WORKING PAPERS

2025-08

08/04/2025

Hukou-Based Discrimination, Dialects and City Characteristics

Thomas Vendryes and Jiaqi Zhan

CEPS
Centre for Economics at Paris-Saclay

Hukou-Based Discrimination, Dialects and City Characteristics*

Thomas Vendryes[†], Jiaqi Zhan[‡]

Abstract

The *hukou* system is one of the most specific as well as consequential institutional features of contemporary China. Linking Chinese citizens' rights with official - and hard to change - status and place of residence, it has far-reaching social and economic implications, especially on internal migration. The consequences of the *hukou* have been a subject of unabated debate, especially as for the discrimination rural migrant workers might face in cities. In this paper, we rely on a series of CHIP (China Household Income Project) surveys from 2007 to 2018, to contribute to this debate by investigating the roles of two sets of factors that have been generally disregarded by the literature so far: at the individual level, the role of the dialect distance, between a migrant's origin and destination areas, and, at the macro level, the influence of destination city's characteristics, such as population, GDP and FDI. Results show that a sizeable part of the *hukou*-related wage gap can be explained by our dialect distance variable, and that the *hukou*-related wage gap also highly depends on destination cities' characteristics.

Keywords: Labor Markets, Wage Discrimination, Rural-Urban Migrants, Hukou, China.

JEL Classification: J31, J61, J71, 015, P23, R23.

*We are very grateful to the participants in the 2022 CES Conference for their comments and suggestions, in particular to Fang Zheng (Singapore University of Social Sciences) for his careful reading of our paper and his insightful discussion.

[†]Université Paris-Saclay, ENS Paris-Saclay, Centre for Economics at Paris-Saclay (CEPS), 91190 Gif-sur-Yvette, France. Email: Thomas.Vendryes@ens-paris-saclay.fr.

[‡]Université Paris 1 Panthéon-Sorbonne, Sorbonne School of Economics, France. Email: jiaqizhan4@gmail.com.

1 Introduction

1.1 The *Hukou* System: History, Functioning and Consequences

The "*hukou*" system is one of the most defining and consequential features of today's China institutional order. Set up in 1958 in the People's Republic to register and control Chinese citizens,[1] the *hukou* defined for each citizen a status ("urban" or "rural")[2] and a place of residence. The status determined what kind of resources were available to the related citizen, and the location where these resources were accessible. During the heyday of the Soviet-style planned and collectivist economy, it was practically impossible for a citizen to live and work outside her *hukou*-registered place of residence, and the *hukou* constituted a critical tool to control and allocate people and resources throughout China's economy and society. With the "reform and opening" era that started under Deng Xiaoping's leadership in 1978, and its dynamics of international opening, privatisation and liberalisation, the functioning and consequences of the *hukou* system have deeply changed: the distinction between the two statuses have gradually diminished, and in some places even suppressed altogether, and Chinese citizens have gradually gained more freedom to live and work outside of their place of official residence. But that does not mean that the *hukou* system is not consequential anymore (Chan and Buckingham, 2008): one's officially registered *hukou* place of residence remains hard to change, especially to get the local *hukou* of attractive localities - like important and dynamic cities - and a vast array of social and economic rights remain tied to one's official, *hukou* registered, place of residence.

Thus, up to today, the *hukou* system still creates constraints on migration and migrants in China. By limiting the rights of migrants in destination areas, it generates migration costs, affecting the geographical reallocation of resources throughout China, and affecting its overall dynamics of development.[3] But some of the most significant consequences of the *hukou* system are in terms of inequalities, segregation and discrimination. Indeed, Chinese citizens with different *hukou* do not have access to the same resources, and this constitutes one critical aspect of the phenomemon

[1] Excellent accounts of the *hukou* system's history, functioning and consequences can be found for example in Wang (2005); Young (2013), and in the works of Kam Wing Chan (e.g. Chan, 2013, 2015; He *et al.*, 2015; Chan *et al.*, 2018; Chan, 2018; Chan and Yang, 2020).

[2] Actually, the two statuses are "agricultural" ("*nongye*") and "non-agricultural" ("*feinongye*"), but we will follow the common usage in the existing literature and use instead the "rural"/"urban" terms.

[3] A series of papers have investigated the consequences of the *hukou*-related migration constraints on the overall dynamics and geographical patterns of China's growth - see for example Whalley and Zhang (2007); Vendryes (2011); Bosker *et al.* (2012); Fields and Song (2013); Zhang (2018); Fan (2019); Tombe and Zhu (2019); Zi (2020).

of "inequality of opportunity" in China (Dai and Li, 2021; Yang *et al.*, 2021; Yu and Liu, 2022). The situation is especially dire for the "*nongmingong*", the rural-to-urban migrant workers. These people live and work in urban areas, but without the local urban *hukou* status. Also called the "floating population" ("*liudong renkou*") because of their lack of stability in destination areas, they constitute one of the mainstays of the Chinese dynamics of growth,[4] their cheap and expendable laborforce having fueled for several decades the development of export-orientated industries. According to national statistics (Chan, 2013, 2021a,b), they represent a population of 376 millions people as of 2020 (up from 155 millions in 2010 - and 20-30 millions around 1990), more than a quarter of the total Chinese population, and more than 40% of the total urban population. Despite their numerical and economic importance, these rural migrant workers do not enjoy the rights, benefits and opportunities linked with a local urban *hukou*: they remain secondary class citizens in their destination cities, even if they have been living there for years (Solinger, 1999; Chan and O'Brien, 2019; Vortherms, 2021). Indeed, most of public services and welfare benefits such as children education, medicare, housing subsidies and social security coverage, are attached to a person's *hukou* officially registered place of residence, and not to their actual physical location. Therefore, urban residents generally have better access for example to education, jobs and housing, especially in big cities like Shanghai, Shenzhen and Beijing. And a urban *hukou* in these big cities, where the wages are higher and urban benefits better, is more difficult to get than in small cities. Rural migrants then often face segregation and discrimination, in many aspects of their daily and working lives (see for example Zhang *et al.*, 2014; Chan and Wei, 2019; Hung, 2022), with far-reaching consequences on many aspects of their lives and perspectives - migration decisions of course (Meng, 2020), but also residential choices at destination (Zhu, 2016a), education of their children (Chang *et al.*, 2019; Chan and Ren, 2020; Yue *et al.*, 2020), to name but a few. Migrants also face persistent prejudice and lack of trust from urban residents (Afridi *et al.*, 2015; Tse, 2016; Luo *et al.*, 2019; Luo and Wang, 2020). All of these factors have eventually a strong negative impact on their subjective well-being and mental health (Bonnefond and Mabrouk, 2019; Zhang and Awaworyi Churchill, 2020; Song and Smith, 2021; Xu *et al.*, 2022).

But from an economics point of view, one of the most important aspects of this discrimination, that has received a very large attention from the literature, is the issue of the discrimination rural migrants face in urban labour markets because of their *hukou* status - such as lower wages, delayed

[4] Imbert *et al.* (2022) provide with a very recent assessment of the consequences of rural-to-urban migration in China on urban production.

payment or wage arrears, absence of written contracts, long working hours, and inadequate social security coverage (Li, 2010). And this is exactly the question our research aims to contribute to: how does the *hukou* status affect the wage gap between urban residents, *ie* local urban *hukou* holders on the one hand, and rural migrants, who do not benefit from the local *hukou* status, on the other? And within this very large and general topic, we aim to investigate in particular two dimensions of heterogeneity that have received little attention in the literature so far: first the difference and distance in terms of dialects, between migrants' origin regions and destination localities, and second the characteristics of destination areas, in terms of GDP, population and FDI.

1.2 Related Literature

The literature on the wage gap, in Chinese urban areas, between rural migrants and official urban residents, is extremely vast, and it is beyond the scope of this article's introduction to propose an exhaustive literature review.[5] This literature revolves around the issue of the discrimination rural migrants face in urban areas, with two main questions: what is the size of the wage gap between these rural migrants and urban residents, and what are its factors and mechanisms? Answers to this latter question aim in particular to measure what is the part of the wage gap that can be explained, and maybe justified, by observable factors like age, experience, etc., and the remaining part, unexplained, that can be linked to pure labour market discrimination, and thus could be considered as unfair.

Several methods have been used in the literature:[6] the Oaxaca (1973) - Blinder (1973) decomposition approach of course (e.g. Alam and He, 2022), but also the wage distribution decomposition method (e.g. Zhu, 2016b) and the model of Brown *et al.* (1980) (e.g. Meng and Zhang, 2001). As for the data, studies on rural-to-urban migrants in China have been impaired by the difficulties to consistently survey them, as they constitute by nature an unofficial and "floating" population, touching on issues that might be politically sensitive. However, starting in the late 1990s, surveys started to be carried out to investigate this population, for example in Shanghai (Meng and Zhang, 2001; Chen and Hoy, 2008), Tianjin (Lu and Song, 2006) and more generally in large Chinese cities (Wang *et al.*, 2015). More general, and potentially more representative surveys of this migrant population, and of their rural and urban counterparts, for the sake of comparison, have

[5]Recent and extensive literature reviews can be found, for example, in the works of Song (2014); Zhang *et al.* (2016); Ma (2018); Cai and Zhang (2021); Alam and He (2022). Liu and Xu (2021) also provide with a meta-analysis of the extensive literature on the *hukou*-related wage gap.

[6]See in particular Ma (2018), who offer a synthetic description of these methods and of the related researches.

become increasingly available and used since the years 2000 (e.g. the China Urban Labour Survey (Lee, 2012), the China 1% Population Survey (Gagnon *et al.*, 2014), the China Family Panel Studies (Abu Bakkar, 2020), the China Employer-Employee Survey (Cheng *et al.*, 2020), and the China Migrants Dynamic Survey (Cai and Zimmermann, 2020; Cai and Zhang, 2021; Chen and Hu, 2021)). But the most widely used series of data have been the CHIP (China Household Income Project) surveys.[7] Carried out regularly since 1988, these surveys were specifically designed to measure and study inequalities in China, their extent and their factors, with a special attention toward migrants, and as such they have been used by a series of papers to measure the wage gap faced by rural-to-urban migrants (Liu, 2005; Démurger *et al.*, 2009; Messinis, 2013; Zhang *et al.*, 2016; Zhu, 2016b; Dreger and Zhang, 2017; Wen *et al.*, 2021; Alam and He, 2022). Based on these various datasets, the objectives of these different researches have been to measure the wage gap between rural migrants and urban residents, and to measure how much of this wage gap can be explained by observable characteristics, and how much remains unexplained, and can be linked to a pure discrimination effect related to the *hukou* status. Besides the usual individual-level characteristics such as age, gender, education, experience etc., other variables have been taken into account, such as the types of industries (monopolistic, competitive, others - see Yue *et al.* (2011)) or firms' ownership (Zhang *et al.*, 2016).

One general caveat in this literature is the selection issue: rural migrant workers self-select into migration, and into migration destinations, jobs, etc. This selection problem has been adressed to some extent in previous articles, using methodologies such as the Heckman two-step method (e.g. Zhang *et al.*, 2016) or the approach of Maddala (1983) (e.g. Ma, 2018). But generally speaking these methods cannot solve for all selection issues - and first and foremost the issue of self-selection of rural workers into migration to urban areas. This probably downwardly biases estimates of *hukou*-related labour discrimination, as the potentially more discriminated rural workers are more likely not to migrate, and as rural-to-urban migrants in all likelihood self-select into destinations, jobs, firms, etc. where they face less discrimination - thus reducing the observed wage gap.

1.3 Position and Contribution

As for our own research presented in this article, we will use a series of CHIP surveys, from 2007 to 2018, to investigate this issue of the extent and sources of the wage gap between rural migrants and

[7] http://www.ciidbnu.org/chip/index.asp?lang=EN

urban residents. In terms of methodology, we will provide with some evidence based on the Oaxaca-Blinder approach, but we will mainly rely on a very simple measure of *hukou*-related discrimination, as done for example by Cheng *et al.* (2020), using a wage equation *à la* Mincer (1974) where a dummy for the *hukou* status is included - and which coefficient should give an indication about the extent of the *hukou*-related wage gap. In this wage equation, we will include the explanatory variables that have been commonly taken into account in the most recent literature. And as done by the existing literature, we will deal as much as we can with the selection issue with the Heckman procedure.

Our contribution is to include two sets of variables, two dimensions of heterogeneity that have received (too) little attention in the literature so far. The first one is at the individual level, and relates to the distance between a migrant's hometown dialect, and the dialect of the destination urban area.[8] We find that this distance plays an important role in the *hukou*-related wage gap: once taken into account, the residual wage gap associated with the *hukou* dummy decreases significantly. The second dimension of heterogeneity is a series of characteristics a the city-level - like GDP, population, FDI, both in levels and growth rates. Indeed, different cities, with different socio-economic structures and dynamics, might display significant differences in terms of labour market functionings, and/or in terms of social and cultural openness towards migrants, affecting the labour markets outcomes of these migrants.[9] And once again, we find significant results, showing that, generally speaking, the *hukou*-related wage gap is wider in bigger and more dynamic urban areas.

