

THE CBR-LRI DATASET:
METHODS, PROPERTIES AND
POTENTIAL OF LEXIMETRIC
CODING OF LABOUR LAWS

Zoe Adams, Parisa Bastani,
Louise Bishop and Simon Deakin

WP 489
March 2017

**THE CBR-LRI DATASET: METHODS, PROPERTIES AND POTENTIAL OF
LEXIMETRIC CODING OF LABOUR LAWS**

Centre for Business Research, University of Cambridge
Working Paper No. 489

Zoe Adams
Centre for Business Research
zla20@cam.ac.uk

Parisa Bastani
Centre for Business Research
pb466@cam.ac.uk

Louise Bishop
Centre for Business Research
lebishop42@gmail.com

Simon Deakin
Centre for Business Research
s.deakin@cbr.cam.ac.uk

March 2017

Abstract

Leximetric data coding techniques aim to measure cross-national and inter-temporal variations in the content of legal rules, thereby facilitating statistical analysis of legal systems and their social and economic impacts. In this paper we explain how leximetric methods were used to create the CBR Labour Index (CBR-LRI), an index and related dataset of labour laws from around the world spanning the period from 1970 to 2013. Datasets of this kind must, we suggest, observe certain conventions of transparency and validity if they are to be usable in statistical analysis. The theoretical framework informing the construction of the dataset and the types of questions which it is designed to answer should be made explicit. Then the choices involved in the selection of indicators, the definition of coding algorithms, and the aggregation and weighting of data to create composite measures, must be spelled out. In addition, primary legal sources should be referenced, and it should be clear how they were used to generate reported values. With these points in mind we provide an overview of the CBR-LRI dataset's main features and structure, discuss issues of weighting, and present some initial findings on what it reveals of global trends in labour regulation.

JEL Codes: C43, C82, J83, J88, K31

Keywords: empirical legal studies, quantitative labour law, leximetrics, index construction, construct validity

Acknowledgments

We thank the many labour law colleagues who have given us feedback on aspects of the CBR-LRI dataset, as it has evolved, and on the methodology underlying it, in particular Gordon Anderson, Graciela Bensusán, Hanneke Bennaars, Adolfo Ciudad Reynaud, Guy Davidov, Valerio De Stefano, Stefan van Eck, Colin Fenwick, Judy Fudge, Elena Gerasimova, John Hamilton, Alexis Hydén, Teun Jaspers, Tove Lundmark Söderberg, Priya Lele, Nikita Lyutov, Jonas Malmberg, Cristina Mangarelli, Sara Martinsson, Richard Mitchell, Viviana Mollica, Pedro Nicoli, Aleksandra Polanska and Mathias Siems. We alone are responsible for the values attributed in the dataset to the laws of individual countries, and for all errors and omissions. We are grateful for funding from the DFID-ESRC Joint Fund on Poverty Alleviation and the Cambridge University Humanities Research Grant Scheme. A version of this paper is forthcoming in the *International Journal of Comparative Labour Law and Industrial Relations*.

Further information about the Centre for Business Research can be found at: www.cbr.cam.ac.uk

1. Introduction

In this paper we present first results from analysis of a new dataset which codes for changes in labour laws around the world over the period 1970-2013. The dataset is based on the CBR Labour Regulation Index ('CBR-LRI') and is one of a number of datasets created through the collaborative efforts of an international team of researchers based principally at the Centre for Business Research in Cambridge. From 2005, the work began of building a statistical picture of changes occurring across a number of areas of labour law, namely the laws governing the definition of the employment relationship, working time, dismissal protection, employee representation, and industrial action. This project was initially a response to the quantification of labour law rules attempted by the research team responsible for the development of the legal origin hypothesis (Botero et al., 2004), and to the related emergence of the World Bank's *Doing Business* indicators (World Bank, 2004). As the project advanced it also addressed issues raised in similar initiatives undertaken by the OECD (Grubb and Wells, 1993; Venn, 2009; OECD, 2004, 2013) and the ILO (2015). In 2007 data covering five countries were published (Deakin, Lele and Siems, 2007). In 2016 a much expanded version of the dataset was completed, covering 117 countries which together represent over 95 per cent of world GDP (Adams, Bishop and Deakin, 2016).

As Saltelli and Funtowicz (2014) argue, stringent criteria of transparency should be adopted if composite indicators are to be used to generate research results and drive policy outcomes. With this point in mind, we have published the dataset in a form which records the score for each individual variable for every country on a year by year basis. This level of disaggregation enables other researchers to see exactly how composite scores are arrived at. We have also published an 800-page codebook which provides an explanation for every single piece of coding and which supplies a reference to the primary legal sources on which every value in the dataset is based (Adams, Bishop and Deakin, 2016). Failure to open up the black box of index construction has, in the past, led to scepticism towards the use of quantitative measures of legal-institutional phenomena (McCrudden, 2006). We aim to address these concerns by adopting an approach based on maximum transparency in the presentation of the coding process, at the same time as making our data fully available to other researchers and research users, on an open-access basis.

Our analysis proceeds in the following steps. In section 2 we briefly review arguments concerning the relevance and validity of quantitative empirical approaches to labour law research. In section 3 we identify some preliminary methodological issues concerning the criteria by which the construction of composite indices should be judged ('construct validity'). In section 4 we

explore the theoretical priors underlying the construction of the dataset, clarifying what exactly it is we are measuring. In section 5 we provide an account of the ‘leximetric’ coding method we adopted. We explain how we used a textual approach based on content analysis to code for differences in the strength of protective labour law regulation, distinguishing this from methods which rely on surveys and expert submissions to generate leximetric data. In this part we also explain in detail the choices underlying the indicators and coding algorithms used in the construction of the dataset. In section 6 we examine the statistical properties of the dataset, looking in particular at issues of weighting and aggregation, and we discuss ways in which it might be used in econometric analysis. Section 7 concludes.

2. Quantifying labour law: justifications and limits

2.1 Arguments for and against empirical legal research in general

Labour law is a field which has been shaped by interactions with the social sciences, and in particular empirical sociology, practically since its inception (Hepple, 1986). Labour lawyers can hardly be unaware of the importance of locating legal analysis in the wider context of developments in the labour market and in the system of industrial relations. This is hard-wired into our field because of the influence of the founders of the discipline. While Kahn-Freund urged labour lawyers to go ‘through’ not ‘round’ the law – in other words, not to neglect conceptual and doctrinal analysis of the kind which forms the core of interpretive legal method (Kahn-Freund, 1966) – his entire analysis of both British and comparative labour law was premised on the need to understand labour law as just one form of regulation among many (Kahn-Freund, 1954). This approach invites, or perhaps even requires, labour lawyers to become familiar with the operation not just of the system of collective bargaining, which was Kahn-Freund’s focus, but with a wider range of institutional forces shaping the way that labour is contracted on both the supply side (such as the household division of labour, the tax-benefit system, vocational education and training, and migration) and the demand side (such as macroeconomic policy, enterprise form, and industrial structure) (Arup et al., 2006).

While it is one thing for labour lawyers to familiarise themselves with what social scientists have to say about phenomena beyond the legal system, it could be argued that it is another thing altogether for them to engage directly in research which uses the models and empirical methods of the social sciences to examine those phenomena. In practice, however, this concern has not stopped labour lawyers engaging in theoretical or empirical social science research, any more than analogous concerns about disciplinary boundaries have prevented

economists and sociologists from analysing the operation and effects of labour laws (for an overview of both types of empirical legal research, see Deakin, 2010). Core techniques of empirical social science research, such as surveys, interview-based fieldwork, dataset construction and statistical analysis of data, are generic methods, not confined to a single discipline; they can be accessed by lawyers just as well as by economists and sociologists, and are increasingly taught to doctoral and other graduate-level students as part of integrated research methods courses.

There is, moreover, a case for saying that the critical questions facing labour law today, as a mode of regulation and as a disciplinary field, cannot be answered except through a combination of interpretive and empirical analyses. If we want to come to a fully-rounded view, for example, on whether the concept of the contract of employment is or is not well suited to a changing technological, organisational and labour market context (as addressed, for example, by Stone and Arthurs, 2013), we cannot confine ourselves to a doctrinal analysis of labour law rules: additional theories and techniques are needed in order for us firstly to conceptualise the nature of causal relations between different legal and socio-economic variables, and then to gain some understanding of how those linkages are playing out in practice.

Those trying to answer these questions increasingly do so through the construction of multi-member teams which cross disciplinary boundaries. Thus labour lawyers are finding themselves working directly with economists or sociologists on particular projects (for example, Gahan et al., 2014), or contributing to edited volumes with a common theme (for example, Stone and Arthurs, 2013). This is happening, we might reasonably infer, because lawyers have knowledge of institutional processes and skills of legal interpretation which usefully complement the data-gathering and statistical skills of social scientists.¹

2.2 Arguments for and against quantitative legal research

We do not think that these arguments in favour of the involvement of labour lawyers in empirical research projects lose any of their force merely because the research is quantitative as opposed to qualitative. Conducting interview-based fieldwork requires researchers to acquire skills which are just as specialised as those needed for quantitative work. (Poteete et al., 2010). Lawyers are trained in textual interpretation, not for research in the field. Thus there is no justification that we can see for labour lawyers prioritising qualitative over quantitative research simply on the basis of their own disciplinary training.

A somewhat different argument about the limits of quantitative approaches to legal research turns on the view that mathematical models assume a rigidly deterministic relationship between variables which is at odds with the open, dynamic nature of social systems in general (Lawson, 1997) and *a fortiori* with the open-textured and indeterminate nature of legal rules, not least those of labour law (Atleson, 1983). This is a more weighty objection to quantitative legal research, but we think that would be going much too far to take it to the point of rejecting all uses of statistical and mathematical approaches to empirical legal analysis. The issue, rather, is to understand what quantitative methods can achieve, and not to push them beyond their inherent limits.

Constructing a labour law index could have a number of purposes. It can, as we aim to show below, clarify trends and patterns which are otherwise obscured by the sheer volume and complexity of legal texts. It can thereby contribute not only to improved understandings of social phenomena, but to better policy.

Index construction is nothing new for social policy. At the end of the nineteenth century, social researchers such as Booth and Rowntree developed a concept of ‘poverty’ based on empirical studies of household expenditure which not only revealed for the first time the huge extent of destitution among the working poor of major British cities, but went on to shape the emerging institutions of the welfare state (Briggs, 1961). Their measures of poverty were ‘constructs’ which abstracted from a more messy and complex reality, but this is true of statistical categories in general. Without such categories, mapping phenomena such as household poverty would be impossible; and without this mapping, it would not have been possible to build the institutions of the welfare state as know them (Williams, 1981).

