



► Skills and employment transitions in Brazil

Authors / Willian Adamczyk, Philipp Ehrl, Leonardo Monasterio





This is an open access work distributed under the Creative Commons Attribution 3.0 IGO License (<http://creativecommons.org/licenses/by/3.0/igo>). Users can reuse, share, adapt and build upon the original work, even for commercial purposes, as detailed in the License. The ILO must be clearly credited as the owner of the original work. The use of the emblem of the ILO is not permitted in connection with users' work.

Translations – In case of a translation of this work, the following disclaimer must be added along with the attribution: *This translation was not created by the International Labour Office (ILO) and should not be considered an official ILO translation. The ILO is not responsible for the content or accuracy of this translation.*

Adaptations – In case of an adaptation of this work, the following disclaimer must be added along with the attribution: *This is an adaptation of an original work by the International Labour Office (ILO). Responsibility for the views and opinions expressed in the adaptation rests solely with the author or authors of the adaptation and are not endorsed by the ILO.*

All queries on rights and licensing should be addressed to ILO Publications (Rights and Licensing), CH-1211 Geneva 22, Switzerland, or by email to rights@ilo.org.

ISBN: 9789220371640 (print)
ISBN: 9789220371657 (web-pdf)
ISBN: 9789220371664 (epub)
ISBN: 9789220371671 (mobi)
ISBN: 9789220371688 (html)
ISSN: 2708-3446

<https://doi.org/10.54394/ZWJU1062>

The designations employed in ILO publications, which are in conformity with United Nations practice, and the presentation of material therein do not imply the expression of any opinion whatsoever on the part of the International Labour Office concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its frontiers.

The responsibility for opinions expressed in signed articles, studies and other contributions rests solely with their authors, and publication does not constitute an endorsement by the International Labour Office of the opinions expressed in them.

Reference to names of firms and commercial products and processes does not imply their endorsement by the International Labour Office, and any failure to mention a particular firm, commercial product or process is not a sign of disapproval.

ILO Working Papers summarize the results of ILO research in progress, and seek to stimulate discussion of a range of issues related to the world of work. Comments on this ILO Working Paper are welcome and can be sent to RESEARCH@ilo.org, liepmann@ilo.org.

Authorization for publication: Richard Samans, Director RESEARCH

ILO Working Papers can be found at: www.ilo.org/global/publications/working-papers

Suggested citation:

Adamczyk, W., Ehrl, P., Monasterio, L. 2022. *Skills and employment transitions in Brazil*, ILO Working Paper 65 (Geneva, ILO).

Abstract

This paper analyses employment transitions and workers' skills in Brazil using a random sample from the universe of formal labour contracts covering the period from 2003 to 2018. We develop a novel procedure to derive a measure of occupational distance and internationally comparable skill measures from occupations' task descriptions in the country under analysis based on machine learning and natural language processing methods, but without usual ad hoc classifications. Our findings confirm that workers who use non-routine cognitive skills intensively experience the highest employment growth rates and wages. Their labour market exit risk is relatively low, occupational and sectoral changes are least common and, in the case of occupational switching, non-routine cognitive workers tend to find occupations that are higher-paid and closer in terms of their task content. Against the same characteristics, routine and non-routine manual workers are worse off in the labour market. Overall, there have been signs of routine-biased technological change and employment polarization since the 2014 Brazilian economic crisis.

About the authors

Willian Adamczyk is a researcher at the National School of Public Administration (ENAP-Brazil) and Pontifical Catholic University of Rio Grande do Sul (PUCRS), willianbadamczyk@gmail.com.

Philipp Ehrl is a professor on the graduate programme in economics at the Catholic University of Brasília (UCB) and a fellow researcher of the CNPq (National Council for Scientific and Technological Development), philipp.ehrl@gmail.com.

Leonardo Monasterio is a researcher at Ipea (Institute for Applied Economic Research- Brazil), professor at IDP-Brasília, and fellow researcher of the CNPq, leonardo.monasterio@gmail.com.

Table of contents

Abstract	01
About the authors	01
<hr/>	
▶ Introduction	06
<hr/>	
▶ 1 Literature Review	08
1.1 Skills and occupations	08
1.2 Career changes, business cycles and skills	09
<hr/>	
▶ 2 Data, variable definitions, and methodology	11
2.1 Data	11
2.2 Occupational distance measure	12
2.3 Skill measures	13
2.4 Model	15
2.4.1 Transition regressions and testable hypotheses	15
2.4.2 Duration analysis	16
<hr/>	
▶ 3 Results	17
3.1 Descriptive statistics	17
3.1.1 Skill scores and workers' characteristics	17
3.1.2 Employment changes and transitions	21
3.2 Propensity to change occupation and sector	25
3.3 Consequences of occupation changes	28
3.4 Duration results	30
3.5 Robustness	31
<hr/>	
▶ 4 Discussion	32
<hr/>	
▶ Conclusion	34
<hr/>	
Annex	35
A.1 Task measurement	35

A.2 Building a dictionary of verbs and skills	36
A.2.1 Introducing TF-RR	38
A.3 Definition of discrete skill groups	39

Additional figures	40
---------------------------	-----------

Additional tables	41
References	43
Acknowledgements	48

List of Figures

Figure 1. Skill scores over the wage distribution in 2018.	18
Figure 2. Employment changes by occupation group, 2003-2018.	22
Figure 3. Employment changes over the wage distribution 2003-2018	23
Figure B1. Skill scores over the wage distribution 2003–2018	40

List of Tables

Table 1. Skill scores by occupations.	17
Table 2. Descriptive statistics by skill groups: 2003–2018.	20
Table 3. Transition matrix between skill groups 2003-2018	24
Table 4. Occupation change, sector change and skill endowments	26
Table 5. Occupation change, sector change, skill endowments and moderators	27
Table 6. Distance after occupation changes	28
Table 7. Wages after occupation changes	29
Table 8. Risk of transition out of the labour market with moderator variables	30
Table A1. 10 most frequent verbs in each skill (in English)	36
Table A2. A random sample of frequency of stems by skill in top_skills_verbs	37
Table A3. Random sample of frequency of stems by CBO code in cbo_verbs_occupations	38
Table C1. Skill scores by skill categories	41
Table C2. Transitions Matrix between skill groups by subperiod 2003–2018	42

▶ Introduction

In the first decade of the 20th century, wage inequality fell in 15 out of 16 Latin American countries. Brazil was a case in point: the Gini index fell from 0.48 to 0.44 between 2002 and 2015 (Messina and Silva 2021; Almeida, Ehrl and Moreira 2021). The reduced urban, gender and racial wage gaps, however, were due primarily to one-off level effects that could hardly be repeated (Ferreira, Firpo and Messina 2017; Firpo and Portella 2019). Furthermore, income redistribution programmes that were simpler to implement and had broad societal support had already been established during a period of solid economic growth. For this reason, commentators such as Holland and Schneider (2017) have argued that perhaps the “easy” phase of income redistribution in Latin America is over and, if so, the political and labour market challenges to reduce inequality are even greater.

For Brazil, the current prospects look all but bright. Firstly, before being hit by the Covid-19 pandemic, the country suffered a historic economic recession and since then has not recovered its GDP level of 2013. Notwithstanding the uncertainty over when the Brazilian economy will get back on track, recent experience from the United States of America shows that a recovery may not necessarily re-establish lost jobs (Jaimovich and Siu 2020; Hershbein and Kahn 2018; Graetz and Michaels 2017). Consistent with the hypothesis that technological change is skill-biased, these papers suggest that firms restructure during a downturn, substituting routine occupations with machines or digital technologies.

This paper studies the employment and wage changes in the formal Brazilian labour market between 2003 and 2018. We are particularly interested in how workers’ skills are related to employment transitions. Our aim is to gain a better understanding of the labour market dynamics in a middle-income country during economic crises and secular (technology-related) changes. Given that these phenomena are mostly studied in developed countries,¹ the present paper will explore whether patterns of job transitions, skill-biased technological change and polarization observed in rich countries are also present in Brazil.

As a first step, we present aggregated statistics regarding employment changes and transitions. Subsequently, worker-level regressions show how skills are related to occupation changes, sector changes, occupational distance, wage changes and labour market exit risk. To tackle these questions, we rely on an administrative registry of all formal employment contracts between 2002 and 2019, the RAIS (Relação Anual de Informações Sociais). Taking a 10 per cent random sample from this database leaves us with over 50 million observations of 6.4 million men and women from all sectors.

We introduce two novel methods to measure skills and occupational distance based on Machine Learning and Natural Language Processing (NLP) techniques. The first technique generates skill intensity scores that: (a) are highly comparable across countries; (b) have modest data requirements; (c) do not rely on cross-country occupation matching and (d) replicate an international standard. We use the frequently used skill categories, non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM) and routine manual (RM), as introduced by Autor, Levy and Murnane (2003). In comparison, existing approaches can be classified into three groups: those derived from a combination of the tasks originating from the United States and an occupational crosswalk (Goos, Manning and Salomons 2014; Arellano-Bover 2020; Ehrl and Monasterio 2021); a discrete classification based on an ad hoc classification of a few occupations (Bachmann, Cim and Green, 2019); or a continuous measure derived from the ad hoc classification of tasks as shares in occupations in the spirit of Spitz-Oener (2006). All methods have their strengths and weaknesses, but the major problem for our study is that they are not adapted to low- and middle-income countries. Our method provides a continuous skill measure that is essentially based on the task descriptions of occupations in the country under consideration, in the present case the Brazilian Classification of Occupations (Classificação

¹ See, for example, Goos (2018), Autor, Levy and Murnane (2003), Acemoglu and Autor (2011), Bachmann, Cim and Green (2019), Carrillo-Tudela et al. (2016), Cortes et al. (2020), and Spitz-Oener (2006).

Brasileira de Ocupações, CBO). As a unique feature, our method replicates rather than relabels existing skill measure definitions. This simple procedure is thus readily applicable to any other country.

The second methodological contribution is a calculation of occupational distance, that is how different any pair of occupations is in terms of the content of their tasks. This variable is then applied to evaluate occupational changes in the context of human capital and productivity losses, as previously done by Gathmann and Schönberg (2010). Again, this technique renders country-specific measures and can be applied whenever a national occupational dictionary with task descriptions is available.

As expected, cognitive skill scores increase over the wage distribution and manual skill scores decrease. Additional metrics also suggest that our skill measures are in line with evidence from other countries. Overall, the results confirm that workers' skills are related to employment transitions. Using moderator variables in our regression reveals substantial heterogeneity across workers' skills with respect to workers' age, tenure and firm size. In particular, we observe that workers with intensive use of NRC skills experience relatively low probabilities of occupational and sectoral switching. Their labour market exit risk is relatively low and, in the case of an occupational change, NRC workers tend to find close, higher-paid occupations. In terms of these characteristics, RM and NRM workers are clearly worse off.

Overall, there have been signs of routine-biased technological change (RBTC) and employment polarization since the 2014 Brazilian economic crisis. Middle-skilled occupations in the manufacturing sector show the largest relative employment losses during the recent recession, from 2013 to 2018. Moreover, when occupations are ranked by their mean wage, we observe employment gains at the lower and upper end of the distribution, while occupations in the middle of the wage distribution suffered employment losses. There is also a clear tendency of middle-wage workers using more and more NRC skills. Further, in line with the RBTC and automation hypothesis, RM skilled workers seem to be more vulnerable than NRM workers. For example, RM workers tend to move to more distant occupations when workers change occupations or sectors. RM workers also show the lowest transition rates to occupations that rely on different, higher-paid skill types and they have the lowest direct job-to-retirement transition probability.

The rest of the paper is organized as follows: section 1 provides a review of the related literature. Section 2 presents the data and methodological approach. In particular, it explains in detail how skill measures and occupational distance are derived, proposes some hypotheses regarding skills and employment transitions and shows the econometric specifications used to test them. Section 3 contains descriptive statistics, regression results and robustness checks. The results are summarized and discussed in Section 4. Section 5 presents the conclusions. The appendix provides further details about the derivation of the skill measures and contains additional tables and figures.

▶ 1 Literature Review

1.1 Skills and occupations

The task-based approach has deepened our understanding of labour markets. The impacts of technological change, innovation, outsourcing and business cycles can be quite distinct according to workers' skills. The task content and skill requirements are also related to the disappearance and emergence of occupations, and employment growth. Due to the heterogeneity across skills, workers with the same formal education can have completely different trajectories during business cycles.

Autor, Levy and Murnane (2003) is the starting point of the contemporary analysis of occupations as a bundle of tasks. In their simple model, "computer capital" would substitute routine tasks (manual or cognitive) and be complementary to non-routine tasks. This hypothesis became known as RBTC. The testable predictions of the model are that an exogenous reduction in the price of computer capital would reduce the demand for labour in routine tasks and increase the demand for labour in non-routine tasks. The analysis of the changes in the tasks of each occupation showed that they did indeed shift in the directions predicted by the model. Autor, Levy and Murnane (2003) also introduce the classification of tasks in a two-by-two matrix: routine vs. non-routine and manual vs. cognitive. In this study, we will follow their typology. Spitz-Oener (2006) backs up the predictions by Autor, Levy and Murnane, (2003) and shows that analyses focused on occupations conceal changes in tasks within each occupation over time. Using keywords to classify tasks, she found similar results to the United States for Germany since 1979, namely there was a general increase in occupational complexity and the demand for analytical and interactive activities has gained importance in the mix of tasks performed by the workers.

Based on these insights and further evidence on the labour market dynamics in the United States, Acemoglu and Autor (2011) developed a model that became the workhorse of task-based empirical analysis by clarifying the concepts of tasks, skills and occupations.² In a nutshell, performed tasks generate products, and occupations are composed of tasks. Skills are the endowments workers require to perform tasks. Both the task composition of occupations and the skill endowments of workers are endogenous and can be changed over time. This framework allows the allocation of tasks by skill to be determined by technology and relative prices. The distribution of tasks across skills occurs in a Ricardian way. The allocation takes place according to the comparative advantages of the skills in performing the tasks. However, the cut-off points for assigning tasks to skills are determined endogenously in the model. Thus, the model's main contribution is to facilitate studies of the less obvious consequences of price changes, endowments and production technologies. Acemoglu and Autor (2011) also update Autor, Levy and Murnane's (2003) skill categories to the more recent O*NET database.

The task-based approach has proven valuable for understanding the current phenomenon of disappearing middle-skill occupations in developed countries. Recent technologies, such as computers and robots, have thus far substituted routine tasks, as the RBTC predicts. In the longer term, machines will not be limited to simple tasks, advancing over more complex, analytical and creative ones.³ As a result, Acemoglu and Restrepo (2017) point out that advances in artificial intelligence and robotization may further increase inequality and polarization.

² The papers by Autor, Katz and Kearney (2008), Goos (2018), Acemoglu and Restrepo (2017) are seminal contributions to this research area.

³ See the discussion around automation predictions in Frey and Osborne (2017), and its overstatements in Arntz, Gregory and Zierahn (2017) and Georgieff and Milanez (2021).