1.4 Paper Outline

The rest of the paper is organised as follows: the next section, section 2 presents the overall methodology and the data, and introduces our research hypotheses, section 3 discusses the results, and section 4 concludes.

[8]A handful of recent papers have tackled issues related to the socio-economic consequences of language and dialects in China (Gao and Smyth, 2011; Chen *et al.*, 2014; Lu *et al.*, 2019; Liu *et al.*, 2020; Zhao *et al.*, 2022) - they will be discussed in more details below, in section 2.1.3.

[9]As for the first dimension of heterogeneity that we investigate, some recent papers have also focused on the heterogeneity, at the city-level, of the constraints and consequences generated by the *hukou* system - in terms of local policies (Chan, 2021a; Wang *et al.*, 2021a; Colas and Ge, 2019), urbanisation externalities (Combes *et al.*, 2015, 2017, 2020) or prejudice (Tse, 2016; Wang *et al.*, 2021b). These papers will also be discussed in more details below, in section 2.1.4.

2 Methodology, Research Hypotheses and Data

2.1 Methodology and Research Hypotheses

In this paper, we thus aim to measure how much of the wage gap between rural-to-urban migrants and urban residents can be explained by their difference in their *hukou* status (local urban *hukou* for the latter, non-local *hukou* for the former), and how this *hukou*-related wage gap depends on two dimensions of heterogeneity that have just begun to receive some attention in this literature: the distance in terms of dialects between a migrants' origin and destination areas, and destination cities' characteristics. To do this, we will rely on two methodologies widely used in the literature - and that are detailed below, respectively in sections 2.1.1 and 2.1.2.. We will first use the Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973), to provide some descriptive evidence on the discrimination faced by rural-to-urban migrants. We will then turn to a simple measure of the *hukou*-based wage gap through the inclusion of a *hukou* dummy in a wage equation *à la* Mincer (1974), that will allow us to study the two dimensions of heterogeneity that we aim to investigate here. The methodological discussion in this section will also allow us to clarify our research hypotheses.

2.1.1 The Oaxaca-Blinder Approach

We will thus first provide evidence based on the Oaxaca-Blinder decomposition method (Oaxaca, 1973; Blinder, 1973)[10], used for example by Alam and He (2022); Liu and Kawata (2022); Zhang *et al.* (2016). The basic idea of this approach is to decompose difference in an outcome - here, the labour income - between two different populations - here, rural-to-urban migrants and urban residents - into two general components: a part that can be "explained" by differences in observable characteristics of the two observed populations, and a part that remains "unexplained" by such differences, and can only be rationalised by a difference in terms of returns to these characteristics, that is to say in terms of discrimination. The former dimension is usually referred to the "composition effect", the latter as the "wage structure effect" (see for example Zhang *et al.*, 2016). Formally, the estimated equation is of the following general form:

$$Y_U - Y_M = \hat{\beta_U}(\bar{X}_U - \bar{X}_M) + (\hat{\beta_U} - \hat{\beta_M})\bar{X}_M \tag{1}$$

[10]See also Jann (2008) for a very practical discussion of the implementation of this approach in Stata.

Where U and M denote the two populations considered - here respectively urban residents and migrants, $Y_U - Y_M$ the difference in the outcome of interest - here the wage gap on urban labour markets, $\bar{X}_U$ and $\bar{X}_M$ the observable characteristics of the two populations that might affect their labour incomes, and finally $\hat{\beta_U}$ and $\hat{\beta_M}$ the estimated returns to these characteristics for respectively urban residents and migrants. Note here that the reference or non-discriminatory returns to observable characteristics are set to be the ones of urban residents, $\hat{\beta_U}$, whereas other choices could be made, affecting the respective importance of the "explained" "composition effect" and of the "unexplained" "wage structure effect".[11] More complex models could also be used, for example the approach of Brown *et al.* (1980) which explicitly takes into account the potential differences between the two considered populations in terms of allocation into different occupations and/or sectors - as done by Ma (2018) in the case of China's urban labour markets. But we have deliberately chosen here to stick to the most simple Oaxaca-Blinder approach, as our objective is mainly to use this methodology and related results for illustrative purposes.

2.1.2 Measuring the *Hukou*-Related Wage Gap

The contribution we want to bring on this issue of *hukou*-related discrimination on Chinese urban labour markets is to investigate the influence on the migrant *vs* urban workers wage gap of two dimensions of heterogeneity: at the individual level, distance in terms of dialect, and at the city level, development and growth (in terms of population, GDP or FDI). To do this, we need a simple measure of the role of the *hukou* status on the wage gap, and see how this measure varies depending on our heterogeneity variables. Formally, we thus start with a very simple labour income equation *à la* Mincer (1974), augmented of a variable indicating the *hukou* status, like:

$$Y_{itc} = \alpha_c + \beta Hukou_{itc} + \gamma X_{itc} + \epsilon_{itc} \tag{2}$$

Where i denotes the observed worker, t the time (year) and c the city. Y is the outcome of interest, *ie* the labour income, α_c the city fixed effect, X_{itc} a set of individual characteristics that are likely to affect labour market outcomes, and ϵ_{itc} is the residual. What we are interested in here is β, the coefficient associated with a dummy indicating the *hukou* status of the worker under consideration. We will use a variable coded as 0 for a local urban resident, formally registered locally in terms of *hukou* status, and 1 for a migrant, without the local *hukou* status. This coefficient $\hat{\beta}$ should

[11] See for example Jann (2008); Alam and He (2022); Zhang *et al.* (2016) for a discussion of this issue.

thus give a simple estimate of the depth of the *hukou*-related wage gap, once taken into account all variables that might affect individual wages. This simple methodology is for example used by Cheng *et al.* (2020).

2.1.3 First Dimension of Heterogeneity: Dialect Distance

We will mainly rely on this simple approach because it allows to take into account very straightforwardly the two dimensions of heterogeneity that we investigate in this research by combining related variables with the *hukou* dummy. Formally, as for the first dimension of heterogeneity, at the individual level, the distance in terms of dialects, the equation we estimate is:

$$Y_{itc} = \alpha_c + \beta_1 Hukou_{itc} \times DialectDistance_{itc} + \beta_2 Hukou_{itc} + \beta_3 DialectDistance_{itc} + \gamma X_{itc} + \epsilon_{itc} \quad (3)$$

Where $DialectDistance_{itc}$ denotes our measure of the distance in terms of dialect for migrant i staying in city c at times t (our measure and the related data are discussed below, in section 2.2.2).

Here, we echo a handful of papers that have investigated the role of language in the behaviours and outcomes of Chinese internal migrants. Indeed, if the impacts of language, cultural distance and social networks and inclusion have been widely studied in the field of international migration studies,[12] only a handful of papers have looked into this issue in China. In a recent paper, Liu *et al.* (2020) find that language (and more precisely dialect distance, as in our own research) does indeed play a role in migration decisions of Chinese rural workers, especially because of its consequences in terms of social integration. This is in line with the results of Chen *et al.* (2014), who find that migrants in Shanghai whose dialects are closer to the Shanghaiese one perform better in the local labour markets, and this is due to mechanisms linked with social identity and integration. In a related paper, Gao and Smyth (2011) show that there are significant economic returns to speaking standard Mandarin for Chinese migrants. A couple of articles (Lu *et al.*, 2019; Zhao *et al.*, 2022) have also shown that dialect distance tends to reduce migrants' access to local health services, negatively affecting their health status - the main causal mechanims being, once again, social capital, cultural identity and social integration. This is also related to the recent literature using experimental economics methods to identify how the *hukou* status and its salience affect individual behaviours, especially in terms of trust and cooperation: experiments show in particular

[12] See for example the literature review of Chiswick and Miller (2015) on the relationship between language and international migration, from an economics point of view.

that when the non-local *hukou* status is salient, local urban residents tend to display less trust and to be less cooperative toward migrants (Afridi *et al.*, 2015; Mobius *et al.*, 2016; Luo *et al.*, 2019; Dulleck *et al.*, 2020; Luo and Wang, 2020).

We thus expect the coefficient $\hat{\beta_1}$ in equation (3) to be negative: for all the aforementioned reasons, a larger dialect distance should increase the *hukou*-related wage gap.

2.1.4 Second Dimension of Heterogeneity: City Characteristics (GDP, Population, FDI)

As for the second dimension of heterogeneity, at the city level, we will take into account the levels and growth rates of variables such as population, GDP or FDI, and estimate the following equation:

$$Y_{itc} = \alpha_c + \beta_1 Hukou_{itc} \times CityVariables_{itc} + \beta_2 Hukou_{itc} + \beta_3 CityVariables_{itc} + \gamma X_{itc} + \epsilon_{itc} \quad (4)$$

Where $CityVariables_{itc}$ denotes our set of variables at the city level (levels and growth rates of population, GDP and FDI - data and variables are detailed below, in section 2.2.3).

There are several channels through which these city characteristics might affect the intensity of the discrimination faced by non-local *hukou* holders in Chinese urban labour markets.

Firstly, changes in the national laws and regulations governing the *hukou* system over the couple of decades have made it more flexible, with in particular a dynamics of devolution of the definition and administration of *hukou* regulations to local authorities (Chan and Buckingham, 2008). Chinese cities have gained much freedom and responsibilities in determining and implementing local *hukou* rules, in particular in setting conditions to get a local *hukou* registration for migrants, and access to social services and public resources for people without a local *hukou*. As a consequence, there has a been a growing heterogeneity in terms of local *hukou* institutions (Chan and Buckingham, 2008; Zhang *et al.*, 2019), of labour market functioning (Wang *et al.*, 2021a) and of *hukou*-related discrimination (Chen and Hu, 2021). And this variation in local *hukou* policies seem to be linked with local characteristics - fiscal capacity for example (Zhang and Li, 2016). In particular, Colas and Ge (2019) show "that larger and more developed cities are more attractive to migrants but tend to set more stringent *hukou* restrictions".

Secondly, a series of papers (Combes *et al.*, 2015, 2017, 2020) have shown that the labour markets gains stemming from agglomeration economies and migration externalities are not equally shared in Chinese cities - and that the extent and distribution of these gains depend on city-level characteristics. In particular, urban residents benefit from strong positive externalities from the presence of migrants (Combes *et al.*, 2017), and they also benefit more than migrants from local agglomeration economies (Combes *et al.*, 2020). Migrants in larger, more developed and more dynamic cities are thus potentially likely to face a larger *hukou*-related wage gap, as they benefit less from these urbanisation externalities than their urban residents counterparts.

Thirdly, it is also possible that a city's level and dynamic of development might affect the behaviors of its residents towards migrants. Indeed, it has been shown in the case of international migration (see for example Mayda, 2006) that the attitudes of native individuals toward migrants depend on individual and country-level characteristics. Even if this issue has not been fully investigated in China, similar mechanisms might be at play. Tse (2016) and Wang *et al.* (2021b) have recently shown that the prejudice of urban residents toward migrants depend on individual as well as local characteristics: for example, richer and more educated urban residents display more negative prejudice (Tse, 2016), and migrant inflows tend to reduce prejudice in small towns, whereas the reverse is true in bigger cities (Wang *et al.*, 2021b). These social integration mechanisms, linked with local characteristics, might well play a role in the functioning of local labour markets for migrants, and thus in the wage gap with urban residents.

For all these reasons, we expect different cities to display different levels of "urban inclusiveness" - as Qiu and Zhao (2019) put it - to affect *hukou*-based discrimination and the related wage gap. We remain somewhat agnostic about the impacts of the specific variables we will use (levels and growth rates of GDP per capita, total population and FDI per capita), but the mechanisms and the literature presented above would tend to make us expect that more developed and dynamic cities will display higher levels of *hukou*-based discrimination. We might thus expect that $\hat{\beta}_1$ in equation (4) to be negative.

2.1.5 The Selection Issue

When comparing rural migrants to urban residents, we will face an issue that is pervasive in the existing literature: the selection bias, especially as for the migrant population. In particular, relative to their urban counterparts, migrants in a given area have self-selected into migration, and into migration to a specific location. The migrants' observations we will use are thus far from a random sample, or even a sample representative of the pool of potential migrants among rural workers. As previous papers in this literature, we cannot solve this general selection issue, but we will at least try to tackle another source of selection bias, in our case selection into work (rather than inactivity) for migrants as well as urban residents, using the Heckman (1976) two-stage procedure.[13] This is far from allowing us to claim that we bypassed selection issues, and that constitutes one of the main caveats when interpreting our results. In particular, if migrants self-select into migration, into different migration destinations, into different sectors and occupations etc. based on their expected labour incomes, it is very likely that the estimates we will find of the *hukou*-related discrimination on labour markets will underestimate its true extent, as migrants very probably favor places and occupations where they face relatively less discrimination.

