SOCIUM · SFB 1342
WorkingPapers No. 29

Armin Müller
John Gibson
Tobias ten Brink

Who is Covered by Unemployment Insurance in China?

De jure coverage, de facto coverage and the dynamics of development

socium
Forschungszentrum
Ungleichheit und Sozialpolitik

Globale
Entwicklungsdynamiken
von Sozialpolitik
SFB 1342

Armin Müller, John Gibson, Tobias ten Brink
Who is Covered by Unemployment Insurance in China? De jure coverage, de facto coverage and the dynamics of development
SOCIUM SFB 1342 WorkingPapers, 29
Bremen: SOCIUM, SFB 1342, 2025

Universität Bremen

SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik /
Research Center on Inequality and Social Policy
SFB 1342 Globale Entwicklungsdynamiken von Sozialpolitik /
CRC 1342 Global Dynamics of Social Policy

Postadresse / Postaddress:
Postfach 33 04 40, D - 28334 Bremen

Websites:
https://www.socium.uni-bremen.de
https://www.socialpolicydynamics.de

[DOI https://doi.org/10.26092/elib/3730]

[ISSN (Print) 2629-5733]
[ISSN (Online) 2629-5741]

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG)
Projektnummer 374666841 – SFB 1342

DFG Deutsche Forschungsgemeinschaft

Armin Müller
John Gibson
Tobias ten Brink

Who is Covered by Unemployment Insurance in China?

De jure coverage, de facto coverage and the dynamics of development

SOCIUM · SFB 1342
No. 29

Armin Müller (armmueller@constructor.university), CRC 1342 "Global Dynamics of Social Policy", Constructor University Bremen

John Gibson (john.gibson@waikato.ac.nz) University Waikato, Hamilton, New Zealand

Tobias ten Brink (ttenbrink@constructor.university), CRC 1342 "Global Dynamics of Social Policy", Constructor University Bremen

Acknowledgements
We are extremely grateful for the many helpful comments we received by members and guests of the CRC (especially Tobias Böger and Aline Grünewald) when we were given the opportunity to present this paper.

Abstract

The long period of fast economic growth in China has created entrenched patterns of dualization in the labour market, and in social security. Chinese law calls for all employment to be formalized and covered by a relatively generous system of five social insurances, including one for unemployment. However, despite stagnation in standard formal employment, the number of persons covered by unemployment insurance continued to increase in the 2010s, and the aim of this study is to explain this puzzle. The case of unemployment insurance is particularly instructive regarding dualization in the labour market, given that only employees at working age participate in the scheme. We rely on data at prefectural-city level from the National Bureau of Statistics, satellite data for night lights and urban land use, event data for labour unrest, and a corpus of regulatory documents labelled via machine learning. The analysis focuses on descriptive statistics and long-differenced regression. The results indicate that beyond classical standard employment, the expansion of unemployment insurance is driven by economic growth and cities' competition for a shrinking pool of migrant labour.

Zusammenfassung

Die lange Zeit der hohen Wirtschaftswachstumsraten in China hat zu fest verankerten Mustern der Dualisierung des Arbeitsmarktes und der sozialen Sicherheit geführt. Das chinesische Recht sieht vor, dass alle Beschäftigungsverhältnisse formalisiert und durch ein relativ großzügiges System von fünf Sozialversicherungen, darunter eine Arbeitslosenversicherung, abgedeckt sind. Trotz der Stagnation der formalen Standardbeschäftigung stieg die Zahl der in der Arbeitslosenversicherung versicherten Personen in den 2010er Jahren weiter an, und das Ziel der vorliegenden Studie ist es, dieses Rätsel zu erklären. Der Fall der Arbeitslosenversicherung ist besonders aufschlussreich in Bezug auf die Dualisierung des Arbeitsmarktes, da nur Arbeitnehmer im erwerbsfähigen Alter an diesem System teilnehmen. Wir stützen uns auf Daten des Nationalen Statistikamtes auf Präfektur- und Stadtebene, auf Satellitendaten über die nächtliche Beleuchtung und die städtische Flächennutzung, auf Ereignisdaten über Arbeitsunruhen und auf einen Korpus von mittels Machine Learning aufbereiteten Gesetzesdokumenten. Die Analyse konzentriert sich auf deskriptive Statistiken und langdifferenzierte OLS-Regressionen. Die Ergebnisse zeigen, dass die Ausweitung der Arbeitslosenversicherung über die klassische Standardbeschäftigung hinaus durch das Wirtschaftswachstum und den Wettbewerb der Städte um einen schrumpfenden Pool von Arbeitsmigranten angetrieben wird.

Contents

1. Introduction

After an unprecedented economic boom, official GDP data suggests the People's Republic of China (PRC) is transitioning from middle-income to high-income status in the mid-2020s, though the objectivity of official GDP figures has been called into question (W. Chen et al., 2019; Martínez, 2022). The boom has been the result of a comparatively successful transition from a command economy to a more market-based economic model. Like in other transition economies (Orenstein, 2001, p. 13), unemployment insurance played an important role in generating social stability and social cohesion when employment guarantees were given up and large numbers of workers in state-owned enterprises (SOEs) were dis-employed during the 1990s (Duckett & Hussain, 2008). Once the transition was completed, the PRC entered a period of quasi-full employment in the 2000s, and scholarly interest in unemployment insurance began to fade.

But the economic transition also has created entrenched patterns of dualization of employment and social security. Many former SOE workers took up jobs in the informal sector of China's cities (Park & Cai, 2011), just like most migrant workers in the emerging industrial centres in the coastal provinces (Y. Chen, 2008). Both groups remained largely excluded from the new social insurance systems for urban employees that replaced previous SOE protections (Duckett, 2001; Frazier, 2010). According to Chinese law, all employment should be contract-based, and covered by five social insurances, including pensions, healthcare and unemployment (Gallagher et al., 2015; Giles et al., 2013). However, official labour market statistics indicate that formal employment has stagnated in the 2010s (Rozelle et al., 2020), while automation of production pushed more and more ageing migrant workers out of industry, and into a service sector dominated by small companies, informal employment and expanding platform-based work (Rozelle & Hell, 2020). Simultaneously, the continuous expansion of higher education has resulted in increasing difficulties for graduates to find stable and formalized employment (Bai, 2006; Mok & Wu, 2016). The problem was further intensified by slower economic growth, the trade war since 2018, and the economic crisis in the wake of Covid 19. In the early 2020s, the PRC is facing the most severe unemployment crisis since the large-scale privatization of SOEs in the 1990s.

However, despite the stagnation of formal employment, coverage of unemployment insurance continued to rise in the 2010s (MoHRSS, 2022). This development appears puzzling, given the institutional coupling of urban social insurance coverage and standard formal employment. This study aims to explain the continuous growth of unemployment insurance coverage in the 2010s. We devise working hypotheses drawing on socio-economic theories of the welfare state and labour mobilization theory, which we test using regression analysis. In terms of data, we rely on a unique dataset of indicators for cities either at the provincial level (the first sub-national level), or the prefectural level (the second sub-national level).[1] The first set of indicators stems from the National Bureau of Statistics (NBS), and we intensively checked these data for reliability, objectivity, and validity. The second set is state-of-the-art satellite data on night lights and land usage, which we compiled for the cities in the dataset. The third set of indicators stems from labour unrest and strike events data compiled by the China Labour Bulletin in Hong Kong. It is the most comprehensive dataset available in the 2010s. Fourth, we compiled a dataset of regulatory documents issued by central, provincial, and prefectural city administrations to regulate unemployment insurance. We extracted relevant regulations via machine-learning and text mining. Finally, we use data from the China Labor-Force Dynamics Survey (CLDS) to get a differentiated analysis of coverage.

Our results indicate that, first, exactly determining *de jure* coverage is complex and some-

1 Cities at the county-level, the third sub-national level, were not considered.

what ambiguous. Beyond central regulation, various subnational levels of government enact diverging regulations pertaining to coverage. Furthermore, subnational governments can interpret the term "employee" in central regulations exclusively in the sense of the Standard Employment Relationship (SER), or more inclusively. Second, *de facto* coverage is overall highest in contract-based, fulltime employment, but features strong regional disparities connected to socio-economic development. Indicators approximating SER employment in official statistics do not include domestic private companies, where formalized employment increased in the 2010s. Quantitative analysis largely confirmed the hypotheses based on socio-economic welfare state theory: coverage increase beyond SER-employment is explained by economic growth, domestic migration, and urbanization, and largely driven by cities' competition for a shrinking pool of migrant labour.

2. Literature review and working hypotheses

This section provides some background information about Chinese unemployment insurance and generates a series of working hypotheses about coverage of unemployment insurance in Chinese cities. It reviews socio-economic theories of the welfare state and adapts them to the context of mainland China. Furthermore, it reviews labour mobilization theory.

2.1 Socio-economic development and welfare state expansion

Approaches linking socio-economic development to the formation and expansion of the welfare state revolve around two central hypotheses: economic growth and development are strong predictors of welfare state development, and countries with comparable development levels will exhibit a certain degree of policy convergence (Flora & Alber, 1981; Obinger, 2019; Wilensky, 1975). Building up on a pioneering study of welfare expenditures at provincial level (Hong & Ngok, 2022), this study applies these principal ideas to sub-national jurisdictions in the PRC – a country of continental dimensions with stark gaps in socio-economic development and strong administrative decentralization.