As long as routine tasks performed by high-skilled workers are substituted, their productivity increases. As a consequence, the remaining workers will be more valuable to firms and should observe growing returns to non-routine skills augmented by capital. At the same time, the supply of dislocated medium-skilled workers rises, lowering the relative wages of routine task-intensive occupations. Finally, low-skilled workers perform NRM labour market polarization: wages and employment increase at two extremes of the skill distribution.⁴

To date, the evidence on job (and wages) polarization in Brazil is mixed. Maloney and Molina (2019) study the impact of technological change in several middle-income countries, including Brazil, and observed some signs of polarization. Yet Ariza and Bara (2020) report that the employment change curve over the wage distribution is flat. According to Firpo and Portella (2019), other factors must have contributed to offset polarization trends. Ferreira, Leite and Litchfield (2008) estimate that the main factor responsible for the fall in wage inequality was the decline in return for experience. Firpo and Portella (2019) suggest that age-biased technical change or skill obsolescence explain this fact. This means that the skills of older people cannot keep up with the new tasks that arise. Almeida, Corseuil and Poole (2017) analyse the impact of computers and the internet on the Brazilian labour market and find evidence in favour of polarization. Herdeiro, Menezes-Filho and Komatsu (2019) identify a more complex temporal dynamic of polarization, as technological advances favoured middle-skilled workers between 1981 and the late 1990s, However, between 2004 and 2015, this trend was reversed and the relative demand for middle-skilled workers began to decline.

In Brazil, some task-based studies exist, however with different focus and skill measurement methods. For example, Ehrl and Monasterio (2019) and (2021) study the (long-term) effects of skill concentration in local labour markets. Neves Jr., Azzoni and Chagas (2017) estimate the skill premium by city size. Albuquerque et al. (2019) and Maciente, Rauen and Kubota (2019) predict the impact of automation on the Brazilian labour market, and Adamczyk, Monasterio, and Fochezatto (2021) on the public sector. Ehrl (2018) analyses the heterogeneous effects of intermediate goods importing on the skill structure within firms. Arellano-Bover (2020) explores how on-the-job learning depends on workers' skills. These papers either apply Maciente's (2013) mapping between O*NET and Brazilian Occupational Classification (Classificação Brasileira de Ocupações - CBO) or define discrete skill measures using occupation groups.

1.2 Career changes, business cycles and skills

Automation, RBTC and other forces cause changes in the task set of occupations prompting some individuals to switch occupations. As is well known, these changes do not always occur without frictions. Recent literature uses individual data and the task-based approach to understand transitions between jobs and their impact on wages and the length of unemployment, among others.

Gathmann and Schönberg (2010) build a model of employment switches that takes into account task-specific human capital, also using Autor, Levy and Murnane's (2003) classification. They tested it on German administrative data at four points in time (1979, 1985, 1991/92, and 1998/99) and found that 40 per cent of wage growth stems from the accumulation of human capital within the same set of tasks. Moreover, they identify that workers have larger wage reductions when they move to occupations with tasks that are more distinct from the ones they previously held. Bachmann, Cim and Green (2019) also examine long-term trends in the German labour market between 1974 and 2018, and identify an accelerating polarization tendency that is more relevant than short-term effects stemming from economic cycles. Again, workers in routine occupations are most likely to lose their jobs and struggle to get out of unemployment. Worse still, these problems of routine workers have become more severe over the decades.

⁴ The polarization of the labour market has been blamed not only for direct economic problems but even for the resurgence of populism (Guriev 2018).

Regarding the United States labour market, Cortes (2016) observes similar results to the papers with German data. The author highlights that workers who migrated from RC to NRC occupations reap gains, while those who moved from RM to NRM occupations are worse off than those who remained in the same occupation class. Likewise, Ross (2017) identifies that a fall in routine tasks in the same occupation leads to falling wages, while an increase in abstract tasks affects wages upwards. Furthermore, the returns from routine tasks fell while abstract tasks rose over time, lending support to the RBTC hypothesis.

Arellano-Bover (2020) show that young OECD workers who faced recessions, despite an education upgrading effect from postponing labour market entry, still had shortfalls in their skills two decades later. And, among these, the ones who suffered the most were those with parents with little schooling. The enduring persistence of these adverse effects lends even more relevance to the issues examined in this study. Additionally, Lazear and Spletzer (2012) and Carrillo-Tudela et al. (2016) show that shifts between occupations are pro-cyclical. When the economy is growing, turnover increases because workers migrate to higher-wage occupations. During a crisis, the gains from occupational changes are lower, and employees tend to avoid change and wait for the economy to recover.

The task-based approach is also helpful to understand the phenomenon of jobless recoveries. As early as 2003, Bernanke (2003) argued that during economic recoveries, the mismatch between workers' skills and available jobs would increase. The current consensus is that those engaged in routine tasks once again tend to lag, accelerating the job polarization process in developed countries (Jaimovich and Siu 2020; Hershbein and Kahn 2018; Graetz and Michaels 2017). However, more research is needed as Graetz and Michaels (2017) find that outside the United States, modern technology is not responsible for jobless recoveries.

Wiczer (2015) presents a search and matching model that takes into account job-specific skills. In his model, wage loss for workers who change occupation is proportional to the skill-distance between occupations. This causes workers to seek employment in occupations similar to those they held and employers to seek workers in the same occupation. This mechanism creates a long tail of unemployment duration that increases the long-term unemployment rate. Wiczer (2015) then estimates the skill-distance between occupations in the United States. His calibrated model explains more than two thirds of long-term unemployment between 1995 and 2013.

As far as we know, in-depth studies on occupational transitions in middle-income countries are rare. Parra and Christian (2021) analyse workers' jobs and occupational switches in Vietnam. Their study is one of the few papers that took transitions between formal and informal labour markets into account.

▶ 2 Data, variable definitions, and methodology

2.1 Data

The main data source for this research is the Annual Social Information Report (Relação Anual de Informações Sociais - RAIS) produced by the Brazilian Ministry of Economics. RAIS is an administrative register of employer-employee data covering the population of formally employed workers in the public and private sectors. It is widely recognized as the most reliable source of information about the formal labour market in Brazil (Dix-Carneiro 2014). We are able to draw a 10 per cent random sample from the population of workers based on the last digit of their PIS (Programa de Integração Social) identification number. This unique number makes it possible to follow workers over time. Our analysis focuses on male and female employees between 20 and 60 years of age with complete information in the selected variables.⁵ Our sample and the present estimates are therefore representative of the entire formal Brazilian labour market.

The informal sector represents almost half of the labour force in Brazil. Of course, it would be ideal if our data also included informal workers, but studies have shown that the informal sector in Brazil acts as a buffer for the unemployed in face of adverse labour demand shocks (Dix-Carneiro and Kovak 2019). Like unemployment rates, informality is counter-cyclical and the transition rates from unemployment to informal jobs are twice as high as those from unemployment to formal jobs (Ulyssea 2020; Dix-Carneiro et al. 2021). In other words, the boundaries between unemployment and the informal sector are quite fluid and more comparable to each other than to formal jobs in the highly regulated formal labour market. Thus, this study deals with the internal dynamics of workers in the formal labour market, and its inflows and outflows, irrespective of whether they are unemployed or working in informal activities.

The data are organized in the form of employment spells. That is, the number of observations per worker-year corresponds to the number of non-overlapping formal employment relations. Labour market entry and exit are permitted at any time during the longest possible observation period from 2003 to 2019. The present analysis takes into account workers' wages, the number of hours worked per week, their occupation, tenure on the job, education level and their employers' size, sector, legal nature and geographical location. Since we are mainly interested in how workers' characteristics relate to employment transitions, the last year is only utilized to calculate change variables (such as occupation change, wage growth and employment distance). The final full sample from 2003 to 2018 contains 50,082,074 observations from 6,358,348 individuals. As usual, the administrative data do not provide information about why an individual leaves the formal labour market. There is, however, a variable that indicates why the labour contract ended, which allows us to calculate the share of retired or deceased workers. Besides, left- and right-censoring of employment trajectories require careful treatment. On the one hand, duration analysis (for the probability of leaving the formal labour market) is adequate for this situation. On the other, to analyse occupation changes, we follow the convention in the literature and restrict the sample to workers with at least two consecutive employment spells. See, for example, Bachmann, Cim and Green (2019) or Cortes et al. (2020). In this consecutive employment sample, we end up with 46,923,670 observations from 5,433,062 individuals.

A key feature of this paper is the derivation of skill measures that are unique and specific to Brazil. To this end, we combine information from the well-known United States O*NET based skill measures with the detailed task descriptions in the Brazilian Occupational Classification (CBO) from the year 2002. The CBO describes the occupations and organizes them in a hierarchy, making it possible to systematize information

⁵ Because our empirical strategy requires that workers' skills can be related to their employment trajectories, only workers with no more than one job at the same time are considered. Due to the particularities and low transition propensities, members of the armed forces are excluded as well.

regarding the workforce according to the occupational characteristics and the nature and content of the work involved. It describes the functions, duties and tasks that make up each occupation, and the content of the work in terms of the set of knowledge, skills and training required for the performance of responsibilities for each occupation (CONCLA, 2019). The details of our skill derivation procedure are explained in Section 2.3.

2.2 Occupational distance measure

In order to find a continuous metric for the distance between a pair of occupations, we proceed in two steps. Firstly, we apply Natural Language Processing (NLP) algorithms to find weights that express the relevance of each word in the Activities Matrix⁶ which describes the tasks inherent to each occupation. Secondly, we chose a distance measure and use the weights from the previous step to derive the occupational distance matrix. Previous studies, such as Gathmann and Schönberg (2010), use a distance metric based on the absolute difference between analytical tasks scores. Instead, we proceed by comparing directly the weights of words that describe each occupation's tasks.

The Term Frequency–Inverse Document Frequency (TF-IDF) method calculates the relative weights of words describing the tasks. The idea of the TF-IDF is that the derived weights increase proportionally to the frequency of occurrence of a word in a given text extract in relation to the number of texts that contain the word. Following NLP jargon, *corpus* is a collection of documents, document is a text object, and text object is a sequence of words. In this subsection, *corpus* is the set of CBO activities descriptions and the documents are these descriptions. Given the set of occupations' activities D , a word w in the activity description, $d \in D$, the calculation of the corresponding weight ($weight_{w,d}$) according to the TF-IDF method is given by

$$weight_{w,d} = f_{w,d} \cdot \log\left(\frac{|D|}{f_{w,D}}\right) \quad (1)$$

where $f_{w,d}$ is the relative frequency of w in d , defined as the number of occurrences of each word relative to the total number of words in the activity description d .⁷ $|D|$ is the total number of activities descriptions and $f_{w,D}$ is the number of activities in D in which w appears. High values of $weight_{w,d}$ imply that w is an important word in d (first term), but not very common in D (second term) in equation (1). In that sense, high TF-IDF weights signal that the word w serves well to discriminate between the task sets of occupations.

A limitation of the TF-IDF method is the incapability to identify the grammatical variations of the same word, especially regarding verbs (Qaiser and Ali 2018). Spellings such as “analysed”, “analyse”, or “analysing” are treated as being different words. In order to avoid this issue, a stemming procedure is carried out before applying the TF-IDF method. Therefore, inflections of the same term are converted to a single common root. In addition, stop-word removal discards phrase-connecting words, such as conjunctions, numbers and special characters that have no semantic value for the analyses.

The resulting vector, $\mathcal{W}_d = (weight_{1,d} \dots weight_{W,d})$, collects the TF-IDF weights of all words from 1 to W , as W being the number of different words that appear in the Activities Matrix. The matrix $\mathcal{W} = (\mathcal{W}_1 \dots \mathcal{W}_D)$ ends up with 2,322 words describing 2,641 occupations. This matrix is 99 per cent sparse since most of the terms are not shared between occupations, which is common in TF-IDF analysis.

⁶ The CBO Activities Matrix contains an average of ten task descriptions for each of the 2,474 different occupations when the most disaggregate 6-digit classification is used. For example, the occupation “truck driver” (code 782510) lists tasks such as “transport container”, “load vehicle”, “deliver cargo”, “handle cargo with safety”, “inspect mechanical parts”, and so on.

⁷ The TF-IDF formula in equation (1) is based on Ramos (2003), however, instead of using the absolute frequency of words, we prefer the relative frequency as it makes the measure more sensitive to terms that define an activity while remaining insensitive to the number of activity descriptions pertaining to occupations.

We apply the cosine similarity measure to derive distances between occupations as vectors in high dimensional sparse matrices. The cosine similarity computes the angle between two vectors, that is the weights of words that compose the activities of two given occupations. Because the cosine measure is insensitive to the absolute length of each document, it works well for sparse matrices.⁸ Based on Aggarwal (2015), the cosine similarity measures for any pair of occupations i and j is defined as

$$\text{Cos}_{ij} = \frac{\sum_{w=1}^W \text{weight}_{w,i} \cdot \text{weight}_{w,j}}{\sqrt{\sum_{w=1}^W \text{weight}_{w,i}^2} \cdot \sqrt{\sum_{w=1}^W \text{weight}_{w,j}^2}} \quad (2)$$

Therefore, we have a symmetric matrix reflecting the bilateral similarity between occupations. Cos_{ij} is a continuous measure in the interval $[0, 1]$, where occupations with value 1 are identical to each other in terms of tasks, and occupations with 0 do not share words describing their tasks. Finally, to have a direct measure for the distance between occupations, we use $\text{OccDist}_{ij} \equiv \log(1 / \text{Cos}_{ij})$ in the following analysis.

2.3 Skill measures

Our aim in this section is to derive a novel methodology to generate skill measures that (a) are highly comparable across countries, (b) have modest data requirements, (c) do not rely on cross-country occupation matching, and (d) replicate an international standard.

Our method of skill measurement has three advantages over the previous ones. The first advantage is that we do not need ad hoc lists of words related to each type of task. Second, the method does not assume that the occupations perform the same activities across countries. Third, our method can measure the skills even of occupations that are not listed in a database such as O*NET. A large number of Brazilians work in occupations that cannot be found in O*NET, such as elevator operators in commercial buildings (CBO 5141-05 - *Ascensorista*) or bus ticket/money collectors (CBO 5112-15 - *Cobrador*).

We have chosen to follow the skills categories introduced by Autor, Levy and Murnane (2003) which were then adapted to the O*NET data by Acemoglu and Autor (2011). This will be the only relevant ad hoc decision in the following methodology. In order to replicate the United States data-based skills, we use the O*NET as an initial stepping stone. The remainder of the procedure relies exclusively on the Brazilian CBO data. The derivation of skill measures is carried out in five steps:

1. Build a dictionary of verbs from the O*NET that are associated with occupations with the highest skills in each of the groups (NRC, RC, NRM, and NRM);
2. Translate these verbs to Portuguese by means of a deep learning translation service;
3. Derive weights of these Portuguese verbs by means of an NLP technique;
4. Apply these weights by skill to the text tasks descriptions of the CBO occupations;
5. Normalize skill measures.

Our choice skill measures are the four categories from Acemoglu and Autor (2011), namely: NRC, RC, NRM and RM.⁹ Having calculated the original task measures of Acemoglu and Autor (2011), we rank occupations

⁸ The Euclidian distance, the most common distance metric in Machine Learning, is not suitable for a high dimensional sparse matrix (Aggarwal, Hinneburg and Keim 2001).

⁹ Minor adaptations and corrections are made to increase the meaningfulness of skills in the Brazilian context. Appendix A lists the task requirements that define these skill categories.

by their skill score and select the top 10 per cent of occupations. Repeating this procedure for all four skill categories defines the selected occupations' elements whose words are then translated into Portuguese. This resulted in around 6,100 tasks.

The conversion of the task file from English to Brazilian Portuguese is done with *DeepL Pro*, a neural machine learning translation service.¹⁰ The task descriptions contain information about contexts that would be undesirable for our purposes. For example: "clean a car" and "design a car" contain verbs associated with different skills, but the noun "car" would have the same weight as the verbs and would harm the classification. Therefore, we prefer to extract only the verbs from the task descriptions.