2.2 Data and Variables

2.2.1 The China Household Income Project (CHIP) Surveys

Our first source of data is a series of waves of the China Household Income Project (CHIP) surveys, for the years 2007, 2008, and 2018. These surveys are the result of a joint effort by Chinese and international institutions and scholars, especially the Beijing Normal University (BNU) and the Australian National University (ANU) with the support of China's National Bureau of Statistics (NBS), to provide micro-level information on Chinese households, with a special focus on income and income distribution.[14] The data are repeated cross-section surveys, which started in 1988, with following waves in 1995, 2002, 2007, 2008, 2013 and 2018. They first covered rural and urban households, in their officially *hukou* registered place of residence, and a rural-to-urban migrant sample was added in 2002. Since then, CHIP surveys have been covering three groups of households: rural households (RHS), urban households (UHS) and rural-urban households (MHS). This coverage, combined with the detailed information gathered at the individual and household levels, have made of the CHIP surveys a prime resource to study the dynamics at play at the micro level during the

[13] Previous papers in this literature have faced the same issue, and taken different approaches. For example, Cheng *et al.* (2020) acknowledge that they cannot fully account for the selection issue. Other papers rely on the Heckman (1976) two-stage procedure, or on related methodologies: Chen and Hu (2021) to deal with the self-selection of migrants into different cities accoring to their entrepreneurial propensity, Ma (2018) to account for the self-selection of migrants into different industries, and Zhang *et al.* (2016) to tackle the specific occupation of family business assistant.

[14] A more detailed presentation of the project, and the data sets themselves, can be found online at the following address: `http://www.ciidbnu.org/chip/index.asp?lang=EN`. Gustafsson *et al.* (2008) and Sicular *et al.* (2020) also present these series of surveys, and provide with an overview of the researches that have been carried out with the various CHIPS surveys waves, and of the related results.

course of China's development, especially in terms of income distribution and inequalities,[15] and the inclusion of the rural-to-urban migrant samples also made it a primary resource for the study of migration, and more specifically *hukou*-related discrimination in urban labour markets.[16] We will also rely on these data for our own research here.

More precisely, we will use the 2007, 2008 and 2018 waves. The first two waves, in 1988 and 1995, did not cover rural-urban migrant, so we cannot use them for the issue at stake here. As for the 2002 and 2013 waves, the sample of rural-urban migrants was chosen among migrants officially registered in destination urban areas, and did not cover the population we are especially interested in, that is to say the "floating population" workers, who do not have a formal status in their destination areas.[17] In the following table of descriptive statistics, Table 1, we show statistics for the CHIP waves from 2007 to 2018, including the 2013 one to illustrate the differences in sampling and thus the characteristics of the rural-urban population for this specific year - making clear why we have chosen not to use it. As we are interested in the wage gap between rural migrants and and urban residents, we will use the urban households and rural-urban households parts of the surveys, not the rural one. To somewhat alleviate selection issues, we will drop from our sample people who have changed their *hukou* registered place of residence - such a change being difficult and rare in most urban areas, and concerning very specific people. Finally, as we are interested in the wage gap between workers in urban areas, we will focus on workers from 16 to 64 years old who earn a positive wage.

Our dependent variable is taken from these series of dataset, and, following most of the literature on this issue (see for example Zhang *et al.*, 2016), is the natural logarithm of the hourly wage of a given worker. We will include in our estimations usual individual characteristics that should play role in labour earnings,[18] such as gender, age (and age squared), experience (and experience squared), years of schooling, marriage status, and ethnicity. Following Yue *et al.* (2011), we also include a variable controlling for the category of industry, between monopoly, competitive and other industries. And following Zhang *et al.* (2016), we also include a variable for the types of ownership of enterprises - state-owned, collective-owned, private-owned, foreign-owned, and individual-owned

[15]The related literature is far too large to be mentioned here, but researches using at least the first waves of the CHIP surveys can be for example found at `https://www.icpsr.umich.edu/web/DSDR/series/243/publications`.

[16]As mentioned in the Introduction, related papers using the CHIP data include for example Démurger *et al.* (2009); Messinis (2013); Zhang *et al.* (2016); Zhu (2016b); Dreger and Zhang (2017); Wen *et al.* (2021); Alam and He (2022).

[17]See Ma (2018, footnote 12).

[18]A description of the different variables is provided in Table 9 in the Appendix.

- and we re-code types of occupations into three categories : white-collar, blue-collar, and services.

Descriptive statistics for all these variables can be found in Table 1 , for each year individually, and for all years together (excluding 2013 - see below). As can be seen, for the waves 2007, 2008 and 2018, we have several thousands observations for urban as well migrant workers. As for 2007 and 2008, these observations are distributed in 15 cities in 9 different provinces. As for 2018, observations come from 126 cities in 15 provinces.

As can be seen from Table 1, and without surprise as it echoes existing literature on the same issue, migrant workers, when compared with their urban counterparts, are on average paid less. They also tend to be younger, less likely to be married, and to have less years of education and experience. As for gender, men tend to be over-represented in the migrants' population in 2007 and 2008, but the reverse is true in 2013, and the difference is almost insignificant in 2018. In terms of ethnicity, both migrants and urban residents are overwhelmingly Han. As for the employment related variables - occupation, industry and ownership distribution - differences are manifest: rural migrants are less likely to be white collar, to enter monopoly industry and stated owned companies. They tend to be concentrated in occupations, industries and sectors with low entry barriers - but also less protection and lower wages. This same Table 1 also shows how the 2013 wave stands out, and why we choose not to use it: the migrants sample is much smaller, less than a third than in other waves, with a much smaller difference in terms of labour incomes with their urban counterparts.

2.2.2 The Dialect Data

As mentioned above, our first aim is to contribute to the understanding of the *hukou*-related discrimination on urban labour markets by taking into account a dimension that has a received (too) little attention so far: the difference in terms of dialect between a migrant's origin and destination localities. The dialect distance data have been manually collected by the authors from the Language Atlas of China database (Lavely and Berman, 2012). To be more specific, Chinese dialects are classified into 10 supergroups based on their phonology, grammar and lexis, including Mandarin, Jin, Wu, Hui, Gan, Xiang, Min, Yue, Hakka and Pinghua. These supergroups are then divided into 20 groups, themselves further divided into 105 subgroups. By comparing the places, within this classification, of the dialects of migrants' origin and destination places, the dialect distance variable is then coded into four levels. It takes the value of zero if the dialects of migrants'

Table 1: Individual Characteristics - CHIP surveys data

	2007				**2008**			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Urban	Migrants	Difference	T-test	Urban	Migrants	Difference	T-test
Hourly wage	14.65	6.43	8.21	29.78***	13.81	6.53	7.27	25.59***
Experience	12.36	4.52	7.84	56.18***	13.78	4.51	9.26	55.65***
Schooling	12.12	9.06	3.06	60.25***	12.43	6.67	5.75	72.73***
Age	40.04	31.07	8.97	52.45***	40.38	30.64	9.73	51***
Gender	56%	59%	-3%	-3.01***	56%	61%	-5%	-4.89***
Marital status	84%	62%	22%	29.52***	83%	54%	29%	34.97***
Ethnicity	98.89%	98.23%	0.66%	3.21***	98.99%	98.66%	0.32%	1.59
Occupation								
White collar	53.49%	6.73%	46.76%	-100***	55.24%	11.6%	43.64%	-80***
Blue collar	16.04%	15.11%	0.93%	-73.8***	16.29%	18.05%	-1.76%	-55.09***
Service	23.16%	35.14%	-11.98%	-43.8***	21.71%	39.97%	-18.26%	-37.6***
Other	7.32%	43.02%	-35.7%	-46.65***	6.76%	30.39%	-23.63%	-67.6***
Industry								
Competitive	31.64%	65.03%	-33.39%	21.43**	28.92%	53.83%	-24.91 %	22.83**
Monopoly	20.70%	4.05%	16.65%	22.83**	21.77%	4.64%	17.13%	18.16***
Other	47.66%	30.65%	17.01%	50.35**	49.31%	41.53%	7.78%	33.13***
Ownership								
State owned	54.50%	9.55%	44.95%	-109.6***	58.4%	13.35 %	41.45%	-94.91**
Collective owned	6.13 %	4.31%	1.82 %	-68.53***	6.41%	4.84%	1.57%	-58.79***
Foreign owned	4.60%	5.09%	-0.49%	-43.53*	4.92%	5.48%	-0.56%	-56.67
Private owned	34.77%	81.04 %	-46.27%	-46.53***	30.27%	76.34%	-46.07%	-78.24***
N obs.	6444	6805			5847	5119		
N cities	15	15			15	15		
	2013				**2018**			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Urban	Migrants	Difference	T-test	Urban	Migrants	Difference	T-test
Hourly wage	18.47	14.53	3.93	7.28***	31.65	24.62	7.02	7.94***
Experience	12.70	7.27	5.43	18.23sym***	22.89	19.80	3.09	1.12
Schooling	12.1	9.36	2.74	31.6***	12.56	10.03	2.52	42.77***
Age	40.63	36.96	3.66	13.31***	42.30	40.17	2.12	11.67***
Gender	56.4%	51.5%	4.8%	3.64**	56.1%	56%	0.1%	0.16 **
Marital status	85.4%	78.63%	6.85%	6.98***	95.6%	95.1%	0.5%	0.92
Ethnicity	98.88%	98.42%	0.46%	2.11***	98.89%	98.76%	0.13%	1.23
Occupation								
White collar	38.26 %	18.3%	19.96 %	-17.72***	58.39%	31.33%	27.06%	-83.34***
Blue collar	19.76 %	28.1 %	-8.34 %	-21.82***	8.71%	19.42%	-10.71%	-56.34***
Service	23.21 %	42.43%	-19.22 %	-29.95***	20.67 %	31.07%	-10.4%	-66.84***
Other	18.77%	11.17 %	7.6%	-74.30**	12.23%	18.18 %	-5.95%	-96.34***
Industry								
Competitive	26.87%	57.8%	-30.93%	68.94***	23.70%	34%	-10.3%	48.64***
Monopoly	17.50%	10.63%	6.87%	98.24***	25.05%	28.41%	-3.36%	28.45***
Other	55.64%	31.57%	-25.93%	58.94***	51.24%	37.6%	13.64%	67.80***
Ownership								
State owned	50.16%	9.31%	40.85%	-34.89***	47.41%	11.82%	35.59%	-24.65***
Collective owned	5.82%	4.48%	1.34%	-37.21***	2.82%	2.12%	0.7%	-36.32***
Foreign owned	3.53%	2.28%	1.25%	-39.88***	2.48%	3.47%	-0.99%	-54.39***
Private owned	40.50%	83.93%	-43.43%	-54.19***	47.29%	82.9%	-35.61%	-59.81***
N obs.	7630	1671			6885	5474		
N cities	126	126			152	152		
	Pooled(2007,2008,2018)							
	(1)	(2)	(3)	(4)				
	Urban	Migrants	Difference	T-test				
Hourly wage	20.50	12.17	8.33	24.75***				
Experience	16.58	9.31	7.26	7.84***				
Schooling	12.37	8.67	3.70	99.42***				
Age	40.56	33.80	7.15	65.61***				
Gender	56.48%	58.82%	-2.33%	-4.51***				
Marital status	84.90%	67.15%	17.75%	40.88***				
Ethnicity	97.2%	97.4%	-0.2%	1.21				
Occupation								
White collar	55.79 %	15.89%	39.9%	-14.51***				
Blue collar	13.48%	17.33%	-3.85%	-45.36**				
Service	21.82%	35.28%	-13.46%	-23.6***				
Other	8.92%	31.51%	-22.59%	-44.21***				
Industry								
Competitive	27.96%	52.11%	-24.14%	44.88***				
Monopoly	22.59%	11.87%	10.72%	64.98***				
Other	49.45%	36.02%	13.43%	41.18***				
Ownership								
State owned	53.14%	11.38%	41.76%	-55.26***				
Collective owned	5.02%	3.77%	1.25%	-76.21***				
Foreign owned	3.94 %	4.69%	-0.75%	-85.39*				
Private owned	37.90%	80.17%	-42.27%	-43.76***				
N obs.	19176	17443						
N cities	152	152						

Source: CHIP surveys data and authors' calculations.

origin and destination places belong to the same subgroup, it takes the value of one if the dialects belong to different subgroups within the same group, it takes the value of two if the two dialects belong to different groups within the same supergroup, and finally it takes the value of three if the dialects belong to different supergroups. A higher value thus indicates a wider linguistic distance between the dialect spoken in the locality where a migrant comes from and the dialect spoken in the city where the same migrant is working. Note that we consider that all urban residents in our data are born in the locality where they are still living, so the value of this dialect distance variable is simply zero for all of them. The dialects classification, as used in this study, is presented in Table 10, in the Annex.