The causal chains linking economic development and welfare state development are long and complex: technological progress changes the methods of production, which in turn has transformative effects on social structure and family relations. Rising life expectancy and lower fertility rates are among the outcomes of this transformation which call for further social policy intervention. In China in particular, filial piety becomes an infeasible mode of securing livelihoods in old age in the context of restrictive birth-planning and rapid aging. Finally, structural change, or the "logic of industrialism", generates pressure for social protection against industrial risks, and economic progress creates the financial resources that allow for social policy intervention by the state. The level and growth over time of GDP is thus a crucial explanatory variable for social policy and welfare state development (Obinger, 2019). This is particularly true for unemployment insurance, although it should be noted that, historically, unemployment insurance was introduced rather late in Western and other economies (Obinger & Schmitt, 2020) since it was more controversial than, for instance, pension insurance. For the statistical indicators of Chinese cities, we therefore expect a connection between economic growth and coverage:

Hypothesis 1: The higher the economic growth rate in a city, the faster the expansion of unemployment insurance coverage.

In the following sections, we delve deeper into the complex causal chains to develop more concrete hypotheses.

2.1.1 Dualization and the Standard Employment Relationship

When societies undergo processes of economic development and modernization, their labour markets tend to develop in a dualized way, i.e. there is a cleavage between a formal, urban sector of employment and an informal sector. The formal sector is characterized by institutionalized partition of labour with stable and specialized roles, which link up to corresponding pathways of formal education or vocational training. By contrast, the informal sector offers broader and less institutionalized roles, which less directly link up to institutionalized pathways of education or vocational training (Middleton et al., 1993). Overall, labour productivity in the formal sector is higher than in the informal sector, so the former produces more resources that can be used for social insurance or other benefits. Among other things, the term dualization points to the notable differences of working and living conditions between the formal and informal sector.

One crucial institutional foundation of work in the formal sector is the SER, which also constitutes an institutionalized dividing line in social protection globally (Hahs & Mückenberger, 2022; Mückenberger & Dingeldey, 2022). There are different definitions for the concept, which usually include the elements of a (1) full-time work relationship based on (2) a contract concluded (3) directly between employer and employee, which is typically (4) permanent rather than fixed-term in nature. Under the planned economy, workers in SOEs checked many of those boxes, but were life-long members of a company rather than hired based on a labour contract (Warner, 1995). Since the 1980s, China's economic growth was largely based on cheap and informal labour (Park & Cai, 2011). But by increasing economic productivity, it also laid the foundation for the expansion of a formal employment sector with mandatory social insurance in urban areas.

In the PRC, the connection between employment status and unemployment insurance is codified by regulatory documents and laws. Like in many transition economies, unemployment insurance played an important role in stabilizing society during reforms of the state sector. Unemployment insurance was introduced for SOE workers in the 1980s, for whom it provided one of three pillars of protection against unemployment in the 1990s and 2000s – alongside continued support from SOEs, and social assistance (Duckett & Hussain, 2008; Solinger, 2005). Since 1999, unemployment insurance was extended to formal employees in the private sector (Guowuyuanling 1999, 258). Existing research points to a strong connection between standard contract-based employment and social insurance coverage, including unemployment insurance (Giles et al., 2013). The China Labour Statistical Yearbook even lists a special category of "on-the-job employees" (*zaigang zhigong*), which is associated with standard employment and social insurance coverage (MoHRSS, 2022). The indicator primarily covers employers in public or mixed (semi-public) ownership, but largely excludes domestic private companies, which are associated more with informal employment. Our second hypothesis reflects patterns of dualization typical for a country transitioning from middle to high income status:

Hypothesis 2: The higher the growth rate of standard (on-the-job) employment, the higher the growth rate of unemployment insurance coverage.

2.1.2 Urbanization and migration

Economic growth in China has been driven by processes of structural change, in which labour shifted from agriculture to industry and services, and people moved from the countryside to the cities (Naughton, 2018). This logic of industrialism strongly facilitated the growth of urban populations and the expansion of urbanized areas into previously rural areas. City governments in coastal China enabled mass migration of cheap labour by creating an implicit consensus between local business and residents on systematically excluding migrants from formal

employment, social protection and permanent local household registration (Y. Chen, 2008; Park & Cai, 2011). The institutional basis for this is the Chinese *hukou* system, which distinguishes between residents with and without a local household registration (*hukou*), the latter mainly including migrant workers from other regions (Chan & Buckingham, 2008).

However, the demographic transition undermined the feasibility of a growth model based on cheap labour (D. Liu, 2015; Y. Zhang et al., 2018), pushing firms to upgrade to more technology- and skill-intensive production (H. Li et al., 2012, 2017). During this process, firms developed a rational interest in unemployment insurance since they competed for skilled employees and needed their employees to invest in coherent skill-sets (like in other countries; see: Mares, 2003). Firms also lobbied local governments to adjust their policies through formal and informal channels (Schubert & Heberer, 2017), and many shifted to more integrative policies towards migrants in the 2010s (J. S.-H. Wang et al., 2023). We argue that these trends facilitated increasing coverage of migrant workers. Statistical indicators are only available for a few years: migrant worker coverage rose from only 3.9% in 2009 to 10.5% in 2014 – 11.5% among intra-provincial migrants and 9.8% among inter-provincial migrants (Tian, 2018 pp. 232).

Formal employment primarily captures the most successful migrant workers, who manage to fully integrate into urban society and live with their families in the destination city. However, the vast majority of migrant workers are seasonal migrants who return home once a year, and may change destination cities frequently (Fan, 2008). Their presence mainly contributes to SER employment of urban residents, but some might be insured too. Many seasonal migrants work in the construction sector (Swider, 2015), and thus contribute to the growth of urban areas, which is the basis for our third hypothesis:

Hypothesis 3: The bigger the expansion of urbanized area in a prefectural or provincial city's territory, the faster the expansion of unemployment insurance coverage.

A minority of migrant workers also manages to establish themselves as permanent residents in one destination city, and an even smaller minority manages to naturalize and acquire a local household registration at the destination place. Those are typically skilled migrants with stable jobs, who often bring their family to live with them at the destination city (H. Wang, 2011). Such migrants are more likely to be insured than seasonal migrants who leave their families at the sending place, which is why we expect the number of permanent residents in a city to affect coverage. The official statistical indicators for permanent residents (*changzhu renkou*) and the registered population (*huji renkou*) are helpful for distinguishing levels of integration. The status of a permanent resident can be assigned by the destination city to migrants who live there for prolonged periods of time, and it does not require changing the household registration. By definition, this category excludes many seasonal migrants, and captures migrants with comparatively stable living conditions who are not in a standard employment relationship.

Hypothesis 4: The higher the growth of the number of permanent residents in a city, the faster the expansion of unemployment insurance coverage.

By contrast, acquiring a local hukou at the destination place represents a deeper level of social integration. The transfer of a household registration is a more complex process, which requires the cooperation of local government and administration in both the sending and destination area. A *hukou* transfer usually requires migrant workers to forgo the right to agricultural land in their native village, and thus a potential fallback option for a livelihood in times of economic crisis. Many migrant workers are sceptical about taking this step (Zhu & Lin, 2011, p. 9). Local governments in sending places in Central and West China, in turn, are assigned subsidies for social insurance and many public services based on the number of the registered population (Müller, 2016b). For them, formalizing outmigration by *hukou* transfer shrinks their fiscal transfer revenue, so they have incentives to retain the regis-

trations of persons who out-migrated. For higher levels of government, the *hukou* population is a more reliable indicator to base accounting on, because *hukou* status is more tightly controlled than permanent residency and cannot easily be inflated by local government to increase their transfer revenue.

Conversely, formalizing in-migration via *hukou* transfer increases the costs of social protection and public service provision in the coastal provinces, where most destination cities are located. This is particularly the case for trans-provincial migration. Coastal provinces receive fewer central subsidies for social protection and public services, leaving a higher fiscal burden on the provincial level. Furthermore, wealthy counties and prefectural cities often increase the generosity of local social benefits, which increases the costs at local level compared to other localities (Müller, 2016a, 2016b). Many destination cities operate point-systems which make certain levels of education or home-ownership a necessary pre-condition for acquiring a local *hukou* (Y. Dong & Goodburn, 2019). But in recent years, more and more cities also facilitate the integration of migrant workers by granting them a formal household registration (J. S.-H. Wang et al., 2023), largely in competition for a shrinking workforce. We thus also expect increases in the registered population to increase coverage, and those should be even stronger predictors of coverage than permanent residency figures:

Hypothesis 5: The higher the growth of the registered population in a city, the faster the expansion of unemployment insurance coverage.

These hypotheses and the corresponding indicators enable us to better grasp the development of dualization and its connection to unemployment insurance coverage: on-the-job employment captures formal employment in the public and semi-public sector, and increases in the registered and permanent population should capture a substantial part of formal employment in the domestic private sector, along with some informal employment. When controlling for those indicators, the growth of urban space captures primarily informal employment, thus providing an indicator for the extent to which dualization is overcome.

2.2 Power resources and labour mobilization

Unemployment insurance is a comparatively contentious branch of social insurance. Theories of power resources interpret social policy as the result of societal power struggles, most notably between interests of capital and labour. Historically, organized labour, workers' parties and business associations influenced the allocation of resources in society through different channels. These theories explained asymmetries in social and political influence of the opposing camps, focusing on indicators regarding voter mobilization in elections and government participation of the respective parties, the strength of unionization and of business associations, as well as strikes, collective bargaining, and labour mobilization more generally (Ebbinghaus, 2019; Esping-Andersen, 1990; Korpi, 1983).[2] Labour mobilization and the balance of power between opponents and proponents were particularly relevant in the case of unemployment insurance, especially before the advent of corporatist arrangements, since it represented the greatest violation of liberal economic principles and protected healthy and employable people. Moreover, it was considered difficult to calculate actuarially in the face of cyclical uncertainties (Öktem, 2022; Paster, 2015).