In this subsection, we use two corpora: verbs that are associated with skills in O*NET and the verbs in the activities descriptions by occupation in CBO. From the 96,900 words in the translated task descriptions of O*NET, 996 verbs have correspondence in the CBO task descriptions. The matrix V contains the weights (as defined below) for the 996 verbs (columns) in the four skill categories $i \in \{NRC, RC, NRM, RM\}$ (rows). Our final skill scores for each occupation are the result of the following matrix product:

$$Skill = V * Occup \quad (3)$$

where the matrix $Occup$ is defined by the raw count of the 996 verbs (rows) in the 2,478 CBO occupations (columns).¹¹

The usual way to identify the weights of terms in a document for V is through the TF-IDF method (see section 2.2). Here, however, TF-IDF has not turned out to be the best option. The reason is simple: if a verb is present in all four skill categories, the value of the IDF is 0, and obviously TF-IDF is equal to 0.¹² For example, if a given verb w_v appears 100 times in total with the following distribution: 90 times in the RM skill category, eight times in NRM, and once both in RC and NRC, the TF-IDF value of the verb w_v will be zero. The same value would be obtained if w_v appeared with equal frequency in the four skills. Thus, relevant information about the distribution of the verb w_v would be lost.

To overcome the distortion that TF-IDF would cause, we replace the IDF with the repeat rate (RR) to measure the relevance of the verbs in the CBO occupation descriptions (NLP documents in this subsection). Although the index was created for other purposes, it fulfils our objectives.¹³ The RR of a verb is the sum of squares of the ratio between the count of verb w_v in each skill and the total count of the same verb. It is a concentration index based on the distribution of a verb between skills. The minimum value of the RR is 0.25 (4×0.25^2) for a verb that is equally distributed among skills and the maximum value of 1 is obtained if the verb is observed in just one skill. The intuition is that if a verb is homogeneously distributed by skills it is not useful for discriminating between skills and vice versa.

¹⁰ DeepL uses a proprietary algorithm with convolutional neural networks. In blind tests, professional human translators considered the DeepL translations better than those by Google, Amazon and Microsoft services (see <https://www.deepl.com/en/quality.html>).

¹¹ The number of CBO occupations differ from the previous section because a few CBO occupations contain verbs for which there is no translation from the O*NET verbs. In these cases, the skill scores are missing. To avoid the artificially resulting gaps in workers' labour market transitions, we replace the missing values with the average skill scores of the corresponding more aggregate 5- or 4-digit occupation classification.

¹² NLP technical terms may create misunderstandings between the use of the term "document" in this subsection and in the previous one. In measuring the distance between occupations, "documents" referred to activities descriptions. In the skills measurement subsection, it refers to the verbs associated with each of the four skills.

¹³ The repetition rate is identical to the Hirschman-Herfindahl index in Economics or the Simpson (1949) index in Biostatistics. Its origins are controversial. Hirschman explicitly defended the merits of its creation (Hirschman 1964), while 19th century authors had already preceded him. According to Good (1982), Alan Turing created it independently at Bletchley Park while working on the decryption of the Enigma machine. Simpson, who also worked there, avoided attributing the authorship to Turing because this could be considered leaking secret information. We decided to choose the term "repeat rate" as a tribute to Alan Turing and because it reflects the intuition of the index. See Rousseau (2018) for a more thorough discussion of the origins of the index.

The main advantage of the TF-RR over TF-IDF is that it is more suitable for asymmetric distributions. If a verb is perfectly distributed across our four skills, then $RR = 0.25$ and $IDF = 0$. In the numeric example above, with a very asymmetric distribution, the RR would equal $0.83 (0.9^2 + 0.08^2 + 0.01^2 + 0.01^2)$ while IDF would still be 0. For the mathematical definition of TF-RR, the reader is referred to the Data Appendix A.A2. Ultimately, the combined TF-RR procedure provides the weights that are used as elements in the matrix V according to the product of TF^{wk} and RR^w . As before, TF is defined as the relative frequency of a verb w_v in the total number of verbs in skill k .

To facilitate the interpretation of the descriptive statistics and regressions, skill scores are standardized such that the sum of the four categories is equal to one in each occupation. The final scores can therefore be interpreted as shares. For example, $NRC = 0.25$ means that a quarter of the average workload in this occupation is executed using NRC skills. A characteristic feature of our skill definition is that the scores differences across occupations are relatively small, see Table 1. The reason is that even occupations with the highest NRC scores require some RC activities (such as making phone calls) or RM skills, like typing on a computer.

2.4 Model

2.4.1 Transition regressions and testable hypotheses

The research question in this paper is: how are skills related to labour market transitions in Brazil? In particular, the following hypotheses will be tested in our regressions and duration analysis:

- [H1] Workers' previous skill endowments affect the probability of transitioning to a different occupation.
- [H2] Workers' previous skill endowments affect the probability of transitioning to a different sector.
- [H3] Previous skill endowments affect the distance between occupations in case of an occupation change.
- [H4] Previous skill endowments affect the wage differential when workers move to a different occupation.
- [H5] Previous skill endowments affect the probability of a worker experiencing a period out of the formal labour market.
- [H6] The size of these marginal effects depends on workers' age, tenure, employer size and the business cycle.

The dependent variables that allow us to assess hypotheses H1 to H6 are defined according to the difference in the employment outcome of worker i in the current period t and the sequential employment spell (which must not necessarily be in year $t + 1$). We use the following variables:

- $\Delta Occ4$ is an indicator variable for whether we observe that a worker has a different 4-digit occupation in the following employment spell.
- $\Delta Occ2$ is defined similarly to the previous definition but based on 2-digit occupations.
- $\Delta Occ4 \uparrow$ is an indicator variable for whether we observe that a worker moves to a different 4-digit occupation with a higher average wage. Average wages are defined as the mean log wage for a given year and 4-digit occupation based on the observations in our sample.
- $\Delta Occ4 \downarrow$ is an indicator variable for whether we observe that a worker moves to a different 4-digit occupation with a lower average wage.
- $\Delta Sect$ indicates whether a worker will move to a different sector.
- $\log(OccDist)$ is the log of the inverse of the occupational distance defined in equation (2).
- $\Delta Wage$ is the difference between the log wage in the next and current employment spell.

- *OFLM* is an indicator variable for whether a worker experiences a period of at least one month out of the formal labour market between the current and the next employment spell.

Our baseline employment transition regressions take the following form

$$y_{it} = \alpha + \beta X_{it} + \sum_k \gamma_k skill_{it}^k + D_s + D_r + D_t + \epsilon_{it} \quad (4)$$

where the vector of control variables X_{it} includes worker i 's number of working hours, level of education, tenure, gender, age and firm size. D_s , D_r , and D_t are dummies for sector, federal state and year, respectively. The coefficients of interest γ_k then indicate how the current skill levels are linked to employment transitions of observably equivalent workers in the near future.

In line with the definitions above, we use NRC, NRM, and RM skills in the regressions, while RC skill is the omitted reference category. In case the dependent variable is an indicator, we apply logit regressions; otherwise, equation (4) is estimated with pooled OLS. Additionally, we also include interaction terms between the worker's skill level and moderator variables such as worker i 's age, tenure, employer size, GDP growth and residence in the federal state. Standard errors are clustered at the occupation level because the explanatory variables of interest - workers' skills - vary only across occupations. This procedure yields more conservative results than clustering at the worker level.

2.4.2 Duration analysis

The most appropriate way to analyse the probability of leaving the formal labour market is via duration analysis. This type of approach estimates the probability of leaving, given that this event has not occurred before, conditional on the set of observable variables (Lancaster 1979). The exit probability at time t upon survival to time t is known as the hazard rate. If the duration time t follows a Weibull distribution, which is the most frequently used in the literature, the hazard rate θ takes the following form

$$\xi Z_{it} = \alpha + \beta X_{it} + \sum_k \gamma_k skill_{it}^k + D_s + D_r + D_t + \epsilon_{it} \quad (5)$$

$$\theta_{it}(t | Z_{it}) = \rho t^{\rho-1} \cdot \exp(\xi Z_{it}) \quad (6)$$

where Z_{it} is composed of the same control variables defined in equation (4) and ρ is a parameter that indicates whether the exit probability has an increasing, decreasing or constant form. The parameters of the duration model are obtained by maximum-likelihood estimation, where standard errors are also clustered at the occupation level.

According to the nature of our worker panel, the survival data includes multiple records per subject with continuous-time measurement and the possibility of multiple exits. The exit event is defined as leaving the formal labour market for at least one month. Because we cannot distinguish whether the individual becomes unemployed, self-employed, or gets a job out of the formal labour market, we consider only exits for individuals that eventually return to the formal labour market. If an individual does not return to the formal labour market, we have no way of knowing the direction of his or her transition into a lower or higher wage, or distant occupation.

► 3 Results

3.1 Descriptive statistics

3.1.1 Skill scores and workers' characteristics

The descriptive statistics section presents aggregate trends in the Brazilian labour market that contribute to the understanding of the subsequent analysis of employment transitions at the worker level. The following results are derived from the full sample of the RAIS data from 2003 to 2018.

First, it is crucial to check whether our novel skill measures produce adequate results. The continuous skill measures' mean values in the most aggregated occupation groups are certainly plausible. There are nine groups, excluding police and military professionals, and taking into account both groups of industrial workers (discrete and continuous processes) in CBO.

Table 1 shows that the groups with the highest wages - professionals, managers and technicians - present the highest scores in NRC skills. These occupations are responsible for abstract and highly specialized tasks, intensive in hard and soft skills since most of the work is done by interacting with other people. Managers and technicians have the highest scores in RC skills as well. Besides, administrative and service occupations mostly require RC and, to a lesser extent, NRC skills. As expected, the highest scores on NRM skills belong to workers in maintenance, service and manufacturing 2 (continuous processes). Manufacturing 1 (discrete processes) and agricultural workers score highest in RM skills as those workers mainly perform repetitive tasks in controlled environments, with little or no necessity for improvisation.

► **Table 1. Skill scores by occupations.**

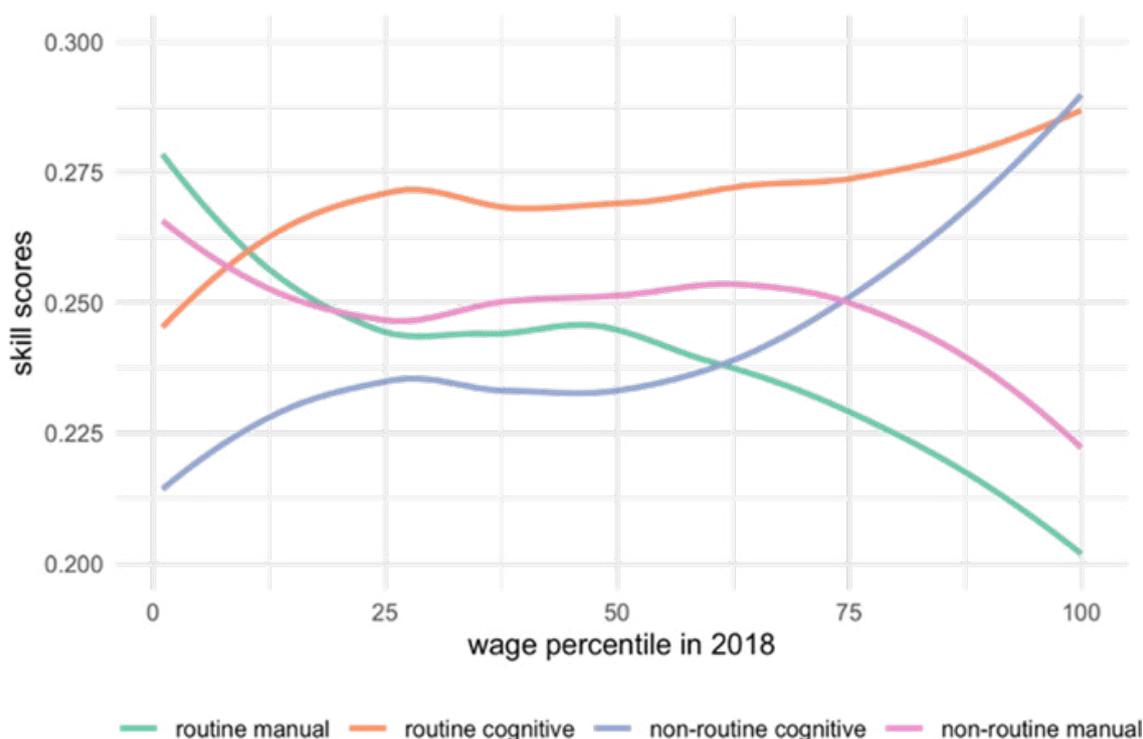
	Manag. (1)	Profes. (2)	Techn. (3)	Admin. (4)	Manuf. 1 (5)	Maint. (6)	Manuf. 2 (7)	Service (8)	Agric. (9)
NRC	0.28 (0.01)	0.29 (0.02)	0.27 (0.02)	0.25 (0.01)	0.21 (0.03)	0.22 (0.02)	0.22 (0.02)	0.24 (0.02)	0.23 (0.03)
RC	0.29 (0.01)	0.28 (0.02)	0.28 (0.02)	0.29 (0.01)	0.25 (0.02)	0.25 (0.01)	0.25 (0.02)	0.27 (0.02)	0.26 (0.02)
NRM	0.24 (0.02)	0.22 (0.01)	0.24 (0.02)	0.24 (0.01)	0.24 (0.03)	0.27 (0.01)	0.27 (0.02)	0.25 (0.01)	0.24 (0.03)
RM	0.19 (0.02)	0.20 (0.02)	0.22 (0.02)	0.22 (0.01)	0.30 (0.05)	0.26 (0.02)	0.27 (0.03)	0.24 (0.03)	0.27 (0.02)

Note: the cells show the mean and standard deviation (in parenthesis) for the skill measures (non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM) and routine manual (RM)) by 1-digit occupations over the entire observation period 2003–2018. Calculations are based on the full sample and apply weights for employment duration and hours worked. The 1-digit occupations are ranked according to their mean wage in descending order, namely: directors and managers; science and arts professionals; medium-level technicians; administrative service workers; workers in the production of industrial goods and services (1), workers in repair and maintenance; workers in the production of industrial goods and services (2); service workers and sellers; and agricultural workers.

Source: elaborated by the authors.

Figure 1 provides another way to compare the plausibility of our skill measures to previous findings from other countries. Following Autor, Katz and Kearney (2008), the figure shows how the skill scores vary across the wage distribution in 2018. Additional information for previous years is provided in the Appendix (see Figure B1). RM and NRM intensive occupations dominate at the lower end of the wage distribution, exhibiting a negative and decreasing slope over the wage distribution. NRC skills show a mirror-inverted pattern, beginning at a low level and rising at an accelerated slope over the upper half of the wage distribution. RC skills are most frequently used in occupations between the 20th and the 80th percentile. While the RC skill intensity is also increasing in wages, it is almost flat for the most part in the middle of the distribution. Regarding the evolution over time, Figure B1 in the Appendix indicates that skill intensities of high- and middle-wage workers remained quite constant. The most pronounced change is that low-wage employment now uses fewer NRC skills, whereas NRM skills are becoming more important. There is also a clear tendency of middle-wage workers to increasingly use NRC skills, in line with the RBTC hypothesis.

► Figure 1. Skill scores over the wage distribution in 2018.



Source: elaborated by the authors.