The resulting distribution of migrants according to our dialect distance variable is presented in Table 2. Badly enough, the CHIP surveys do not allow to get this information for the waves 2013 and 2018, and as can be seen in Table 2, there is surprisingly little variation as for this variable of dialect distance in 2007, as almost all migrants (98,58% of the sample) belong to a different linguistic supergroup from the one of the city they are working in. We do not have a satisfying explanation for this somewhat puzzling distribution in 2007, but we chose to focus then on the information available for 2008, for which there is more variation, which seems more realistic. As for 2008, then, 30.31% of migrant workers speak a dialect that belong to the same subgroup as the dialect spoken in the urban area they are working with - *ie*, their mother tongue is very close, if not completely similar, to the local dialect of the city they are in. Conversely, almost half of migrants (46.3%) natively speak a dialect that belongs to a different supergroup than the local urban dialect - probably impairing their communication and integration possibilities in the local society and labour market. The remainder of the migrants are distributed in the two intermediate categories - people speaking a dialect of a different subgroup within the same group (8.14%), or of a different group within the same supergroup (15.26%).

2.2.3 City Characteristics Data

As for the variables related to city characteristics that might influence the intensity of *hukou*-related labour discrimination, we simply rely of statistics provided by the National Bureau of Statistics, as for local GDP, population and FDI. GDP and FDI are normalized per capita, and for each of the three variables, we use both their (ln-transformed) levels and their growth rates (over the last 5 years), to investigate the possible impacts of their levels as well as their dynamics of development.

Table 2: Dialect Distance Distribution

	2007		2008	
	(1)	(2)	(3)	(4)
	Urban	Migrants	Urban	Migrants
Same subgroup	100%	0%	100%	30.31%
Same group, different subgroup	0%	0%	0%	8.14%
Same supergroup, different group	0%	1.42%	0%	15.26%
Different supergroups	0%	98.58%	0%	46.30%
N obs.	6444	6805	5847	5119
N cities	15	15	15	15

Source: CHIP surveys (2007 and 2008), Language Atlas of China (Lavely and Berman, 2012) and authors' calculations.

Note here that due to data access limitations, we do not use data at the city level, but we proxy them by the same information at the province-level. We lose in terms of precision, but we hope that, as the cities surveyed for the CHIP data are generally main cities in their respective provinces, this proxy is still satisfyingly informative. Descriptive statistics are presented in Table 3. Unsurprisingly, they show that there is an important variation in terms of levels and dynamics of development from a locality to another.

3 Results

3.1 Evidence on the *Hukou*-Related Wage Gap

To begin with, we provide two elements of descriptive evidence on the *hukou*-related wage gap in Chinese urban labour markets.

3.1.1 Oaxaca-Blinder Decomposition

Firstly, we carry out an Oaxaca-Blinder decomposition exercise on the CHIP data for the 2007, 2008, 2013 and 2018, and then all years pooled together (excluding 2013, which stands out as peculiar - see below). As discussed in section 2.1.1, we estimate the following relationship:

$$Y_U - Y_M = \hat{\beta_U}(\bar{X}_U - \bar{X}_M) + (\hat{\beta_U} - \hat{\beta_M})\bar{X}_M \tag{1}$$

Where U and M denote respectively urban residents and migrants, $Y_U - Y_M$ their wage gap, $\bar{X}_U$ and $\bar{X}_M$ relevant observable characteristics of the two population, and $\hat{\beta_U}$ and $\hat{\beta_M}$ the estimated

Table 3: Descriptive Statistics for City Characteristics

	2007				**2008**			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
(ln) GDP per capita	0.90	0.51	-0.24	1.83	1.07	0.47	-0.02	1.91
(ln) FDI per capita	2.72	1.30	-0.26	4.82	2.88	1.23	-0.11	4.92
(ln) Population	8.67	0.51	5.66	9.17	8.68	0.51	5.67	9.19
GDP per capita growth rate	1.21	0.24	0.76	1.74	1.05	0.23	0.61	1.58
FDI per capita growth rate	1.24	0.64	-0.41	9.31	1.07	0.74	-0.42	8.9
Population growth rate	8.27	4.99	1.00	16.00	8.47	4.27	1.00	15.00
N obs.	6444	6805			5847	5119		
N cities	15	15			15	15		
	2013				**2018**			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
(ln) GDP per capita	1.49	0.38	0.79	2.29	1.89	0.36	1.17	2.71
(ln) FDI per capita	2.83	1.07	0.94	5.23	3.65	1.04	2.01	5.88
(ln) Population	8.59	0.54	5.75	9.23	8.60	0.57	5.87	9.42
GDP per capita growth rate	0.80	0.22	0.44	1.29	0.48	0.09	0.23	0.75
FDI per capita growth rate	0.68	0.25	-0.71	3.13	1.65	0.76	0.49	3.85
Population growth rate	9.42	4.22	1.00	17.00	5.85	3.21	1.00	13.00
N obs.	7630	1671			6885	5474		
N cities	126	126			152	152		
	Pooled(2007,2008,2018)							
	(1)	(2)	(3)	(4)				
	Mean	Std. Dev.	Min	Max				
(ln) GDP per capita	1.29	0.63	-0.24	2.71				
(ln) FDI per capita	3.08	1.26	-0.26	5.88				
(ln) Population	8.65	0.53	5.66	9.42				
GDP per capita growth rate	0.92	0.37	0.23	1.74				
FDI per capita growth rate	1.33	0.75	-0.42	9.31				
Population growth rate	7.52	4.41	1.00	16.00				
N obs.	19176	17443						
N cities	152	152						

Source: National Bureau of Statistic (NBS) and authors' computations.

returns to these characteristics for respectively urban residents and migrants.

Results are presented in Table 4. They of course confirm that there is a *hukou*-related wage gap in Chinese urban labour markets, to the detriment of migrant workers. They also show that a good part of this wage gap can be "explained" by observable characteristics: for example, most notably, migrant workers have less education and less experience, which decreases their hourly wage when compared with their urban counterparts. Also, in line with the existing literature (in particular Yue *et al.*, 2011; Zhang *et al.*, 2016), migrant workers tend to concentrate in occupations (blue collar, services), industries (competitive ones) and firms (private owned ones), that offer less protection and lower wage when compared with their urban counterparts, who are more likely to be in white-collar occupations, in state-owned firms in monopolistic industries. But a sizeable part of this wage gap remains "unexplained", especially in 2007 and 2008, indicating the importance of discrimination issues against migrant workers in China's labour markets, and maybe that the intensity of this discrimination is decreasing, as the "unexplained" component is not significant anymore in 2018.

The 2013 CHIP wave also clearly stands out: the same Oaxaca-Blinder decomposition exercise indicates that the "unexplained" component is significantly positive - meaning that, after accounting for individual characteristics, migrant workers would earn more than their urban counterparts. The discrimination would thus play the other way around... This does not seem realistic, and this result is probably related to the very specific selection of the sample of migrant workers in 2013: as noted in section 2.2.1, for that survey, migrants were selected among the ones officially registered in destination urban areas. We thus prefer to drop the 2013 wave from our following estimations, as it seems not to cover the migrants we are especially interested in, those who do not have a formal status in their destination areas.

3.1.2 Mincerian Wage Regression Result

Secondly, as discussed in section 2.1.2, we estimate a wage equation *à la* Mincer (1974), with a variable indicating the *hukou* status:

$$Y_{itc} = \alpha_c + \beta Hukou_{itc} + \gamma X_{itc} + \epsilon_{itc} \tag{2}$$

Table 4: Oaxaca-Blinder Decomposition Results

	(1)	(2)	(3)	(4)	(5)
	2007	2008	2013	2018	Pooled(2007,2008,2018)
Explained	-0.369***	-0.369***	-0.284***	-0.253***	-0.488***
Experience	-0.100***	-0.100***	-0.067***	0.000	-0.001***
Schooling	-0.153***	-0.134***	-0.190***	-0.181***	-0.241***
Gender	0.009***	0.009**	0.003	-0.000	0.006***
Age	0.062***	0.075***	-0.002	-0.006**	-0.069***
Marital status	-0.038***	-0.047***	-0.002	-0.003	-0.039***
Ethnicity	-0.000	-0.001	-0.001	0.000	-0.000
Ownership	-0.039**	-0.043***	-0.003	-0.020**	0.020
Occupation	-0.091***	-0.125***	-0.012***	-0.043***	-0.142***
Industry	-0.018**	-0.004	-0.011	-0.001	-0.019***
Unexplained	-0.254***	-0.203***	0.089**	0.032	-0.013*
Experience	0.056**	0.070***	0.000	-0.004*	-0.002*
Schooling	-0.008	-0.493***	-0.419***	-0.450***	-0.217**
Gender	-0.058***	0.010	0.084***	0.063***	0.003
Age	0.112	-0.189**	-0.327**	-0.368***	0.238***
Marital status	-0.067	-0.002	0.021	-0.028	0.013
Ethnicity	-0.189	-0.117	0.280**	0.109	0.129
Ownership	-0.083**	-0.047***	-0.006	0.137***	-0.031
Occupation	0.173***	0.071**	0.041	-0.011	-0.103***
Industry	0.006	-0.052*	-0.018	-0.078***	-0.057*
Constant	-0.195	0.546***	0.433	0.662***	0.015
N obs.	12712	10709	8755	12338	35759

Notes: ***, **, *denote statistical significance at the 1%, 5%, 10% level.

Where i denotes the observed worker, t the time (year) and c the city. Y is the log hourly wage, α_c the city fixed effect, X_{itc} a set of individual characteristics, and ϵ_{itc} is the residual. As our *hukou* dummy is equal to one for rural migrants, and zero for urban residents, we expect $\hat{\beta}$ to be negative, and to provide with a simple measure of the *hukou*-related wage gap.

Results are presented in Table 5, for each year individually (2007, 2008, 2013, 2018) and pooled together (once again, without the year 2013, which appears peculiar). Each time, we present the results of estimating equation (2) with only our *hukou* dummy, without any control variables, then with these control variables included, and finally with also the Heckman two-step procedure to control for selection into employment.

These results first confirm the existence and the width of the wage gap. For example, the first column, for the year 2007, shows that without any control variables taken into account nor any correction for the selection issue, the coefficient associated with the *hukou* dummy in equation (2) is negative and statistically very significant. Its magnitude would imply that the hourly wage of migrants is around 46% lower than the one of their urban residents counterparts.[19] When individual controls and the Heckman correction are taken into account, this *hukou*-related wage gap very significantly decreases - which is not surprising given that the characteristics of migrants presented in Table 1 and discussed in section 2.2.1 show that most of these characteristics tend to reduce the labour income of migrant workers compared with their urban residents counterparts: indeed, they are on average younger, less educated, less experienced, and they tend to end up in occupations that pay less - blue collar or service and not white collar, in private owned firms, in competitive industries. Once all these factors taken into account, the impact of the *hukou* dummy, measuring the extent of *hukou*-based discrimination, decreases very significantly. In 2007 for example, the coefficient of the *hukou* dummy in column (3) implies that the *hukou* status of migrants decreases, on average, their hourly wage income by almost 22% - which is very significant.

The same is true in 2008, but as one can see also - and that echoes the results of our Oaxaca-decomposition exercise in the previous section (3.1.1), the results are reversed for the 2013 survey: taken into account only the controls (with the Heckman correction), results would imply that the

[19]To compute and interpret these coefficients for the *hukou* dummy, we rely on the results of Halvorsen and Palmquist (1980) and Kennedy (1981). Percentage changes g in the dependent variable of interest (in our case: hourly wage) due to a change in the dummy variable (in our case: *hukou* status) are computed with the following formulae: $g = 100 \times (\exp(\hat{\beta} - \hat{Var}(\beta)/2) - 1)$, where $\hat{\beta}$ denotes the estimated value of the coefficient of interest, and $\hat{Var}(\beta)$ its estimated variance.

hukou status of migrants would actually increase their hourly wage compared with their urban residents' counterparts. This seems highly unrealistic, and is in all likelihood due to the very specific selection of the CHIP migrants sample in 2013 (see section 2.2.1). That further confirms our choice not to use this specific CHIP wave for our present investigation. As for 2018, coefficients associated with the *hukou* dummy are consistently negative, even if it is not significant when both the individual controls and the Heckman correction are taken into account. The order of magnitude of the coefficient of the *hukou* dummy for 2018, when individual controls are taken into account (column (2)) indicates that on average migrant workers were earning 3.7% less than their urban counterparts (everything else held equal) that specific year. This is statistically significant, but much smaller that in 2007 and 2008. That might indicate that the *hukou*-related discrimination has indeed decreased over time - or that there is also a different selection of the migrants' sample. We still choose to keep the 2018 data, and results for the pooled data (using the CHIP surveys of 2007, 2008 and 2018 - but not 2013) are displayed in the last three columns of Table 5. They show that, indeed, on average over the period, migrant workers are paid less (14.3%) than their urban counterparts because of their *hukou* status.