Extending the techniques of index construction to aspects of the legal system does not seem to us to be a step too far. There is no reason in principle for regarding legal rules and institutions as beyond the scope of statistical categorisation. The issue is whether the process of building an index can be

justified using generally accepted statistical conventions, including those relating to ‘construct validity’, which we explore in more detail below (see sections 3-5).

Index construction is also undertaken in order to facilitate econometric analysis, that is, statistical analysis which tests economic hypotheses and claims. We do not hold the view that econometrics has the answers to everything or that it is always the best method to use. It is one technique among many and care should be taken to use it appropriately.

Econometric analyses are often presented on the basis that they disclose general relationships between variables which interact in a predictable and determinate way. In practice, however, complex phenomena such as inequality and productivity are affected by multiple factors which interact in different ways according to the context being considered. There is a danger, then, in treating isolated statistical correlations as evidence of general trends (Poteete et al., 2010: 9-10).

This danger is heightened in the charged normative context of labour law rules, which, by their nature, involved the distribution of power and resources, and so are politically contested. It would be wrong to ignore the risk of econometric analysis being used to lend false scientific weight to what are, in essence, political claims about the desirability or otherwise of particular labour law rules.

Our view on this issue is that it is no more appropriate to reject quantitative analysis out of hand than it is any other empirical method. The issue, in each case, is whether the method used is appropriate for answering the question which is being addressed. Where data are scarce, hard to retrieve and non-comparable, quantitative analysis may be inappropriate. Even where good data are available, statistical associations may represent spurious correlations which are a feature of the way data are collected rather than any real underlying phenomenon. In complex multivariate environments, regression techniques may simply not be up to the task of identifying causal relationships (Poteete et al, 2010: 13-15).

However, other methods also have their limitations. Doctrinal analysis of legal rules, while it may elucidate the meaning of texts and provide some indication of how they will be applied by legal actors, tells us next to nothing of the way those rules shape social and economic relationships beyond the sphere of operation of the legal system. Qualitative interviewing, by its nature, is often non-replicable (interviews in a given time and place may be so context-specific that they cannot be repeated) and hence may lack the condition of external validity which should ideally characterise empirical research (Poteete et al.,

2010: 12). Multiple-methods approaches combining quantitative and qualitative techniques, and overcoming the limitations of each, may currently be the state of the art in the social sciences, but using more than one method is no guarantee of success in a project. Thus it seems to us likely that quantitative approaches, either singly or in combination with other methods according to the needs of particular projects, will help us obtain a better understanding of the social and economic operation of labour law systems.

A final objection to our project could be that it is pointless to seek replicate the attempts of previous researchers, including those working for international organisations, at index construction, given that several such indices already exist. As we have seen, these include the indices developed by the legal origin research team (Botero et al., 2004), the World Bank (World Bank, 2008) and the OECD (Grubb and Wells, 1993; OECD, various years; Venn, 2009).

One of the motivations of our project was not simply to replicate these earlier studies (although replication is an important and currently under-valued aspect of social science research) but to improve on them. These other indices are, in various ways, partial in their coverage of labour law systems, by reference to the scope of rules covered, and to the periods of time coded for. They also, as we shall see (section 4, below), have built into them theoretical assumptions which are open to question. We could have confined our analysis to a critique of these other indices, but this would not have offered a good defence to the argument that they are the best that there is. Thus in this paper we show how an index can be, and has been, constructed, in ways which seek to address the limitations of these previous attempts.

3. Leximetric data coding: elements of construct validity

We constructed the CBR-LRI dataset using a methodology which has come to be known as ‘leximetrics’. The term ‘leximetrics’ can be used to refer to any quantitative analysis of law (Cooter and Ginsberg, 2003) but it also refers more precisely to the process of translating legal materials, principally texts of statutes, decrees and judgments, into a form which can be used in statistical analysis (Lele and Siems, 2007; Adams and Deakin, 2015).

Leximetric data coding seeks to capture in quantitative form features of legal systems which are not naturally expressed in that form. In the social science literature on index construction, a quantitative measure of this kind is known as a ‘construct’ which aims to capture an underlying ‘concept’, and there is a growing literature discussing the conditions for ‘construct validity’, that is to say, the requirements which such indices must meet to be regarded as reliable measures (Strauss and Smith, 2009). The *Handbook on Constructing*

Composite Indicators prepared jointly by the OECD and the European Commission in 2008 suggests that developing an index is akin to building a computational model and as such ‘owes more to the craftsmanship of the modeller than to universally accepted scientific rules for encoding’ (OECD, 2008: 14). The *Handbook* makes a series of recommendations on index construction which emphasise the importance of arriving at an appropriate theoretical framework, which is used to inform the selection of indicators and the weighting scheme through which they are aggregated into a composite measure or series of measures. It also stresses the need for methodological issues ‘to be addressed transparently prior to the construction and use of composite indicators in order to avoid data manipulation and misrepresentation’ (OECD, 2008: 15).

In the approach we have followed, the process of dataset construction consists of the following steps:

- (i) identification of a *concept* which represents the underlying phenomenon of interest (here, ‘labour regulation’);
- (ii) development of a *construct* which provides a basis for measuring the concept;
- (iii) identification of *indicators* or *variables* which express aspects of the construct in numerical terms;
- (iv) development of a *coding algorithm* which sets out a series of steps to be taken in assigning numerical values to the primary source material;
- (v) identification of a *measurement scale* which is embedded in the algorithm;
- (vi) allocation of *weights*, where necessary or relevant, to the individual variables or indicators;
- (vii) aggregation of the individual indicators in an *index* which provides a composite measure of the underlying phenomenon of interest.

Items (i)-(ii) are dealt with in our discussion of the theoretical framework (section 4, below), (iii)-(v) in our account of the coding methodology (section 5, below), and issues (vi)-(vii) in our discussion of the statistical properties of the dataset and its uses in econometric analysis (section 6, below).

4. Theoretical framework

Datasets coding for legal or institutional data necessarily rest upon theoretical assumptions of some kind; the collation and organisation of data do not take place except in response to research questions which are generated by theories or models of the world, or by way of reference to pressing policy concerns. The CBR-LRI is a response to the questions raised in labour economics and the economics of law concerning the impact of laws and regulations on labour market outcomes and, more generally, on the economic performance of firms, sectors and nations.

Broadly speaking, the theoretical literature sees a number of possible outcomes of labour regulation on the economy, depending on the assumptions made in the underlying model. Neoclassical analyses see worker-protective labour laws as an interference with freedom of contract and, hence, as a distortion of market outcomes, which are likely to lead to involuntary unemployment and a number of other negative effects flowing from the misallocation of society's resources. By contrast, new-institutionalist approaches see legal regulation as countering market failures which are inherent in the operation of the labour market, including asymmetric information, and transaction costs arising from the open-ended and incomplete nature of the employment contract (Williamson, Wachter and Harris, 1975). Post-Keynesian approaches see labour regulation as having a number of potentially positive macroeconomic effects, including those relating to the management of aggregate demand (Dutt, 1984). The view that labour regulations may induce firms to manage labour more efficiently and, relatedly, to invest in productivity-enhancing improvements, is supported by both new-institutionalists (Bartling et al., 2013) and post-Keynesians (Stockhammer, 2015).

The CBR-LRI does not assume that any one of these theories is necessarily more correct, in the sense of being more likely to be vindicated by evidence, than any of the others. It treats the claims they make as hypotheses for empirical testing, any of which could turn out to be supported by statistical analyses of the one-way or possibly two-way relationships which exist between legal and economic variables.

This does not mean that the index is not informed by theory in any way. The core assumption implicit in its design is that regulations governing the employment relationship and collective labour relations are capable of affecting labour market outcomes in various ways. The law, together with related regulations with the capacity for normative effect such as collective agreements, is not just a filter or proxy for other social forces, but an independent causal agent which can change the behaviour of economic and social actors and alter

structural or aggregate outcomes in the economy. (for a defence of this position, see Deakin et al., 2016).

We should, however, make an important clarification at this point, which is that the CBR-LRI index does not measure ‘costs’, as Botero et al. (2004) aim to do, nor the ‘strictness’ of rules, as in the case of the OECD index (Venn, 2009), nor their ‘rigidity’, as in the case of the World Bank (various years). We are coding for the presence and qualities of publicly-enunciated legal and regulatory norms. We cannot assume, in the absence of other evidence, that these norms always impose costs on firms; they may do, or they may not. Nor can we assume that labour laws are effective, in practice, in protecting workers.

Since most labour law rules take effect as minimum standards, which may be improved on through contracting or collective bargaining, their direct influence on most firms may be limited; It may also be the case that legislative standards replicate norms which the parties would have contracted for anyway (Deakin and Sarkar, 2008). Legal rules may have a number of dynamic effects across the economy as a whole, but it is not clear a priori that these will always be harmful from the point of view of economic efficiency. Labour law rules may improve efficiency by enabling parties to employment contracts to overcome transaction costs and related barriers to exchange (Bartling et al., 2013). The use of the terms ‘strictness’ and ‘rigidity’ in the OECD and World Bank indices suggests that these measures have built into them an assumption that labour law rules necessarily have harmful effects. This makes them of questionable use for testing (as opposed to assuming) claims concerning the social and economic impacts of labour laws.

5. Coding methodology

5.1 Choice of indicators

Five areas of labour law are coded in the CBR-LRI, producing five sub-indices. These are:

(A) the law governing the definition of the employment relationship and different forms of employment, including the regulation of the parties’ choice of legal form, and the rules relating to part-time, fixed-term and temporary agency work;

(B) the law on working time;

(C) the law relating to dismissal;

(D) the law governing employee representation; and

(E) the law relating to industrial action.

These five sub-indices approximately correspond to the categories analysed by Botero et al. (2004) in their dataset, and were chosen, in part, to facilitate comparison between the two indices, although with the difference that the CBR index incorporates a time series, whereas Botero et al. (2004) has no longitudinal element. As we shall explain in more detail below (section 3.3), the individual coding algorithms in the CBR-LRI are not exactly the same as those used by Botero et al. (2004), but this does not, in itself, make a comparison inappropriate.

The five sub-indices, in addition to respecting the choice made by Botero et al. (2004) in their leading study, reflect the accepted subdivision of the area into individual and collective labour law (sub-indices A-C and D-E respectively), and, within those categories, broadly follow the ordering of materials presented in labour law textbooks (such as Deakin and Morris, 2012). This ordering has, we suggest, an inherent logic to it, reflecting the interdependence of the different subdivisions of worker-protective labour law rules, in the way they regulate the decision-making powers of employers.

