirically with what weight each indicator is factored into this additive aggregate measure, regardless of its theoretical importance.

These two philosophies thus deal differently with the three fundamental challenges of index construction. These are *conceptualization*, *measurement*, and *aggregation*. The piece by Munck and Verkuilen (2002) has become the standard reference that unpacks these three challenges most systematically and develops guidelines for best practice. They argue that to address the challenge of conceptualization, one begins by identifying the relevant attributes of the concept to be measured. This concept specification should avoid both minimalist and maximalist definitions. Next, these conceptual attributes should then be organized hierarchically using a graphical representation – a so-called “concept tree”. Here it is important to identify potential conceptual redundancies and conflation; if present, these should be eliminated.

To solve the challenge of measurement, according to Munck and Verkuilen (2002) we must then select valid indicators that properly capture the conceptual attributes. Validity is achieved when there is no systematic measurement error inherent in the construction and content of the indicator so that it measures what it is supposed to measure. It is like a gun that must be pointed correctly to hit the target. To ensure validity, it is best to use multiple indicators and ensure cross-national equivalence. Also, we should use indicators that can be cross-checked based on multiple sources. The additional

standard of reliability must also be met. Measures are reliable when there is no random measurement error, especially when multiple researchers code the same indicators across many cases. This is like some wind that may send the bullet off target in different ways when shooting multiple times. Next, we must select and justify the measurement level of these indicators. Metric indicators use and function like linear numbers, while nominal indicators represent parallel or flat categorical distinctions. Ordinal indicators are in between as they are also categorical but ordered in a way so that higher values represent “more” of the measured attribute without functioning like numbers that have linear spacing between them – ordinal indicators are non-linear. The measurement level must also be valid and reliable. At the same time, it must be replicable – coding rules, the coding process, and disaggregated data must be made transparent and publicly available.

Finally, to meet the challenge of aggregation, Munck and Verkuilen (2002) argue that we must first select the level of aggregation. In doing so, we should balance the goal of having a parsimonious one-dimensional measure with the concern that measures of various attributes may contain multiple dimensions that an aggregate score would hide. Aggregation always implies the loss of some information, and this loss should be minimized. Next, the aggregation rule must be properly specified. Here we must ensure that the theoretical concept structure – how the various attributes on different levels relate to each other – is consistent with the mathematical formula we use to calculate the aggregate measure. This should again be made fully transparent and replicable.

Thus understood, the framework by Munck and Verkuilen (2002) applies elements of both philosophies of index building, but it is closer to the ontological and semantic approach than the latent variable approach. This reflects the evolution of scientific practice in the social sciences. While the latent variable approach has dominated quantitative work for decades, the ontological and semantic approach has evolved from its more qualitative roots into a much more broadly accepted template also among quantitative researchers. Overall, the state of the art in comparative political science has progressed to a point where in many research areas – ranging from the measurement of democracy (REF V-DEM) to the comparative analysis of immigration-related laws – combine these two strategies to simultaneously leverage their specific strengths or to tailor each approach or variations and mixes thereof to specific measurement problems.

4. QUANTIFYING IMMIGRATION LAW: INDEXING TOOLS

In this section, I critically discuss how two specific quantitative measures of immigration law – Helbling et al. (2017) and Peters (2017) – have approached and addressed these challenges of index building. Each of them applies one of the two different basic philosophies of concept formation. Overall, I hope to show how valid and reliable legal indicators and composite indices are constructed in practice in this field of study. Against this background, the next section presents more recent typologies that have emerged in the field of immigration law and shows how they have overcome the limitations of the classical “national models approach”. Note that I do not aim to provide a comprehensive overview of existing quantitative legal indicators and databases in the field of immigration law. There are many other pieces that already provide such information (Boucher and Gest 2014, 2016; Bjerre et al. 2015; Helbling 2013, 2016; Goodman 2015, 2019; Solano and Huddleston 2021; Schmid 2021).

The first application I want to present is the *Immigration Policies in Comparison Database* (IMPIC). I believe that it represents one of the most ambitious and sophisticated resources to measure immigration law. In its current version at the time of writing, it covers data across the OECD from 1980 to 2010. To construct this database, Helbling and colleagues (2017) start with a comprehensive multi-dimensional conceptualization of immigration law. Focusing on *immigration* rather than integration, the first dimension distinguishes four broad “policy fields” that will be familiar from the first section of this chapter: labor immigration, family reunification, asylum, and co-ethnics (Helbling et al. 2017: 83). Orthogonal to this first dimension, the second dimension distinguishes between two different “modus operandi,” separating *regulations* from *control mechanisms*. Regulations refer to “binding legal provisions that create or constrain rights” (Dreher 2002, cited in Helbling et al. 2017: 85) in each policy field. By contrast, control mechanisms “monitor whether the regulations are adhered to” (ibid.). They are not distinguished along policy fields but instead apply across them, and they also include elements that refer to the treatment of irregular immigrants.