These theories are applicable in the PRC, albeit to a limited extent. Labour unions and sectoral associations are typically extensions of the party and/or the state, rather than autonomous bottom-up organizations. Unions do not act as an independent representation of workers interests, but rather as a mediator between management, workers and government.[3] However,

2 For literature on social movements more generally, see: (Rucht, 2023)

3 In a similar way, entrepreneurs and business associations seek to influence government through informal networks, rather than confronting it openly (Schubert & Heberer, 2017).

both the presence of unions and party cells has been found to positively impact the coverage of social insurance more generally (Z. Dong et al., 2016; Gao & Rickne, 2017; Rickne, 2013). More importantly, the new working classes actively influenced the recent social fabric of the PRC (Butollo & ten Brink, 2012; Elfstrom, 2021). This can be seen in the tens of thousands of social protests inside and outside companies since the 2000s, which the government considers to be "unofficial mass events" (including mobilization events, violent protests, and strikes) – and which the government is determined to defuse. In this study, we therefore expect an impact of labour mobilization on the recent expansion of unemployment insurance (Carter et al., 2013; Ciccia & Guzmán-Concha, 2018; Niedzwiecki & Pribble, 2023).

Hypothesis 6: Cities with more frequent and more intense labour unrest expand unemployment insurance coverage more quickly.

3. Methods and data

Our research design focuses on observational data from Chinese prefectural and provincial cities. The level of statistical aggregation – the second of three sub-national levels – matches our level of analysis, because unemployment insurance is typically pooled at the level of prefectural or provincial city. Furthermore, the gap between registered and permanent population in census data indicates that most variation is between cities, rather than between provinces (Z. Li et al., 2024). This points to prefectural and provincial cities being the most appropriate level of analysis.

We conduct regression analyses using a two-wave panel with the growth rate of the number of people insured in unemployment insurance as the dependent variable (long differences). This parametrization avoids potential distortions inherent to conventional coverage rates in case there are errors in the data (see also: C. Li & Gibson, 2015). The timespan of analysis ranges from 2012, the earliest year for which data on insurance coverage at prefectural city level are available, till 2019, to avoid the distortions caused by the Covid 19 pandemic. The core of the dataset was compiled from time series from the National Bureau of Statistics (NBS), covering 288 prefectural and provincial cities and more than 90% of China's population as of 2012. To optimize data quality, we implemented extensive triangulation of statistical indicators.[4] In the regression analysis, we only consider highly significant results to maximize robustness against random measurement error. To provide a more differentiated view of core indicators, we contextualize NBS data with data from the CLDS. For the analysis, we limit the impact of outliers by winsorizing the indicator for the registered population with a threshold of 0.05.

Furthermore, we use satellite-detected nighttime lights (NTL) data to supplement reported GDP data as a measure of city economic growth. These NTL data are popular in research where GDP data are either unavailable or manipulation is suspected (Martínez, 2022) because they are outside of political control and are comparable over space irrespective of varying statistical capacity and economic informality (which can undermine administrative data). We chose a NTL data source that should be best for our setting – NASA's Black Marble near-nadir annual composites (Gibson, 2021; Román et al., 2018; X. Zhang & Gibson, 2022). For the final analysis, we create a light-corrected GDP indicator by averaging over the growth rates of GDP and lights. The outcome has a correlation with the original GDP growth rates of 0.96, whereas the correlation between original GDP and luminosity growth is 0.69. The independence of the errors in night lights and official GDP data cancels out errors in each of the original indicators, and thus mitigates potential bias in the calculations.

4 NBS data is transferred upwards through the administrative hierarchy by Chinese cities, but not all local indicators form consistent time series. Therefore, we conducted extensive data triangulation and consistency checks, excluding indicators where we perceived problems with validity, reliability, and objectivity.

For the rate of urban area growth of each city, we use land cover data provided by the Copernicus Climate Change Service (C3S, 2019), which is the earth observation component of the EU Space Programme. These global land cover maps categorize the land surface into 22 classes[5] at a spatial resolution of 10 arc-seconds (corresponding to about 250 meters at China's latitude). The estimates are updated yearly, with images obtained from the Sentinel 3 satellites, and are consistent with earlier global annual land cover data for 1992 to 2015 produced by the European Space Agency (ESA) Climate Change Initiative (CCI). We calculate the sum of built-up area by city in 2012 and in 2019 to derive the rates of change. For the analysis, we limit the impact of outliers by winsorizing the indicator with a threshold of 0.05.

To capture the intensity of labour mobilization in different cities, we draw on strike data from the China Labour Bulletin in Hong Kong. It is the richest and most comprehensive collection of labour unrest data in the 2010s, provides event descriptions and classifications regarding the nature of demands, and the type and scope of the event. As the occurrence of protest events strongly fluctuates from year to year, we aggregated the event count over the period of observation. However, this way, we could not calculate a growth rate, for data is not available for the eight years before 2012, including a major strike wave in Guangdong in 2010 (Butollo & ten Brink, 2012), that preceded a massive expansion of coverage there. Instead of growth rates, we use three event counts: the total number of recorded protest events, the number of strikes, and the number of protest events with more than 100 participants.

Finally, we compiled a corpus of relevant regulatory documents to extract information on sub-national coverage rules and pooling arrangements via machine learning text classification, and other text mining techniques. Relying on Beijing University's PKUlaw database (*falü fagui shujuku*) we collected 1,935 relevant documents from the central government, all 31 provincial jurisdictions, and 175 prefectural cities. We split the documents into a total of 48,136 sentences, from which 7,472 sentences were selected randomizing across documents and sentences. We coded whether these sentences contained coverage regulations and trained an ensemble of machine learning classifiers (logistic regression, bagging, boosting, and transformers) to predict the relevance of the remaining sentences and documents. Then, we analysed the documents and sentences in descending order of predicted relevance until relevant information ceased to come up. The methodological appendix provides an in-depth description of our approach.

4. Results

This section presents the results of our analysis. The first sub-section discusses *de jure* coverage in the context of China's multilevel polity. It was not possible to reconstruct clear and unambiguous coverage regulations for all prefectural jurisdictions, which precluded the creation of a quantitative indicator for *de jure* coverage expansion at local level. Rather, on-the-job employment is the best indicator for *de jure* coverage, but it has a certain margin of error. The second sub-section analyses how *de facto* coverage relates to the SER, developed over time, and differs regionally. The third sub-section discusses the results of long-differenced regression analysis. For robustness, we test the model on a data set in which influential outliers are removed,[6] and on subsets of the data split along the median of light-corrected GDP. The analysis is conducted for one larger period, rather than multiple periods, because the use of satellite data in panel models requires long intervals.

5 One of these 22 classes is urban area, based on artificial and impermeable surfaces.

6 We removed the city of Lhasa due to an extreme value that would become an influential outlier and distort the calculations.

4.1 *De jure* coverage of unemployment insurance

As this section will show, *de jure* coverage is fragmented, with provincial and prefectural jurisdictions all enacting their own coverage regulations. However, it was not possible to unambiguously reconstruct de jure coverage for all the prefectural jurisdictions for several reasons. First, there were no documents available for some prefectural jurisdictions (see: Appendix A2.4). Second, most regulatory development occurred after the enactment of unemployment insurance in 1999. We originally expected broad and thorough regulatory change at local level to coincide with the expansion and changing structure of coverage of unemployment insurance. But our results point to limited development in formal regulations. Rather, we found ambiguous definitions of who is considered an "employee," or under which conditions staff in public service units is required to participate. We see this as enabling flexible policy implementation at local level. However, since this flexibility is not always formalized in the regulatory documents pertaining to unemployment insurance, we were unable to construct a meaningful quantitative indicator for *de jure* coverage expansion to be used in the following sections.

In principle, social insurance should have very broad coverage. The 1995 Labour Law clarified that all labour relationships (*laodong guanxi*) should (*yingdang*, rather than must) be formalized with a labour contract, and that all labourers should participate in social insurance (Renda 1995, §16 & §72). The 2007 Labour Contract Law tightened the connection between employment and labour contract; and asked for social insurance participation to be specified in labour contracts, and for compliance to be monitored by local governments among other things (Renda 2007, §17 & §74). So, social insurance should cover all dependent work relationships, but standard contract-based employment is at the core, and on-the-job employment is the statistical indicator most suitable for measuring it.

For unemployment insurance, *de jure* coverage is slightly ambiguous. On the one hand, coverage is defined narrowly in central regulations. A first system was introduced by central decree in 1986 along with contract-based employment (Guofa 1986, 77). It primarily targeted SOE workers. In 1999, the State Council extended the system to urban enterprises and public service units. Urban enterprises include those in state, collective, foreign, private (*siying*) or other ownership. Provincial governments can decide to extend coverage to social organizations (*shehui tuanti*), private non-company organizations (*minban feiqiye danwei*) and household companies with employees (*you gugong de geti gongshanghu*) (Guowoyuanling 1999, 258, §2 & §32). On the other hand, however, the Ministry of Labour and Social Security (MoLSS)[7] called for the term "employee" (*zhigong*) to be interpreted broadly, including informal employees, workers with rural and/or external household registration,[8] and foreigners (Laoshebufa 1999, 9). In 2005, the MoLSS specifically called for extending coverage among flexible employees (*linghuo jiuye renyuan*) where the SOE reforms led to declining numbers of contributors (Laosheshiyesihan 2005, 1).[9] Subnational levels thus could for example include flexible workers either by mentioning them explicitly in their documents, or by implicitly treating them as employees.

Due to such ambiguities, determining whether a specific group is or is not de jure covered in a specific province and year is not always straightforward. Provincial-level governments could include different groups of employers and employees by explicit choice, or implicitly by interpretation of terms like "employee" or "company." Regarding employer types, provincial coverage regulations were overall quite homogeneous. They nearly always explicitly included social and private non-company organizations;

7 In 2008, the MoLSS was merged with the Ministry of Personnel to form the Ministry of Human Resources and Social Security (MoHRSS).