3.2 The Role of Dialect Distance

After this preliminary set of evidence on the extent of the *hukou*-related discrimination, we then turn to the investigation of the impact of dialect distance in this wage gap, by estimating equation (3) introduced and discussed in section 2.1.3:

$$Y_{itc} = \alpha_c + \beta_1 Hukou_{itc} \times DialectDistance_{itc} + \beta_2 Hukou_{itc} + \beta_3 DialectDistance_{itc} + \gamma X_{itc} + \epsilon_{itc} \quad (3)$$

Where $DialectDistance_{itc}$ denotes our measure of the distance in terms of dialect for migrant i staying in city c at times t (see section 2.2.2 for details). Actually, for reasons discussed in section 2.2.2, we will focus on the year 2008, for which we have the relevant information, with enough variation. And then, practically, as both our independent variables of interest, $Hukou$ and $DialectDistance$ are categorical variables, our OLS estimates of equation (3) will actually compare the levels of the wage for five different categories of people: urban resident, for which the *hukou* dummy as well as the $DialectDistance$ variable are both nil, and four categories of migrants, for which the value of the *hukou* dummy is one, and with a variable $DialectDistance$ taking integer values between 0 and 3. The first category, urban resident, for which both variables are nil, are the

Table 5: OLS Regression Results

	2007			2008			2013			2018			Pooled(2007,2008,2018)		
	(1)	(2)	(3)Heckman	(1)	(2)	(3)Heckman	(1)	(2)	(3)Heckman	(1)	(2)	(3)Heckman	(1)	(2)	(3)Heckman
Hukou	-0.615***	-0.242***	-0.244***	-0.582***	-0.197***	-0.194**	-0.221***	0.082***	0.066	-0.290***	-0.038**	-0.039	-0.508***	-0.155***	-0.156***
	(0.011)	(0.016)	(0.016)	(0.012)	(0.017)	(0.081)	(0.023)	(0.024)	(0.272)	(0.016)	(0.017)	(0.348)	(0.007)	(0.010)	(0.028)
Experience		0.026***	0.027***		0.022***	0.020***		0.018***	0.018		0.013***	0.012		0.014***	0.014***
		(0.002)	(0.002)		(0.002)	(0.008)		(0.003)	(0.032)		(0.001)	(0.019)		(0.001)	(0.001)
Experience2		-0.0004***	-0.0004***		-0.0004***	-0.0004*		-0.0004**	-0.0003		-0.0003***	-0.0003**		-0.0004***	-0.0004***
		(0.0001)	(0.0001)		(0.0002)	(0.0001)		(0.0001)	(0.0001)		(0.0002)	(0.0001)		(0.0001)	(0.0001)
Schooling		0.043***	0.046***		0.018***	0.031***		0.052***	0.070		0.053***	0.052		0.032***	0.044***
		(0.002)	(0.003)		(0.001)	(0.010)		(0.003)	(0.045)		(0.003)	(0.056)		(0.001)	(0.005)
Gender		0.184***	0.204***		0.181***	0.211***		0.217***	0.235		0.220***	0.260		0.195***	0.256***
		(0.011)	(0.014)		(0.012)	(0.057)		(0.014)	(0.170)		(0.014)	(0.364)		(0.007)	(0.027)
Age		0.018***	0.019***		0.031***	0.041**		0.042***	-0.001		0.054***	0.054*		0.035***	0.044***
		(0.004)	(0.004)		(0.004)	(0.021)		(0.007)	(0.001)		(0.006)	(0.124)		(0.003)	(0.008)
Age2		-0.0001***	-0.0001***		-0.0001***	-0.001***		-0.0001***	-0.001		-0.001***	-0.001***		-0.001***	-0.001***
		(0.0002)	(0.0002)		(0.0001)	(0.0002)		(0.0001)	(0.0002)		(0.0001)	(0.0002)		(0.0001)	(0.0001)
Marital status		0.095***	0.078***		0.061***	0.026		0.060*	0.029		0.074***	0.091		0.078***	0.013
		(0.017)	(0.019)		(0.016)	(0.082)		(0.025)	(0.001)		(0.023)	(0.497)		(0.011)	(0.038)
Ethnicity		0.061	0.062		0.161***	0.160		-0.024	-0.018		0.031	0.031		0.067***	0.067
		(0.040)	(0.044)		(0.057)	(0.244)		(0.040)	(0.455)		(0.037)	(0.467)		(0.026)	(0.067)
Occupation															
Ref: White collar															
Bluecollar		-0.258***	-0.256***		-0.326***	-0.302***		-0.116***	-0.108		-0.120***	-0.119		-0.251***	-0.239***
		(0.017)	(0.018)		(0.018)	(0.083)		(0.021)	(0.248)		(0.023)	(0.309)		(0.011)	(0.031)
Service		-0.250***	-0.250***		-0.289***	-0.274***		-0.207***	-0.200		-0.252***	-0.252		-0.280***	-0.275***
		(0.017)	(0.016)		(0.016)	(0.077)		(0.022)	(0.247)		(0.020)	(0.264)		(0.010)	(0.029)
Other		-0.181***	-0.179***		-0.316***	-0.290***		-0.107***	-0.100		-0.165***	-0.165		-0.251***	-0.237***
		(0.019)	(0.018)		(0.018)	(0.090)		(0.021)	(0.246)		(0.024)	(0.300)		(0.012)	(0.032)
Ownership															
Ref: State owned															
Collective owned		-0.030	-0.029		-0.095***	-0.088		-0.121***	-0.114		-0.236***	-0.236		-0.093***	-0.086*
		(0.025)	(0.025)		(0.024)	(0.119)		(0.033)	(0.373)		(0.027)	(0.602)		(0.017)	(0.050)
Foriegn owned		0.157***	0.157***		0.066**	0.064		0.153***	0.159		0.060	0.060		0.106***	0.106**
		(0.024)	(0.026)		(0.027)	(0.126)		(0.040)	(0.472)		(0.039)	(0.065)		(0.017)	(0.052)
Private owned		-0.079***	-0.079***		-0.066**	-0.063		-0.048**	-0.046		-0.040**	-0.040		-0.071***	-0.067**
		(0.014)	(0.014)		(0.014)	(0.68)		(0.020)	(0.224)		(0.010)	(0.260)		(0.009)	(0.027)
Industry															
Ref: Competitive															
Monopoly		0.111***	0.111***		0.097***	0.093		0.171***	0.172		0.015	0.014		0.079***	0.080***
		(0.019)	(0.018)		(0.020)	(0.090)		(0.023)	(0.265)		(0.019)	(0.247)		(0.011)	(0.030)
Other		0.042***	0.042***		0.037***	0.034		0.051***	0.049		-0.021	-0.022		0.023***	0.022
		(0.013)	(0.013)		(0.013)	(0.065)		(0.019)	(0.217)		(0.018)	(0.243)		(0.008)	(0.025)
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes
R2	0.267	0.403		0.246	0.378		0.139	0.304		0.122	0.251		0.394	0.491	
N obs.	12921	12712	13085	10891	10709	10784	9238	9134	9123	12339	12338	12356	36151	35759	36225
Mill's Ratio			0.254**			2.71*			7.48			-9.912			1.96***
			(0.14)			(1.35)			(12.31)			(163.84)			(0.49)

Notes: ***, **, *denote statistical significance at the 1, 5, 10 level. Error terms are clustered at city level reference category.

Results are displayed in Table 6, and a graphical illustration of these results is proposed in Figure 1. For the sake of space and readability, we report only the estimated coefficients for our main variables of interest, *hukou* and dialect distance, and not for all other individual characteristics. In Table 6, the first three columns simply recall, for comparison, the results of estimating equation (2) on the 2008 CHIP data. The following columns show the results when estimating equation (3), that is to say when we take into account the dialect distance, on the same data. Results are quite striking: the pure *hukou*-related wage gap decreases significantly - the coefficients displayed on the very last column, when all individual characteristics and the Heckman procedure are taken into account, indicate that migrant workers who speak the same dialect as their urban counterparts earn on average 11% less, whereas migrants who speak a dialect that belongs to a different supergroup earn on average 21%. The *hukou*-related wage gap roughly doubles when one moves from one extreme to the other in terms of dialect distance. Figure 1 proposes a graphical illustration of these results. The y-axis reports the (ln) hourly wage. The blue triangle represents the level of this variable for urban workers with the local *hukou*. They are all considered as speaking the local dialect as their mother tongue - and so our variable "Dialect Distance" is set to zero for all of them. The red circles show, similarly, the levels of the wage for different groups of migrants, depending on our measure of dialect distance. For migrants who speak the same (or almost the same) dialect as their urban counterparts, the level of the (ln) hourly wage is lower - but what is visually striking is how this wage gap widens for migrants who speak a dialect from a different supergroup or group.

Table 6: *Hukou*-Related Wage Gap & Dialect Distance

	Pure Effect			**Dialect Distance**	
	(1)	(2)	(3)Heckman	(4)	(5)Heckman
Hukou=1 × Dialect Distance=0	-0.582***	-0.197***	-0.194**	-0.115***	-0.117***
	(0.012)	(0.017)	(0.081)	(0.022)	(0.042)
Hukou=1 × Dialect Distance=1				-0.100***	-0.112***
				(0.033)	(0.046)
Hukou=1 × Dialect Distance=2				-0.212***	-0.216***
				(0.025)	(0.052)
Hukou=1 × Dialect Distance=3				-0.244***	-0.239***
				(0.022)	(0.038)
Individual Characteristics	No	Yes	Yes	Yes	Yes
Occupation	No	Yes	Yes	Yes	Yes
Industry	No	Yes	Yes	Yes	Yes
Ownership	No	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Mill's Ratio			2.71*		0.098***
			(1.35)		(0.212)
R2	0.405	0.384	0.252	0.388	
N obs.	10891	10709	10784	9145	9142

Notes: ***, **, *denote statistical significance at the 1, 5, 10 level. Error terms are clustered at city level

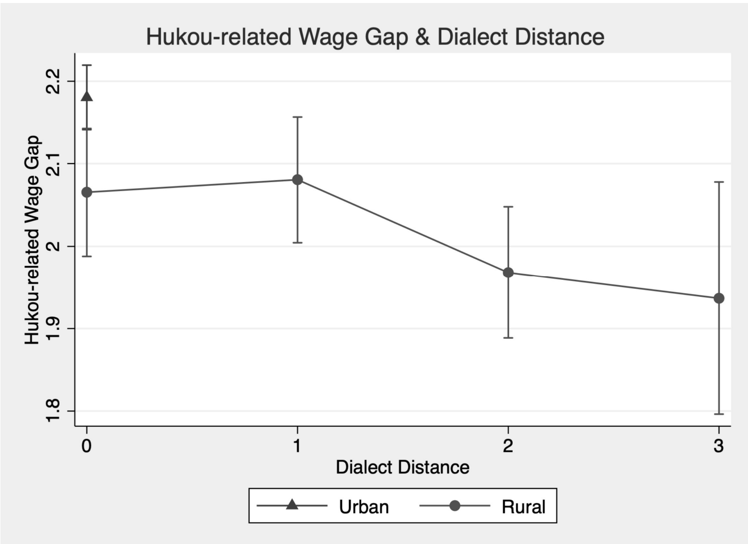

Figure 1: *Hukou*-Related Wage Gap & Dialect Distance

To give a probably too crude crude interpretation, that would mean that linguistic issues, the fact that different Chinese workers, coming from and living in different localities, do not have the same mother tongue, might explain roughly half of the wage gap that is usually associated with the *hukou* status. It thus seems quite important to account for this linguistic dimension when studying the functioning of labour markets in China - an issue that probably too few papers have investigated so far (Gao and Smyth, 2011; Chen *et al.*, 2014; Liu *et al.*, 2020).

This raises also the question of the mechanisms at play. This impact of the dialect distance on labour earnings might be due to the fact that it decreases labour productivity: a worker that is not fully familiar with the local dialect might see her ability to interact and work impaired. That would echo in some way the results of Gao and Smyth (2011) who show that standard Mandarin (or "Putonghua"), which is the national official language in China, has a very significant positive impact on migrants workers' labour earnings - on possible mechanism being that fluency in Mandarin would allow migrant workers to expand the set of occupations they can undertake at destination.

But the impact of this dialect distance on labour earnings might also be due to discrimination mechanisms, as, firstly, the fact of not being fluent in the local dialect reveals that someone is not a "local", leading to discriminatory attitudes, and, secondly, because it can also reduce the possibilities of building networks, reducing social integration in the locals' community. This is one of the main conclusions of Chen *et al.* (2014) on the specific case of Shanghai: "Speaking the local dialect is a way for migrant workers to integrate into the local society and also to reduce transaction costs in the labor market."

Thus, determining the causality mechanisms at play behind this impact of the dialect distance on labour earnings is important: does it manifest a lower productivity, or a higher discrimination - or something else altogether? Mechanisms at play, their "fairness" and the policy implications can be very different.