It is informative that the skill intensities over the wage distribution are similar to the United States experience. Autor, Katz and Kearney (2008), for example, report that NRC skills are monotonically rising, NRM skills are markedly decreasing, while RC skills are most frequent in middle-wage occupations. These patterns are as expected and are not related to taking the United States task descriptions in the O*NET as a starting point for our method. The observed patterns in Figure 1 reflect that NRC skills require a large investment in human capital (such as a university degree). Consequently, the return to these skills is highest and occupations with high NRC requirements are located at the top of the wage distribution.

The pattern in Figure 1 and the order of relative wages are in line with other studies of the Brazilian labour market. Controlling for observed and unobserved characteristics in a fixed-effects model, Detoni, Freguglia and Corseuil (2020) find higher returns for NRC intensive occupations, followed by RC, RM, and NRM skills. Sulzbach (2020) shows the increasing value of tasks related to cognitive skills since 2003, while

routine skills saw a decreasing wage compensation. This evidence is in line with RBTC, that is advancing automation substitutes previous routine tasks performed by humans, and therefore degrades their wages (Acemoglu and Autor 2011).

In order to evaluate how the socioeconomic characteristics of workers relate to their skills, we define discrete skill groups following Goos and Manning (2007) and Cortes (2016), among many others. These categories are listed in Data Appendix 1. Table 2 shows the descriptive statistics by comparing occupational and socioeconomic characteristics of each skill group between 2003 and 2018.¹⁴ NRC workers receive higher wages, followed by RC, RM, and NRM workers. This ordering remains stable through 2018, but all occupational groups experienced real gains over our observation period. In particular, NRM and RM occupations gained 33.8 per cent and 29.9 per cent higher real wages than in 2003. This catching up may not be a sign of high demand but rather be influenced by the important adjustments of the minimum wage, which grew by 76.5 per cent in real terms since 2003.

Regarding gender, men dominate the RM occupations in both years, scoring 83 per cent on average. On the other hand, women increased their participation in the three remaining categories and are predominant in NRC, RC, and NRM occupations. The most relevant effect of the increased female labour market participation can be seen in NRM jobs, which had 58 per cent of its positions occupied by men in 2003, decreasing to 47 per cent in 2018. These NRM occupations encompass service and cultural services workers, technicians in transport services and agricultural producers.

¹⁴ The values in Table 2 refer to the consecutive employment sample that will be used for most of the regression analysis below. Notwithstanding, the full sample yields very similar patterns.

► **Table 2. Descriptive statistics by skill groups: 2003–2018.**

	2003				2018			
	NRC (1)	RC (2)	NRM (3)	RM (4)	NRC (5)	RC (6)	NRM (7)	RM (8)
wage per hour (BRL)	126.34 (164.91)	55.70 (77.16)	31.90 (36.26)	39.71 (40.43)	150.28 (179.78)	65.30 (76.86)	42.68 (36.71)	51.57 (44.09)
male	0.49 (0.50)	0.44 (0.50)	0.58 (0.49)	0.83 (0.37)	0.46 (0.50)	0.40 (0.49)	0.47 (0.50)	0.83 (0.38)
age (years)	36.08 (8.82)	32.52 (8.81)	34.98 (8.97)	33.47 (8.77)	39.76 (9.32)	37.91 (9.89)	41.31 (9.73)	40.21 (9.53)
tenure (month)	87.27 (86.98)	64.90 (74.35)	53.97 (59.07)	47.23 (55.98)	92.27 (96.14)	74.34 (86.21)	72.24 (78.65)	64.21 (71.70)
occupation change	0.17 (0.38)	0.15 (0.36)	0.15 (0.36)	0.17 (0.38)	0.10 (0.30)	0.11 (0.31)	0.09 (0.29)	0.11 (0.32)
occ. change (2-digit)	0.14 (0.35)	0.13 (0.34)	0.10 (0.30)	0.12 (0.32)	0.08 (0.28)	0.09 (0.29)	0.05 (0.22)	0.08 (0.28)
sector change	0.06 (0.23)	0.07 (0.26)	0.07 (0.26)	0.10 (0.30)	0.05 (0.22)	0.07 (0.25)	0.07 (0.25)	0.09 (0.28)
log occ. distance	2.21 (0.92)	2.35 (0.90)	2.40 (0.87)	2.49 (0.98)	2.22 (0.88)	2.30 (0.91)	2.28 (0.84)	2.51 (0.95)
employment share (in %)	12.8	31.8	17.8	37.6	16.7	32.3	20.8	30.1
observations	270k	670k	374k	791k	482k	931k	598k	878k

Note: the cells show the mean and standard deviation (in parenthesis) for selected employment transition indicators in 2003 (columns (1) to (4)) and in 2018 (columns (5) to (8)) for each of the discrete skill groups: non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM) and routine manual (RM). Inflation is adjusted by the official consumer price index (IPCA) to 2018 prices. Calculations apply weights for employment skill duration and hours worked and are based on the consecutive employment sample.

Source: elaborated by the authors.

Likewise, the evolution of workers' age reflects, at least partly, the demographic transition in Brazil (IBGE 2018). All skill groups exhibit a higher average age in 2018 in comparison to 2003. Additionally, the most pronounced increase in average age occurs between 2013 and 2018, suggesting that firms primarily dismissed younger, less experienced workers during the economic recession. Workers in non-routine, cognitive-intensive occupations have longer employment tenure, increasing from 87.3 months in 2003 to 92.3 in 2018. RM workers are most prone to employment changes, that is to say that they present the lowest average tenure. Analogous to age, tenure shows a crescent trend that has accelerated since 2013 for all skill groups.

The variable "occupation change" in Table 2 shows the fraction of workers that moved into a different 4-digit occupation.¹⁵ We can see that occupational changes became less frequent over the years. In 2003, 17 per cent of NRC workers transitioned to another occupation, whereas only 10 per cent did in 2018. The same pattern is seen for RM, RC, and NRM workers. The wide definition that considers occupation group

¹⁵ Recall that, according to the definition of our consecutive employment sample, the variable occupation change is equal to one in year t if a worker is observed in a different occupation in either year t or $t + 1$. The skill group refers to a worker's occupation before moving to the different one.

changes based on a higher level of aggregation (the 2-digit CBO classification) shows the same declining pattern. In both cases, the change occurs gradually over time and does not seem to be influenced by the business cycle. The sector change variable presents a stable pattern over time, ranging from 5 per cent for NRC workers to 10 per cent for the RM skill group. The log occupational distance suggests that, without accounting for other observable characteristics, RM workers move the farthest away from the previous occupation. The mean occupational distance in 2003 for RM workers was 2.49, increasing slightly to 2.51 in 2018. NRM and RC workers decrease their mean occupational distance. NRC workers present a stable, relatively low occupational distance.

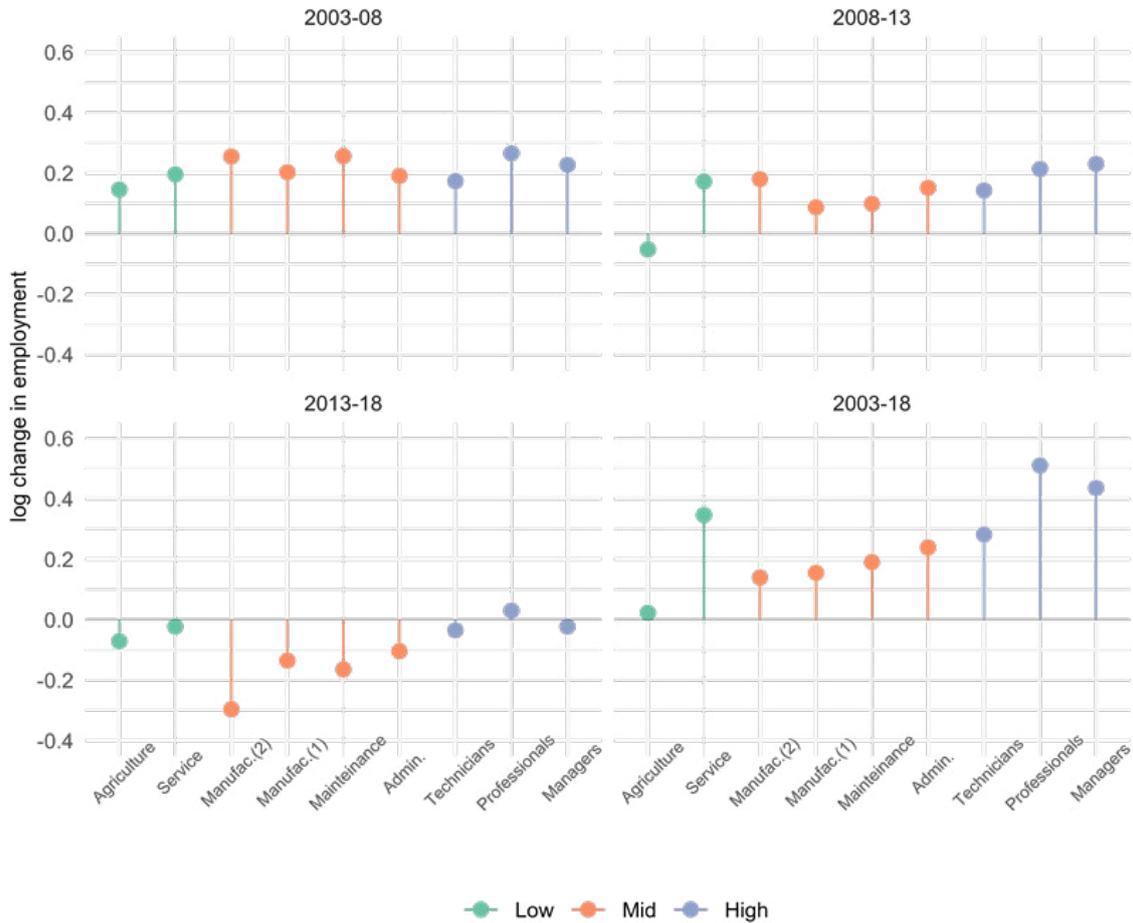
3.1.2 Employment changes and transitions

Figure 2 shows the employment change by occupational groups between 2003 and 2018, in total, and for three sub-periods. Over the last two decades, the largest relative change has been observed for the professional group, with a growth of 51 per cent relative to 2003, followed by 44 per cent growth for managers and 35 per cent in service occupations. In general, all aggregated groups increased the number of formal employees, in line with the 34 per cent increase of our 10 per cent random sample, which grew from 1.8 million in 2003 to 2.4 million full-time equivalent workers in 2018.

The three graphs for the subperiods in Figure 2 show that the Brazilian economy experienced a period of rapid GDP expansion between 2003 and 2009, with a mean growth rate of 4.2 per cent. The 2008 financial crisis did little harm to the economy, as GDP growth decelerated only slightly to a mean of 3.3 per cent between 2008 and 2013, before the economy plunged into a severe recession from 2014 onward (IBGE 2019b). The business cycle helps to explain the patterns of employment change shown in Figure 2. Between 2013 and 2018, occupations in manufacturing of continuous processes (manufacturing 2) suffered the strongest retraction of 30 per cent in employment, followed by maintenance workers and manufacturing of discrete processes (manufacturing 1). The recession was severe in sectors such as construction and transformation, falling to 85 per cent and 70 per cent of their production level in 2014, respectively (IFI 2018). Between 2003 and 2013, in contrast, all 1-digit occupation groups grew.

Concerning the whole period, agriculture seems to be an outlier, and in fact this sector is usually excluded. Nevertheless, agriculture is an important component of the Brazilian economy and we aim to be as inclusive as possible in the present analysis. The relative stability of agricultural employment can be explained by the fact that Brazil is among the most productive and competitive agro-exporters worldwide thanks to state-driven innovation, large properties and continuous expansion of its agricultural frontier over the last decades that has made land available for commercial agriculture (Hopewell 2016). Interestingly, back in 1970, Brazil was a net importer of agricultural goods. Agriculture is directly responsible for (only) 5.3 per cent of GDP but employs 15 per cent of the country's labour force (IBGE, 2019a, 2019b).

► Figure 2. Employment changes by occupation group, 2003-2018.



Note: the 1-digit occupations are ranked according to their mean hourly wage in 2003. See Appendix Table 1 for the precise definition of the occupation groups. Calculations are based on the consecutive employment sample and apply weights for employment duration and hours worked.

Source: elaborated by the authors.

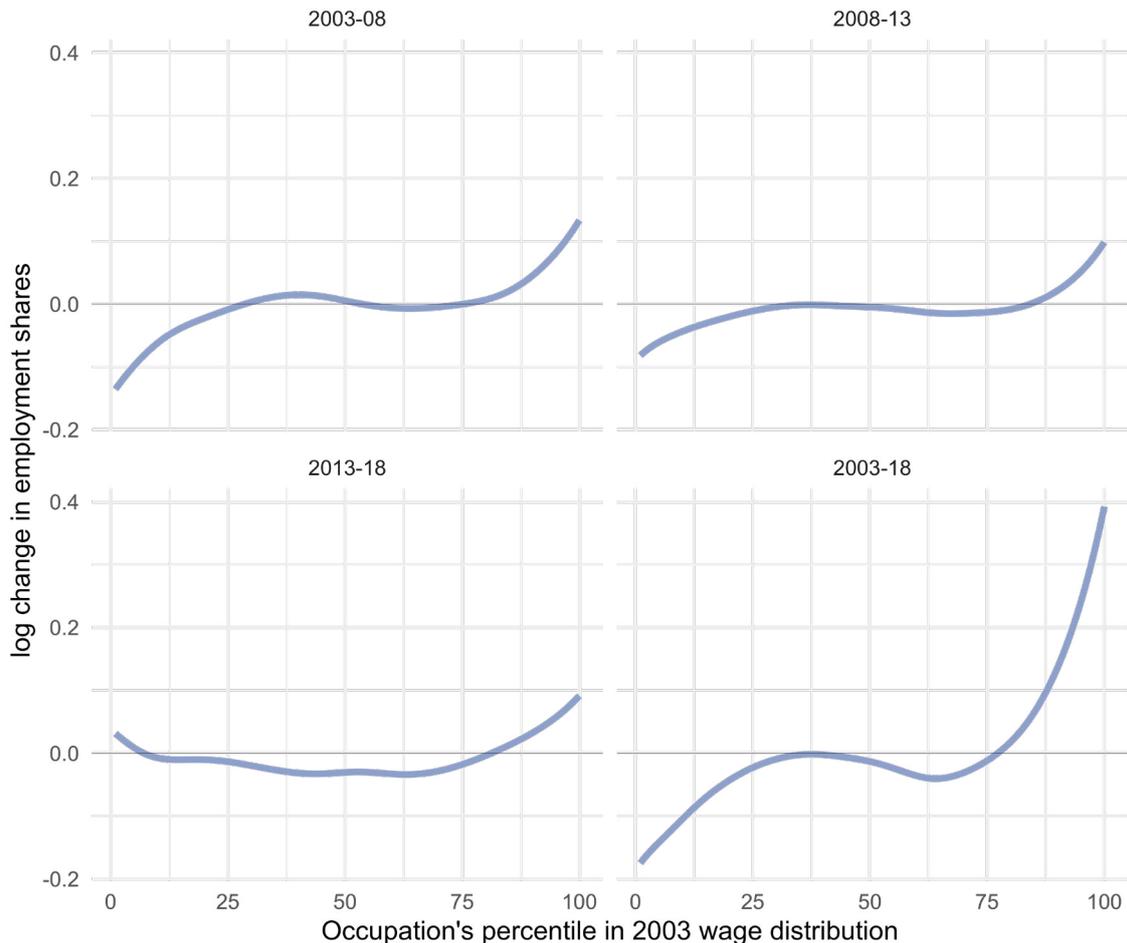
Jobs in agriculture and services (such as food service workers, security guards, cleaners, hotel and housekeepers and sales attendants) are often referred to as “low-skill” occupations; manufacturing, maintenance, and administrative workers as “medium-skill” occupations; and managers, professionals and technicians as “high-skill” workers¹⁶ (Autor 2019). Using these categories, Figure 2 shows that low-skill occupations grew on average between 2003 and 2013 and fell slightly over the period 2013-2018. The high-skill occupations followed the same pattern but grew at higher rates. Middle-skill occupations, in contrast, present a pattern of steadily declining growth, culminating in a heavy contraction of up to -20 per cent in 2013-2018. In sum, Figure 2 shows the first sign of employment polarization with low-skill and high-skill occupations growing relatively more than medium-skill occupations.