8 The State Council merely specified that migrant workers were exempt from premium payment in 1999.

9 The 2010 Social Insurance Law merely specified that all employees should participate in unemployment insurance (Renda 2010, §44).

Global Dynamics of Social Policy CRC 1342
socium Research Center on Inequality and Social Policy

and few jurisdictions explicitly mentioned civil servants, township-village enterprises, and employees in army organizations or other groups respectively. The inclusion of household companies with employees was initially not mentioned in twelve of the thirty-one jurisdictions, but they gradually seem to have been included there as small companies.[10] In a similar way, some provinces explicitly included flexible and self-employed workers[11] on a voluntary basis, such as Hubei and Yunnan (Hubei Province government 2002, §22; Yunnan Province government 2006, §40); whereas others like Fujian arguably included them by interpretation.[12] We found direct and indirect evidence of those groups' inclusion in thirteen provincial jurisdictions, but others may have included them without producing regulatory evidence (cf.: Giles et al., 2013, p. 129). During the period of observation, there were few explicit changes in *de jure* coverage, rendering it quasi time-invariant during the period of observation and thus unsuitable for quantitative analysis in a panel model.

In 27 out of 31 provincial jurisdictions, provincial regulations merely constitute an umbrella for the subordinated jurisdictions. There, prefectural cities can have their own coverage regulations, and sometimes also counties and county-level cities. Given regulatory ambiguity, it is not always clear whether coverage regulations at prefectural level go beyond the provincial standards or not, or whether all inclusion is formalized. However, we expect governments that enact special regulations to put a particular emphasis on the groups targeted. For example, our dataset includes seven cities with specific provisions for flexible employees (cf.: Tian, 2018, pp. 150–179): Nanjing and Suqian in Jiangsu Province, Harbin (Heilongjiang Province), Jilin (Jilin Province), Ordos (Inner Mongolia) and Xiamen (Fujian Province).

In summary, while *de jure* coverage is embedded in the fragmented multi-level nature of China's polity, the ambiguity of core concepts makes it flexible as well. This explains the absence of major change in coverage regulations during the expansion of coverage in the 2010s, which the following section analyses.

4.2 *De facto* Coverage

The previous section analysed laws and regulations defining *de jure* coverage of unemployment insurance. This section focuses on de facto coverage, on the connection to the SER, on the development of coverage over time, and on regional differences. To illuminate the connection between unemployment insurance coverage and the SER, we use data from the 2016 CLDS. Overall, unemployment insurance covered 2,200 of 14,011 economically active respondents, which equals a coverage rate of 15.7%.[13] Table 1 provides a more in-depth look at different status groups in 2016: coverage was highest by far among employees with full-time, contract-based employment, followed by those in part-time employment. However, substantial shares of informal full-time employees and of

10 During the Covid crisis, household companies with employees were entitled to subsidies from unemployment insurance funds in most of these provinces.

11 Self-employed persons are referred to by various terminologies, including free-lancers (*ziyou zhiye*), self-employed (*zimou zhiye*), individual economic organizations (*geti jingji zuzhi*), and household companies without employees (*wu/wei gugong de geti gongshanghu*).

12 In Fujian, the indirect evidence of their inclusion is a letter of the provincial Bureau of Labour and Social Security (BoLSS) indicating which districts of Fuzhou City should enact a lower cost rate for household companies' and flexible workers' premiums because these workers opposed higher premiums (Fujian Province BoLSS 2007). The provincial regulations enacted in the previous year did not mention these groups at all. The central government mandated substantial cost rate decreases in the 2010s, arguably eliminating the need for group-specific rates during the extension to the private sector (see also: Tian, 2018, pp. 236-251).

13 The 218 CLDS respondents who failed to report on unemployment insurance coverage are treated as uncovered here. CLDS coverage is substantially lower than the coverage rate of the economically active population of 22.4% for 2016 suggested by NBS figures. These differences may be due to sampling issues in the CLDS; and/or the NBS data under-estimating the total number of economically active persons (Holz, 2013).

Table 1: Unemployment insurance coverage by employment status in 2016

Status	Employees: 5686 (40.58%)				Employers: 287 (2.05%)	Self-employed: 1709 (12.20%)
Contract	Contract: 2847 (50.07%)		No contract: 1710 (30.07%)			
Fulltime	Full-time: 2766 (97.15%)	Part-time: 81 (2.85%)	Full-time: 1555 (90.94%)	Part-time: 155 (9.06%)		
Unemployment insurance	Covered: 1631 (58.97%)	Covered: 23 (28.40%)	Covered: 273 (17.56%)	Covered: 9 (5.81%)	Covered: 34 (11.85%)	Covered: 56 (3.28%)

Note: Data has been taken from the 2016 CLDS. There was a total of 14,011 economically active respondents, the status group of agricultural workers was omitted here. Among the status group employees, 1,129 did not answer the question regarding the labour contract. Respondents who failed to report on unemployment insurance coverage were counted as not covered.

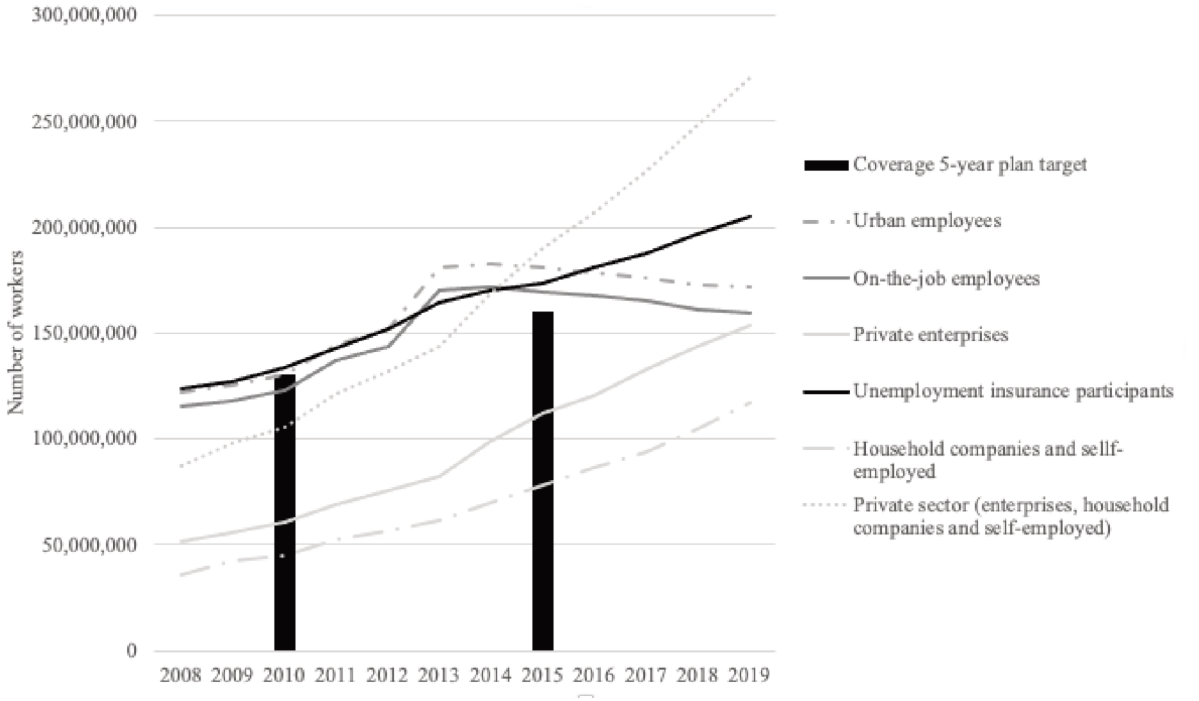
employers were covered, while coverage was lowest in informal part-time and self-employment. While there was no direct indicator for permanent employment, coverage was higher among formal employees whose labour contract had been signed before 2015, which excludes one-year contracts. In this group, 840 of 1281 full-time employees were covered (65.57%), and 16 of 34 part-time employees (47.06%). So overall, unemployment insurance coverage was clearly higher in employment relationships approximating the SER. It reflects a state of largely dualized coverage, but with some significant inroads of coverage in the informal sector.

How does this cross-section relate to national-level developments? Figure 1 illustrates the development of urban employment and other economic activity between 2008 and 2019. The indicator on-the-job employees refers to contract-based employment with social insurance, and has strong overlaps with the SER. Representing a conventional concept of formal employment (see also: Mok & Wu, 2016), the indicator focuses on employers that are state-owned, collective, foreign-invested or of mixed ownership, but excludes domestic private companies, which are traditionally associated more with informal employment. On-the-job employment also excludes various forms of a-typical employment – such as during probation time, or informal employment (MoHRSS, 2022). According to 2016 CLDS data, in organizations associated with SER (on-the-job) employment, coverage was 69.3% for those with full-time, contract-based employment, up from 62.3% in 2014; and 54.4% for all employees, up from 45.9% in 2014. The "urban employees" indicator goes beyond on-the-job employment by also covering a-typical employment in the same types of organisations.

Figure 1 illustrates the process of coverage expansion. Until 2012, the number of persons enrolled with unemployment insurance was slightly above the "on-the-job employees," indicating that coverage was focused on SER employment. In the 2010s, the number of on-the-job employees began to stagnate and decrease, while coverage continued growing, pointing to increasing coverage of the private and informal sector. Coverage also surpassed the targets specified in the five-year plans, which required 160 million persons by 2015, and 180 million by 2020 (Guofa 2012, 17; Renshebufa 2016, 63).