Altogether we code for 40 individual indicators. We could have coded for more rules: for example, more detail on shift-work and nightwork in the working time sub-index, and more detail on the laws governing discriminatory treatment of trade union members in the employee representation sub-index.

Coding is a resource-intensive process and has to stop somewhere. The issue comes down to this: is the choice of indicators sufficiently plausible for us to believe that the index scores are a good proxy for the underlying variable of interest, the concept we are attempting to measure, labour ‘regulation’? The complexity of that concept, and the multi-faceted nature of the individual rules which feed into it, are such that any measure of the kind we have developed would be only an approximation of the underlying reality. But this is a feature of all social science datasets, whether they rely on content analysis, surveys, fieldwork, or other modes of data collection. The issue is not whether the index is a precise match for social reality, but whether it is a good enough approximation, given resource constraints.

5.2 Algorithms and scales

5.2.1 Algorithm design: binary versus graduated coding

Having defined the scope of the index through the identification of the individual indicators and their grouping into sub-indices, our next step was to draw up protocols or algorithms for each indicator, to ensure that the laws of individual countries were coded as far as possible in a consistent way. There are essentially three approaches which can be taken to this part of the data coding process. The first is to use dummy variables which indicate the presence or absence of a particular rule. Hence a value of 0 can be given if, for example, there is no right to reinstatement for unjust dismissal, and a value of 1 if the law in the country in question makes reinstatement the normal remedy for an unfair termination. This approach, which may be characterised as ‘binary coding’, is that taken by Botero et al. (2004) when defining most of the variables in their index. For example, their variable ‘alternative employment contracts’ is constructed out of four sub-variables, three of which are dummy variables in this sense:

- (1) a dummy variable equal to one if part-time workers enjoy the mandatory benefits of full-time workers, (2) a dummy variable equal to one if terminating part-time workers is at least as costly as terminating full-time workers, (3) a dummy variable equal to one if fixed-term contracts are allowed only for fixed-term tasks, and (4) the normalised maximum duration of fixed-term contracts. (Botero et al., 2004: 1348.)

The second approach is to find a law which can be expressed in numerical terms, such as the fourth sub-variable set out above which is described in years and months, and to normalise the score using a standard scale, such as 0 to 1. The third approach is to use graduated scores which express the strength with which the law of a given country regulates employer power or, conversely, protects employee interests.

The CBR index makes use of all three approaches but predominantly uses the third. Graduated coding allows for a more fine-grained analysis, which is capable of capturing a greater variety of cross-national differences than is the case with the use of dummy variables.

5.2.2. Retrieving primary data

The primary data on which the values in the CBR-LRI are based are the laws and regulations of various countries. To complete the coding, original materials were accessed at source, that is, in the form of statutes and other legislative materials (decrees, orders and ordinances), and, where relevant, case law. These primary legal texts were read, wherever possible, in their original language or in an official English translation. For most countries it was possible to retrieve texts from online sources, including the ILO's NATLEX database of labour legislation, and country-level legal databases. Although accessing repealed statutes and superseded versions of labour codes was not always straightforward, it was generally possible to find original texts or to reconstruct their contents by consulting textbooks and secondary sources. Textbooks were also consulted to clarify accepted interpretations of legal texts. In cases of uncertainty over the meaning of an original or translated text, or where no official translation was available, advice was sought from country-level experts.

It may be argued that to centralise the coding process in this way is likely to result in misunderstandings or misinterpretations of local laws, and that a better approach would be to work with a country expert in each case and to use their codings. This approach may work with a relatively small number of countries (see, for example, Lele and Siems, 2007; Siems, 2008), but it becomes problematic in resource terms when rolled out on the scale of a dataset coding for over 100 countries.

There is also the issue of introducing potential inconsistencies into the codings. As Holger Spamann explains, in the context of his re-coding of La Porta et al.'s index of shareholder protection laws, involving foreign lawyers in the coding

does not solve the problem of *classifying* the foreign law, i.e., of providing a workable variable definition and ensuring its consistent application. In fact, it may make these matters worse. First of all, the introduction of an intermediary between the foreign law and the quantitative researcher creates the novel problem that the intermediary may misunderstand the *question*. Second, even if the question is understood correctly, there is the risk that each foreign lawyer will fill in gaps in the definitions of terms used in a question according to her own priors... Without some feedback loop that ensures that the many gap-filling needs will be addressed centrally and hence uniformly for all countries, inconsistent coding is almost certain to occur. This requires that one person, or a group of closely communicating persons, centralizes the coding process. (Spamann, 2006: 18-19)

Thus where we consulted experts in the laws of particular countries we did so in order to clarify our understanding of the underlying law, not for the purpose of getting their assessment of the appropriate score for their country: in order to maintain consistent coding, all the scores in the index were arrived at by the central team.

5.2.3 Coding for different sources of labour law: court decisions, statutes, and collective agreements

Although in civil law countries court decisions are not regarded as a source of law, they were taken into account in the coding because they often, in practice, produce a rule or norm which is as important as a statutory provision. Statutory law was coded in the year in which it came into force and case law was coded in the year in which judgments were reported. Statutes passed but not yet in force or decisions which are unpublished or unavailable were not coded.

In line with the approach also adopted by Botero et al. (2004), sources such as administrative regulation and collective agreements were included where they could be regarded as ‘functional equivalents’ to statutes or court decisions. However, these sources are not coded unless they are designed to have some external binding force on the enterprise. Thus sector-level collective agreements are coded where the standards they contain have an *erga omnes* effect as a result, for example, of extension legislation.

5.2.4 Mandatory and default rules

The index takes into account both mandatory and default rules, but with a reduction in the score to indicate the non-binding nature of the latter. Default rules are rules that apply unless the parties to an agreement (collective or individual) contract out of them. Where the conditions for contracting out are strict, a score closer to 1 is given. Where opting out is straightforward, a score closer to 0.5 or below is given.

5.2.5 Federal systems

In some of the countries coded labour law does not operate in a uniform way, for instance if it is a federal state. In these cases, the law for the commercially or industrially dominant sub-unit of that state, (the unit where the most significant firms are based) is coded, or, failing that, the federal-level law, or in appropriate cases, a mixture of the two.

5.2.6 Focus on particular worker and enterprise types

The dataset in principle codes for the law as it applies to an indeterminate (or ‘permanent’) employment relationship, except in the case of the variables in the first sub-index (different forms of employment) where the focus is on the regulation of the parties’ choice of legal form (employment versus self-employment) and in the regulatory costs and benefits attached to employment relationships of different types (part-time, fixed-term and temporary agency work). Where laws differ in their effects according to the size and location of the enterprise, the coding is based on the rules which apply in the default or standard case, that is, enterprises or workplaces at or above generally-applicable size thresholds. Where different standards are set for different groups of workers, such as white-collar and blue-collar workers, the dataset codes for the minimal or less protective standards. We do not report different results according to firm size or worker category but the coding algorithms could be applied to generate separate sets of scores along these lines for particular countries, and this is a matter which we can in return in future.

5.2.7 Text-based versus opinion-based coding, and the gap between law ‘in the books’ and ‘law in action’

Could the coding of the CBR-LRI have been done more efficiently by avoiding a text-based analysis altogether, and relying on expert comments and opinion polling, as is the case with some other indices? The values contained in the World Bank’s Labour Market Regulation dataset are based largely on returns made by law firms and other business actors. The scores are a hybrid of these actors’ understandings of the law and their perceptions of how it operates in practice. The OECD’s Employment Protection Indicators appear to be based on a similar mix of opinion-based and text-based sources, with the latter apparently becoming more prominent as a source of codings over time. From 2013 the OECD has published online explanations for its scores based on legal texts (OECD, 2013). This documentation does not explain how earlier versions of the index were coded, so it is not possible to determine which factors influenced the scores provided for earlier periods.

One advantage of using opinion-based coding is that some account can be taken of the way the law works in practice. The CBR-LRI, as it depends entirely on a text-based analysis, can only capture the ‘law in the books’, not ‘law in action’. A practical reason for not using opinion-based coding in the case of the CBR-LRI, however, is that it would be hard to get meaningful estimates of the perceived strength of the law in practice for the historical periods, going back to the 1970s, which the dataset codes for.

There are other problems with opinion-based coding. The OECD relies on returns from governmental officials, and the World Bank on information provided by law firms and other commercial entities. This is, in each case, a narrow knowledge base to draw from. Widening the constituency of respondents to include other actors, such as NGOs and trade unions, or academic experts, could be done, but efforts would have to be made to ensure that the samples chosen were representative. Unless underlying polling data are made available, all such data suffer from a transparency problem and run the risk of subjectivity bias. Thus it is not at all clear that opinion-based coding produces results which are more accurate than those based on textual analysis.

The text-based approach adopted in the construction of the CBR-LRI means that its potential value as a measure of the wider social and economic significance of labour laws could be criticised as limited by comparison to surveys (where they are appropriately conducted) of the law in practice. More generally, labour lawyers and others may be sceptical of the relevance of the scores reported in the CBR-LRI, given the high likelihood that labour law rules are not translated fully into practice, in particular for smaller enterprises, and in the context of national systems with limited regulatory capacity.

We acknowledge this problem, but we think that the answer is not to mix up text-based coding with polling or survey evidence. When this happens, as appears to have been the case with the OECD and World Bank datasets, it becomes impossible to disentangle how far a given score is caused by the content of the rule and how far it is determined by weaknesses in the enforcement or observance of the law. It is essential in the construction of an index of this kind not to mix up law and practice (Spamann, 2006: 10).

Nor is it appropriate for the scores in the index to be adjusted downwards in an impressionistic way, based on an intuitive sense of the extent to which labour law rules operate well or badly in a given country setting. In the course of presenting the dataset at workshops and conferences we have sometimes been told that the score we have accorded to a given country's laws does not accord with the impression that a lawyer or social scientist from that country has of the strength of that law in practice. While a score should be changed if it turns out not to reflect an accurate understanding of the underlying law, or because the algorithm has been misapplied, we do not think that it is appropriate to alter values on the basis of subjective impressions, no matter how far they may reflect the experiences of those working in a particular system, since unless this type of adjustment is made on a systematic basis, it will simply import new biases and inconsistencies into the coding.

A better solution is to adjust the scores in the CBR index using other data sources, in particular those which provide information on the legitimacy of legal rules and on the effectiveness of legal institutions in practice, to get a fully-rounded picture of the functioning of labour laws in a given country. Thus the World Bank's Rule of Law index, part of its wider set of governance-related datasets (Kaufmann et al., 2008), can be used to adjust the scores in the CBR-LRI. This is available on a longitudinal basis, as are the indicators of legal and political rights developed by the Freedom House NGO (Freedom House, 2016). Separating legal and institutional data in this way makes it possible to disentangle the effects of labour law rules in and of themselves from the wider context of respect for the rule of law and observance of law in a particular country setting. This is not possible with the hybridised approach used by the World Bank and OECD.