Thus far, one may note the similarity to the conceptual synthesis developed above, as I partly drew on the IMPIC conceptualization to distinguish the overarching function of *control* from other aspects. What follows is also similar: two “locus operandi” account for the fact that “states regulate and control immigration not only at their borders, but also within their territories” (Helbling et al. 2017: 85). Accordingly, for both regulations and control mechanisms, IMPIC differentiates between laws that operate *externally* and laws that operate *internally*. Moreover, there are several sub-dimensions

of regulations. *External regulations* consist of *eligibility* requirements, which stipulate the criteria for immigrants to qualify for a certain entry track, and *conditions*, which define additional requirements that need to be fulfilled (ibid.). I have referred to these attributes in my conceptual synthesis as *entry*. *Internal regulations* consist of *security of status*, which comprises all policies that regulate the duration of specific permits and the access to long-term settlement, and the *rights associated* with specific permits across entry tracks, which mainly comprise the access to the labor market that makes residence economically viable (see Bjerre et al. 2015: 7 for a justification why these specific rights are considered, while other more general immigrant rights are not). These attributes therefore relate to *stay* and *rights* in the conceptual synthesis. Only the aspect of rights somewhat blurs the boundary that IMPIC draws between immigration and integration and reminds us that the two are necessarily intertwined and relevant for each other once we aim to gauge immigration law comprehensively.

Entry tracks – the most basic unit that constitutes immigration law – are then grouped by IMPIC along the broad policy fields. For instance, for labor immigration, low-skilled, high-skilled, self-employed, and unspecified entry tracks are distinguished. Various ordinal indicators are then used to measure the relevant field-specific internal and external regulations features of each entry track. For instance, one indicator for labor immigration is skill-based discrimination itself. If an entry track only allows highly skilled immigrants to enter, the entry track receives a very restrictive score – if it allows more immigrants to enter, it is less restrictive depending on how many or whether all skill-levels are allowed. [EXAMPLE OF ANOTHER INDICATOR NECESSARY?] In each policy field, the indicators across existing entry tracks are then averaged using arithmetic means applying equal weights to yield aggregate scores for internal and external regulations and for the policy field more broadly. These scores can then be combined to further with other policy fields to gauge more general levels of immigration law *restrictiveness*. If desired, one could also include the indicators that measure external and internal control mechanisms to derive an even more comprehensive measure.

The IMPIC Database carefully solves each challenge of index building. This is not surprising as Helbling and colleagues (2017) have been guided by the framework of Munck and Verkuilen (2002). Regarding conceptualization, it constitutes an extraordinarily complete effort as it develops an overarching multi-dimensional grid that manages to capture the hierarchical ontology and architecture engrained in national immigration laws without redundancies or conceptual conflation. This also makes the project rather maximalist, but this is an advantage for a database that is explicitly designed to be used by various researchers for varying purposes. By focusing on the notion of *restrictiveness* as the

generic concept to be measured, the project can also avoid more normative or potentially ambiguous terms such as *openness* or *inclusiveness*, although they are implied as the negative pole of restrictiveness.

Regarding measurement, IMPIC also rigorously derives its ordinal measurements for each indicator. This measurement level is typical for legal indicators in the realm of immigration as most laws can be characterized by “more or less of something” considering a specific concept rather than something that can be captured with a linear number or with flat nominal categories. [HERE I WILL EXPLAIN IN MORE DETAIL HOW THE ORDINAL SCALING IN IMPIC WORKS.]

In sum, the indicators in IMPIC appear valid. Various other aspects of validity – ranging from content validity to comparisons with findings from other databases (external validity) to the correspondence between conceptualization and empirical dimensionality using latent variable models (internal validity) – have also been explicitly tested and confirmed in a separate analysis by Schmid and Helbling (2016). Meanwhile, reliability in measurement has been mainly ensured by relying on individually employed national experts filling out a questionnaire that is then used to score each indicator for each country and year under investigation. Finally, the additive aggregation schemes are well-justified and, like any other aspect and decision of the project, is made fully transparent and replicable.