The plans also emphasized small and medium enterprises, typically associated with private ownership and informal employment. CLDS data indicates that overall coverage of the private sector, including private companies and self-employment, grew from 11.3% in 2014 to 15.5% in 2016. In domestic private companies alone, it grew from 19.9% to 29.0%. Full-time, contract-based employees' coverage grew from

Figure 1: Urban Employment and Unemployment Insurance Participation



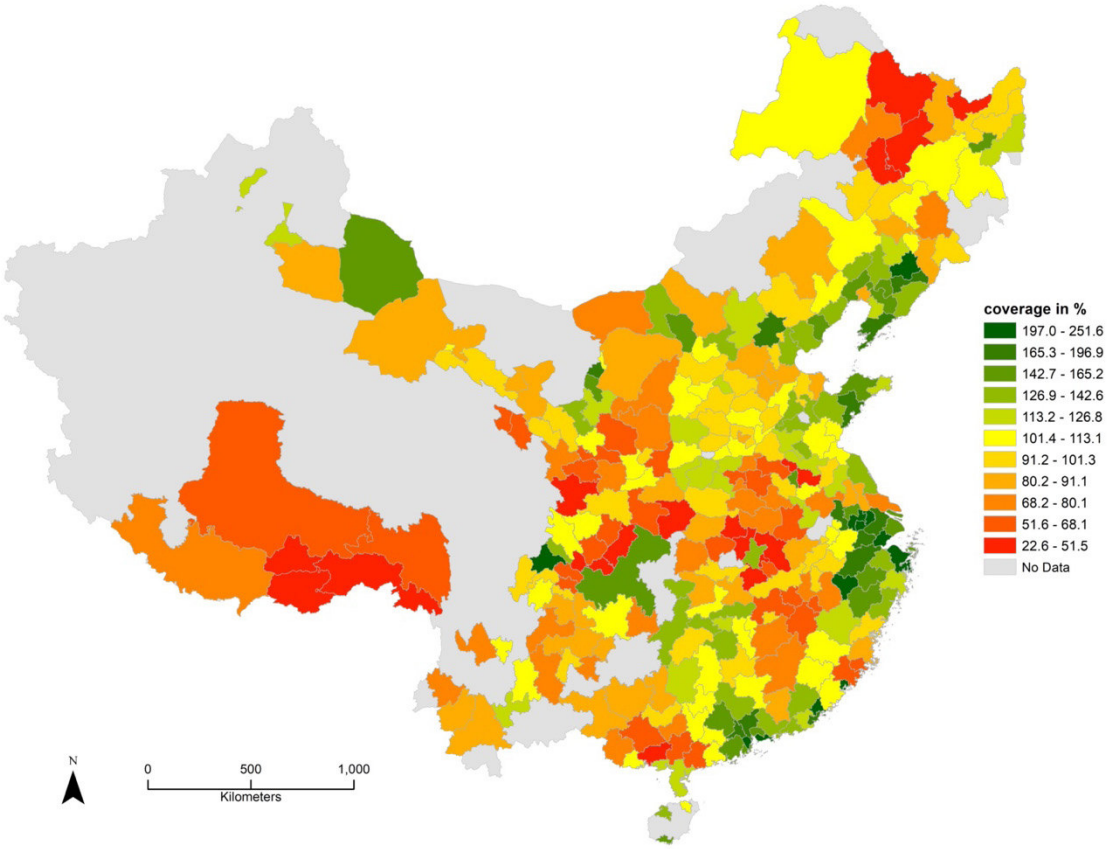
Sources: China Statistical Yearbook (various years); China Labour Statistical Yearbook (various years); reporting of employment in private enterprises ended in 2017, later figures are estimates.

39.5% to 47.9%. For comparison, coverage among those without a labour contract grew from 4.94% to 6.67%. According to CLDS data, formal employment in the domestic private sector grew from 37.2% in 2012 to 55.1% in 2016, explaining part of this trend. But at the same time, the Phi coefficient between having a labour contract and being covered decreased from 0.5 to 0.44, passing from an intermediary to a weak connection, and thus indicating increasing coverage of informal employment as well. These figures illustrate that the coverage expansion in the domestic private sector was in part facilitated by growing formal employment there, while coverage of informal employment and self-employment was also growing. These results point to a weakening linkage between coverage and the SER, and to weakening dualization.

Dualization was weakened in part because coverage was incomplete among on-the-job employees as well. This is further confirmed when taking a cross-sectional perspective on coverage. Map 1 illustrates an indicator we call SER-coverage in China's prefectural and provincial cities. It represents the number of insured persons divided by the number of on-the-job employees as of 2019. With a generous margin of error, the indicator corresponds to the relationship between *de facto* and *de jure* coverage. Yellow jurisdictions feature coverage around 100%, green ones feature higher-than-expected coverage, and red ones feature lower-than-expected coverage. There are clusters of elevated coverage in the Pearl River Delta, the Yangzi Delta, and around the Shandong Peninsula and the Bohai Gulf. Here, we can expect most on-the-job employees and much of the private sector to be covered. Conversely, there is a belt of mostly low coverage in the agricultural inland provinces, stretching from Jiangxi via Hubei and Henan to Shaanxi and Gansu in the North-West. Coverage is also low in the mountainous South-West. It is remarkable that even coverage of SER employees strongly depends on the uneven levels of socio-economic development. In much of central China, even full-time, contract-based employment in the public sector does not ensure unemployment insurance coverage.

To analyze the process that led to the 2019 levels of coverage and weakened dualization,

Map 1: SER-Coverage 2019



Note: Coverage of formal employees was calculated by dividing the total number of insured by the number of on-the-job employees, which are most closely associated with contract-based standard employment and social insurance coverage.

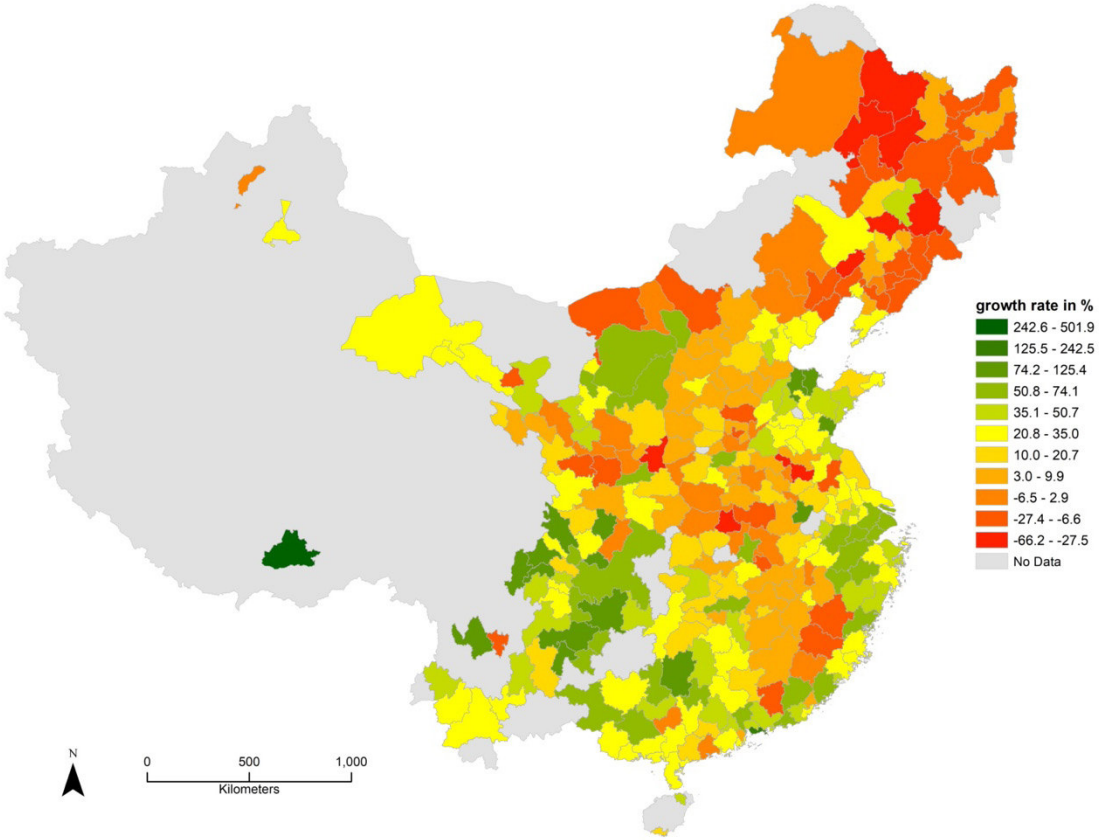
we shift to looking at the growth rate of the number of insured persons between 2012 and 2019. The national growth rate was 19.47%,[14] which is roughly represented by the transition from light orange to yellow. Growth rates de-emphasize the historical legacies of uneven development in two ways. First, they focus on changes between 2012 and 2019. Second, they quantify growth relative to the starting point by measuring it as a percentage. The growth rates are also higher in the coastal provinces than in the agrarian inland provinces. However, growth rates are mostly low in the North-East, reflecting economic stagnation and outmigration from that region, while de-emphasizing the historically higher level of socio-economic development. Conversely, growth rates are higher in the South-West, which has developed more dynamically from a lower starting point in the 2010s, which is arguably due to tourism and the relocation of industry from the coast. There are only two extreme outliers: Shenzhen (242.54%), which saw a massive coverage increase following a strike wave in Guangdong in 2010 as noted above; and Lhasa (501.94%), which arguably saw quick urbanization and social insurance expansion other parts of the country experienced in previous decades. To avoid distortions, we use two-year averages for the quantitative analysis below and exclude Lhasa from the analysis.