6. Statistical properties of the dataset

6.1 Weighting and aggregation

The data in the CBR-LRI are presented in such a way that it is possible to identify the score for each individual indicator by year and country. Thus researchers can analyse the data at various levels of aggregation, and can choose to create composite indices from the individual indicators, as they wish. It is open to them to assign weights to individual indicators or categories of indicators and choose from various methods of aggregation to arrive at a composite measure. In this section we discuss some of the weighting options.

If the scores from individual indicators are aggregated or averaged without any weighting at all, the implicit assumption is that each indicator is of equal weight or 'importance' in arriving at the overall measure. This assumption can be analysed by reference to the theoretical priors and design principles underlying the construction of the coding algorithms.

From the point of view of these theoretical priors, our suggestion is that researchers should only depart from the assumption that all the indicators are of equal weight if there is a compelling reason to do so. This is because the indicators have been designed to capture *distinct* but *interlocking* aspects of the way labour law rules work in practice.

The principle that each indicator is distinct from all the others is another way of saying that they have been designed so as not to overlap, thereby avoiding double counting. Thus the rule governing the length of the qualifying period needed to obtain basic protection against unjust dismissal, and the rule setting out the normal remedy for unjust dismissal, contribute to worker protection in

two distinct ways. This justifies the design decision of allocating these two distinct rules to two separate indicators (CBR indicators 18 and 21 respectively).

Conversely, a single indicator may be used to capture two rules which are distinct in their form but largely equivalent in their effects. Thus CBR indicator 18 ‘measures the period of service required before a worker qualifies for general protection against unjust dismissal’. In some countries this rule is referred to as a ‘probation period’ and in others as a ‘qualifying period’. For our purposes, both, potentially, have the same effect, namely of delaying the point at which the individual worker is protected against unjust dismissal. We are concerned here with substance, not form. Thus these rules are coded by reference to a single variable. Coding them separately would give rise to the risk of double counting for the same rule.

The principle underlying the grouping of indicators into sub-indices is that the rules they code for interlock to achieve labour law protection. Thus the sub-index on dismissal law is the sum of individual rules which interlock to achieve worker protection in the area of termination of employment. The degree of protection a worker enjoys is the result of numerous inter-related rules: the length of any probation or qualifying period, the length of notice to be given prior to dismissal, the need for notification of dismissal, the rules governing procedural and substantive fairness, the rules governing redundancy selection, compensation and re-employment, and the remedies for dismissal. These rules do not operate in isolation but form linked parts of the systemic whole, ‘dismissal law’. Because they work together to produce an overall level of protection for workers, it is meaningful to group them together to form a single sub-index. Similarly, the individual sub-indices are distinct from each other but sum to form, we would suggest, a more or less coherent view of the overall regulatory impact of labour law as a whole in a given country.

Thus the possibility of aggregation is built into the design of the index. Combinations of indicators other than those suggested by the grouping of the indicators into the five sub-categories are entirely possible, and it is open to researchers to suggest others, but they would need to be justified by an underlying theory of how labour law rules interlock in practice.

Does the design of the index also justify equal weighting? If each indicator represents a rule which contributes to the systemic effect of the whole, and if we have avoided double counting, we should operate on a presumption of equal weighting, unless we have grounds for departing from that view.

The presumption of equal weighting may be departed from where certain variables of interest are more complex to express than others, and so are captured in a greater number of individual indicators. Thus across the index as a whole, there are slightly more indicators in some sub-indices than others. This problem can be remedied by presenting composite scores for the index as a whole in the form of an average of the five sub-indices, as opposed to the average of the 40 indicators. In practice, as the numbers of individual indicators do not vary greatly across the sub-indices, the different magnitudes are very small.

There may also be a case for adjusting the weights attached to certain individual indicators within sub-indices. For the most part, one indicator is sufficient to express the relevant underlying variable, but sometimes this is not the case. Thus in sub-index A (different forms of employment), only one indicator is used to express the variable relating to the choice between self-employment and subordinated labour, but two are used to express the rules relating to part-time work and agency work respectively, and three are used to express the rules relating to the use of fixed-term contracts. Again, this problem could be solved through simple re-weighting to achieve parity between the different categories (self-employment, part-time work, fixed-term employment and agency work). Quantitatively, however, the effects of this kind of reweighting are very small, so the problem is not a very significant one in practice.

A separate issue is whether the weights attached to individual indicators, or of particular sub-indices, should be altered, in order to reflect the significance which a given area of law might be assumed to have within a particular national legal system. Thus it could be the case, for example, that sub-index (D), employee representation, should be given proportionately more weight in the case of Germany than sub-index (E), strike law, given the pivotal importance attached to the institution of codetermination (as captured in sub-index (D)) in the German system. Similar conjectures could be suggested for every country. However, that is all they are: conjectures. Without a clear justification for introducing such weights, the likelihood is that applying this approach would introduce a high degree of subjectivity and arbitrariness into the dataset. This is not to say that some re-weighting could not be done on a cross-country or time-variant basis if relevant data to justify that weighting were available for a particular country or period. This is a decision to be made by each researcher and justified accordingly. As the dataset is reported without weights, it can be straightforwardly adapted or adjusted according to particular weighting schemes in future, where a case can be made for them.

6.2 Covariance and latent structure of data

A somewhat different approach to understanding the dataset is to examine it from the point of view of its internal statistical properties, that is, to see whether a statistical analysis can reveal patterns or structures which are not apparent from a consideration of the dataset's theoretical priors. A first step is to undertake a correlation analysis to determine the degree of covariance across particular indicators or groups of indicators. Table 1 reports correlations, for the dataset as a whole (all countries and all years), between the five sub-indices and two versions of the aggregate score, one based on an equal weighting of the 40 indicators, and which takes the average of the sub-indices (this addresses the point, discussed in subsection 6.1 above, concerning the unequal number of indicators in the different sub-indices). The Table shows that there is a positive correlation between each of sub-indices and the overall score, and that all these coefficients fall more or less in the range between 0.6 and 0.7.

Table 1: Labour law, all sub-indices, correlation matrix

	DFE	Working time	Dismissal	Employee representation	Industrial action	All	All reaveraged
DFE	1.0000						
Working time	0.3099	1.0000					
Dismissal	0.3426	0.3060	1.0000				
Ee rep	0.3544	0.2132	0.2874	1.0000			
Ind action	0.2318	0.1382	0.2294	0.3148	1.0000		
All	0.7009	0.5569	0.6800	0.6676	0.6232	1.0000	
Reaveraged	0.7052	0.5840	0.6569	0.6939	0.5941	0.9980	1.0000

Table 1 is reported here only as an exercise in illustration, as an 'overall' correlation between all countries and time periods does not, in itself, tell us very much. If, for example, labour law in one country in 1990 turned out to be correlated with that in another country in 2005, we would need to do further work to identify the possible channels at work, which could include trade, globalization, or harmonisation initiatives. This is a matter for future research.

A correlation analysis tells us whether two variables have a linear relationship with one another and whether they both lie on the same side of the mean (if the correlation is positive) or opposite sides (if it is negative). What would we be learning if we saw that there was a high and positive correlation between the scores for variables 18 (qualifying period) and 21 (reinstatement)? This would mean that across a group of countries or for an individual country as the case might be, a shorter qualifying period tended to go hand in hand with a more worker-protective remedy. We could then describe the two rules as mutually reinforcing modes of achieving worker protection. What would we conclude if,

on the other hand, we observed negative correlations between two variables? This would not be grounds for thinking that the coding algorithm had been badly designed. It would simply show that in a group of countries, or in a single country, a policy decision had been made to trade off a wide scope of protection (if the qualifying period was short) against a less worker-protective remedy (if compensation was favoured over reinstatement). Either outcome is possible, depending on the policy choices made in particular countries.

Nor does the presence of negative correlations between two variables mean that aggregating them is unsound. The two rules would, in principle, be offsetting each other. The aggregate score for a country where this is occurring should, if the coding is to be consistent, be lower than it would be in a country where both indicators were on the same side of their mean. For example, it is possible that policy makers in a given country might have decided for strong remedies for employees unjustly dismissed but to ration access to this protection by imposing a lengthy qualifying or probation period. In that case the overall score for dismissal protection would, correctly, be lower than it would be if a strong remedy were combined with a short qualifying or probation period.

Thus the results reported in Table 2, while of interest, are only the starting point for further analysis. The preponderance of positive correlations tells us that, across all countries and periods, labour law rules tend to reinforce each other, but it is entirely possible that when we drill down to the level of individual regions or countries, or to certain time periods, we would see different patterns, which could reveal trade-offs of various kinds. This is a matter for future work.

A further step would be to undertake a statistical analysis designed to reveal latent features of the dataset, such as factor analysis or principal component analysis. These techniques can be used to identify individual variables, or combinations of variables, which capture overall variances in the data for a particular sub-index or for the dataset as a whole. The purpose of doing this is to identify a core set of indicators that better captures the variance in the underlying phenomenon of interest, in the sense of doing so more parsimoniously, thereby making it possible to get a clearer result when regressing the scores in the index against potential outcomes variables such as indicators of employment or productivity (OECD, 2008).

A drawback with this approach is that components of variables identified in this way may make no theoretical sense. In other words, the components might not represent any tangible law or policy, but could be an artificial result of the way the dataset has been constructed. Such components would be drawing on laws whose association was a purely statistical artefact which bore no relationship to

the way rules interacted with each other in practice. Where that was the case, the component would have limited (if any) usability in formal modelling.

It may be that principal component analysis can be used in future to refine the study of the dataset. This issue is best addressed at the stage of econometric analysis of particular countries or groups of countries, rather than at the level of dataset design. We confine our discussion here to making the point that the dataset has been constructed in such a way as to capture a series of linkages across indicators, which are represented by their grouping into the five sub-indices. If associations of variables arrived at through factor analysis or principal component analysis are to be used in econometric studies, they should, we suggest, be justified on theoretical grounds. The presentation of components or clusters without reference to underlying theories runs the risk of giving undue attention to groupings of indicators which may be nothing more than statistical artefacts.

6.3 Trends in labour regulation revealed by the dataset

The longitudinal dimension of the CBR-LRI makes it possible to capture dynamic effects which are not observable in the cross-sectional dataset constructed by Botero et al. (2004). Figure 1 illustrates trends over time in the form of density estimations, for the index as a whole. It shows that for each of four years chosen at roughly equal intervals (1970, 1986, 2000 and 2013), there is a non-skewed distribution of scores, and that there has been an increase in protection as indicated by the rightward shift in the distribution over time.