3.3 The Role of City Characteristics

We now turn to the second dimension of heterogeneity we are interested in, related to local development levels and dynamics, in terms of GDP per capita, population and FDI per capita. As explained in section 2.1.4, the equation we estimate is:

$$Y_{itc} = \alpha_c + \beta_1 Hukou_{itc} \times CityVariables_{itc} + \beta_2 Hukou_{itc} + \beta_3 CityVariables_{itc} + \gamma X_{itc} + \epsilon_{itc} \quad (4)$$

Where $CityVariables_{itc}$ denotes our set of variables at the local level (levels and growth rates of population, GDP and FDI). Results are presented in Tables 7 and 8, respectively for the levels and the growth rates of our dependent variables of interest (GDP per capita, population and FDI per capita), and graphical illustrations of these results are proposed respectively in Figures 2 and 3. For the sake of space and readability, we report in each table - and in related figures - only the estimated coefficients for our main variables of interest, *hukou* and city characteristics, and not for all other individual characteristics. For the same reason, we show the results per year and all years pooled (excluding 2013 for the reasons explained above), but directly with all individual controls. In the first four columns of Tables 7 and 8, the results of Table 5, without city characteristics are recalled, for comparison.

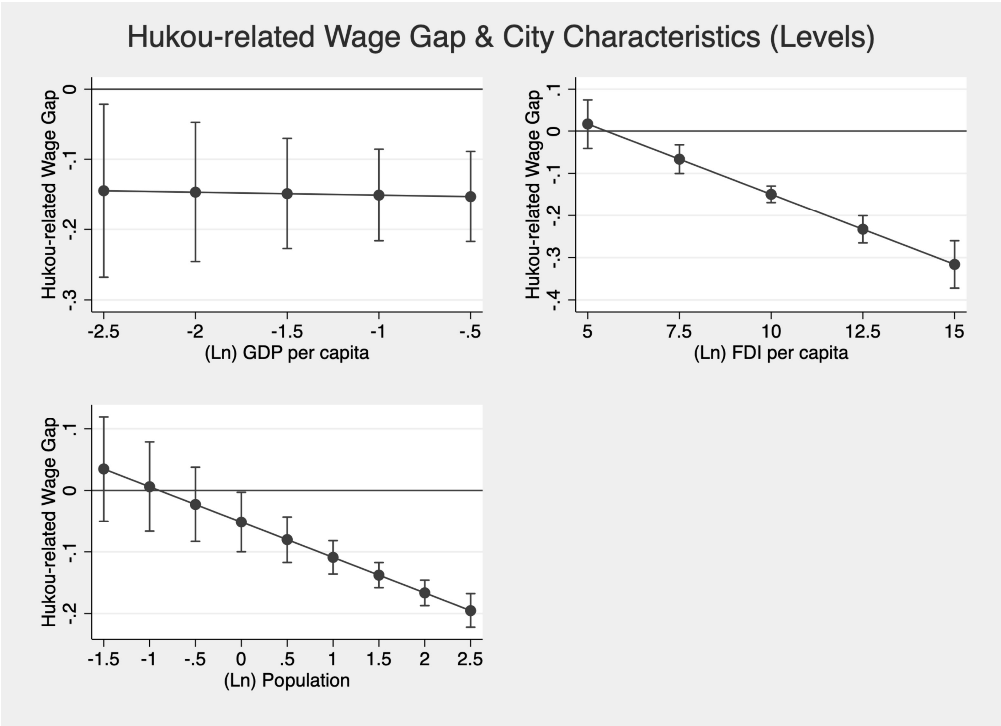

Figure 2: Discrimination Intensity & City Factors (Levels)

As can be seen from these two tables, most of the coefficients associated with the interaction between the *hukou* status dummy and our city characteristics variables are significant. In Figures 2 and 3, these results are represented graphically, based on the fourth column for each variable, *ie* with data for all years (2007, 2008, 2018) pooled together, and, in each case, for a range of values of the city characteristics that correspond to what we observe in the data. They show the estimated *hukou*-based wage gap, with all individual controls taken into account, for different levels of the city characteristics. It is then quite clear that the width of the wage gap does indeed depend on these city characteristics. In particular, it increases with population and (ln) FDI per capita

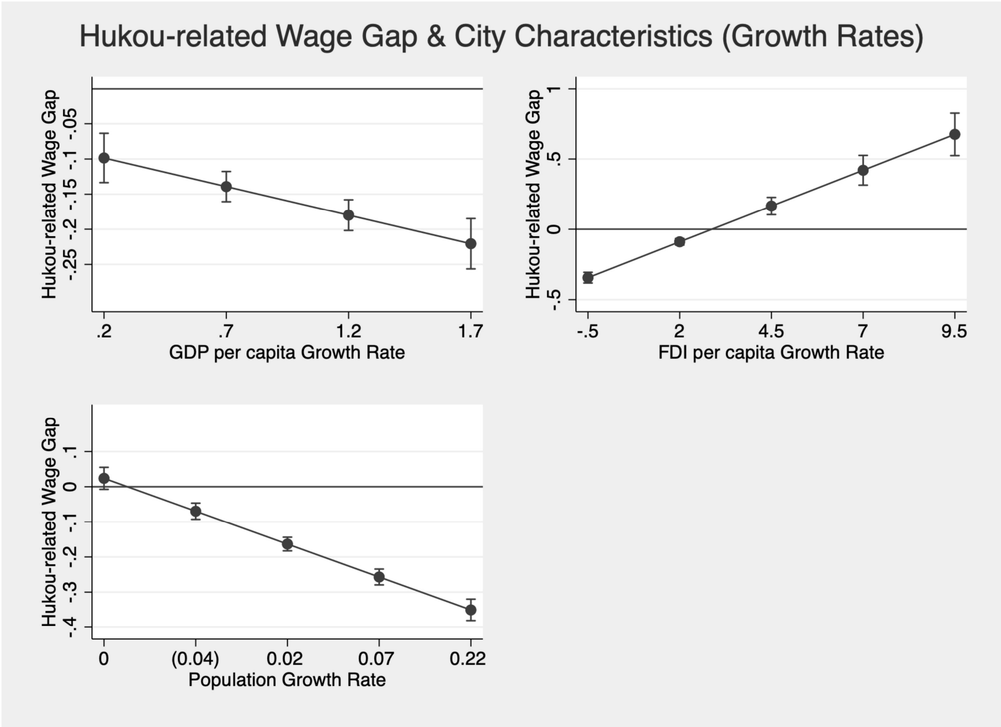

Figure 3: Discrimination Intensity & City Factors (Growth Rates)

(Figure 2). It is actually nil in places that have the lowest population and level of FDI per capita in our sample, and steadily increases afterward, to become quite large in the places that are at the top of the distribution of these variables in our sample. In a parallel way (Figure 3), this wage gap increases in places that are more dynamic in terms of population and GDP per capita. These results seem in line with the existing literature. As discussed above (section 2.1.4), more developed and dynamic localities tend to have more stringent *hukou* policies (e.g. Colas and Ge, 2019), to generate more agglomeration externalities, that disproportionately benefit urban residents (Combes *et al.*, 2020), and their residents tend to display more prejudice toward migrants (Wang *et al.*, 2021b). But our results do not allow to discriminate between these potential causality mechanisms.

Results related to the impact of the (ln) GDP per capita (Figure 2) and the growth rate of FDI per capita (Figure 3) seem harder to reconcile with these cauality mechanisms. We do not find a significant impact of the level of the (ln) GDP per capita on the *hukou*-related wage gap when we pool all years together (see column (4) for the GDP per capita variable in Table 7), and it even turns out positive when all variables are taken into account (see the last column in Table 7), whereas it is - as could be expected - negative when the impact is estimated for each year separately (see columns (1)-(3) for the GDP per capita variable in Table 7). We do not have a satisfying explanation for this lack of robustness of the results for the impact of GDP per capita. It might be at least partly due to the fact that, as mentioned above (section 2.2.3), we do not use, because of data access issues, information at the city-level, but at the province level - and this lack of precision of the information might affect the lack of precision of the results. Further investigation, with more granular data, would be necessary here.

The results for the growth rate of FDI per capita (see Table 8 and Figure 3) are also harder to rationalise: they indicate that the *hukou*-based wage gap is actually decreasing when the growth rate of FDI per capita is higher - whereas this same wage gap is increasing when the level of FDI per capita is higher, as discussed above. Once again, these results and their apparent contradiction might stem from the imprecision of our data, that are at the province and not city level. Or maybe these two pieces of evidence could be reconcile if FDI growth has been on average faster in places where the level of FDI was relatively low. But a deeper investigation would be necessary - especially to clarify how FDI affect local labour markets.

Moreover, we have intrepreted so far our results as illustrating causality mechanisms going from city characteristics to the *hukou*-related wage gap. But causality mechanisms could, at least to some extent, run the other way around: from the *hukou*-related wage gap to the level and dynamics of development of a locality. If rural-to-urban migrants are indeed one the main factors in Chinese development, in particular because they constitute a reserve of cheap labour, cities were they are more discriminated and less paid relative to urban residents could be able to attract more activities, more investment, and eventually more population. Here again, further research is needed to clarify the causality mechanisms at play - and their direction.

4 Conclusion

The aim of this research was first to provide with some evidence on the discrimination faced by internal migrants on China's urban labour markets and the related wage gap, based on data from several waves of the China Household Income Project (CHIP) from 2007 to 2018. Through a simple Oaxaca-Blinder (Oaxaca, 1973; Blinder, 1973) decomposition, and with a similarly simple Mincerian wage regression including a specific dummy for rural-to-urban migrant workers, without a local urban *hukou*, we confirm the existing - and large - literature on this issue, by showing that the hourly wage of these migrants is indeed significantly lower than the one of their urban counterparts. The *hukou* system is thus without doubt a source of discrimination on China's urban labour markets.

But beyond this evidence, our second and most important series of objectives was to contribute to the literature on this issue by investigating two dimensions of heterogeneity in the intensity of this *hukou*-based labour discrimination that have received too little attention so far: at the individual level, the language issue, and at a more aggregate level, the impact of city characteristics.

As for the first dimension, combining the CHIP surveys data with data on Chinese dialects (Lavely and Berman, 2012), we show that when we include in estimations of the *hukou*-based wage gap a categorical variable measuring the linguistic distance between the native dialect of a migrant and the dialect of the same migrant's destination area, this dialect distance variable captures a good part of the *hukou*-related wage gap: to give rough orders of magnitude, for migrants who speak the same dialect as their urban residents counterparts, the wage gap associated with their non-local *hukou* status is divided by two. The impact is thus important, and might affect our understanding of *hukou*-related discrimination and inequalities in China. Indeed, if the dialect distance decreases a migrant's productivity in a local urban labour market, from an economics point of view, the fact that her wage decreases as a consequence could be considered as normal or fair. However, if the effect of this dialect distance is not due to productivity mechanisms, but for example to social integration issues, playing against migrants, then this wage gap could be seen as unfair - and this is the direction to which point the handful of papers that have had a look into this issue (Gao and Smyth, 2011; Chen *et al.*, 2014; Liu *et al.*, 2020), echoing equally recent experimental evidence showing the influence of migrant/*hukou* status salience on migrants' and urban citizens' behaviors and interactions.[20]

As for the second dimension of heterogeneity, pertaining to city characteristics, combining the CHIP surveys data with information on the level and dynamics of development at the local level, we show that the size of the *hukou*-related wage gap, *ie* the intensity of the *hukou*-based discrimination, does indeed depend on local characteristics: for example, cities that are more populous, more exposed to FDI, and more dynamic in terms of GDP per capita and population growth rates, display wider *hukou*-related wage gap. This could be explained by the fact that local

[20] Afridi *et al.* (2015); Mobius *et al.* (2016); Luo *et al.* (2019); Dulleck *et al.* (2020); Luo and Wang (2020).

hukou policies differ from a city from another (see for example Colas and Ge, 2019), that migrants and urban residents do not benefit in the same way from local agglomeration externalities (see for example Combes *et al.*, 2020) and/or, finally, that urban residents' attitudes toward migrants vary from city to city (see for example Wang *et al.*, 2021b).

Actually, in both cases, for both dimensions of heterogeneity, the evidence we present in this paper, constrained by the information available in the data we use, is far from definitive, and in particular we cannot discriminate between the different potential causality mechanisms behind the results we eventually find. Our more modest objective from this point of view is to contribute to the opening of these axes of research, by showing that, indeed, there is heterogeneity as for the *hukou*-based wage gap, at the individual as well as local levels, and that this issue of momentous significance from a social and economic point of view would deserve more attention.

Moreover - and this also is meant to open avenues for future research, our results are imperfect, and could be made more precise and convincing. In terms of data, for example, we proxy city-level characteristics by province-level statistics. A more granular analysis would be necessary to confirm and deepen the results presented here. And, perhaps more importantly, as virtually all papers investigating this issue of the *hukou*-based discrimination, we do not fully account for selection, and especially for the selection of the rural-to-urban migrants we observe into migration, into specific migration destinations, etc. Very probably, the results we find are thus a lower bound of the discrimination Chinese migrant workers actually face on urban labour markets.