As it is informed by Munck and Verkuilen (2002), the IMPIC Database and its approach tend towards the ontological and semantic approach to concept formation. Another notable and encompassing dataset in the field of immigration law – the one introduced by Peters (2017) – shows how the latent variable approach to concept formation works. The dataset covers 19 countries until 2010 with varying starting times for the observations. In some countries, the coverage extends back to the late 18th century – like for the USA and France – while other countries are observed from the 19th century and some from the 20th century onwards (for details see *ibid.* 45). Alongside Western democracies, also Latin American as well as East Asian countries and two autocracies in the Persian Gulf (Saudi Arabia and Kuwait) are included.⁴ Hence, this dataset boasts impressive historical depth as well as geographical and political diversity, which helps broaden the scope of Peter’s core argument and finding – that as the openness to international trade has liberalized immigration laws have become more restrictive – while at the same time keeping the number of cases low enough for data collection be manageable.

⁴ Shin expands this dataset (ADD DETAILS).

The key common denominator, and the scope condition of the argument, is that these states are relatively rich and need low-skilled labor. Peters therefore focuses on low-skill immigration specifically, defining immigration policy as “the laws that policymakers pass to regulate the number of low-skill immigrants entering (and potentially leaving) the country in a given year” (ibid. 243). She explains that she ultimately wants to capture the intent of policymakers: whether they want to encourage or discourage low-skill immigration. To do so, Peters identifies “policies that either specifically target low-skill immigrants, such as agricultural workers programs, or that target all immigrant groups and do not exclude low-skill immigrants” (ibid.). This identification was done inductively after the sample had been determined. Evidence from primary and secondary sources were consulted to derive *twelve different categories of immigration laws* that used similar criteria: nationality-based discrimination, skill-based discrimination, quotas, recruitment, work prohibitions, family reunification, refugee policy, asylum policy, access to citizenship, other immigrant rights, deportation, and enforcement.

For each category, Peters (2017: 253) then develops and employs a single composite indicator with an ordinal coding scheme from 1 to 5, with greater openness receiving higher values. For instance, for the access to citizenship 1 means that citizenship can only be obtained by birth to a native parent while 5 denote configurations of citizenship laws that combine territorial birthright citizenship (citizenship is obtained automatically by children born to non-citizen parents in the territory) with a “fairly easy process” of naturalization such as one without language requirements that can occur after a residence requirement of five years or less (ibid. 267). After an initial score is determined in a country in its starting year in the dataset, policy changes lead to the modification of the score on the scale according to their magnitude, ranging from the most minor changes (coded as a change of 0.1) to the most drastic changes (coded as a change of 4; ibid. 253). For example, a 0.25 reduction in the overall citizenship score was applied to Great Britain when the country started demanding from immigrants that they know English and must demonstrate an understanding of British society in 2002. A further introduction of a formal citizenship test led to another 0.25 deduction in 2005, yielding an overall score of 3.25. Britain had started out with a score of 4.5 in 1792 because both birth and residence in the territory led to subjecthood in the empire. While many research assistants helped compile the sources and refine the coding scheme, Peters coded the data herself when the final scheme and coding rules had been determined. A reliability test showed that the values assigned by a second coder across eight countries correlated with the coding of Peters at 0.9 (ibid. 252). This is a very high positive correlation

(assumedly using Pearson's correlation coefficient) that indicates strong inter-coder agreement and thus reliability. Therefore, the coding rules lead to replicable data.

The guiding notion for aggregation is that the resulting one-dimensional measure should reflect how each policy category affects the number of immigrants, which she assumes to be the intent driving the making of immigration laws (Peters 2017: 285). Peters argues that a latent variable model – in this case, more specifically, a principal component analysis – can best determine how the categories are factored in. Other weighting schemes are deemed more arbitrary (ibid. 286). The model shows that there are two main empirical dimensions in the data, one related to immigration policy proper and another to immigrant rights. Peters (ibid.) then uses the first dimension of the model, labeling it conceptually as *Immigration Policy Openness*.

Obviously, Peters' approach is very different from that of Helbling and colleagues. All three challenges of index-building – from the conceptualization to the measurement and its aggregation – are solved inductively by Peters and much more deductively by Helbling and colleagues. The two projects reflect the two contrasting index building philosophies. When evaluated according to Munck and Verkuilen (2002), which places more emphasis on deduction, Peters' inductive approach appears more problematic. In the following, I unpack some of these problems. Regarding conceptualization, most importantly, one does not know why there are exactly twelve immigration policy categories. Peters (ibid. 252) concedes that certain policy categories could be further disaggregated. For instance, one category covers all kinds of immigrant rights (short of the access to citizenship, which is a separate category) without distinguishing these rights further. This points to the lack of the conceptual hierarchy that Munck and Verkuilen (2002) would like to see. If rights were disaggregated, they would still enter the list on the same conceptual level. This is typical for the two-level latent variable approach but also shows why index building methodology is moving beyond and increasingly embraces multi-dimensional and multi-level constructs. One could then also ask why asylum and refugee policies are not grouped into one overarching category of humanitarian admissions rather than having two categories for this policy field while having only one for family reunification. This shows that the index would be stronger if its rather maximalist conceptual approach would be matched by an appreciation of the potentially complex and multi-level relations between the various immigration policy categories.