When we look at areas of law over time we can observe, again, an increase in overall levels of protection, but variations over time in the relative degrees of protection provided by different regulatory instruments (Figure 2). Thus in 1970 the working time category of laws had a higher score than the other four, which had a more or less similar profile. By 2013, the industrial action category had become the area with the lowest score. This suggests that, over time, there has been something of a move away from collective labour law to individual labour law as the preferred mode of protection, but the difference is not huge and is relative, not absolute, as the level of protection in the industrial action category had been increasing over time, just not as quickly as in the case of the other groupings of variables.

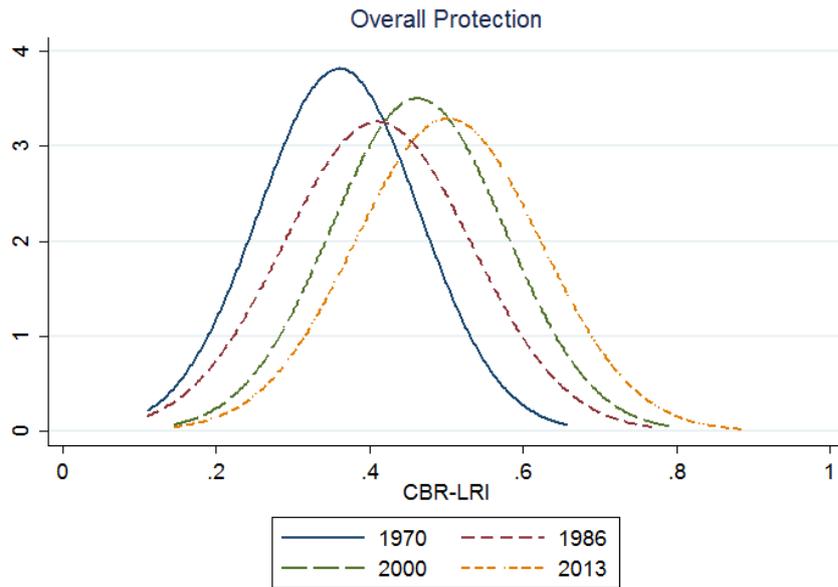


Figure 1. CBR-LRI, trends in overall labour protection, 1970-2013, kernel density estimations

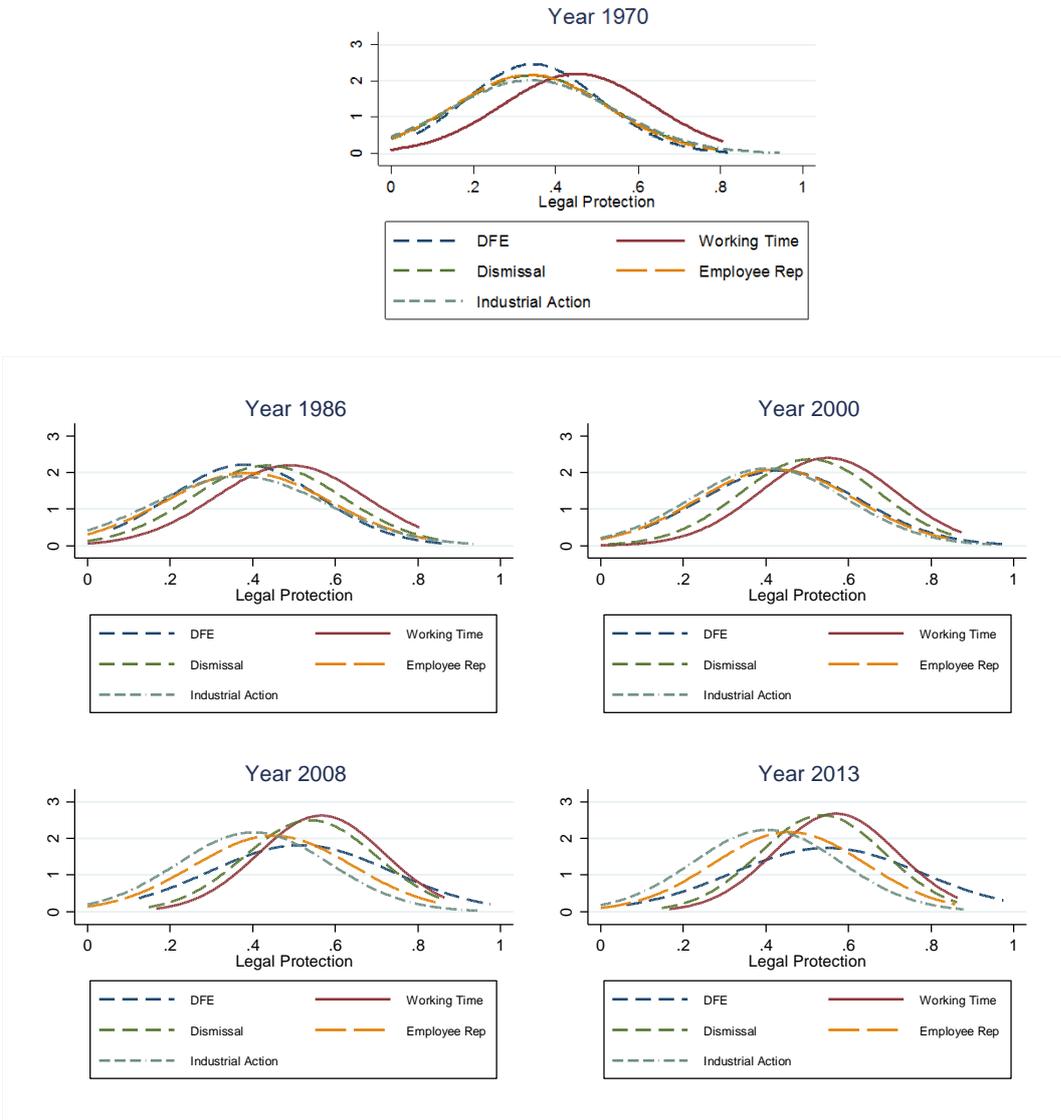


Figure 2. Trends in labour law by sub-categories of legal regulation, kernel density estimations, 1970-2013

The dataset can also be used to compare trends in regions across time. Figure 3 shows that in 1970 the most protective region was Latin America, but that by 2013 it had been displaced in this respect by the European Union. North America was by some distance the least protective region throughout the period, reflecting the very low scores attributed to the USA; Canada had somewhat higher, but still below average, scores. Asia was more protective than North America in each of the four years but below Africa. Each of the five regions experienced increasing protection over time.

Finally, we can compare the position of common law and civil law countries. Botero et al's cross-sectional data for the late 1990s found that 'legal origin matters for several areas of labor law', with civil law origin systems reporting higher scores than common law ones. From Figure 4 we can see that the CBR-LRI data report a similar gaps between the common law and civil law at both the beginning and of the period covered by the dataset, with the civil law remaining more protective, on average, than the common law, throughout the period covered by the dataset.

Reporting averages for countries by region, level of development and legal origin may give a misleading picture, in so far it might be read as implying a degree of homogeneity across these categories which is simply the result of averaging. When we look at the experience of individual countries, we see, firstly, enormous heterogeneity across systems, and also over time. Figure 5 shows the time trends for six developed countries. From this it can be seen that France, Sweden and Germany had much higher scores than the United States throughout the period between 1970 and 2013. This chart also shows that some countries experience considerable shifts in the level of worker protection over time; the UK, in particular, underwent a major transformation in its labour law regime, towards becoming less worker-protective, in the 1980s, and then experienced a more incremental revival of protection in the 1990s and 2000s. Figure 6 presents data for the five BRICS countries, and indicates, again, considerable heterogeneity of experience as well as shifts over time, with the increase in regulation in China after 2008 a marked trend.

Detailed analysis of the trends in the dataset is a matter for future research, but it is already clear that it may prompt some reconsideration of the way in which developments in labour law are described. In particular, we do not observe the clear declines in protection which might have been expected from accounts of the impact on labour law of trends which include the rise in new forms of employment. We do not see, for example, a reduction in the protection afforded to workers in different forms of employment over time. On the contrary, as the numbers employed in part-time, fixed-term and agency work have been rising, countries have responded by passing laws to protect workers in these forms of

work, by, for example, tightening the conditions under which they may be used, or requiring that they be treated in an equal or proportionate way with workers in so-called ‘standard’ (that is, full-time, indeterminate-duration) employment. This trend is particularly visible in the case of European Union countries, in part as a result of the adoption of the directives on part-time work and fixed-term employment in the 1990s and of the directive on temporary agency work in 2008. However, the trend is not confined to the global north, as China and South Africa are among middle-income countries to adopt laws more tightly regulating fixed-term and temporary agency work since the mid-2000s.