4.3 Hypotheses tests

Table 2 illustrates the results of the hypothesis tests on two-wave panel data via long-differenced regression of the growth rates. It features regular standard errors, which were almost always larger than robust standard errors (HC1).

14 The figure represents the median of the two-year averages of the growth rates in the sample.

Map 2: Growth Rates of Insurance Participation 2012 - 2019



Calculations with robust standard errors are provided in the appendix for comparison.

Hypothesis 1: The higher the economic growth rate in a city, the faster the expansion of unemployment insurance coverage.

The core hypothesis of socio-economic theory is confirmed.[15] As a stand-alone, the light-corrected GDP growth indicator explains 12.1% of variation in the growth rates of the number of insured (Model 1). For 100% growth in light-corrected GDP, the number of insured grows by 28.9%. This coefficient shrinks and loses some of its significance when other independent variables are introduced.

Hypothesis 2: The higher the growth rate of formal (on-the-job) employment, the higher the growth rate of unemployment insurance coverage.

As noted above, the indicator of on-the-job employees provides a very good – though not perfect – proxy for formal urban employment with social insurance. All other things being equal, an increase in on-the-job employees of 100% in a city is connected to an increase of 30.8% in the number of insured in Model 2.[16] The corresponding effect of light-corrected GDP shrinks to 18.1%. Explained variance rises to 16.9%. The indicator's omission of some formal employment in domestic private companies arguably contributes to the significance of the intercept in

15 Robustness checks indicate the indicator's significance is stronger in the subset of cities below the median of light-corrected GDP growth.

16 Robustness checks prove the indicator significant in both subsets of cities above and below the median of light-corrected GDP, but more strongly so in the subset below the median.

Table 2: Regression analysis with long differences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Intercept)	4.396+	9.438***	0.487	1.160	-1.708	-0.546	-4.518	-3.128	-3.669
	(2.585)	(2.790)	(4.153)	(3.912)	(3.908)	(3.884)	(4.282)	(4.214)	(4.217)
GDP (light-corrected)	0.289***	0.181***	0.193***	0.146**	0.107*	0.110*	0.125*	0.123*	0.129*
	(0.045)	(0.051)	(0.051)	(0.048)	(0.050)	(0.049)	(0.049)	(0.050)	(0.050)
On-the-job employees		0.308***	0.289***	0.173*	0.244***	0.193**	0.160*	0.160*	0.154*
		(0.074)	(0.073)	(0.072)	(0.069)	(0.071)	(0.072)	(0.074)	(0.073)
Urban space growth			0.216**	0.243***	0.268***	0.266***	0.294***	0.286***	0.293***
			(0.075)	(0.071)	(0.071)	(0.070)	(0.071)	(0.071)	(0.071)
Population 1: permanent				0.557***		0.322**	0.221+	0.264*	0.254*
				(0.091)		(0.116)	(0.124)	(0.121)	(0.121)
Population 2: registered					1.856***	1.202**	1.196**	1.204**	1.205**
					(0.293)	(0.372)	(0.370)	(0.371)	(0.371)
Protest events							0.059*		
							(0.028)		
Protest events (strikes)								0.156	
								(0.100)	
Large protest events									0.177+
									(0.095)
Num.Obs.	287	287	287	287	287	287	287	287	287
R2	0.124	0.175	0.198	0.292	0.298	0.317	0.328	0.323	0.325
R2 Adj.	0.121	0.169	0.190	0.282	0.288	0.305	0.314	0.308	0.311

Note: Significance levels: +=0.1, *=0.05, **=0.01, ***=0.001

Model 2, and its own loss of significance when adding the permanent population in Model 4.

We formulated 3 hypotheses regarding urbanization and immigration:

Hypothesis 3: The bigger the expansion of urbanized area in a prefectural or provincial city's territory, the faster the expansion of unemployment insurance coverage.

Hypothesis 3 is confirmed. If the winsorized urban space indicator grows by 100%, that leads to a 21.6% increase in unemployment insurance coverage (Model 3). The effect becomes stronger and more significant when the population indicators are introduced, pointing to increasing coverage in the informal sector and a weakening of dualization. The winsorized urban space indicator reflects the growth of urbanized areas and, indirectly, seasonal labour migrants, who are unlikely to have stable and formalized employment. Some reports raise doubts regarding seasonal migrant workers' entitlements to unemployment benefits, citing receipt of income replacement being conditional on either 12 months of continuous premium payment – *de facto* excluding seasonal migrants – and/or a local household registration (X. Wang & Sun, 2014; Yu & Yang, 2020). However, seasonal migrants are entitled to one-time payments (Guowoyuanling 1999, 258, §21),[17] and household registration requirements should have been abolished since 2008. Local governments can easily adjust the benefit structure to seasonal migrant workers' needs. We would thus expect coverage of informal and flexible employment to grow in cities that are expanding.[18]

Hypothesis 4: The higher the growth of the number of permanent residents in a city, the faster the expansion of unemployment insurance coverage.

Hypothesis 4 captures a different group of migrants, who settle down at the registration place and often bring their families with them. In Model 4, all other things being equal, if the number of permanent residents increases by 100%, the number of insured persons increases by 55.7%. This effect is highly significant, including in cities above and below the median of light-corrected GDP. Concurrently, the coefficient for on-the-job employment shrinks and significance drops. This indicates that migrants who become permanent residents at the destination place are very likely to both, get into SER-like employment and be covered by unemployment insurance. Such migrants are more likely to fulfil the requirements for regular unemployment benefits in the event of job loss. CLDS data indicates that the median unemployment benefit in 2016 was 800 RMB per month, equivalent to 34.29% of the median monthly net salary.[19] While this is significantly lower than unemployment benefits in some industrialized economies, it nevertheless temporarily reduces the pressure to relocate or accept a job unrelated to one's skills to generate immediate income, thus binding migrants and supporting the preservation of their work-related skills.[20]

17 As of 2016, such one-time payments amounted to 2.6% of total expenditures. (Tian, 2018: 244).

18 Robustness checks show the effect was more significant in cities below the median of light-corrected GDP.

19 For 2010 to 2016, Tian estimates an average replacement rate of about 30% in the private sector, and below 20% in the public and semi-public sector, i.e. for the category of urban employees. (Tian, 2018 pp. 196f).

20 In the 2016 CLDS sample, less than 1% of those who had not worked in the 12 months before the survey received unemployment insurance benefits. This may under-represent the total number of beneficiaries because most either found another job or exhausted their entitlement within 12 months. The average duration of benefit receipt in 2016 was 8.5 months nationwide, with 30.1% of recipients getting benefits for 3 months or less, and only 25.2% for more than a year (Tian, 2018 pp. 200-203). Strict conditionality requirements also severely limit the share of beneficiaries among the registered unemployed, which mostly oscillated between 20% and 25% in the 2010s (Tian, 2018 pp. 181-190). For example, the duration of benefit receipt is conditional on the duration of premium payment, with more than 5 years of premium payment required for a duration of more than 12 months (Guowuyuanling 1999, 58, §17). Apart from this, the unemployment rate is subject to political targets featuring in cadre evaluation systems, which may limit the number of people allowed to register as unemployed.

Hypothesis 5: The higher the growth of the registered population in a city, the faster the expansion of unemployment insurance coverage.

Growth of the registered population points to migrants who become fully naturalized in the destination city. The winsorized indicator produces a very strong and highly significant effect in model 5. All other things being equal, if the number of registered residents grows by 100%, the number of insured persons grows by 185.6%.[21] So, for most residents who acquire a *hukou*, two persons are being insured. The coefficient decreases substantially in model 6, indicating that many of the additional insured persons are permanent residents. While permanent and *hukou* population are highly correlated, a comparison of Models 4 to 6 illustrates the absence of multicollinearity problems, like dramatic changes in explained variance. Overall, model 6 explains 30.5% of the variation in coverage growth.

Finally, we analysed the impact of labour unrest on unemployment insurance coverage.

Hypothesis 6: Cities with more frequent and more intense labour unrest expand unemployment insurance coverage more quickly.

We focus here mainly on the total number of protest events, actual strikes, and the number of large events with more than 100 participants. All three indicators have positive coefficients, suggesting a positive influence on coverage, with the effects of large events with more than 100 participants being the strongest. Explained variance is highest in model 7. However, the significance of the indicators is low or very low, which indicates little robustness to measurement error. This is a problem with indicators for which the probability of events being captured and the level of detail of the description depend on the strength of the NGO's network in specific locations. Given that outliers – like the Pearl River Delta – here may be more accurate than the average observation, winsorization does not promise much improvement. The results are thus inconclusive, and more research is needed to confirm our hypotheses.

5 Discussion: dualization and coverage

Our results indicate that the degree to which labour market dualization is reflected in unemployment insurance coverage is gradually decreasing. First, dualization is not strictly institutionalized in *de jure* coverage, the ambiguity of which allows flexible interpretation in a local context. Central regulations set a minimum standard and leave the choice to include some groups to provincial governments, though variation in provincial coverage regulations remains limited. Final decisions about coverage are mostly made at the prefectural level, though some jurisdictions also pool at provincial or county level. But local governments need not necessarily adjust regulatory documents pertaining to unemployment insurance to include additional groups; they can also interpret terms like "employee" or "company" narrowly or broadly, leading to differences in inclusiveness that are not necessarily formalized. Overall, the expansion of coverage in the 2010s was not accompanied by substantial regulatory change.

Second, evidence of dualization in prefectural cities is weaker than expected because coverage at prefectural level may be substantially higher or lower than SER-employment, as the SER-coverage indicator reveals. Even though coverage is institutionally coupled to SER-like employment, this coupling weakened in the 2010s. In some prefectures, formal and informal employees in the private sector were covered in large numbers; whereas in others, even much of the staff in public service units was not covered. The analysis of *de facto* coverage reveals huge variation in SER-coverage, and thus additional complexity in the way in which dualization manifests itself.