¹⁶ While this may be the right classification for the majority of those workers, some occupations, such as production managers, directors, and supervisors obviously require a high degree of NRC skills. The primary sector in Brazil and other heavily exporting nations receive substantial R&D investment, and consequently there are numerous researchers and engineers employed in agriculture-related companies. These examples illustrate that a more parsimonious classification of skills (like the discrete one we use alternatively) does not do justice to all workers and occupations.

Figure 3 describes the changes in employment shares along with the wage distribution over the last two decades, following the seminal work of Autor, Katz and Kearney (2008). The total change and the breakdown for three sub-periods allow us to analyse whether Brazil experienced employment polarization similar to developed countries. Only the most recent period, 2013 to 2018, shows employment growth in occupations at the top and bottom of the wage distribution and a retraction in medium-income occupations. This is additional evidence of wage polarization since the left and right tails of the wage distribution are increasing relative to its middle.

Over the last observation period (2013-2018), the Brazilian formal labour market resembles the change in the United States labour market between 1990 and 2000. In both countries, employment losses are most pronounced between the 40th and 60th wage percentile. The magnitude of the most extreme aggregate changes in a single percentile in Brazil, range from +8 per cent to -3.5 per cent, while those reported by Autor, Katz and Kearney (2008) over a period twice as long are between -13 per cent and +16 per cent.

► **Figure 3. Employment changes over the wage distribution 2003-2018**



Note: The figure shows the employment changes in 4-digit occupations ranked according to their 2003 mean wage using a locally weighted smoothing for the entire observation period, as well as three different subperiods, as indicated in the above four graphs. Wages are deflated to 2018 prices using the IPCA. Calculations are based on the entire RAIS sample and apply weights for employment skill duration and hours worked.

Source: elaborated by the authors.

The polarization pattern is not seen when we consider other time intervals, like 2003 to 2008, 2008 to 2013, or the full period. Although high-skilled employment is continually expanding, the low-skilled occupations experienced negative growth rates and changes in the middle of the skill distribution were close to zero. Jaimovich and Siu (2020) and Hershbein and Kahn (2018) show that in the United States, routine-biased technical change and jobless recoveries go hand in hand. Increasing automation of routine and manual tasks can accelerate these trends. That is, Brazilian firms may have used the recent recession as an “opportunity” to displace and substitute middle-skill workers. It remains to be seen whether the employment polarization in Brazil is a transitional phenomenon related to the economic downturn or the beginning of a secular trend that arrived later than in developed countries.

Lastly, we can look at transitions between occupations in different skill groups, as well as transitions into or out of the labour market and into retirement. Table 3 shows a transition matrix based on the discrete skill classification. The share of workers who remained in the same classification in 2003 and 2018 range from 27 per cent for NRM to 34.2 per cent in RM occupations. In general, routine skilled workers are less likely to move into another occupational group. Transitions from manual to cognitive occupations and vice-versa are least common.

► **Table 3. Transition matrix between skill groups 2003-2018**

2003 \ 2018	2018					
	NRC	RC	NRM	RM	OFLM	Retired
	(1)	(2)	(3)	(4)	(5)	(6)
NRC	28.4	10.2	2.5	3.9	46.2	8.9
RC	10.6	30.3	4.6	5.1	44.3	5.1
NRM	3.1	7.4	26.7	9.1	48.6	5.1
RM	2.8	5.1	4.7	34.2	49.9	3.3
OFLM	15.7	32.9	22.9	27.1		1.4

Note: the cells show the percentage of workers from a given skill category (non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM), and routine manual (RM) in 2003 (column 1) that are either out of the formal labour market (OFLM), employed in the same or in another skill group (row 1) in 2018, or retired in or before 2018. Per definition, the row sum is equal to 1. Calculations are based on the entire sample and apply weights for employment skill duration and hours worked.

Source: elaborated by the authors.

Although the observation period spans 16 years, the share of transitions out of the formal labour market is strikingly high. About 46 per cent of the NRC workers in 2003 were outside the formal labour market in 2018. Another 8.9 per cent of all NRC workers observed in 2003 moved into retirement by or before 2018. So overall, the effects of aging, self-employment (in the growing gig-economy) and increasing unemployment, among others, contribute to the fact that 55 per cent of all NRC workers are not in the formal labour market 16 years later. The high turnover is only slightly affected by transition periods within a calendar year, because the RAIS data allow us to observe the individual over the entire year and not only at a specific date.

The differences in the job-to-retirement rates are also remarkable. Among the RM skilled, only 3.3 per cent moved from formal employment directly into retirement. For RM and NRM workers, the rate is higher (5.1 per cent) but still much below that of the most skilled NRC workers. Two more observations from Table 3 confirm that the RM workers are the most vulnerable group. Their transition rate out of the formal labour market is highest at 49.4 per cent and, despite being the largest group in terms of the aggregate employment share (30-37 per cent), only 27 per cent of those who were not observed (OFLM) in 2003 but entered the labour market in one of the following years were in a RM occupation in 2018. This picture is in line with RM workers being hardest hit by the economic recession (see Figure 2). In general, the transition matrix reveals a high job market turnover. This is a known problem in the Brazilian economy that deteriorates productivity and wages. Rocha, Pero and Corseuil (2019) indicate that the annual turnover rate at the firm level in Brazil fluctuates around 45 per cent, ranging from 38 per cent to 64 per cent. The high turnover negatively impacts productivity by reducing specific learning in organizations.

3.2 Propensity to change occupation and sector

This section analyses how skill endowments influence the probability of workers moving into different occupations or sectors. That is, we aim at providing answers to the testable hypotheses H1, H2 and H6 listed in section III.D. In the next section, the consequences of these changes on wages and occupational distance are explored further.

Table 4 presents the results from the logit regressions where the binary dependent variables are the two definitions of occupation change and if a worker migrates to a different sector. The upper part of the table reports the marginal effects of the continuous skill variables. The lower part shows the predicted change probabilities for workers with average characteristics in each discrete skill group.¹⁷

¹⁷ These regressions are based on the consecutive employment sample. This choice disregards occupation changes that occur after a period out of the labour market of more than one month. In this way, the estimations will neither be biased due to right censoring of employment spells nor confounded by the impacts that the worker experiences out of the labour market, such as unemployment stigma or educational upgrade. In even columns, the sample is additionally restricted to workers who change sectors.

► Table 4. Occupation change, sector change and skill endowments

dep. var.:	ΔOcc4		ΔOcc2		sector change
sample:	CE (1)	CE & $\Delta\text{Sect.}$ (2)	CE (3)	CE & $\Delta\text{Sec.}$ (4)	CE (5)
marginal effects					
NRC	0.541 (0.345)	0.340 (0.685)	0.304 (0.343)	0.494 (0.831)	0.119 (0.163)
NRM	0.448 (0.288)	-0.904* (0.511)	0.061 (0.300)	-1.307* (0.739)	0.482*** (0.174)
RM	0.386 (0.277)	0.351 (0.561)	0.095 (0.248)	0.259 (0.602)	0.187 (0.120)
N	39,086,764	5,220,179	39,086,764	5,220,179	39,086,764
LL	-18,319,130	-3,092,822	-15,527,561	-3,384,799	-13,856,998
Pseudo R2	0.049	0.058	0.051	0.054	0.098
predicted probabilities					
NRC	0.160*** (0.005)	0.623*** (0.017)	0.128*** (0.004)	0.530*** (0.015)	0.089*** (0.002)
RC	0.187*** (0.009)	0.684*** (0.020)	0.153*** (0.005)	0.581*** (0.016)	0.120*** (0.003)
NRM	0.187*** (0.006)	0.667*** (0.015)	0.132*** (0.006)	0.522*** (0.013)	0.133*** (0.004)
RM	0.219*** (0.008)	0.688*** (0.018)	0.158*** (0.008)	0.568*** (0.017)	0.164*** (0.004)

Note: the upper part of the table presents the marginal effects of skill scores on the probability of either moving to a different occupation or sector applying logit regressions according to equation (4). The lower part of the table presents the predicted probabilities for workers from each of the four discrete skill groups (non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM), and routine manual (RM)) with average characteristics. The dependent variables in columns (1) to (4) assume the value 1 if an occupation change between the current and the next employment spell is observed. Columns (1) and (2) use the 4-digit occupation classification, whereas columns (3) and (4) are based on the wider (2-digit) occupational classification. The estimation in columns (2) and (4) are similar to columns (1) and (3), respectively, but the sample is further restricted to workers who change sectors. In column (5), the dependent variable indicates whether the worker switches to a different sector and the consecutive employment (CE) spell sample is used as in columns (1) and (3). The estimations control for workers' tenure, gender, age and age squared, education, number of hours worked, type of employment contract, firm size and legal nature, as well as year, sector, and federal state fixed effects. Standard errors in parenthesis are clustered at the occupation level. * denotes significance at ten, ** at five and *** at 1 per cent level.

Source: elaborated by the authors.

The first column of Table 4 shows that higher NRC skills are positively related to the probability of moving to a different job. This relation is stronger than for any other type of skill, yet none of the marginal effects is statistically significant. Once the observable individual characteristics are taken into account, an average worker in a typical NRC job has an occupation change probability of 16 per cent. RC and NRM workers change occupations at slightly higher rates (19 per cent), whereas RM workers are most likely to switch occupations (22 per cent). Note that these predicted probabilities are higher and of a different order than the unconditional observed probabilities in Table 2. That is, characteristics such as workers' age, tenure, sectoral affiliation, employment contract type and firm size result in the lowest occupational and sectoral

change probabilities for NRC workers. In contrast, NRM skills seem to be best employed in the same occupation, especially when one moves to a different sector.¹⁸

More precisely, column (2) shows that workers who use 10 per cent more NRM skills have a 9 per cent lower occupation change probability, *ceteris paribus*. By conditioning the sample based on sectoral change, we observe a jump in predicted probabilities to above 62 per cent for all skill groups. The differences in predicted probabilities across skill groups are less pronounced in this case. As in the previous estimations, NRC workers are most likely to maintain their previous occupation in the new sector, and RM workers are more likely to switch jobs as they change sector. However, column (5) shows that NRM skills seem to boost sectoral switching the most. So, in this case, the work activity of NRM (predominantly service jobs) seems to be much more relevant than the sector in which it is applied. The high occupational and sectoral change probabilities of RM workers seem to stem from their unfavourable labour market position, including low wages, higher job instability and unemployment rates.

► **Table 5. Occupation change, sector change, skill endowments and moderators**

dep. var.:	occupation change				sector change			
	tenure (1)	age (2)	ΔGDP_s (3)	firm size (4)	tenure (5)	age (6)	ΔGDP_s (7)	firm size (8)
NRC @low	0.284 (0.404)	0.377 (0.361)	0.549 (0.339)	0.676 (0.487)	0.046 (0.231)	0.128 (0.178)	0.169 (0.161)	0.355* (0.206)
NRC @ high	0.632* (0.332)	0.770** (0.351)	0.580* (0.350)	0.417 (0.334)	0.256* (0.156)	0.219 (0.151)	0.156 (0.162)	-0.046 (0.190)
NRM @low	0.425 (0.329)	0.519* (0.301)	0.496* (0.286)	0.935** (0.411)	0.744*** (0.239)	0.560*** (0.198)	0.577*** (0.177)	0.928*** (0.265)
NRM @ high	0.425 (0.285)	0.438 (0.309)	0.500* (0.296)	0.081 (0.278)	0.292** (0.131)	0.567*** (0.151)	0.546*** (0.175)	0.203 (0.155)
RM @low	0.429 (0.322)	0.418 (0.298)	0.390 (0.271)	0.531 (0.411)	0.308* (0.170)	0.255* (0.133)	0.233** (0.118)	0.429*** (0.146)
RM @ high	0.233 (0.275)	0.366 (0.275)	0.434 (0.285)	0.283 (0.242)	0.092 (0.121)	0.198* (0.111)	0.233* (0.120)	0.031 (0.138)
LL	-18,288k	-18,314k	-18,321k	-18,319k	-13,844k	-13,862k	-13,863k	-13,859k
Pseudo R2	0.050	0.049	0.049	0.049	0.099	0.098	0.098	0.098

Note: the table shows the marginal effects of the skill measures (non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM), and routine manual (RM)) evaluated at the lowest and highest tertile of the moderator variable's distribution. These marginal effects are derived from logit regressions according to equation (4) with interactions between the skill measures and a moderator variable, as indicated in the second row. The dependent variables assume the value 1 if an occupational or sectoral change between the current and the next employment spell is observed. The estimations control for workers' tenure, gender, age and age squared, education, number of hours worked, type of employment contract, firm size, legal nature, as well as year, sector, and federal state fixed effects. The consecutive employment (CE) spell sample is used and the number of observations is equal to 39,086,764 in all estimations. Standard errors in parenthesis are clustered at the occupation level. * denotes significance at ten, ** at five, and *** at one percent level.

Source: elaborated by the authors.

Table 5 reveals some heterogeneity across skills for the moderator variables of age, tenure, GDP growth, and firm size. Higher tenure seems to increase the probability of NRC workers assuming different occupations and sectors. For example, column (2) indicates that NRC workers in the lower tertile of the age distribution have a marginal effect on the occupation change probability equal to 0.37, which increases to 0.77

¹⁸ When we include fewer control variables in the estimations in Table 4 (only age, gender and year fixed effects), the coefficients tend to be lower but in the same order relative to each other. Thus, controlling for observable worker characteristics seems to disclose the differences among skills more clearly. Restricting the sample to male workers tends to magnify the coefficients' size. For example, apart from the significant coefficients in Table 4 the NRC coefficients in columns (1) and (3) become significant at the 10 per cent level.

in the upper age tertile. In other words, when an average worker aged 28 increases their NRC skill set by 10 per cent, the occupation change probability increases from 16 per cent to 19.7 per cent, while the probability for a worker aged 43 would increase to 24.7 per cent. This observation seems to be in line with the age-biased technical change in Brazil, highlighted in Firpo and Portella (2019) whereby some of the older NRC workers, whose skills are particularly complemented by new technologies, would rather switch to a different occupation and constantly face the need for adjustments in their job. Contrarily, higher tenure and higher manual skills seem to decrease the propensity for sectoral switching. GDP growth does not seem to change the marginal effects of skills in a relevant way. Lastly, firm size appears to decrease the marginal effects of either skill type, both regarding occupational and sectoral changes.