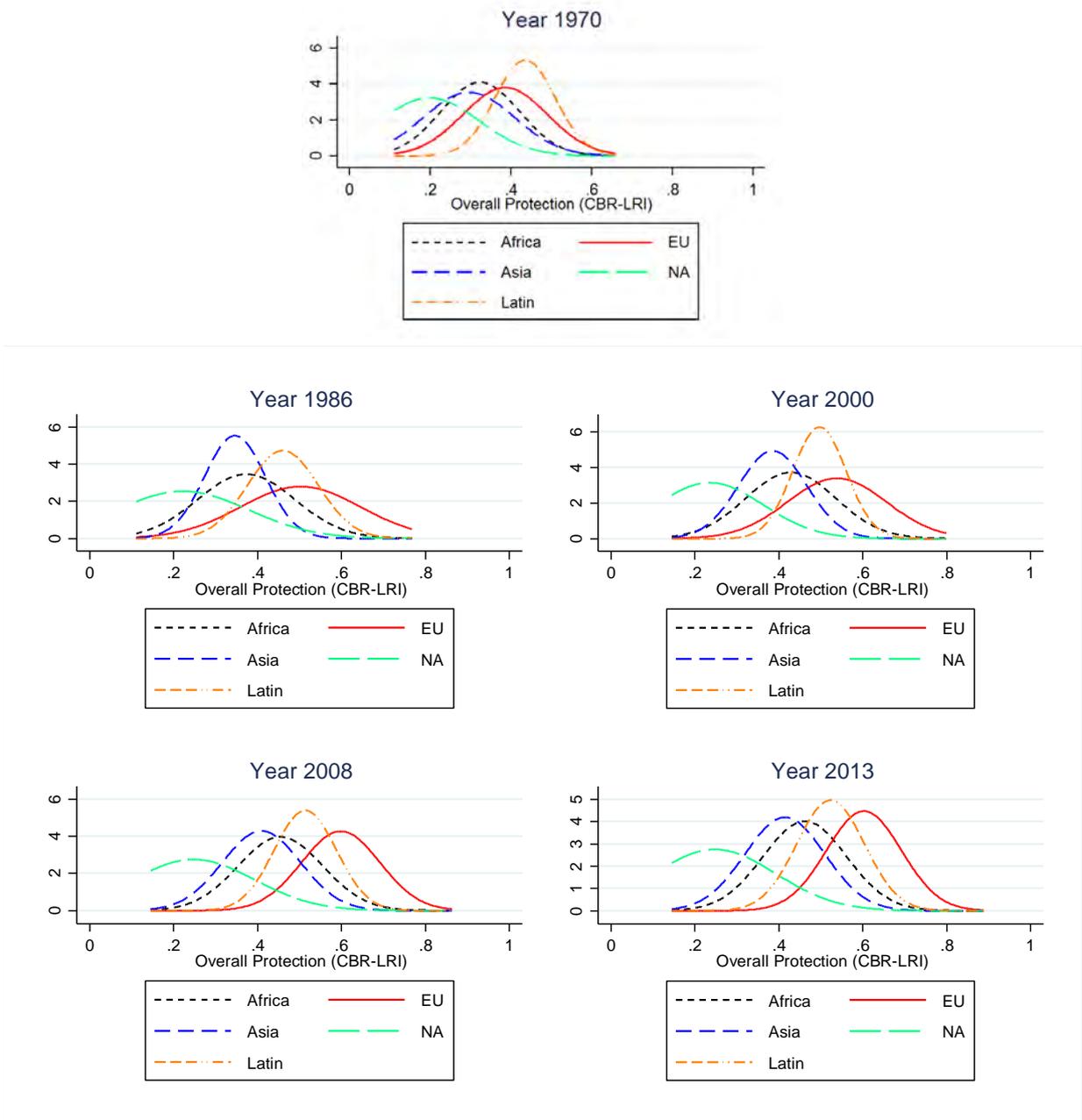


Figure 3. Trends in labour law in different regions, kernel density estimations, 1970-2013

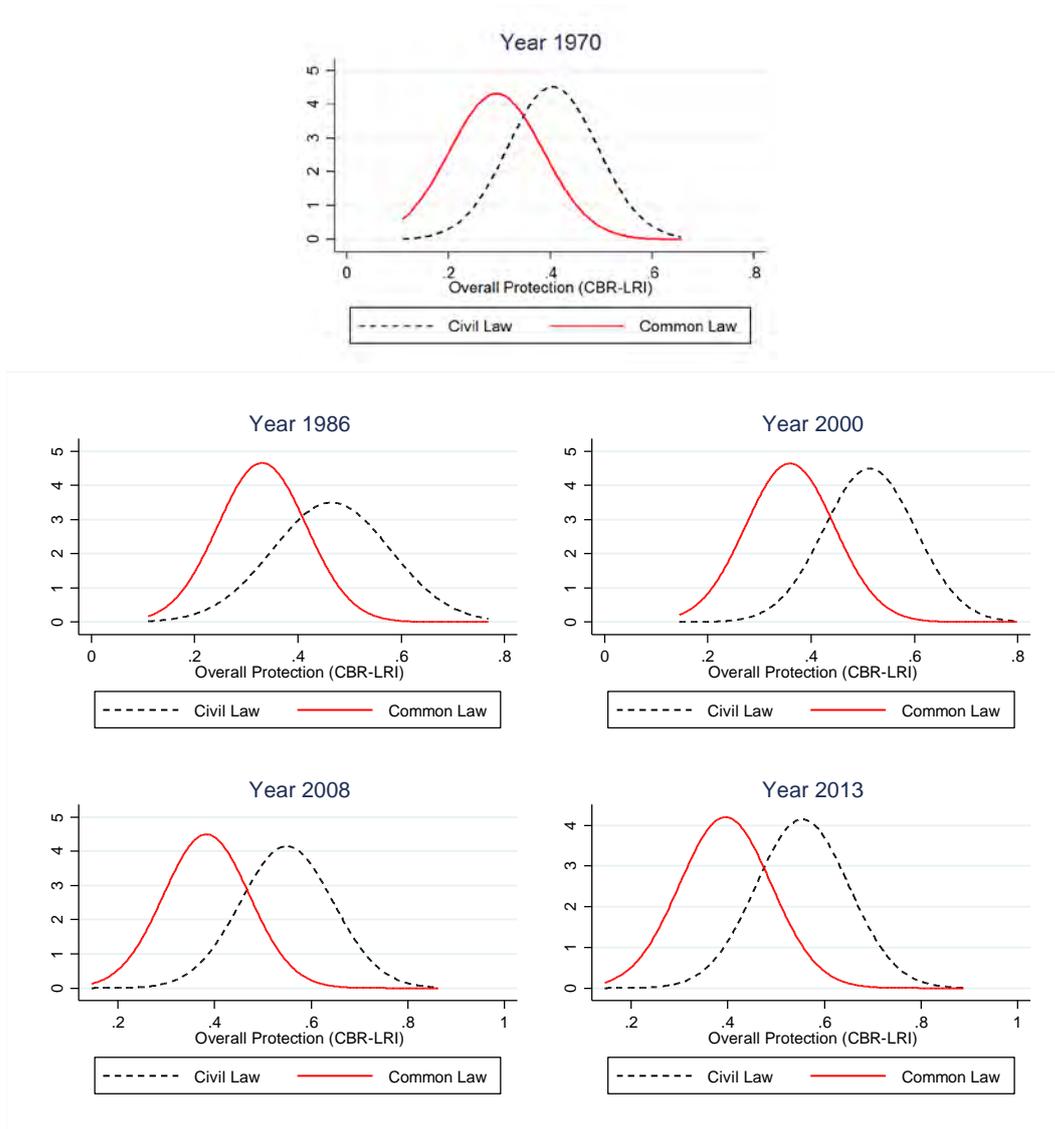


Figure 4. Trends in labour law by country grouping according to legal origin, kernel density estimations 1970- 2013

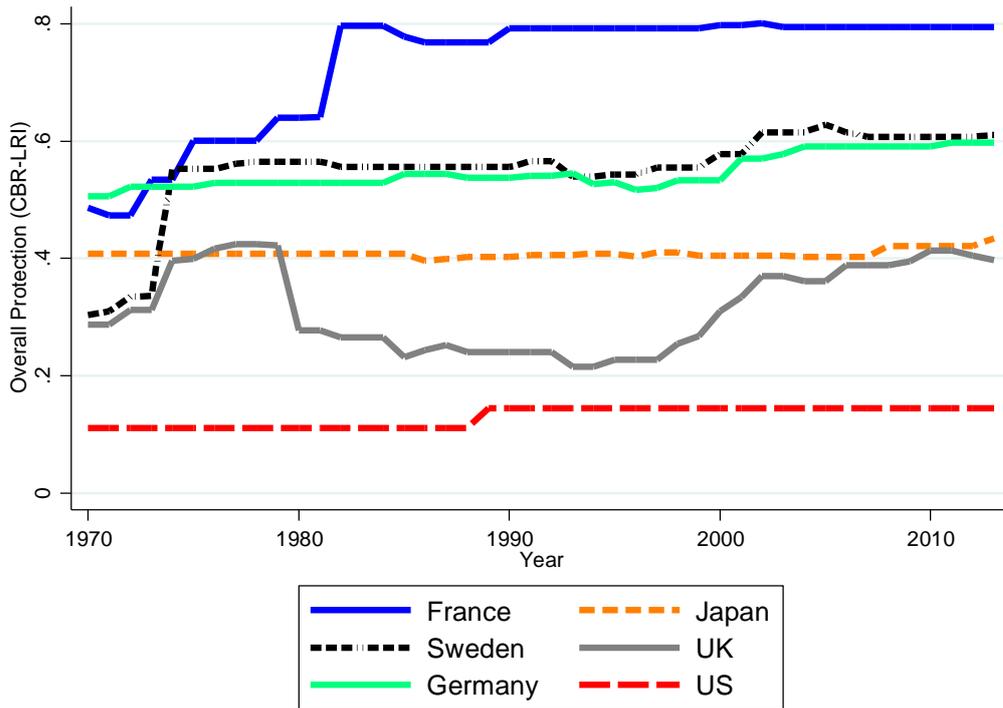


Figure 5a. Trends in labour regulation (overall protection) in six OECD countries, 1970-2013