All in all, we hope to have provided in this research results confirming the *hukou*-related wage gap in China's urban labour markets, and, more importantly, showing the importance of the heterogeneity of this wage gap, especially depending on dialect distance and on city characteristics. And we hope to have contributed to the interest for and advancement of these research questions - on which further research is certainly needed!

Table 7: Discrimination Intensity & City Characteristics (Levels)

	OLS Results				GDP per capita				FDI per capita				Population				All			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	2007	2008	2018	Pooled	2007	2008	2018	Pooled	2007	2008	2018	Pooled	2007	2008	2018	Pooled	2007	2008	2018	Pooled
Hukou	-0.242***	-0.197***	-0.038**	-0.155***	-0.466***	-0.365***	-0.106***	-0.155***	0.341***	0.253***	0.619***	0.183***	-0.242***	0.060	-0.018	-0.051**	-1.182**	-1.621***	0.346	1.431***
	(0.016)	(0.017)	(0.017)	(0.010)	(0.031)	(0.032)	(0.023)	(0.015)	(0.083)	(0.088)	(0.151)	(0.056)	(0.038)	(0.040)	(0.047)	(0.025)	(0.475)	(0.516)	(0.321)	(0.146)
(ln)GDP per capita					0.358***	0.281***	0.202***	0.482***									1.002***	-3.524***	2.355	-0.152
					(0.019)	(0.021)	(0.042)	(0.065)									(0.233)	(0.849)	(1.939)	(0.147)
Hukou × (ln)GDP per capita					-0.163***	-0.136***	-0.169***	-0.004									-0.343***	-0.507***	-0.084	0.208***
					(0.020)	(0.022)	(0.039)	(0.011)									(0.092)	(0.103)	(0.079)	(0.025)
(ln)FDI per capita									0.150***	0.119***	0.094***	0.118**					-0.257**	1.661***	-1.045	0.188***
									(0.008)	(0.009)	(0.020)	(0.058)					(0.101)	(0.359)	(0.931)	(0.065)
Hukou × (ln)FDI per capita									-0.060***	-0.046***	-0.062***	-0.033***					0.062*	0.120***	-0.037	-0.126***
									(0.008)	(0.009)	(0.014)	(0.005)					(0.037)	(0.040)	(0.028)	(0.012)
(ln)Population													-0.318***	-0.183***	-1.943***	-1.299***	0.075***	-0.064	-0.730	-1.340***
													(0.021)	(0.022)	(0.473)	(0.160)	(0.021)	(0.047)	(0.521)	(0.353)
Hukou × (ln)Population													0.000	-0.145***	-0.013	-0.058***	-0.076***	-0.209***	-0.017	-0.064***
													(0.020)	(0.021)	(0.026)	(0.013)	(0.021)	(0.022)	(0.027)	(0.013)
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ownership	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.403	0.378	0.251	0.491	0.408	0.386	0.253	0.468	0.407	0.385	0.254	0.468	0.405	0.386	0.252	0.468	0.408	0.391	0.253	0.470
N obs.	12712	10709	12338	35759	12703	10726	12337	35766	12703	10726	12337	35766	12703	10726	12337	35766	12703	10726	12337	35766

Notes: ***, **, *denote statistical significance at the 1, 5, 10 level. Error terms are clustered at city level

Table 8: Discrimination Intensity & City Characteristics (Growth Rates)

	OLS Results				GDP per capita				FDI per capita				Population				All			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	2007	2008	2018	Pooled	2007	2008	2018	Pooled	2007	2008	2018	Pooled	2007	2008	2018	Pooled	2007	2008	2018	Pooled
Hukou	-0.242***	-0.197***	-0.038**	-0.155***	-0.679***	-0.455***	-0.031	-0.082***	-0.042***	-0.337***	-0.021	-0.293***	-0.054**	0.006	0.069**	0.024	-0.255**	0.379***	-0.025	0.037
	(0.016)	(0.017)	(0.017)	(0.010)	(0.054)	(0.050)	(0.081)	(0.021)	(0.024)	(0.023)	(0.036)	(0.016)	(0.026)	(0.029)	(0.032)	(0.016)	(0.105)	(0.118)	(0.083)	(0.033)
GDP per capita growth rate					-0.857***	-0.634***	1.503***	-0.307***									-0.747***	-0.782***	-3.975**	-0.214**
					(0.032)	(0.047)	(0.360)	(0.082)									(0.108)	(0.157)	(1.674)	(0.086)
Hukou × GDP per capita growth rate					0.363***	0.242***	-0.016	-0.081***									0.037	-0.361***	0.226	-0.123***
					(0.044)	(0.045)	(0.162)	(0.020)									(0.067)	(0.077)	(0.174)	(0.021)
FDI per capita growth rate									-0.036***	-0.243***	-0.315***	-0.061***					-0.124***	-0.202***	-0.857**	-0.033***
									(0.018)	(0.016)	(0.075)	(0.010)					(0.019)	(0.037)	(0.392)	(0.011)
Hukou × FDI per capita growth rate									0.14***	0.127***	-0.011	0.102***					0.080***	0.081***	-0.002	0.057***
									(0.014)	(0.014)	(0.019)	(0.009)					(0.018)	(0.016)	(0.020)	(0.010)
Population growth rate													0.042***	0.050***	0.046***	0.039***	-0.003	-0.053***	0.043	0.036***
													(0.002)	(0.003)	(0.011)	(0.004)	(0.006)	(0.019)	(0.030)	(0.004)
Hukou × Population growth rate													-0.022***	-0.023***	-0.018***	-0.023***	-0.015***	-0.033***	-0.021***	-0.021***
													(0.002)	(0.003)	(0.005)	(0.002)	(0.003)	(0.005)	(0.005)	(0.002)
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ownership	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.403	0.378	0.251	0.491	0.408	0.385	0.252	0.493	0.408	0.388	0.252	0.494	0.410	0.389	0.253	0.496	0.410	0.391	0.253	0.497
N obs.	12712	10709	12338	35759	12703	10726	12337	35766	12703	10726	12337	35766	12703	10726	12337	35766	12703	10726	12337	35766

Notes: ***, **, *denote statistical significance at the 1. 5, 10 level. Error terms are clustered at city level

A Appendix

Table 9: Variables Definition

Variable	Source	Definition
(ln)Wage	CHIP Surveys	Log form of hourly wage
Experience	CHIP Surveys	Working experience at current occupation
Experience2	CHIP Surveys	Square of Experience
Schooling	CHIP Surveys	Years of education
Age	CHIP Surveys	Age in the surveying year
Age2	CHIP Surveys	Square of Age
Gender	CHIP Surveys	0 for female and 1 for male
Marriage	CHIP Surveys	0 single and 1 for married
Ethnicity	CHIP Surveys	Ethical minority; 0 for other and 1 for Han
Occupation	CHIP Surveys	0 for white collar; 1 for blue collar; 2 for services and 3 for other
Industry	CHIP Surveys	0 for competitive; 1 for monopoly and 2 for other
Ownership	CHIP Surveys	0 for state owned ; 1 for collective owned;2 for foreign owned; 3 for private owned
Dialect distance	Language Atlas of China	Dialect similarity level between migrants' origin and destination places
(ln)GDP per capita	NBS	Log form of GDP per capita for surveyed year
(ln)FDI per capita	NBS	Log form of FDI growth rate for surveyed year
(ln)Population	NBS	Log form of total population for surveyed year
GDP per capita growth rate	NBS	GDP per capita growth rate in last five years
FDI per capita growth rate	NBS	FDI per capita growth rate in last five years
Population growth rate	NBS	Population growth rate in last five years

Table 10: Dialects Classification

Supergroup	**Group**	**Subgroup**
Mandarin	Jinlu Mandarin	Changhui; Shiji; Baotang
	Beijing Mandarin	Jingcheng
	Dongbei Mandarin	Hafu; Jishen; Heisong
	Jianghuai Mandarin	Tairu; Songcao; Huangxiao
	Zhongyuan Mandarin	Xuhuai; Shangbu; Xinbeng; Yanhe; Guanzhong; Luoxiang; Zhenkai; Luosong; Nanlu; Qinlong; Hezhou; Jincheng; Longzhong; Fenhe; Nanjiang
	Xinan Mandarin	Ebei; Huguang; Guiliu; Chuanqian; Xishu; Chuanxi; Yunnan
	Jiaoliao Mandarin	Qinglai; Denglian
	Lanyin Mandarin	Jincheng; Hexi; Yinwu; Beijiang
Gan Dialect	Huaiyue	/
	Yiliu	/
	Fuguang	/
	Yingyi	/
	Changdu	/
	Jicha	/
	Yiliu	/
	Datong	/
	Leizi	/
Hui Dialect	Yanzhou	/
	Shenzhan	/
	Jiyi	/
	xiuyi	/
	Qiwu	/
Jin Dialect	Ganxin	Cizhang
	Shangdang	Changye
	Wutai	/
	Bingzhou	/
	Dabao	/
	Zhanghu	/
	Zhiyan	/
Hakka dialect	Yuetai	Meihui; Longhua
	Yuebei	/
	Tonggui	/
	Yuxin	/
	Quannan	/
	Tingzhou	/
Min Dialect	Yuezhong	/
	Yuedong	Houguan; Funing
	Yuenan	Datian; Chaoshan; Quanzhang

	Yuebei	Jianzhong; Jianyang
	Shaojiang	Shaowu
	Leizhou	/
	Qiongwen	Wanning; Wenchang
	Puxian	/
Pinghua Dialect	Xiangnan	/
Wu Dialect	Taihu	Shanghai; Pilin; Sujiahu; Linsa; Yongjiang; Hangzhou
	Jinqu	/
	Taizhou	/
	Oujiang	/
Xiang Dialect	Dongquan	Dongqi
	Loushao	Xiangshuang; Lianmei; Xinhua; Wushao
	Changyi	Yueyang; Changzhutan; Yiyuan
	Yongquan	Daojiang
	Hengzhou	Hengyang; Hengshan
	Chenxu	/
Yue Dialect	Guangfu	/
	Gaoyang	/
	Siyi	/
	Wuhua	/
	Yongxun	/
	Qinglian	/
	Goulou	/

Source: The Language Atlas of China (2012), author's collection.

References

Abu Bakkar, S. (2020). Identity-based Earning Discrimination among Chinese People. *IZA Journal of Development and Migration*, **11** (1), 1–42.

Afridi, F., Li, S. X. and Ren, Y. (2015). Social identity and inequality: The impact of China's hukou system. *Journal of Public Economics*, **123** (C), 17–29.

Alam, M. A.-U. and He, X. (2022). Is the discrimination against migrant workers tending toward zero in urban china? *International Studies of Economics*, **17** (1), 65–81.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, **8** (4), 436–455.

Bonnefond, C. and Mabrouk, F. (2019). Subjective well-being in China: direct and indirect effects of rural-to-urban migrant status. *Review of Social Economy*, **77** (4), 442–468.

Bosker, M., Brakman, S., Garretsen, H. and Schramm, M. (2012). Relaxing Hukou: Increased labor mobility and China's economic geography. *Journal of Urban Economics*, **72** (2), 252–266.

Brown, R. S., Moon, M. and Zoloth, B. S. (1980). Incorporating Occupational Attainment in Studies of Male-Female Earnings Differentials. *Journal of Human Resources*, **15** (1), 3–28.

Cai, S. and Zhang, X. (2021). *Anatomy of the Wage Gap between Local and Migrant Workers in Urban China: New Evidence from Matched Data*. Tech. Rep. 3933758, SSRN.

— and Zimmermann, K. F. (2020). *Social Assimilation and Labor Market Outcomes of Migrants in China*. CEPR Discussion Papers 15496, C.E.P.R. Discussion Papers.

Chan, A. T. and O'Brien, K. J. (2019). Phantom services: Deflecting migrant workers in china. *The China Journal*, **81** (1), 103–122.

Chan, K. W. (2013). China: Internal migration. In *The Encyclopedia of Global Human Migration*, Blackwell Publishing Ltd Oxford, UK.

— (2015). Five decades of the chinese hukou system. In R. R. Iredale and F. Guo (eds.), *Handbook of Chinese Migration*, Edward Elgar Publishing.

— (2018). China's hukou system at 60: Continuity and reform. In R. Yep, J. Wang and T. Johnson (eds.), *Handbook on Urban Development in China*, Edward Elgar.

— (2021a). *Internal Migration in China: Integrating Migration with Urbanization Policies and Hukou Reform*. Policy Note 16, KNOMAD.

— (2021b). What the 2020 chinese census tells us about progressin hukou reform. *ChinaBrief*,

21 (15).

— and Buckingham, W. (2008). Is china abolishing the hukou system? *The China Quarterly*, **195**, 582–606.

—, Cai, F., Wan, G. and Wang, M. (2018). *Urbanization with Chinese Characteristics: The Hukou System and Migration*. Routledge.

— and Ren, Y. (2020). *Children of Migrants in China*. Routledge.

— and Wei, Y. (2019). Two systems in one country: The origin, functions, and mechanisms of the rural-urban dual system in china. *Eurasian Geography and Economics*, **60** (4), 422–454.