3.3 Consequences of occupation changes

Tables 6 and 7 show how skills relate to the distance between occupations and to wages when workers switch occupations according to hypotheses H3, H4 and H6. Note that these findings complement our previous observations from the Transitions Matrix between skill groups in Table 3. The difference is that the regressions in Table 6: (a) make workers more comparable with each other, (b) apply the continuous skill measures, and (c) consider only job-to-job transitions. Similar to the observations in Table 5, the skill measures' coefficients are insignificant in the baseline specification because the apparently substantial heterogeneity between workers leads to high standard errors. Once we use interaction regressions or distinguish whether workers move to higher or lower-paying occupations in Tables 6 and 7, we uncover some significant heterogeneity among workers.

► **Table 6. Distance after occupation changes**

dep. var.:	log occupational distance						
	sample:	$\Delta Occ4.\uparrow$	$\Delta Occ4.\downarrow$	$\Delta Sect.$	$\Delta Occ4.$		
var. mediator	(1)	(2)	(3)	tenure	age	GDP	firm size
NRC	-4.345 (2.950)	-4.812* (2.893)	0.021 (2.393)	-1.642 (2.452)	7.702*** (2.858)	-3.458 (2.606)	3.588 (2.954)
NRM	-2.933 (3.261)	-6.489** (2.519)	-1.989 (2.391)	-2.696 (2.659)	3.735 (2.421)	-3.993 (2.666)	2.099 (2.249)
RM	2.241 (2.255)	-0.761 (2.397)	3.842** (1.811)	2.671 (1.989)	10.724*** (2.181)	1.226 (2.134)	5.957*** (2.106)
var				0.040*** (0.011)	0.209*** (0.043)	0.026 (0.025)	2.188*** (0.780)
var * NRC				-0.069*** (0.019)	-0.351*** (0.076)	-0.07 (0.048)	-3.591** (1.471)
var * NRM				-0.046*** (0.017)	-0.239*** (0.057)	-0.008 (0.034)	-3.027*** (0.963)
var * RM				-0.052*** (0.014)	-0.294*** (0.053)	-0.03 (0.029)	-2.385** (0.984)
N	4231257	3371208	3536374	7602465	7602465	7602465	7602465
F	38	19	26	32	35	33	33
R2	0.07	0.059	0.07	0.061	0.061	0.059	0.06

Note: the table presents the estimated coefficient of skill measures (non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM), and routine manual (RM)) and their interaction effects with a moderator variable as indicated in the third row applying OLS regressions according to equation (4). The second line indicates whether the sample consists of workers changing occupations (columns (4) to (7)) or occupations and sectors (column (3)) or changing to an occupation with higher or lower average wage in columns (1) and (2), respectively. The dependent variable is the log of the occupational distance. The estimations control for workers' tenure, gender, age and age squared, education, number of hours worked, type of employment contract,

firm size and legal nature, as well as year, sector and federal state fixed effects. Standard errors in parenthesis are clustered at the occupation level. * denotes significance at ten, ** at five and *** at one per cent level.

Source: elaborated by the authors.

The results in columns (1) to (2) in Table 6 indicate that having more non-routine skills decreases the occupational distance. Once we account for interactions among skills and age, tenure and firm size, the majority of marginal interaction effects also become significant. For example, columns (4) and (7) in Table 6 show that the pure effects of age and firm size on occupational distance are positive, while their interaction with skills is negative. That is, having more non-routine skills allows older, more tenured workers to find closer occupations. The same is true for workers that have worked in large companies. It is indeed intuitive that workers who have accumulated more experience on the job or in the labour market would want to utilize those abilities in a similar occupation whenever feasible. The business cycle does not seem to affect occupational distance, and neither shows any heterogeneous effect across skill types. Finally, column (3) shows that workers with more RM skills move to more distant occupations once they change occupation and sector.

Regarding wage changes following a job-to-job transition, Table 7 indicates that only the NRC skills show a positive relation, independent of whether the worker moves to an occupation with higher or lower *average* wage. When workers change sectors and occupations, the differences between skill groups are insignificant. Extending those regressions with the four interaction terms used previously, Table 7 shows that neither firm size, tenure nor age show meaningful or significant estimates. Only regional GDP growth seems to magnify the wage changes of non-routine skills. In line with the observed employment changes and demand for skill, we observe that during expansionary periods wage increases are higher for NRC workers, whereas the marginal return to RM skills is almost zero.

► **Table 7. Wages after occupation changes**

dep. var.:	log wage change						
sample:	$\Delta\text{Occ4.}\uparrow$	$\Delta\text{Occ4.}\downarrow$	$\Delta\text{Sect.}$	$\Delta\text{Occ4.}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
var. mediator				tenure	age	GDP	firm size
NRC	0.587*** (0.177)	0.385* (0.214)	0.296 (0.314)	0.039 (0.261)	-0.059 (0.385)	0.118 (0.258)	-0.295 (0.433)
NRM	0.183 (0.149)	-0.226 (0.202)	-0.030 (0.310)	0.058 (0.296)	-0.027 (0.325)	-0.1 (0.274)	0.004 (0.442)
RM	0.161 (0.150)	0.302 (0.198)	0.285 (0.255)	0.178 (0.221)	-0.358 (0.298)	0.242 (0.221)	0.067 (0.387)
var.				-0.001 (0.001)	-0.009 (0.007)	-0.014** (0.006)	-0.093 (0.111)
var. * NRC				0.005** (0.002)	0.009 (0.012)	0.026** (0.010)	0.248 (0.172)
var. * NRM				-0.002 (0.002)	0.002 (0.009)	0.029*** (0.009)	0.008 (0.163)
var. * RM				0.002 (0.002)	0.019** (0.009)	0.004 (0.007)	0.089 (0.158)
N	4,231,257	3,371,208	3,536,374	7,602,465	7,602,465	7,602,465	7,602,465
F	341	186	320	307	279	293	266
R2	0.032	0.04	0.061	0.035	0.034	0.035	0.034

Note: the table presents the estimated coefficient of skill measures (non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM), and routine manual (RM)) and their interaction effects with a moderator variable as indicated in the third row applying OLS regressions according to equation (4). The second line indicates whether the sample consists of workers changing occupations (columns (4) to (7)) or occupations and sectors (column (3)) or changing to an occupation with higher or lower average wage in columns (1) and (2), respectively. The dependent variables are the log wage change. The estimations control for workers' tenure, gender, age and age squared, education, number of hours worked, type of employment contract, firm size and legal nature, as well as year, sector and federal state fixed effects. Standard errors in parenthesis are clustered at the occupation level. * denotes significance at ten, ** at five and *** at one per cent level.

Source: elaborated by the authors.

It is important to note that the present approach does not identify marginal wage returns as Gonzaga and Guanziroli (2019) derive from Mincerian wage regressions. Evidence from Brazil and the United States indicates that NRC wages receive the highest marginal returns, particularly in agglomerations (Ehrl and Monasterio 2019). The present approach only allows us to check whether otherwise observably equivalent workers instantly receive a systematically higher or lower wage in a novel or different occupation, depending on their current skill level. Another caveat is that the present estimates are conditional on changing jobs, inside or in other company. RAIS captures changes in job titles and wages annually, even if the worker remains with the same employer. However, the sample may not be representative of the entire population due to selection bias, as the probability of job changes is related to skills, as shown in Table 5.

3.4 Duration results

This section presents how skills are related to the formal labour market exit probability given that the worker re-enters at a later time. That is, the following duration analyses test and largely confirm hypotheses H5 and H6. The reader should bear in mind that the period out of the labour market may either be unemployment, voluntary time-out or in the informal sector. Temporary illness or injury leave is not accounted as a labour market exit.

► **Table 8. Risk of transition out of the labour market with moderator variables**

moderator var.:	dep. var.: hazard rate				
	(1)	age (2)	tenure (3)	Δ GDP (4)	firm size (5)
NRC	0.730 (1.085)	-5.668** (2.444)	-1.485 (2.202)	0.731 (1.024)	1.433 (1.831)
NRM	2.099* (1.145)	-1.662 (2.756)	2.91 (2.362)	2.053* (1.052)	6.945*** (1.776)
RM	1.930** (0.834)	-1.96 (2.082)	2.588 (1.822)	1.856** (0.800)	2.916** (1.309)
var		-0.120*** (0.037)	-0.024 (0.044)	-0.007 (0.018)	0.687 (0.471)
var * NRC		0.192*** (0.060)	0.098 (0.066)	0.000 (0.034)	-0.471 (0.915)
var * NRM		0.113** (0.057)	-0.053 (0.075)	0.01 (0.031)	-2.620*** (0.712)
var * RM		0.116** (0.048)	-0.041 (0.056)	0.017 (0.023)	-0.583 (0.699)
N	50,082,074	50,082,074	50,082,074	50,082,074	50,082,074
LL	-7,255,895	-7,256,886	-7,187,585	-7,255,807	-7,247,139
AIC	14,512,013	14,513,998	14,375,397	14,511,844	14,494,504
BIC	14,513,759	14,515,775	14,377,191	14,513,653	14,496,282

Note: the table presents the estimated coefficients of the skill measures (non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM), and routine manual (RM)) from equation (6) and its extension with interaction terms using the moderator variables specified in the second line according to the duration model in equation (6). The estimations are based on the complete sample and include the workers' tenure, gender, age and age squared, education, number of hours worked, type of employment contract, firm size and legal nature, as well as year, sector and federal state fixed effects. Standard errors in parenthesis are clustered at the occupation level. * denotes significance at ten, ** at five and *** at one per cent level.

Source: elaborated by the authors.

Table 8 displays the estimated coefficients ξ according to the model specified in equation (6). That is, variables with positive coefficients increase a worker's hazard rate, *ceteris paribus*. Thus, the first column indicates that workers who use manual skills intensively have a greater exit risk than the reference category of RC workers. For a different interpretation of the results, recall that the hazard rate representation can be derived as $\exp(\xi)$. So, for example, the elasticity of the hazard rate with respect to RM skill is equal to $6.9 = \exp(1.93)$. Occupations that require 10 p.p. more RM skills have a 69 per cent higher hazard rate.

Column (3) reveals substantial interaction effects between skill endowments and the age profiles of workers. On the one hand, higher age, *per se*, and having higher NRC skills seems to decrease the exit probability. On the other hand, the combination of being more specialized in either NRC or manual skills and being older increases the labour market exit risk. This result could be due to older, more experienced workers having stronger preferences for self-employment or retirement. The business cycle and tenure, however, are apparently unrelated to the labour market exit risk of employees with different skill types. Lastly, column (5) shows that manual skilled workers are more protected from losing their job when they work in large firms. There seems to be no interaction between NRC skills and firm size.

3.5 Robustness

Our findings are robust with regard to the definition of our skill measures. First, there are no qualitative modifications regarding Acemoglu and Autor's (2011) choice of O*NET task elements of each skill measure. Our two minor changes were necessary to correct one obvious problem of O*NET and one adaptation to the Brazilian labour market (see Data Appendix A.A1). The methodology for measuring skill rests on only one ad hoc choice: the percentile of occupations for each skill group that feeds the NLP algorithm that extracts the verbs matched to the CBO. The results presented above are based on the top 10 per cent of occupations. Selecting only the occupations closest to the top increases the risk of leaving out verbs representing each skill. On the other hand, too low a cut-off point would include occupations with verbs not characteristic of the four skills we considered. Notwithstanding, the qualitative results hold with different cut-off points at, for example, 5 per cent and 15 per cent.

The results are also robust to the choice of the English-to-Brazilian Portuguese translation algorithm. We ran initial tests with the "Google translate" function from Google Calc. Overall, the results are robust, but DeepL Pro yields the best translations because it considers the context of the occupation description more accurately.

Moreover, the regression results hold when: (a) the sample is restricted to full-time employees; (b) workers who change jobs more than three times over the observation period are excluded; (c) when instead of $\square\text{Occ4}$, the wider definition of occupation change ($\square\text{Occ2}$) is used; and (d) the definition of sector and occupation changes is conditional upon moving to a different firm. As mentioned previously, the findings are also robust against changes in the set of control variables, as well as to restricting the sample to male employees.

▶ 4 Discussion

After two decades of falling inequality and rising wages, the recent economic recession, aggravated by the Covid-19 pandemic, brings uncertainty as to whether the improvements in the Brazilian labour market, particularly the reduction in inequality, can be sustained. The Brazilian economy may suffer a slow and jobless recovery with wage and employment polarization. This scenario should prompt policymakers to understand the interplay between skills and occupational transitions to increase workers' welfare. To sum up the main findings from the different areas of our analysis, we provide a brief characterization of workers by the relative intensity of skills.

NRC workers clearly receive the highest wages, have higher tenure and are equally divided between men and women. They show the highest probability of direct job to retirement transition and have the lowest predicted probability of occupation switching. Sectoral switching for these workers is also the least common. Their new occupation is relatively close to the previous one unless it is in a different sector and, consequently, wages tend to increase. Their formal labour market exit risk is relatively low, particularly for young employees.

NRM workers receive the lowest wages and are mainly employed in maintenance occupations. They have the highest probability of switching sectors while remaining in the same occupation and, when they change occupations, the distance between jobs is small. Upon changing occupations, their wages tend to increase, particularly when the economy is expanding. Labour market exit risk is highest, particularly for older employees.

RM workers' wages and tenure are relatively low and the share of men in these occupations is above 80 per cent. They show the largest occupational distances in new occupations and have the highest predicted occupational and sectoral switching probabilities. They are primarily employed in manufacturing and agricultural jobs, and their labour market exit risk is high, particularly for older employees. Their probability of job-to-retirement transition is the lowest.

RC workers' wages are higher than in manual skilled occupations but still well below NRC jobs. The RC-intensive occupations are dominated by women workers. (Un)conditional probabilities of occupational and sectoral switching for these workers are intermediate, in line with our choice of RC as the reference category in the regressions. When compared to other workers, their occupational distances and wage changes are also intermediate. The skills of RC workers apply to a wide array of occupations but, above all, in administrative and service jobs.

In general, our data and skill measures confirm that NRC workers receive the highest wages, while NRM workers receive the lowest. All types of workers experienced rising real wages over the period 2003–2018, but their ordering remained stable. RC and RM intensive occupations are more present in the middle of the wage distribution, which helps to explain the polarization trends.

The debate about whether the Brazilian labour market is polarizing in a similar way to developing economies has been inconclusive thus far. We present additional evidence of polarization conditional upon the time frame analysed. The pattern of relative loss of employment in middle-skill occupations emerged only since 2013, during years marked by a severe recession, especially in the industrial and construction sectors. As the labour market grew, we observed decreases in low-wage occupations, a constant middle and employment increases at the right of the wage distribution. Polarization may become more evident if the Brazilian economy does not go through another period of sustained growth.

Additionally, displaced routine workers move far in the occupational space and are likely to suffer productivity losses. This suggests that their skills are less and less valuable over time. In comparison, when the better off NRC workers move, they end up in similar and higher paid occupations. These workers are well protected when they age, since they exhibit the higher propensities to transition into retirement. The welfare losses of RM workers are more alarming because, on average, their pre-crisis living standards were already worse.

▶ Conclusion

We have provided evidence that skill endowments are related to the probability of a worker leaving the formal labour market and transitioning to a different occupation or sector. Workers' skills also seem to play a fundamental role in explaining the distance in terms of tasks between two occupations and related wage changes. By and large, we document that NRC workers are in the most privileged position in all the analysed areas. In particular, their skills protect them relatively well against job losses, even during an economic recession. Our estimations also reveal large heterogeneity among workers of the same skill type. Studies that focus more on differences between sectors and education, etc., are likely to uncover further subtle repercussions of workers' skills.

Our assessment of the Brazilian labour market covers two decades of profound changes and economic crisis. We aim at comprehensively documenting the relationship between skills and employment transitions in this period. There is still a large relevant research agenda for low and middle-income countries, where the lack of knowledge about such channels is even more severe. Future research can improve upon differentiating the diverse mechanisms by which different shocks – related to technology, outsourcing or demand – impact the labour market in developing countries.