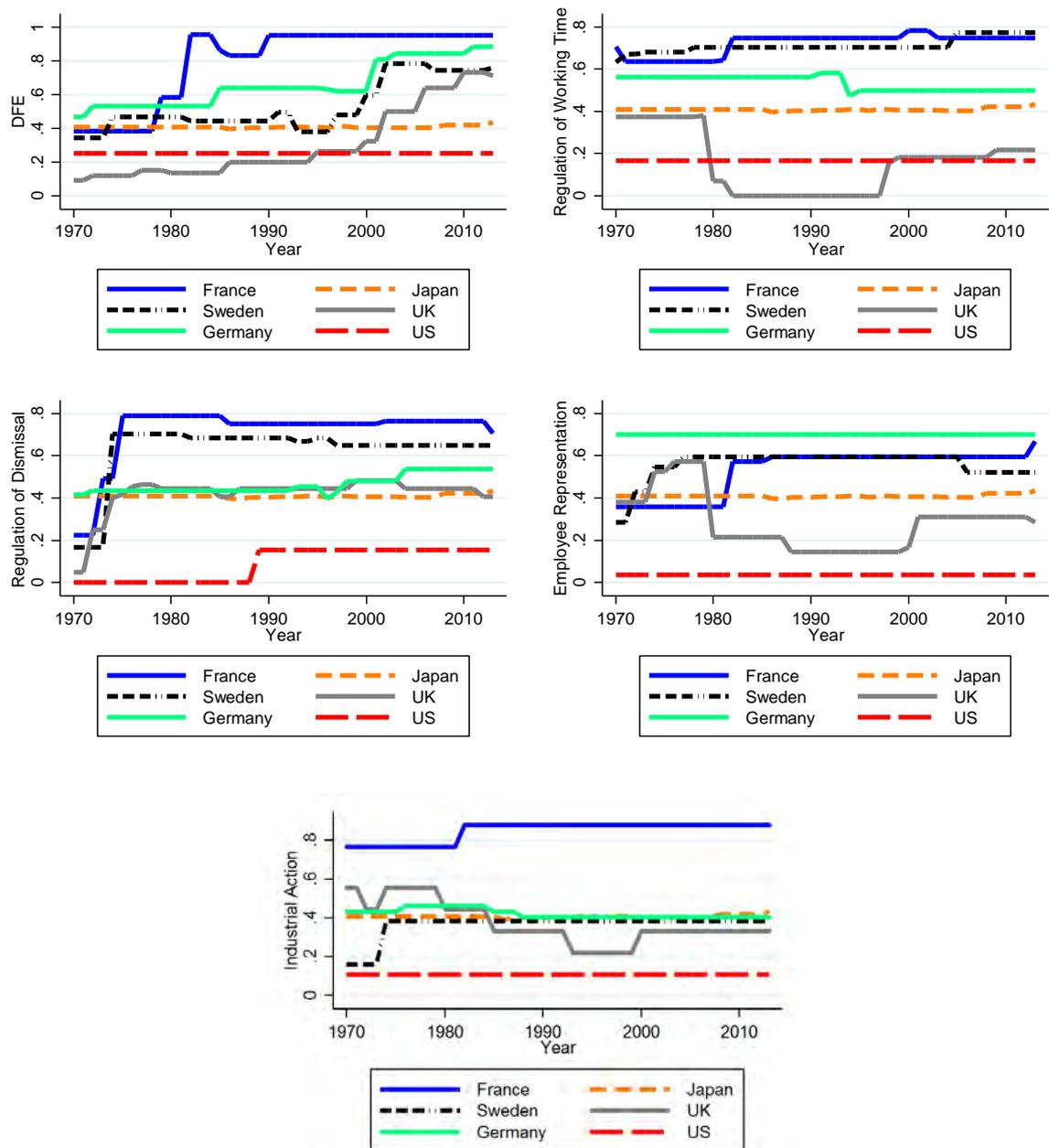


Figure 5b. Trends in labour regulation (DFE, Working time, Dismissal, Employee Rep, and Industrial Action) in six OECD countries, 1970-2013

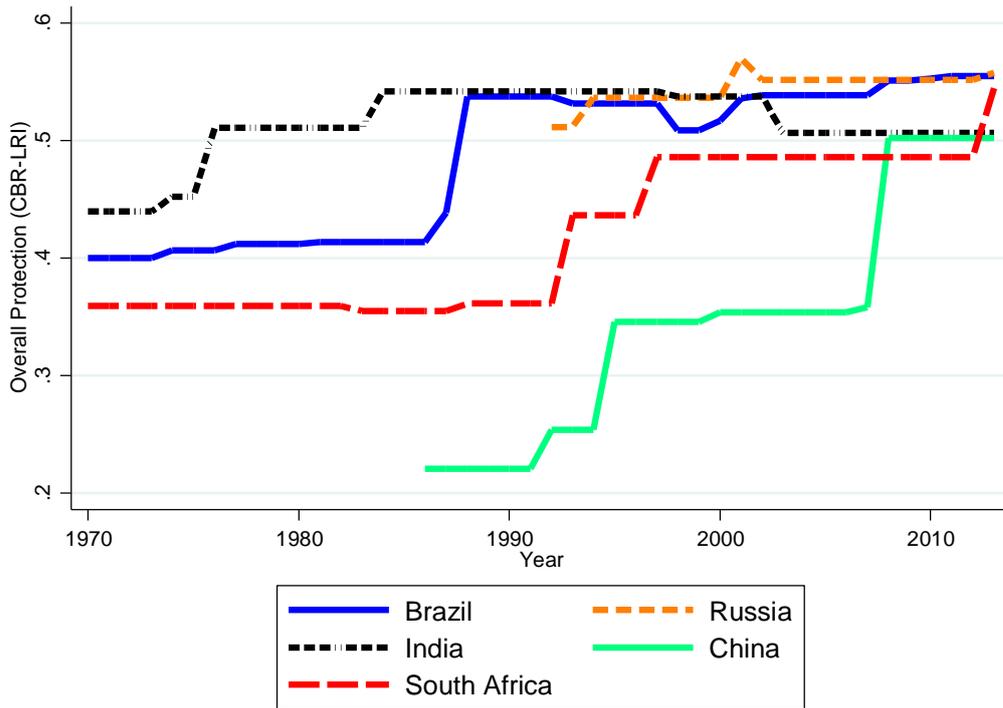


Figure 6a. Trends in labour regulation (overall protection) in five BRICS countries, 1970-2013

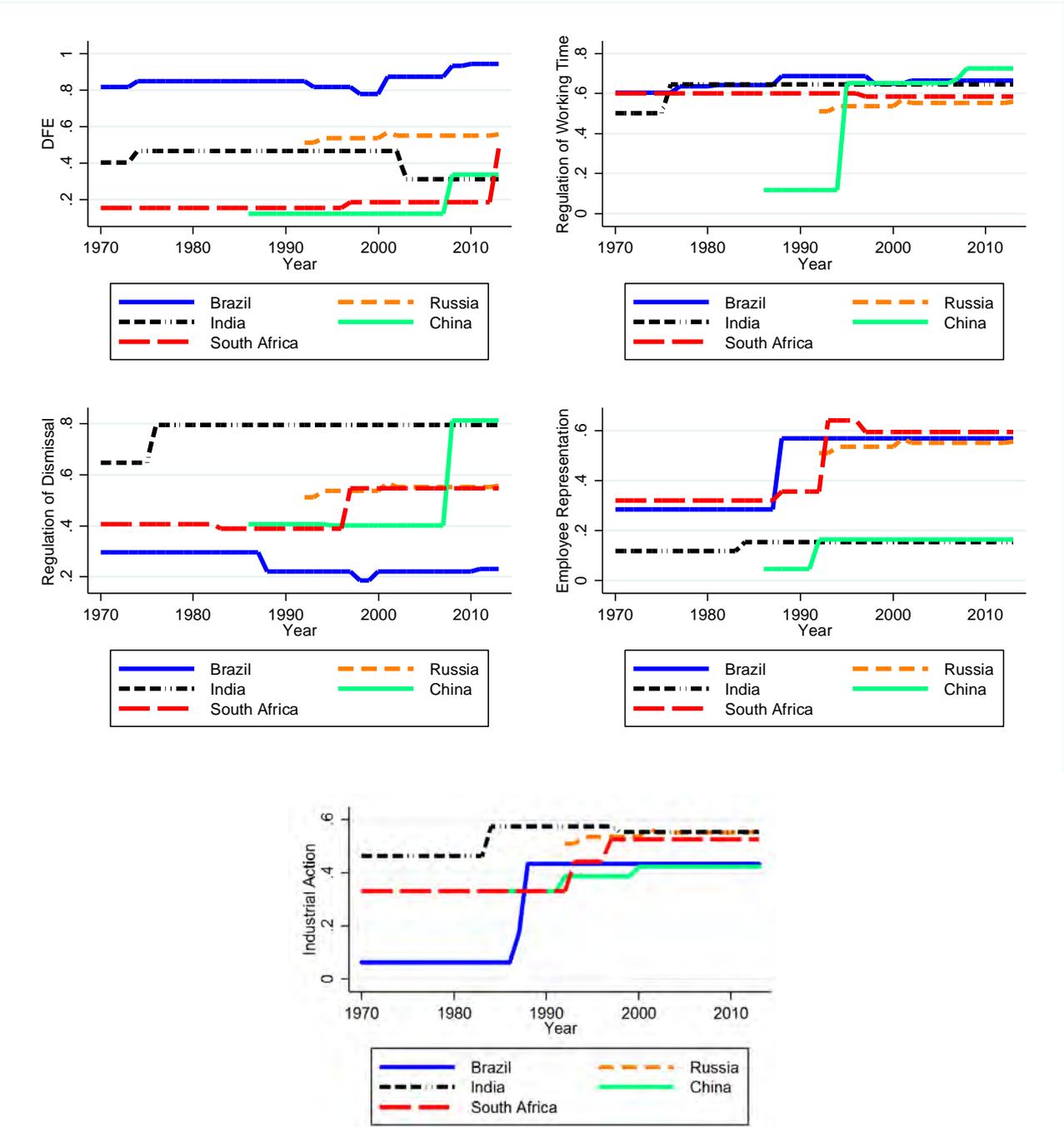


Figure 6b. Trends in labour regulation (DFE, Working time, Dismissal, Employee Rep, and Industrial Action) in five BRICS countries, 1970-2013

Nor do the data show that labour laws across the world have been in decline as a result of the adoption since the onset of the global financial crisis of policies of structural adjustment or fiscal consolidation. Even within the EU, where policies of fiscal consolidation have been pursued since the onset of the Eurozone crisis in 2008, the deregulatory impact of these policies on labour laws has been confined to the subset of countries which have received debt relief and have, as a consequence, been required to loosen controls over dismissals and over the use by employers of different forms of employment.

Some caveats are in order. First, to repeat a point already made, the dataset only reports changes in formal law. It does capture cross-country differences in the effectiveness of laws in practice; nor is it reporting trends in the wider economy, such as falling union density and collective bargaining coverage, which may have had a significant impact on the relative bargaining power of labour and capital. Nor does the CBR-LRI capture changes occurring in other areas of law, such as shareholder protection laws, which have empowered capital at the expense of labour. Studying potential interactions between labour regulation and these wider economic and legal developments is a matter for future research.

6.4 Potential uses of the CBR-LRI in econometric analysis

The most distinctive feature of the CBR-LRI, when compared to other indices, is the lengthy, continuous and internally consistent times series which it provides. This makes it possible to analyse relationships in a way which is not possible with static, cross-sectional analyses, and to distinguish between the short-run and long-run effects of changes in regulation.

Because, as we have seen, the CBR-LRI measures de jure regulation, its use in isolation from other data may give a misleading picture of the practical operation of labour laws. Thus it should be used in conjunction with data on the legitimacy or effectiveness of laws in practice, such as those derived from the World Bank's governance indices, the rule of law project of the World Justice Project, or rule of law data from think tanks such as Freedom House.

While new opportunities for econometric analysis arise, certain constraints also need to be borne in mind. Where data in long time series are non-stationary, co-integration techniques should generally be deployed to deal with the possibility of spurious regressions (for discussion, see Deakin and Sarkar, 2008). In addition, the results obtained from time-series analysis of the CBR-LRI are likely to be sensitive to the length of the lags chosen to capture the delayed effects of legal change. Thus the choice of lag may need careful justification.

The same point applies to the selection of controls and to use of any instrumental variables.