The challenge of measurement reveals an issue in Peters' approach that is related to its two-level methodology. She uses composite scales for each category rather than more fine-grained and pre-defined categories that operationalize the category more precisely (which would need another lower

level of conceptual abstraction). This makes the measurement much more complex and open-ended. Yet, it also makes it very versatile and flexible, allowing to capture any sort of policy change. Indeed, this shows that there is a trade-off between two desirable properties of useful measurement tools: the clarity of their operationalization and the flexibility to account for any change in legal substance that is not anticipated by the index builder. Flexibility becomes increasingly important as global datasets are constructed. They need to be fit to adequately capture realities that are functionally equivalent across different political systems. This is why Peters' approach, if appropriately incorporated into multi-dimensional concepts, could be very useful for future index-building projects (see Schmid 2022).

Peters' approach is most problematic in how it solves the challenge of aggregation. Recall that latent variable models imply that the aggregation scheme is additive, and that the empirical correlations between the indicators define with what weights they are combined. This is not a fundamental issue when the latent variable model retrieves a single statistically consistent dimension based on very high positive correlations of the indicators. But this is not the case in Peters' data. In contrast to the other policies, policies regarding refugees, asylum seekers, and family reunification are negatively related to the latent variable (ibid. 287). Using this latent variable as an index therefore means that while the other policies are correctly valued, *more open immigration* for refugees, asylum seekers, and family migrants count towards *lower values in the resulting index* – all else being equal. Nevertheless, while this leads to concept-measure inconsistency – as the statistical direction and the conceptual meaning of indicators do not align – the problem may not be serious as validity tests with other aggregation techniques such as standardized and equally weighted averages of the indicators estimated to be most important lead to highly correlated aggregate measures (ibid. 286).

[PARAGRAPH NEEDED TO CONCLUDE ON PETERS INDEX.]

[THE CHAPTER COULD CONCLUDE HERE WITH ANOTHER FITTING PARAGRAPH CONTRASTING IMPIC AND PETERS.]

[BUT: It would be useful to discuss further immigration-related measures that highlight another key challenge pertinent in this field: the definition of the unit of analysis. Also, the conceptual labels may need more discussion and selectiveness by skill could serve as a prime example and put into relation to the notions of restrictiveness and openness.]

5. CATEGORIZING IMMIGRATION LAW: TYPOLOGICAL TOOLS

Here I want to discuss Koopmans' ICRI, my own typology in Schmid (2020) and the one in CROSSROADS (Boucher and Gest 2016). I would end with voices that are "against regime types" (Schultz et al. 2021). The key message is here is that we need to discuss more rigorously our overall conceptual labels we attach to immigration law, whether they are single dimensions, and whether we need more than one dimension to capture these notions, which leads to typologies. The different typologies I discuss also differ in how, methodologically speaking, they arrive at these types in the end. I also want to discuss potential normative issues – they can be well related to ideal types and their empirical approximations.

6. MAPPING VARIATION IN IMMIGRATION LAW

This section explains how – once we have measures of immigration law – we can descriptively map immigration law across space and time. It does so by unpacking trend analyses that use IMPIC data and Peters data and shows why the two do not agree. It then briefly contrasts the findings by Boucher and Gest (2016) and Schmid (2020) and also explains why they differ. If there is space, I also want to outline some more sophisticated statistical techniques that could be applied in the future.

7. EXPLAINING VARIATION IN IMMIGRATION LAW

This section describes methods that we can use to identify explanatory factors that explain variations in the cross-national and cross-temporal patterns identified in descriptive analyses. It distinguishes between X-centric and Y-centric approaches, and some key sub-categories thereof. It also discusses the issue of valid causal inference for both observational and experimental data. I do so again by describing two prominent examples in the literature (Koopmans et al. 2012 and Peters 2017).

8. ESTIMATING EFFECTS OF IMMIGRATION LAW

This section describes how we can estimate the effects of various aspects of immigration law on various social and individual-level outcomes. It does so by contrasting the gold standard for causal inference – experimental and quasi-experimental designs – to more traditional observational applications. I again discuss two pertinent examples (Hainmueller et al. 2015 and Helbling and Leblang 2019).

9. CONCLUSION

The conclusion ends with a call for comparative legal scholars and political scientists to collaborate. It proposes areas they can work on together while maintaining a sensible division of labor.