In line with previous findings, the data indicate that (routine) manual workers are heavily impacted by economic downturns, even in the long run. To minimize such impacts, efforts to stimulate human capital accumulation for this population group are highly recommended. Studies based on the task-approach found that training programmes have the potential to increase workers' performance on non-routine analytical and soft skills (Tamm 2018; Görlitz and Tamm 2016). The impact of those programmes on employment or wages, however, does not seem as promising. On the one hand, international evidence shows that success is not trivial and depends on the details of the design and implementation of these programmes. On the other hand, training measures tend to be more successful in low- and middle-income countries (Kluve et al. 2017), particularly when firms are involved in defining the content of the courses (Petterini 2016; O'Connell et al. 2017).

Lastly, we document evidence that supports the RBTC and the fundamental changes that computerization and automation technologies bring to the labour market. These technologies displace middle-skill workers, intensive in routine activities, in favour of both low-skill jobs not worth automating and high-skill workers that are complemented by automation. Studies for automation in the Brazilian labour market consider that at least 45 per cent of employment, most of it intensive in low-skill manual occupations, is susceptible to automation in the near future (Albuquerque et al. 2019; Kubota and Maciente 2019; Adamczyk 2021). However, at least until now, automation technologies remain prospects, not showing clear trends of employment destruction and massive technological unemployment, as referred to by Autor (2015) or Willcocks (2020).

Annex

A.1 Task measurement

The skill requirements by occupation are estimated by means of a novel technique suited to the Brazilian labour market. Previous studies for Brazil (Ehrl and Monasterio 2019; Arellano-Bover 2020; among others), calculated skills by occupation based on a mapping between Brazilian and United States occupations and the task scores from the O*NET. Kubota and Maciente (2019) avoided CBO-O*NET mapping and discrete occupational categories. The authors have built an ad hoc dictionary of keywords associated with automation, based on the classification of activities according to their routine requirements and cognition as proposed for Germany in Spitz-Oener (2006).

Acemoglu and Autor (2011, p. 1163) is our starting point for converting O*NET task requirements to operational measurements. Their classification includes the element "4.A.3.a.3 Controlling Machines and Processes (not including computers or vehicles)" in the RM task measure. There are two major problems with replicating this in our study: (a) "4.A.3.a.3 Controlling Machines and Processes" explicitly excludes the handling of machines and computers, but the occupations that achieve the highest importance scores are airline pilots and computer programmers.¹⁹ This obvious contradiction may have gone unnoticed by Acemoglu and Autor (2011); (b) in the Brazilian economy, RM tasks are much more physical than in the United States since its capital endowments are relatively lower. A direct application of Autor and Acemoglu's classification would not take this into account. Therefore, we chose to change their measure for the routine tasks to correct these shortcomings and limitations. Our paper makes the following change to the manual routine activities measure: instead of "Controlling Machines and Processes", we have included the element 1.A.3.a.3 Dynamic Strength. Therefore, our measure of task requirements by skill is as follows:

1. Non-routine cognitive (NRC)

- 4.A.2.a.4 Analysing data/information
- 4.A.2.b.2 Thinking creatively
- 4.A.4.a.1 Interpreting information for others

2. Routine cognitive (RC)

- 4.A.4.a.4 Establishing and maintaining personal relationships
- 4.A.4.b.4 Guiding, directing and motivating subordinates
- 4.A.4.b.5 Coaching/developing others' routine cognitive
- 4.C.3.b.7 Importance of repeating the same tasks
- 4.C.3.b.4 Importance of being exact or accurate
- 4.C.3.b.8 Structured v. unstructured work (reverse)

¹⁹ The following occupations are among those that O*Net ranks highest in this element and are related to airplanes, computers and vehicles: first place 3-2011.00 Airline Pilots, Copilots, and Flight Engineers; fourth place 51-4041.00 Machinists; thirty-sixth place 51-9162.00 Computer Numerically Controlled Tool Programmers.

3. Routine manual (RM)

4.C.3.d.3 Pace determined by speed of equipment

4.C.2.d.1.i Spend time making repetitive motions

1.A.3.a.3 Dynamic strength

4. Non-routine manual (NRM)

4.A.3.a.4 Operating vehicles, mechanized devices or equipment

4.C.2.d.1.g Spend time using hands to handle, control or feel objects tools or controls

1.A.2.a.2 Manual dexterity

1.A.1.f.1 Spatial orientation

A.2 Building a dictionary of verbs and skills

After recalculating the task measures, we selected the occupations that reached the top 10 per cent in each skill. The elements associated with each of these occupations composed the file to be translated into Portuguese. This resulted in 6,100 tasks.

Table A1 lists the top ten most frequent verbs in each skill. A visual inspection of these verb lists suggests that the procedure we created for this study makes sense. Even without an ad hoc verb list, we have obtained verbs that intuitively correspond to skills.

► **Table A1. 10 most frequent verbs in each skill (in English)**

Non routine cognitive	Routine cognitve	Non routine manual	Routine manual
develop	keep	repair	adjust
evaluate	set up	inspect	operate
prepare	determine	install	determine
maintain	check	maintain	cut
advise	realize	determine	assemble
participate	ensure	replace	secure
conduct	evaluate	secure	clean
analyze	provide	operate	maintain
determine	monitor	clean	check
get	get	perform	start

Source: elaborated by the authors.

► Table A2. A random sample of frequency of stems by skill in top_skills_verbs

skill	word_stem	n
routine_cognitive	dirig	19
routine_cognitive	projet	18
nonroutine_manual	aplic	18
nonroutine_cognitive	rastr	2
nonroutine_manual	discut	1
routine_cognitive	selet	1
routine_manual	pent	2
nonroutine_cognitive	descrev	6
routine_manual	conceb	2
nonroutine_manual	sent	1
nonroutine_cognitive	ler	10
(...)	(...)	(...)

Source: elaborated by the authors.

The same process of filtering verbs was done with respect to the CBO activities descriptions. This resulted in a list of more than 132,000 thousand verbs listed in task descriptions of all 2,487 CBO occupations. We named this file `cbo_verbs_occupations`. Table A3 presents a sample of this dataset. We selected only the verbs that both files `cbo_verbs_occupations` and `top_skills_verbs` had in common. There were 996 verbs that fulfil this restriction.

► Table A3. Random sample of frequency of stems by CBO code in `cbo_verbs_occupations`

CBO code	word_stem	number
261815	ambient	1
239430	express	2
848415	equip	4
821230	acident	1
612715	grad	1
524315	segu	1
622210	armazen	1
352320	registr	6
318120	consult	1
223204	public	2
(...)	(...)	(...)

Source: elaborated by the authors.

A.2.1 Introducing TF-RR

Our aim is to use the data from `top_skills_verbs` to evaluate the importance of each skill in each CBO occupation. A simple count of the frequency of verbs associated with each skill is not a good choice. After all, the verbs in the `top_skills_verbs` contain a count of the frequency of verbs per skill, and therefore assigning a verb to a single skill would lose information.

In general terms, our measure of skill by occupation of CBO is the result of the following matrix product.

$$Skill = W * Occup$$

Where:

- W : Matrix l skills by J verbs based on `top_skills_verbs` dataset;
- $Occup$: Matrix M verbs by N CBO occupations based on `cbo_verbs_occupations` dataset;
- $Skill$: Matrix with measure of the intensity I skills by N occupations.

In our case:

- $J = M = 996$: number of verbs j and m ;
- $l = 4$: $i \in \{rc, nrc, rm, nrm\}$;
- $N = 2487$: number of CBO occupations.

$$tf_{ij} = \frac{\text{raw count of verb } j \text{ in skill } i}{\text{raw count of verbs in skill } i}$$

$$rr_j = \sum_{i=1}^l \left(\frac{\text{raw count of verb } j \text{ in skill } i}{\text{raw count of verb } j \text{ in all skills}} \right)^2$$

$$1 / l \leq rr \leq 1$$

$$tf-rr_{ij} = tf_{ij} * rr_j$$

Our matrix W_{ij} in eq A1 is:

$$\mathbf{W} = \begin{bmatrix} tf-rr_{11} & tf-rr_{12} & \dots & tf-rr_{1j} \\ tf-rr_{21} & tf-rr_{22} & \dots & tf-rr_{2j} \\ tf-rr_{31} & tf-rr_{31} & \dots & tf-rr_{3j} \\ tf-rr_{41} & tf-rr_{41} & \dots & tf-rr_{4j} \end{bmatrix}$$

where $tf-rr_{ij}$ is the $tf-rr$ of skill i , term j . If skill i has no verb j , then $tf-rr=0$.

The elements of the *Occup* (Eq. A1) matrix are much simpler.

$$\mathbf{Occup} = \begin{bmatrix} Occup_{11} & Occup_{12} & \dots & Occup_{1N} \\ Occup_{21} & Occup_{22} & \dots & Occup_{2N} \\ \dots & \dots & \dots & \dots \\ Occup_{j1} & Occup_{j1} & \dots & Occup_{jN} \end{bmatrix}$$

where $Occup_{jn}$ is the raw count of verb j in the description of CBO occupation n .

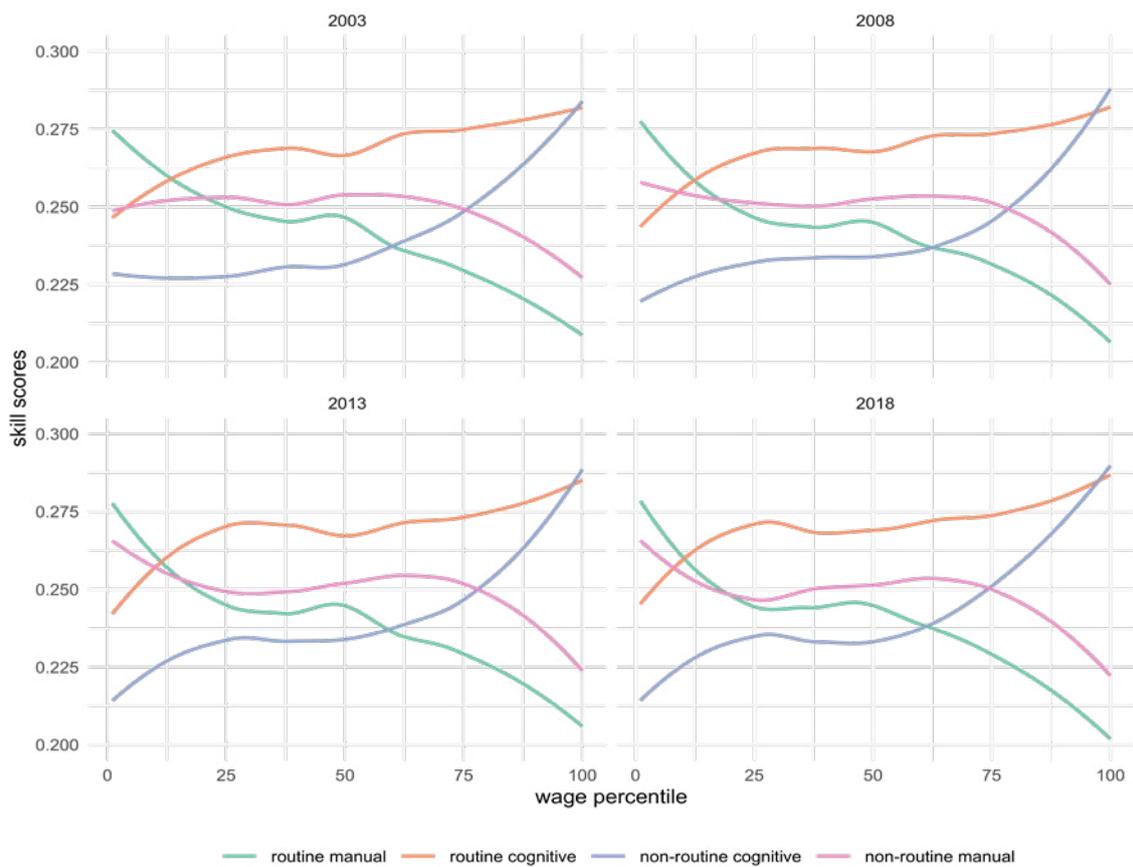
A.3 Definition of discrete skill groups

Our classification of discrete skill groups according to CBO 2002 1- and 2-digit occupations is similar to the one by Cortes et al. (2020) with minor adaptations according to the skill values from the O*NET:

1. Non-routine cognitive (NRC): managers and professionals, science technicians encompassing the 1- and 2-digit occupations [1, 2, 30-32];
2. Routine cognitive (RC): teachers, advanced technicians in administration and others, administrative services and sales [33, 35, 39, 4, 52];
3. Non-routine manual (NRM): technicians in transport services, in cultural services, service workers, agricultural producers [34, 37, 51, 61];
4. Routine manual (RM): workers in agriculture, production, craft, repair [62-64, 7-9].

Additional figures

► Figure B1. Skill scores over the wage distribution 2003–2018



Source: elaborated by the authors.

Additional tables

► **Table C1. Skill scores by skill categories**

	NRC (1)	RC (2)	NRM (3)	RM (4)
NRC	0.28 (0.02)	0.25 (0.02)	0.23 (0.02)	0.22 (0.02)
RC	0.28 (0.02)	0.29 (0.01)	0.26 (0.01)	0.25 (0.02)
NRM	0.24 (0.02)	0.24 (0.01)	0.26 (0.02)	0.26 (0.02)
RM	0.21 (0.02)	0.22 (0.02)	0.24 (0.02)	0.27 (0.03)

Note: the cells show the mean and standard deviation (in parenthesis) for the skill scores (rows) over the entire observation period 2003–2018 for the skill categories (columns) non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM), and routine manual (RM). Calculations and based on the entire sample and apply weights for employment spell duration and hours worked.

Source: elaborated by the authors.

► **Table C2. Transitions Matrix between skill groups by subperiod 2003–2018**

		2008					
		NRC	RC	NRM	RM	OFLM	retir.
2003		(1)	(2)	(3)	(4)	(5)	(6)
	NRC	51.4	13.6	2.5	5.1	25.1	2.3
	RC	8.8	55.2	4.2	5.8	24.5	1.4
	NRM	2.6	8.9	46.3	11.5	29.0	1.7
	RM	2.2	5.3	3.9	60.7	26.7	1.2
	OFLM	11.4	31.0	20.5	36.1		0.9
		2013					
		NRC	RC	NRM	RM	OFLM	retir.
2008							
	NRC	54.2	11.2	2.1	4.5	25.6	2.5
	RC	9.4	53.3	4.8	5.8	25.1	1.6
	NRM	2.6	8.9	48.7	10.3	27.9	1.6
	RM	2.4	5.4	4.5	58.9	27.7	1.1
	OFLM	12.4	32.3	21.9	33.0		0.3
		2018					
		NRC	RC	NRM	RM	OFLM	retir.
2013							
	NRC	52.8	9.4	2.0	2.9	30.3	2.6
	RC	7.3	51.3	4.6	3.9	31.1	1.7
	NRM	1.9	6.1	49.7	5.3	35.5	1.6
	RM	1.9	4.4	4.8	52.2	35.6	1.1
	OFLM	16.1	34.3	24.4	24.9		0.4

Note: the cells show the percentage of workers from a given skill category (non-routine cognitive (NRC), routine cognitive (RC), non-routine manual (NRM) and routine manual (RM)) in the initial year (column 1) that are either out of the formal labour market (OFLM), employed in the same or in another skill group five years later, or retired in or before that year. Per definition, the row sum is equal to 1. Calculations are based on the entire sample and apply weights for employment skill duration and hours worked.