Notwithstanding these caveats, the CBR-LRI has already been used in a number of exploratory econometric studies, using relatively small samples compared to the current extent of the dataset, to explore the relationship between labour regulation, on the one hand, and relevant economic variables, including employment, productivity and inequality, on the other. The techniques so far used include difference-in-differences models (Acharya et al., 2014), autoregressive distributive lag approaches (Deakin and Sarkar, 2008), vector autoregression and vector error correction models (Deakin and Sarkar, 2011), fixed-effects and random-effects models (Deakin, Fenwick and Sarkar, 2014), and dynamic panel data analysis (Deakin, Malmberg and Sarkar, 2014). It is not our intention to review this emerging body of evidence here; it is too early to say what the current, expanded form of the dataset may be able to tell us. Our view is that there is nothing in the design of the dataset that makes it inherently more or less likely to favour a particular normative position on the desirability, from either an efficiency or a fairness perspective, of labour law protections.

7. Conclusion

Whether worker-protective labour laws cause unemployment, how far they improve employee well-being, and whether they contribute to rising or falling inequality, are pressing policy issues. Without good data on variations in the content of labour law rules across countries and over time it will be very difficult to get reliable answers to these questions. The task of collating data on legal systems in a form appropriate for statistical analysis is far from straightforward. In this paper we have sought to show that coding techniques can be used to construct legal datasets which are transparent and capable of being externally validated. The problems inherent in the construction of datasets based on composite indices can be, if not entirely overcome, then at least significantly minimised, if researchers take a number of steps: these include being clear about theoretical priors, making coding choices explicit, reporting primary sources in full, and setting out the bases for aggregation and weighting of data.

In this paper, we have described the process underlying the construction of a dataset which follows these precepts, the CBR-LRI. The dataset codes for the labour laws of 117 countries over the period 1970-2013. It provides a continuous time series of changes in labour law rules covering five areas: different forms of employment, working time, dismissal, employee representation, and industrial action. It can be used in time series analysis to

explore issues which until now have been largely inaccessible to empirical research.

Notes

1. Our own project is interdisciplinary in this sense: Adams, Bishop and Deakin are lawyers; Bastani is an economist. The dataset was constructed as part of a series of wider projects involving a number of lawyers, sociologists and economists. See <http://www.cbr.cam.ac.uk/datasets/>

References

- Acharya, V., Baghai-Wadji, R. and Subramanian, K. (2014) 'Labor laws and innovation' *Journal of Law and Economics*, 56: 997-1073.
- Adams, Z., and Deakin, S. (2015) 'Quantitative labour law', in A. Ludlow and A. Blackham (eds.) *New Frontiers in Empirical Labour Law Research* (Oxford: Hart).
- Adams, Z., Bishop, L. and Deakin, S. (2016) 'CBR Labour Regulation Index (Dataset of 117 Countries)', in J. Armour, S. Deakin and M. Siems (eds.) *CBR Leximetric Datasets* <http://dx.doi.org/10.17863/CAM.506> (University of Cambridge Data Repository), accessed 10 December 2016.
- Arup, C., Gahan, P., Howe, J., Johnstone, R, Mitchell R. and O'Donnell, A. (eds.) (2006) *Labour Law and Labour Market Regulation: Essays on the Construction, Constitution and Regulation of Labour Markets* (Sydney: Federation Press).
- Baker, D., Glyn, A., Howell, D. and Schmitt, J. (2005). 'Labour market institutions and unemployment: a critical assessment of cross-country evidence', in D. Howell (ed.) *Fighting Unemployment: The Limits of Free Market Orthodoxy* (Oxford: OUP).
- Bartling, B., Fehr, E. and Schmidt, K. (2014) 'Use and abuse of authority: a behavioral foundation of the employment relation' *Journal of the European Economic Association*, 11: 711-742.
- Botero J., Djankov S., La Porta R., Lopez-de-Silanes F. and Shleifer A. (2004) 'The regulation of labor' *Quarterly Journal of Economics*, 119: 1340-82.
- Briggs, A. (1961) *Social Thought and Social Action: A Study of the Work of Seeborn Rowntree* (London: Longmans).
- Cooter, R., and Ginsberg, T. (2003) 'Leximetrics: why the same laws are longer in some countries than others' working paper NO LE03-012 (2003, University of Illinois College of Law.

- Deakin, S. and Wilkinson, F. (1999) 'Labour law and economic theory: a reappraisal', in G. De Geest, J. Siegers and R. van den Bergh (eds.) *Law and Economics and the Labour Market* (Aldershot: Elgar).
- Deakin, S. and Wilkinson, F. (2005) *The Law of the Labour Market: Industrialization, Employment, and Legal Evolution* (Oxford: Oxford University Press).
- Deakin, S., Lele, P., Siems, M. (2007) 'The evolution of labour law: calibrating and comparing regulatory regimes', *International Labour Review*, 146: 133-162. Deakin,
- Deakin, S. and Sarkar, P. (2008) 'Assessing the long-run economic impact of labour law systems: a theoretical reappraisal and analysis of new time series data' *Industrial Relations Journal*, 39: 453-487.
- Deakin, S. (2010) 'Labor and employment laws', in P. Cane and H. Kritzer (eds.) *The Oxford Handbook of Empirical Legal Research* (Oxford: OUP).
- Deakin, S., and Morris, G. (2012) *Labour Law* 5th. ed. (Oxford: Hart).
- Deakin, S., and Sarkar, P. (2011) 'Indian labour law and its impact on unemployment, 1973-2006: a leximetric study' *Indian Journal of Labour Economics*, 53: 607.
- Deakin, S., Fenwick, C., and Sarkar, P. (2014) 'Labour law and inclusive development: the economic effects of industrial relations laws in middle-income countries', in M. Schmiegelow (ed.) *Institutional Competition between Common Law and Civil Law: Theory and Policy* (Frankfurt: Springer).
- Deakin, S., Malmberg, J. and Sarkar, P. (2014) 'How do labour laws affect unemployment and the labour share of national income? The experience of six OECD countries, 1970–2010' *International Labour Review*, 153:1.

Deakin, S., Gindis, D., Hodgson, G., Huang, K., and Pistor, K. (2016) 'Legal institutionalism: capitalism and the constitutive role of law' *Journal of Comparative Economics* online
<http://dx.doi.org/10.1016/j.jce.2016.04.005>.

Dutt, A. (1984) 'Stagnation, income distribution and "monopoly power"' *Cambridge Journal of Economics*, 8: 25.

Freedom House, *Freedom in the World 2016*, available at <https://freedomhouse.org/report-types/freedom-world>, accessed 11 December 2016.

Gahan, P., Mitchell, R., Cooney, S., Steward, A. and Cooper, B. (2012) 'Economic globalization and convergence in labor market regulation: an empirical assessment' *American Journal of Comparative Law*, 40: 703.

Grubb, D. and Wells, W. (1993) 'Employment regulation and patterns of work in EC countries' *OECD Economic Studies*, 21: 7-58.

Hepple, B. (1986) 'Introduction' in B. Hepple (ed.) *The Making of Labour Law in Europe* (London: Mansell).

Kahn-Freund, O. (1954) 'Legal framework', in A. Flanders and H. Clegg (eds.) *The System of Industrial Relations in Britain* (Oxford: Blackwell).

Kahn-Freund, O. (1966) 'Reflections on legal education' *Modern Law Review*, 29: 121.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2008) 'Governance Matters VII: aggregate and individual governance indicators for 1996-2007' World Bank Policy Research Working Paper No. 4654 (2008).

Lawson, T. (1997) *Economics and Reality* (London: Routledge).

- Lele, P., and Siems, M. (2007) 'Shareholder protection: a leximetric approach' *Journal of Corporate Law Studies*, 7: 17.
- McCrudden, C. (2006) 'Legal research and the social sciences' *Law Quarterly Review*, 122: 632.
- OECD (1994) *OECD Jobs Study, Evidence and Explanations, Part I: Labour Market Trends and Underlying Forces of Change* (Paris: Organization for Economic Cooperation and Development).
- OECD (2004) *Employment Outlook* (Paris: Organization for Economic Cooperation and Development).
- OECD (2008) *Handbook on Constructing Composite Indicators. Methodology and User Guide* (Paris: OECD).
- OECD (2013) *Employment Outlook* (Paris: Organization for Economic Cooperation and Development).
- OECD (2016) 'Detailed description of employment protection legislation, 2012-13, OECD countries', available at <http://www.oecd.org/els/emp/All.pdf>, accessed 11 December 2016.
- Pesaran, M. H., Shin, Y., and Smith, R. (1999) 'Pooled mean group estimation of dynamic heterogeneous panels' *Journal of the American Statistical Association*, 94: 621-634.
- Pesaran, M. H. and Smith, R. (1995) 'Estimating long-run relationships from dynamic heterogeneous panels' *Journal of Econometrics*, 68: 79-113.
- Poteete, A., Janssen, M., and Ostrom, E. (2010) *Working Together: Collective Action, The Commons, and Multiple Methods in Practice* (Princeton, NJ: Princeton University Press).

- Saltelli, A., and Funtowicz, S. (2014) 'When all models are wrong' *Issues in Science and Technology* 30 (2) (Winter), <http://issues.org/30-2/andrea/>, accessed 10 December 2016.
- Siems, M. (2008) 'Shareholder protection around the world (Leximetric II)' *Delaware Journal of Corporate Law*, 33: 111.
- Spamann, H. (2006) 'On the insignificance and/or endogeneity of La Porta et al.'s 'Anti-Director Rights Index' under consistent coding' John M. Ohlin Center for Law, Economics and Business, Harvard University, Discussion Paper No. 7/2006.
- Stone, K., and Arthurs, H. (eds.) (2013) *Rethinking Workplace Regulation: Beyond the Standard Contract of Employment* (New York: Russell Sage Foundation).
- Strauss, M., and Smith, G. (2009) 'Construct validity: advances in theory and methodology' *Annual Review of Clinical Psychology*, 5: 1-25.
- Stockhammer, E. (2015) 'Determinants of the wage share: a panel analysis of advanced and developing economies' *British Journal of British Industrial Relations* online, doi: 10.1111/bjir.12165.
- Supiot, A. (2010) *L'esprit de Philadelphie: la justice sociale face au marché total* (Paris: Seuil).
- Venn, D. (2009) 'Legislation, collective bargaining and enforcement: Updating the OECD employment protection indicators' OECD Working Paper, available at www.oecd.org/els/workingpapers/, accessed 10 December 2016.
- Williams, K. (1981) *From Pauperism to Poverty* (London: Routledge and Kegan Paul).

Williamson, O., Wachter, M., and Harris, J. (1975) 'Understanding the employment relation: understanding the economics of idiosyncratic exchange' *Bell Journal of Economics and Management Science*, 6: 250.

World Bank (2008) *Doing Business in 2008* (Washington, DC: International Bank for Reconstruction and Development).

World Bank (2016) *Doing Business 2017: Equal Opportunities for All* (Washington, DC: International Bank for Reconstruction and Development).