— and Yang, X. (2020). Internal migration and development: A perspective from china. In *Routledge Handbook of Migration and Development*, Routledge Handbook of Migration and Development, pp. 567–584.

Chang, F., Jiang, Y., Loyalka, P., Chu, J., Shi, Y., Osborn, A. and Rozelle, S. (2019). Parental migration, educational achievement, and mental health of junior high school students in rural china. *China Economic Review*.

Chen, J. and Hu, M. (2021). City-level hukou-based labor market discrimination and migrant entrepreneurship in china. *Technological and Economic Development of Economy*, **27** (5), 1095–1118.

Chen, Y. and Hoy, C. (2008). Rural migrants, urban migrants and local workers in shanghai: Segmented or competitive labour markets? *Built Environment*, **34** (4), 499–516.

Chen, Z., Lu, M. and Xu, L. (2014). Returns to dialect: Identity exposure through language in the chinese labor market. *China Economic Review*, **30**, 27–43.

Cheng, H., Hu, D. and Li, H. (2020). Wage differential between rural migrant and urban workers in the people's republic of china. *Asian Development Review*, **37** (1), 43–60.

Chiswick, B. R. and Miller, P. W. (2015). International migration and the economics of language. In B. R. Chiswick and P. W. Miller (eds.), *Handbook of the Economics of International Migration*, *Handbook of the Economics of International Migration*, vol. 1, *5*, North-Holland, pp. 211–269.

Colas, M. and Ge, S. (2019). Transformations in China's Internal Labor Migration and Hukou System. *Journal of Labor Research*, **40** (3), 296–331.

Combes, P.-P., Démurger, S. and Li, S. (2015). Migration externalities in Chinese cities. *European Economic Review*, **76** (C), 152–167.

—, — and — (2017). Productivity Gains from Agglomeration and Migration in the People's

Republic of China between 2002 and 2013. *Asian Development Review*, **34** (2), 184–200.

—, —, — and Wang, J. (2020). Unequal migration and urbanisation gains in china. *Journal of Development Economics*, **142**, 102328, special Issue on papers from "10th AFD-World Bank Development Conference held at CERDI, Clermont-Ferrand, on June 30 - July 1, 2017".

Dai, X. and Li, J. (2021). Inequality of Opportunity in China: Evidence from Pseudo Panel Data. *China Economic Review*, **68** (C).

Démurger, S., Gurgand, M., Li, S. and Yue, X. (2009). Migrants as second-class workers in urban china? a decomposition analysis. *Journal of Comparative Economics*, **37** (4), 610–628.

Dreger, C. and Zhang, Y. (2017). *The Hukou Impact on the Chinese Wage Structure*. IZA Discussion Papers 10720, Institute of Labor Economics (IZA).

Dulleck, U., Fooken, J. and He, Y. (2020). Hukou Status and Individual-Level Labor Market Discrimination: An Experiment in China. *ILR Review*, **73** (3), 628–649.

Fan, J. (2019). Internal geography, labor mobility, and the distributional impacts of trade. *American Economic Journal: Macroeconomics*, **11** (3), 252–88.

Fields, G. S. and Song, Y. (2013). *A Theoretical Model of the Chinese Labor Market*. IZA Discussion Papers 7278, Institute of Labor Economics (IZA).

Gagnon, J., Xenogiani, T. and Xing, C. (2014). Are migrants discriminated against in chinese urban labour markets? *IZA Journal of Labor & Development*, **3** (1), 1–23.

Gao, W. and Smyth, R. (2011). Economic returns to speaking 'standard mandarin' among migrants in china's urban labour market. *Economics of Education Review*, **30** (2), 342–352.

Gustafsson, B., Shi, L. and Sicular, T. (2008). *Inequality and Public Policy in China*. Cambridge University Press.

Halvorsen, R. and Palmquist, R. (1980). The Interpretation of Dummy Variables in Semilogarithmic Equations. *American Economic Review*, **70** (3), 474–475.

He, S., Li, S.-M. and Chan, K. W. (2015). Migration, communities, and segregation in chinese cities: Introducing the special issue. *Eurasian Geography and Economics*, **56** (3), 223–230.

Heckman, J. (1976). The common structure of statistical models of truncation: Sample selection and limited dependent variables and a sample estimatitor for such models. *Annals of Economic and Social Measurement*, **5**, 475–492.

Hung, J. (2022). Hukou system influencing the structural, institutional inequalities in china: The multifaceted disadvantages rural hukou holders face. *Social Sciences*, **11** (5).

Imbert, C., Seror, M., Zhang, Y. and Zylberberg, Y. (2022). Migrants and Firms: Evidence from China. *American Economic Review*, **112** (6), 1885–1914.

Jann, B. (2008). The blinder–oaxaca decomposition for linear regression models. *The Stata Journal*, **8** (4), 453–479.

Kennedy, P. E. (1981). Estimation with correctly interpreted dummy variables in semilogarithmic equations. *American Economic Review*, **71** (4), 801–801.

Lavely, W. and Berman, L. (2012). Language Atlas of China. Harvard Dataverse: `https://doi.org/10.7910/DVN/QPUONU`.

Lee, L. (2012). Decomposing wage differentials between migrant workers and urban workers in urban china's labor markets. *China Economic Review*, **23** (2), 461–470.

Li, S. (2010). The economic situation of rural migrant workers in china. *China Perspectives*, **2010** (2010/4).

Liu, Y., Jiao, Y. and Xu, X. (2020). Promoting or preventing labor migration? revisiting the role of language. *China Economic Review*, **60**, 101407.

— and Kawata, K. (2022). Labour market and wage inequalities between permanent rural migrants and urban-born residents in china. *Applied Economics*, **54** (9), 996–1010.

— and Xu, Z. (2021). *A Meta-Analysis of the Hukou-Based Wage Discrimination in China*. Working Paper 3993049, SSRN.

Liu, Z. (2005). Institution and inequality: the hukou system in china. *Journal of comparative economics*, **33** (1), 133–157.

Lu, S., Chen, S. and Wang, P. (2019). Language barriers and health status of elderly migrants: Micro-evidence from china. *China Economic Review*, **54**, 94–112.

Lu, Z. and Song, S. (2006). Rural–urban migration and wage determination: The case of tianjin, china. *China Economic Review*, **17** (3), 337–345, symposium on Inequality, Market Development, and Sources of Growth in China under Accelerating Reform.

Luo, J., Chen, Y., He, H. and Gao, G. (2019). Hukou identity and fairness in the ultimatum game. *Theory and Decision*, **87** (3), 389–420.

— and Wang, X. (2020). Hukou identity and trust—Evidence from a framed field experiment in China. *China Economic Review*, **59** (C).

Ma, X. (2018). Labor market segmentation by industry sectors and wage gaps between migrants and local urban residents in urban china. *China Economic Review*, **47**, 96–115.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press.

Mayda, A. M. (2006). Who Is Against Immigration? A Cross-Country Investigation of Individual Attitudes toward Immigrants. *The Review of Economics and Statistics*, **88** (3), 510–530.

Meng, L. (2020). Permanent migration desire of chinese rural residents: Evidence from field surveys, 2006–2015. *China Economic Review*, **61**, 101262.

Meng, X. and Zhang, J. (2001). The two-tier labor market in urban china: Occupational segregation and wage differentials between urban residents and rural migrants in shanghai. *Journal of Comparative Economics*, **29** (3), 485–504.

Messinis, G. (2013). Returns to education and urban-migrant wage differentials in china: Iv quantile treatment effects. *China Economic Review*, **26**, 39–55.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Human behavior and social institutions, National Bureau of Economic Research.

Mobius, M., Rosenblat, T. and Wang, Q. (2016). Ethnic discrimination: Evidence from China. *European Economic Review*, **90** (C), 165–177.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, **14** (3), 693–709.

Qiu, L. and Zhao, D. (2019). Urban inclusiveness and income inequality in China. *Regional Science and Urban Economics*, **74** (C), 57–64.

Sicular, T., Li, S., Yue, X. and Sato, H. (eds.) (2020). *Changing Trends in China's Inequality: Evidence, Analysis, and Prospects*. Oxford University Press.

Solinger, D. J. (1999). Citizenship issues in china's internal migration: Comparisons with germany and japan. *Political Science Quarterly*, **114** (3), 455–478.

Song, Q. and Smith, J. P. (2021). The Citizenship Advantage in Psychological Well-being: An Examination of the Hukou System in China. *Demography*, **58** (1), 165–189.

Song, Y. (2014). What should economists know about the current chinese hukou system? *China Economic Review*, **29**, 200 – 212.

Tombe, T. and Zhu, X. (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review*, **109** (5), 1843–72.

Tse, C. W. (2016). Urban residents' prejudice and integration of rural migrants into urban china. *Journal of Contemporary China*, **25** (100), 579–595.

Vendryes, T. (2011). Migration constraints and development: Hukou and capital accumulation in china. *China Economic Review*, **22** (4), 669–692.

Vortherms, S. A. (2021). Hukou as a case of multi-level citizenship. In *The Routledge Handbook* *of Chinese Citizenship*, 1st edn., Routledge, pp. 132–142.

Wang, F., Milner, C. and Scheffel, J. (2021a). Labour market reform and firm-level employment adjustment: Evidence from the hukou reform in China. *Journal of Development Economics*, **149** (C).

Wang, F.-L. (2005). *Organizing Through Division and Exclusion: China's Hukou System*. Stanford University Press.

Wang, H., Guo, F. and Cheng, Z. (2015). A distributional analysis of wage discrimination against migrant workers in china's urban labour market. *Urban Studies*, **52** (13), 2383–2403.

Wang, X., Zhang, Y., Wang, S. and Zhao, K. (2021b). Migrant Inflows and Online Expressions of Regional Prejudice in China. *Public Opinion Quarterly*, **85** (1), 123–146.

Wen, L., Paudel, K. P., Chen, Y. and He, Q. (2021). Urban segregation and consumption inequality: Does hukou conversion matter in China? *Review of Development Economics*, **25** (4), 2298–2322.

Whalley, J. and Zhang, S. (2007). A numerical simulation analysis of (Hukou) labour mobility restrictions in China. *Journal of Development Economics*, **83** (2), 392–410.

Xu, L., Wang, L. and Nygaard, C. (2022). Locational decisions and subjective well-being: an empirical study of Chinese urban migrants. *Applied Economics*, **54** (27), 3180–3195.

Yang, X., Gustafsson, B. and Sicular, T. (2021). Inequality of opportunity in household income, china 2002–2018. *China Economic Review*, **69**, 101684.

Young, J. (2013). *China's Hukou system: Markets, migrants and institutional change*. Palgrave Macmillan.

Yu, Y. and Liu, C. (2022). Pre-market discrimination or post-market discrimination: research on inequality of opportunity for labor income in china. *Empirical Economics*, pp. 1–23.

Yue, A., Bai, Y., Shi, Y., Luo, R., Rozelle, S., Medina, A. and Sylvia, S. (2020). Parental migration and early childhood development in rural china. *Demography*, pp. 1–20.

Yue, X., Li, S. and Sicular, T. (2011). High incomes in monopoly industries: A discussion. *Social Sciences in China*, **32** (2), 178–196.

Zhang, J., Wang, R. and Lu, C. (2019). A quantitative analysis of hukou reform in chinese cities: 2000-2016. *Growth and Change*, **50** (1), 201–221.

Zhang, J. X. (2018). A general equilibrium analysis of china's urbanization. *Journal of the Asia Pacific Economy*, **23** (3), 340–356.

Zhang, L. and Li, M. (2016). Local fiscal capability and liberalization of urban hukou. *Journal* *of Contemporary China*, **25** (102), 893–907.

—, Sharpe, R. V., Li, S. and Darity, W. A. (2016). Wage differentials between urban and rural-urban migrant workers in china. *China Economic Review*, **41**, 222–233.

Zhang, M., Zhu, C. J. and Nyland, C. (2014). The institution of hukou-based social exclusion: A unique institution reshaping the characteristics of contemporary urban china. *International Journal of Urban and Regional Research*, **38** (4), 1437–1457.

Zhang, Q. and Awaworyi Churchill, S. (2020). Income inequality and subjective wellbeing: Panel data evidence from china. *China Economic Review*, **60**, 101392.

Zhao, Q., Xu, S., Aziz, N., He, J. and Wang, Y. (2022). Dialect culture and the utilization of public health service by rural migrants: Insights from china. *Frontiers in Public Health*, **10**.

Zhu, P. (2016a). Residential segregation and employment outcomes of rural migrant workers in China. *Urban Studies*, **53** (8), 1635–1656.

Zhu, R. (2016b). Wage differentials between urban residents and rural migrants in urban china during 2002–2007: A distributional analysis. *China Economic Review*, **37**, 2–14, special Issue on Human Capital, Labor Markets, and Migration.

Zi, Y. (2020). *Trade Liberalization and the Great Labor Reallocation*. CEPR Discussion Papers 14490, C.E.P.R. Discussion Papers.