Source: elaborated by the authors.

References

- Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings". In *Handbook of Labor Economics*, 4:1043–1171.
- Acemoglu, Daron, and Pascual Restrepo. 2017. "Robots and Jobs: Evidence from US Labor Markets". National Bureau of Economic Research Working Paper No. 23285. <https://doi.org/10.3386/w23285>
- Adamczyk, Willian Boschetti. 2021. "Ensaio Sobre as Tecnologias de Automação No Mercado de Trabalho Brasileiro". PhD thesis, Escola de Negócios - Pontifícia Universidade Católica do Rio Grande do Sul.
- Adamczyk, Willian Boschetti, Leonardo Monasterio, and Adelar Fochezatto. 2021. "Automation in the Future of Public Sector Employment: The Case of Brazilian Federal Government". *Technology in Society* 67: 101722.
- Aggarwal, Charu C. 2015. *Data Mining: The Textbook*. New York: Springer.
- Aggarwal, Charu C, Alexander Hinneburg, and Daniel A Keim. 2001. "On the Surprising Behavior of Distance Metrics in High Dimensional Space". In *International Conference on Database Theory*, 420–34. New York: Springer.
- Albuquerque, Pedro Henrique Melo, Cayan Atreio Portela Bárcena Saavedra, Rafael Lima de Moraes, and Yaohao Peng. 2019. "The Robot from Ipanema Goes Working: Estimating the Probability of Jobs Automation in Brazil". *Latin American Business Review* 20 (3): 227–48.
- Almeida, Rita, Carlos Henrique Corseuil, and Jennifer Poole. 2017. "The Impact of Digital Technologies on Routine Tasks: Do Labor Policies Matter?" SSRN Scholarly Paper 6560. World Bank Policy Research Working Paper.
- Almeida, Rubiane Daniele Cardoso, Philipp Ehrl, and Tito Belchior Silva Moreira. 2021. "Social and Economic Convergence Across Brazilian States Between 1990 and 2010." *Social Indicators Research* 157: 225–46.
- Arellano-Bover, Jaime. 2020. "The Effect of Labor Market Conditions at Entry on Workers' Long-Term Skills". *Review of Economics and Statistics*, 1–45.
- Ariza, John, and Josep Lluís Raymond Bara. 2020. "Technological Change and Employment in Brazil, Colombia and Mexico: Which Workers Are Most Affected?" *International Labour Review* 159 (2): 137–59.
- Arntz, Melanie, Terry Gregory, and Ulrich Zierahn. 2017. "Revisiting the risk of automation". *Economics Letters* 159: 157-160.
- Autor, David H. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation". *Journal of Economic Perspectives* 29 (3): 3–30.
- . 2019. "Work of the Past, Work of the Future". *AEA Papers and Proceedings* 109 (May): 1–32.
- Autor, David H, Lawrence F Katz, and Melissa S Kearney. 2008. "Trends in US Wage Inequality: Revising the Revisionists". *Review of Economics and Statistics* 90 (2): 300–323.
- Autor, David H, Frank Levy, and Richard J Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration". *The Quarterly Journal of Economics* 118 (4): 1279–1333.

Bachmann, Ronald, Merve Cim, and Colin Green. 2019. "Long-Run Patterns of Labour Market Polarization: Evidence from German Micro Data". *British Journal of Industrial Relations* 57 (2): 350–76.

Bernanke, Ben S. 2003. "The Jobless Recovery". Global Economic and Investment Outlook Conference. Carnegie Mellon University, Pittsburgh, Pennsylvania. European Central Bank Working Paper, v. 1192, n. 4, pp 1674-707.

Carrillo-Tudela, Carlos, Bart Hobijn, Powen She, and Ludo Visschers. 2016. "The Extent and Cyclicalities of Career Changes: Evidence for the UK". *European Economic Review* 84 (May): 18–41.

Comissão Nacional de Classificação - CONCLA. 2019. "Classificação Brasileira de Ocupações – CBO". Instituto Brasileiro de Geografia e Estatística, Available at: <https://concla.ibge.gov.br/classificacoes/por-tema/ocupacao/classificacao-brasileira-de-ocupacoes.html>.

Cortes, Guido Matias. 2016. "Where Have the Middle-Wage Workers Gone? A Study of Polarization Using Panel Data". *Journal of Labor Economics* 34 (1): 63–105.

Cortes, Guido Matias, Nir Jaimovich, Christopher J Nekarda, and Henry E Siu. 2020. "The Dynamics of Disappearing Routine Jobs: A Flows Approach". *Labour Economics* 65: 101823.

Detoni, Otávio Florentino, Ricardo Freguglia, and Carlos Henrique Corseuil. 2020. "Prêmio Salarial Associado Às Competências Dos Trabalhadores No Brasil: Uma Análise Com Dados Em Painel (2003-2013)". In 48 Encontro Nacional de Economia. ANPEC.

Dix-Carneiro, Rafael. 2014. "Trade Liberalization and Labor Market Dynamics". *Econometrica* 82 (3): 825–85.

Dix-Carneiro, Rafael, Pinelopi K Goldberg, Costas Meghir, and Gabriel Ulyssea. 2021. "Trade and Informality in the Presence of Labor Market Frictions and Regulations". Working Paper 28391. Working Paper Series. National Bureau of Economic Research.

Dix-Carneiro, Rafael, and Brian K. Kovak. 2019. "Margins of labor market adjustment to trade". *Journal of International Economics* 117 (C): 125–42.

Ehrl, Philipp. 2018. "Task Trade and Employment Patterns: The Offshoring and Onshoring of Brazilian Firms". *Journal of International Trade & Economic Development* 27 (3): 235–66.

Ehrl, Philipp, and Leonardo Monasterio. 2021. "Spatial Skill Concentration Agglomeration Economies". *Journal of Regional Science* 61 (1): 140–61.

———. 2019. "Skill Concentration and Persistence in Brazil". *Regional Studies* 53: 1544–54.

Ferreira, Francisco H. G., Sergio P. Firpo, and Julian Messina. 2017. "Ageing Poorly? : Accounting for the Decline in Earnings Inequality in Brazil, 1995-2012". World Bank Working Paper. <https://doi.org/10.1596/1813-9450-8018>.

Ferreira, Francisco H. G., Phillippe G. Leite, and Julie A. Litchfield. 2008. "The Rise and Fall of Brazilian Inequality: 1981–2004". *Macroeconomic Dynamics* 12 (S2): 199–230. <https://doi.org/10.1017/S1365100507070137>.

Firpo, Sergio, and Alysson Portella. 2019. *Decline In Wage Inequality In Brazil: A Survey*. World Bank, Washington, DC. <https://doi.org/10.1596/1813-9450-9096>.

Frey, Carl Benedikt, and Michael A Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114: 254–80.

Gathmann, Christina, and Uta Schönberg. 2010. "How General Is Human Capital? A Task-Based Approach". *Journal of Labor Economics* 28 (1): 1–49.

Gonzaga, Gustavo, and Tomás Guanziroli. 2019. "Returns to Experience Across Tasks: Evidence from Brazil". *Applied Economics Letters* 26 (20): 1718–23.

Good, I. J. 1982. "Diversity as a Concept and Its Measurement: Comment". *Journal of the American Statistical Association* 77 (379): 561–63.

Goos, Maarten. 2018. "The Impact of Technological Progress on Labour Markets: Policy Challenges". *Oxford Review of Economic Policy* 34 (3): 362–75.

Goos, Maarten, and Alan Manning. 2007. "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain". *The Review of Economics and Statistics* 89 (1): 118–33.

Goos, Maarten, Alan Manning, and Anna Salomons. 2014. "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring". *American Economic Review* 104 (8): 2509–26.

Görlitz, Katja, and Marcus Tamm. 2016. "The Returns to Voucher-Financed Training on Wages, Employment and Job Tasks". *Economics of Education Review* 52: 51–62.

Graetz, Georg, and Guy Michaels. 2017. "Is Modern Technology Responsible for Jobless Recoveries?" *American Economic Review* 107 (5): 168–73.

Guriev, Sergei. 2018. "Economic Drivers of Populism". *AEA Papers and Proceedings* 108 (May): 200–203.

Herdeiro, Renato., Naércio Menezes-Filho, and Bruno Komatsu. 2019. "Explicando a Evolução Dos Salários Relativos Por Grupos de Qualificação No Brasil". In *47 Encontro Nacional de Economia*. ANPEC.

Hershbein, Brad, and Lisa B. Kahn. 2018. "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings". *American Economic Review* 108 (7): 1737–72.

Hirschman, Albert O. 1964. "The Paternity of an Index". *The American Economic Review* 54 (5): 761–62.

Holland, Alisha C., and Ben Ross Schneider. 2017. "Easy and Hard Redistribution: The Political Economy of Welfare States in Latin America". *Perspectives on Politics* 15 (4): 988–1006.

Hopewell, Kristen. 2016. "The Accidental Agro-Power: Constructing Comparative Advantage in Brazil". *New Political Economy* 21 (6): 536–54.

IBGE, Instituto Brasileiro de Geografia e Estatística. 2018. "Projeções Da População Do Brasil e Unidades Da Federação Por Sexo e Idade – 2010-2060".

———. 2019a. "Censo Agropecuário de 2017".

———. 2019b. "Sistema de Contas Nacionais. Tabela 6 – Produto Interno Bruto, Produto Interno Bruto Per Capita, População Residente e Deflator – 1996-2018".

IFI, Instituição Fiscal Independente. 2018. "Relatório de Acompanhamento Fiscal, Nº 35, Dezembro de 2019".

Jaimovich, Nir, and Henry E. Siu. 2020. "Job Polarization and Jobless Recoveries". *The Review of Economics and Statistics* 102 (1): 129–47.

Kluge, Jochen, Susana Puerto, David Robalino, Jose Manuel Romero, Friederike Rother, Jonathan Stoeterau, Felix Weidenkaff, and Marc Witte. 2017. "Interventions to Improve the Labour Market Outcomes of Youth: A Systematic Review of Training, Entrepreneurship Promotion, Employment Services and Subsidized Employment Interventions". *Campbell Systematic Reviews* 13 (1): 1–288.

Kubota, Luis Claudio, and Aguinaldo Nogueira Maciente. 2019. "Propensão à Automação Das Tarefas Ocupacionais No Brasil". *Boletim Radar* 61 (December).

Lancaster, Tony. 1979. "Econometric Methods for the Duration of Unemployment". *Econometrica*, 939–56.

Lazear, Edward P., and James R. Spletzer. 2012. "Hiring, Churn, and the Business Cycle". *American Economic Review* 102 (3): 575–79.

Maciente, Aguinaldo Nogueira. 2013. "The Determinants of Agglomeration in Brazil: Input-Output, Labor and Knowledge Externalities". Ph.D. Thesis, Department of Agricultural; Consumer Economics; University of Illinois at Urbana-Champaign.

Maciente, Aguinaldo Nogueira, Cristiane Vianna Rauen, and Luis Claudio Kubota. 2019. "Tecnologias Digitais, Habilidades Ocupacionais e Emprego Formal No Brasil Entre 2003 e 2017". *Boletim Mercado de Trabalho - Conjuntura e Análise* 66 (April).

Maloney, William F., and Carlos Molina. 2019. *Is Automation Labor-Displacing in the Developing Countries, Too? Robots, Polarization, and Jobs*. World Bank.

Messina, Julian, and Joana Silva. 2021. "Twenty Years of Wage Inequality in Latin America". *The World Bank Economic Review* 35 (1): 117–47.

Neves Jr., Edivaldo C., Carlos R. Azzoni, and Andre Chagas. 2017. "Skill Wage Premium and City Size". 2017_19. University of São Paulo (FEA-USP).

O'Connell, Stephen D, Lucas Ferreira Mation, Joao Bevilacqua Teixeira Basto, and Mark A Dutz. 2017. "Can Business Input Improve the Effectiveness of Worker Training? Evidence from Brazil's Pronatec-MDIC". 8155. Evidence from Brazil's Pronatec-MDIC (July 31, 2017). World Bank Policy Research Working Paper.

Petterini, Francis Carlo. 2016. "Uma Avaliação Econômica Do Plano Setorial de Qualificação (PLANSEQ)". *Economia Aplicada* 20 (3): 173–94.

Qaiser, Shahzad, and Ramsha Ali. 2018. "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents". *International Journal of Computer Applications* 181 (1): 25–29.

Ramos, Juan. 2003. "Using Tf-Idf to Determine Word Relevance in Document Queries". In *Proceedings of the First Instructional Conference on Machine Learning*, 242:29–48. 1. ICML.

Rocha, Leandro Pereira da, Valéria Lúcia Pero, and Carlos Henrique Corseuil. 2019. "Turnover, Learning by Doing, and the Dynamics of Productivity in Brazil". *Economia* 20 (3): 191–210.

Ross, Matthew B. 2017. "Routine-Biased Technical Change: Panel Evidence of Task Orientation and Wage Effects". *Labour Economics* 48 (October): 198–214.

Rousseau, Ronald. 2018. "The Repeat Rate: From Hirschman to Stirling". *Scientometrics* 116 (1): 645–53.

Samaniego de la Parra, Brenda, and Christian. 2021. "Estimating Labour Market Transitions from Labour Force Surveys the Case of Viet Nam", ILO Working Paper.

Simpson, Edward H. 1949. "Measurement of Diversity". *Nature* 163 (4148): 688–88.

Spitz-Oener, Alexandra. 2006. "Technical Change, Job Tasks, and Rising Educational Demands: Looking Outside the Wage Structure". *Journal of Labor Economics* 24 (2): 235–70.

Sulzbach, Vanessa Neumann. 2020. "Essays on Job Polarization in the Brazilian Labor Market". Ph.D. thesis, UFRGS.

Tamm, Marcus. 2018. "Training and Changes in Job Tasks". *Economics of Education Review* 67 (December): 137–47.

Ulyssea, Gabriel. 2020. "Informality: Causes and Consequences for Development". *Annual Review of Economics* 12 (1): 525–46.

Wiczer, David G. 2015. "Long-Term Unemployment: Attached and Mismatched?" Federal Reserve Bank of St. Louis. Working Paper 2015-042A.

Willcocks, Leslie. 2020. "Robo-Apocalypse Cancelled? Reframing the Automation and Future of Work Debate". *Journal of Information Technology* 35 (4): 286–302.

Acknowledgements

The authors thank Janine Berg, Verónica Escudero, Hannah Liepmann, Aguinaldo Maciente, two anonymous referees and seminar participants at the UCB and ILO for valuable comments and suggestions. Financial support by the ILO is gratefully acknowledged.

The responsibility for opinions expressed in this article rests solely with its authors, and publication does not constitute an endorsement by the International Labour Office of the opinions expressed in it.

► Advancing social justice, promoting decent work

The International Labour Organization is the United Nations agency for the world of work. We bring together governments, employers and workers to improve the working lives of all people, driving a human-centred approach to the future of work through employment creation, rights at work, social protection and social dialogue.

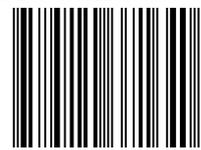
Contact details

Research Department (RESEARCH)

International Labour Organization
Route des Morillons 4
1211 Geneva 22
Switzerland
T +41 22 799 6530
research@ilo.org
www.ilo.org/research



ISBN 9789220371640



9 789220 371640