Third, the quantitative analysis of coverage growth rates at prefectural level reveals precise-

21 The effect is more significant in cities below the median of light-corrected growth.

ly how dualization is weakened. It shows how SER (on-the-job) employment influences coverage compared to economic growth, migration, urbanization, and labour mobilization. Coverage increases beyond on-the-job employment in successful destination cities, which grow both economically and spatially, and integrate a substantial share of the newly arrived – first as permanent residents, and then as registered population. Coverage of those groups increases with the level of integration; and it reaches beyond the boundaries of the traditional public and semi-public sector captured by on-the-job employment, into an emerging sphere of formalized employment in domestic private companies, and into the sphere of informal employment.

6. Conclusion

This study aimed at shedding light on the puzzle of how unemployment insurance coverage in China could keep growing in the 2010s while traditional indicators pointed to stagnating standard employment. To some extent, growth in formalized employment in the domestic private sector may have been missed by the traditional indicators. More generally, we found growth in the number of insured persons beyond traditional formal employment to be driven by economic growth, immigration, and urbanization. Our results confirm socio-economic theories of welfare state development. Chinese cities compete for migrant workers, and unemployment insurance serves the dual purpose of offering some income replacement when employment is interrupted, as well as support for skill development and subsidies for struggling companies in times of economic crisis.

Chinese unemployment insurance primarily protects labour market insiders. By contrast, actual unemployment is concentrated among higher education graduates who never managed to access or consolidate standard employment, and among aging migrant workers leaving manufacturing or construction. New instruments are needed to tackle this form of dualization. Unemployment insurance funds accumulated huge surpluses during the 2010s, which could fund new instruments to tackle the unemployment problem. This would, however, require the institutionalization of redistribution between labour market insiders and outsiders.

References

Bai, L. (2006). Graduate unemployment. *The China Quarterly, 185*(1), 128–144.

Bluhm, R., & Krause, M. (2022). Top lights: Bright cities and their contribution to economic development. *Journal of Development Economics, 157*, 102880. https://doi.org/10.1016/j.jdeveco.2022.102880

Butollo, F., & ten Brink, T. (2012). Challenging the atomization of discontent: Patterns of Migrant-Worker Protest in China during the Series of Strikes in 2010. *Critical Asian Studies, 44*(3), 419–440. https://doi.org/10.1080/14672715.2012.711978

Carter, J., Bédard, M., & Bista, C. P. (2013). *Comparative review of unemployment and employment insurance experiences in Asia and worldwide*. ILO.

Chan, K. W., & Buckingham, W. (2008). Is China abolishing the hukou system? *The China Quarterly, 195*, 582–606.

Chen, W., Chen, X., Hsieh, C.-T., & Song, Z. (2019). A Forensic Examination of China's National Accounts. *Brookings Papers on Economic Activity, 1*, 77–127.

Chen, X. (Ed.). (2013). *Zhongguo chengshi tongji nianjian (China city statistical yearbook)*. China Statistics Press.

Chen, X. (Ed.). (2020). *Zhongguo chengshi tongji nianjian (China city statistical yearbook)*. China Statistics Press.

Chen, Y. (2008). "Chaîne d'intérêts" et absorption des migrants en ville. In *La société chinoise vue par ses sociologues*. SciencesPo. Les Presses.

Ciccia, R., & Guzmán-Concha, C. (2018). Protest and the Politics of Unemployment Insurance: Reforming Welfare States in Times of Austerity. In *Social Welfare Responses in a Neoliberal Era* (pp. 88–118). Brill.

Dong, Y., & Goodburn, C. (2019). Residence Permits and Points Systems. *Journal of Contemporary China*, 1–20.

Dong, Z., Luo, Z., & Wei, X. (2016). Social insurance with Chinese characteristics: The role of communist party in private firms. *China Economic Review, 37*, 40–51.

Duckett, J. (2001). Political interests and the implementation of China's urban health insurance reform. *Social Policy & Administration, 35*(3), 290–306.

Duckett, J., & Hussain, A. (2008). Tackling unemployment in China: State capacity and governance issues. *The Pacific Review, 21*(2), 211–229.

Ebbinghaus, B. (2019). Machtressourcentheorie und Korporatismusansatz. In H. Obinger & M. G. Schmidt (Eds.), *Handbuch Sozialpolitik* (pp. 117–138). Springer VS.

Edin, M. (2003). State capacity and local agent control in China. *The China Quarterly, 173*, 35–52.

Elfstrom, M. (2021). *Workers and Change in China: Resistance, Repression, Responsiveness*. Cambridge University Press.

Elvidge, C. D., Baugh, K. E., Zhizhin, M., & Hsu, F.-C. (2013). Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proceedings of the Asia-Pacific Advanced Network, 35*(0), 62. https://doi.org/10.7125/APAN.35.7

Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Princeton University Press.

Fan, C. C. (2008). *China on the move*. Routledge.

Flora, P., & Alber, J. (1981). Modernization, democratization, and the development of welfare states in Western Europe. In F. Peter & A. Heidenheimer (Eds.), *The development of welfare states in Europe and America* (pp. 37–80). Transaction.

Frazier, M. W. (2010). *Socialist insecurity*. Cornell University Press.

Gallagher, M. E., Giles, J., Park, A., & Wang, M. (2015). China's 2008 Labor Contract Law: Implementation and implications for China's workers. *Human Relations, 68*(2), 197–235.

Gao, Q., & Rickne, J. (2017). Inequality in social insurance participation and generosity: Do firm characteristics matter? *Social Policy & Administration, 51*(5), 755–775.

Gibson, J. (2021). Better Night Lights Data, For Longer*. *Oxford Bulletin of Economics and Statistics, 83*(3), 770–791. https://doi.org/10.1111/obes.12417

Gibson, J., & Li, C. (2017). The erroneous use of China'sS population and per capita data. *Journal of Economic Surveys, 31*(4), 905–922. https://doi.org/10.1111/joes.12178

Giles, J., Wang, D., & Park, A. (2013). Expanding social insurance coverage in urban China. *Research in Labor Economics, 37*, 123–179.

Goldblatt, R., Heilmann, K., & Vaizman, Y. (2020). Can Medium-Resolution Satellite Imagery Measure Economic Activity at Small Geographies? Evidence from Landsat in Vietnam. *The World Bank Economic Review, 34*(3), 635–653. https://doi.org/10.1093/wber/lhz001

Hahs, J., & Mückenberger, U. (2022). Segmenting and equalizing narratives in the ILO's standard-setting

practice. *International Labour Review*, *161*(4), 635–655. https://doi.org/10.1111/ilr.12344

Holz, C. A. (2013). Chinese statistics: Classification systems and data sources. *Eurasian Geography and Economics*, *54*(5–6), 532–571. https://doi.org/10.1080/15387216.2014.908312

Hong, I., & Ngok, K. (2022). The industrialisation thesis, revisited. *Social Policy & Administration*, *56*(6), 874–890. https://doi.org/10.1111/spol.12836

Klein, M. C., & Pettis, M. (2021). *Trade wars are class wars: How rising inequality distorts the global economy and threatens international peace: with a new preface*. Yale University Press.

Korpi, W. (1983). *The democratic class struggle*. Routledge & Kegan Paul.

Li, C., & Gibson, J. (2015). City scale and productivity in China. *Economics Letters*, *131*, 86–90. https://doi.org/10.1016/j.econlet.2015.04.001

Li, H., Li, L., Wu, B., & Xiong, Y. (2012). The End of Cheap Chinese Labor. *Journal of Economic Perspectives*, *26*(4), 57–74. https://doi.org/DOI: 10.1257/jep.26.4.57

Li, H., Loyalka, P., Rozelle, S., & Wu, B. (2017). Human Capital and China's Future Growth. *Journal of Economic Perspectives*, *31*(1), 25–48. https://doi.org/10.1257/jep.31.1.25

Li, Z., Li, C., Gibson, J., & Deng, X. (2024, February 6). *Two decades of inter-city growth in China*. AARES Conference, Australia National University. https://www.aares.org.au/Events/AARES-2024/program/

Lichtenberg, E., & Ding, C. (2009). Local officials as land developers: Urban spatial expansion in China. *Journal of Urban Economics*, *66*(1), 57–64. https://doi.org/10.1016/j.jue.2009.03.002

Liu, D. (2015). Has the Chinese economy passed the Lewis turning point? *Journal of the Asia Pacific Economy*, *20*(3), 404–422. https://doi.org/10.1080/13547860.2015.1054167

Liu, S., & Schmitt, U. (2014). China's Urbanization and Land: A Framework for Reform. In World Bank & Development Research Center for the State Council of the PRC (Eds.), *Urban China: Towards Efficient, Inclusive, and Sustainable Urbanzation.* (pp. 163–196). World Bank Group.

Mares, I. (2003). The Sources of Business Interest in Social Insurance: Sectoral versus National Differences. *World Politics*, *55*(2), 229–258. https://doi.org/10.1353/wp.2003.0012

Martínez, L. R. (2022). How Much Should We Trust the Dictator's GDP Growth Estimates? *Journal of Political Economy*, *130*(10), 2731–2769. https://doi.org/10.1086/720458

Middleton, J., Ziderman, A., & van Adams, A. (1993). *Skills for productivity: Vocational education and training in developing countries*. Oxford University Press.

MoHRSS. (2022). *Zhongguo laodong tongji nianjian (China Statistical Yearbook of Labor)*. Zhongguo tongji chubanshe (China Statistics Press).

Mok, K. H., & Wu, A. M. (2016). Higher education, changing labour market and social mobility in the era of massification in China. *Journal of Education and Work*, *29*(1), 77–97. https://doi.org/10.1080/13639080.2015.1049028

Mückenberger, U., & Dingeldey, I. (2022). Introduction: Worldwide patterns of legal segmentation in employment law. *International Labour Review*, *161*(4), 511–534. https://doi.org/10.1111/ilr.12355

Müller, A. (2016a). *China's new public health insurance*. Routledge.

Müller, A. (2016b). Hukou and Health Insurance Coverage for Migrant Workers. *Journal of Current Chinese Affairs*, *45*(2), 53–82.

Naughton, B. (2018). *The Chinese economy: Adaptation and growth*. The MIT Press.

Niedzwiecki, S., & Pribble, J. (2023). Social Policy Expansion and Retrenchment in Latin America: Causal Paths to Successful Reform. *Journal of Social Policy*, 1–21. https://doi.org/10.1017/S0047279423000090

Obinger, H. (2019). Sozio-ökonomische Theorie des Wohlfahrtsstaates. In H. Obinger & M. G. Schmidt (Eds.), *Handbuch Sozialpolitik* (pp. 97–116). Springer VS.

Obinger, H., & Schmitt, C. (2020). Total War and the Emergence of Unemployment Insurance in Western Countries. *Journal of European Public Policy*, *27*(12), 1879–1901.

Öktem, K. G. (2022). Causal Mechanisms in the Introduction and Development of Unemployment Insurance in Turkey. In J. Kuhlmann & F. Nullmeier (Eds.), *Causal Mechanisms in the Global Development of Social Policies* (pp. 103–139). Springer International Publishing. https://doi.org/10.1007/978-3-030-91088-4_4

Orenstein, M. A. (2001). *Out of the red*. The University of Michigan Press.

Park, A., & Cai, F. (2011). The informalization of the Chinese labor market. In S. Kuruvilla, C. K. Lee, & M. E.

Gallagher (Eds.), *From iron rice bowl to informalization* (Vol. 14, pp. 17–35). ILR Press.

Paster, T. (2015). *Bringing Power Back In: A review of the literature on business and the welfare state*. *15*(03).

Rickne, J. (2013). Labor market conditions and social insurance in China. *China Economic Review*, *27*, 52–68.

Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., Schultz, L., Bell, J., Stokes, E. C., Pandey, B., Seto, K. C., Hall, D., Oda, T., Wolfe, R. E., Lin, G., Golpayegani, N., Devadiga, S., Davidson, C., Sarkar, S., … Masuoka, E. J. (2018). NASA's Black Marble nighttime lights product suite. *Remote Sensing of Environment*, *210*, 113–143. https://doi.org/10.1016/j.rse.2018.03.017

Rozelle, S., & Hell, N. (2020). *Invisible China: How the urban-rural divide threatens China's rise*. The University of Chicago Press.

Rozelle, S., Xia, Y., Friesen, D., Vanderjack, B., & Cohen, N. (2020). Moving beyond Lewis: Employment and wage trends in China's high- and low-skilled industries and the emergence of an era of polarization. *Comparative Economic Studies*, *62*, 555–589.

Rucht, D. (2023). *Social Movements: A Theoretical Approach* (1st ed.). Oxford University Press. https://doi.org/10.1093/oso/9780198877400.001.0001

Schubert, G., & Heberer, T. (2017). Private entrepreneurs as a 'strategic group' in the Chinese polity. *The China Review*, *17*(2), 95–122.

Solinger, D. (2005). Path dependency reexamined. *Comparative Politics*, *38*(1), 83–101.

Swider, S. C. (2015). *Building China*. ILR Press, an imprint of Cornell University Press.

Tian, D. (2018). *Zhongguo jiji de shiye baoxian zhengce (The active policies of unemployment insurance in China)*. Shehui kexue wenxian chubanshe.

Wang, H. (2011). *Dangqian nongmingong shehui rongru de zhuyao zhang'ai jiqi zhidu genyuan fenxi (Obstructions to the Social Integration of Migrant Workers and their Institutional Roots)* (32; Development Research Center of the State Council - Research Report).

Wang, J. S.-H., Zhu, Y., Peng, C., & You, J. (2023). Internal Migration Policies in China: Patterns and Determinants of the Household Registration Reform Policy Design in 2014. *The China Quarterly*.

Wang, X., & Sun, W. (2014). Discrepancy between registered and actual unemployment rates in China: An investigation in provincial capital cities. *China & World Economy*, *22*(4), 40–59.

Warner, M. (1995). *The management of human resources in Chinese industry*. Macmillan.

Wilensky, H. (1975). *The welfare state and equality*. University of California Press.

Yu, S., & Yang, Y. (2020, August 18). The Chinese economy's two-speed recovery from coronavirus. *Financial Times*. https://www.ft.com/content/e0e2940a-17cb-40ed-8d27-3722c9349a5d

Zhang, X., & Gibson, J. (2022). Using Multi-Source Nighttime Lights Data to Proxy for County-Level Economic Activity in China from 2012 to 2019. *Remote Sensing*, *14*(5), 1282. https://doi.org/10.3390/rs14051282

Zhang, Y., Shao, T., & Dong, Q. (2018). Reassessing the Lewis Turning Point in China: Evidence from 70,000 Rural Households. *China & World Economy*, *26*(1), 4–17. https://doi.org/10.1111/cwe.12226

Zhu, Y., & Lin, L. (2011, January 1). *The mobility patterns of rural-urban migrants and their social protection in China*. https://www.ids.ac.uk/download.php?file=files/dmfile/ResearchReport01FINAL.pdf

Appendix 1: Laws and regulations

Table A1.1: Central documents

Enacted by	Year	No.	Abbreviation	Title
State Council	1986	77	Guofa	*Guoying qiye zhigong daiye baoxian zanxing guiding* (Preliminary regulations regarding "waiting-for-work" insurance for employees in state-owned enterprises)
People's Congress	1995		Renda	*Laodong fa* (Labour Law)
State Council	1999	258	Guowuyuanling	*Shiye baoxian tiaoli* (Unemployment Insurance Regulation)
MoLSS	1999	9	Laoshebufa	*Shehui baoxian fei zhengjiao zanxing tiaoli xuanchuan tigang* (Propaganda outline for the Preliminary Regulations for Social Insurance Premium Collection)
MoLSS	1999	29	Laoshebufa	*Guanyu shiye danwei canjia shiye baoxian youguan wenti de tongzhi* (Notification regarding problems of public service units participating in unemployment insurance)
MoLSS	2005	1	Laosheshiyesihan	*2005 nian shiye baoxian gongzuo yaodian* (Unemployment insurance work: essentials for 2005)
People's Congress	2007		Renda	*Laodong hetong fa* (Labour Contract Law)
People's Congress	2010		Renda	*Shehui baoxian fa* (Social Insurance Law)
State Council	2012	17	Guofa	*Shehui baozhang shi'er wu guihua gangyao* (Outline for the 12th Five-year Plan for Social Insurance)
MoHRSS	2016	63	Renshebufa	*Renli ziyuan he shehui baozhang shiye fazhan shisan wu guihua gangyao* (Outline for the 13th Five-year Plan for Human Resources and Social Insurance)

Table A1.2 Sub-national documents

Enacted by	Year	No.	Abbreviation	Title
Hubei Province government	2002	235	Hubei Sheng Renmin Zhengfu ling	*Hubei Sheng shiye baoxian shishi banfa* (Implementation method for unemployment insurance in Hubei Province)
Yunnan Province government	2006			*Yunnan Sheng shiye baoxian tiaoli* (Unemployment Insurance Reglation of Yunnan Province)
Fujian Province BoLSS	2007	363	Minlaoshehan	*Guanyu baosong kaizhan "qiye zhigong yanglao baoxian, shiye baoxian kuamian zhengjiao xuanchuan" huodong qingkuang de han* (Letter regarding the situation of "Premium collection extension propaganda for company employees' pension and unemployment insurance)
Qinghai Province government	2012	1	Qingzheng	*Qinghai Sheng shiye baoxian jijin shengji tongchou banfa* (Measure for Provincial Pooling of Unemployment Insurance Funds in Qinghai Province)
Jiujiang City government	2012	198	Jiufutingzi	Jiujiang Shi shiye baoxian shiji tongchou zanxing banfa (Preliminary measures for city-level pooling of unemployment insurance funds in Jiujiang City)

Appendix 2: Methods and data

Our research design focuses on observational data, with Chinese prefectural and provincial cities as the unit of analysis. We conduct regression analyses with the growth rate of the number of people insured in unemployment insurance as dependent variable (long differences). This parametrization avoids potential distortions inherent to conventional coverage rates in case there are errors in the data (C. Li & Gibson, 2015). The timespan of analysis reaches from 2012, the earliest time when data on insurance coverage at prefectural city level is available, till 2019, to avoid the distortions induced by the Covid 19 pandemic. The core of the dataset was compiled from time series from the National Bureau of Statistics, which covered 288 prefectural and provincial cities and more than 90% of the Chinese population as of 2012. To optimize data quality, we implemented extensive triangulation of statistical indicators. We also use cutting-edge satellite data for more objective, valid and reliable estimates of economic development and urban growth. To capture the intensity of labour mobilization in different cities, we draw on strike data from the China Labour Bulletin. Finally, we analysed a large corpus of regulatory documents for a detailed reconstruction of de jure coverage, relying on machine-learning and text mining.

A2.1 National Bureau of Statistics data

The core of the data set consists of indicators for Chinese cities compiled by the National Bureau of Statistics (NBS), and published mainly in the China City Statistical Yearbook (CCSY; X. Chen, 2013, 2020). The level of statistical aggregation – the second of three sub-national levels – matches our level of analysis, because unemployment insurance is typically pooled at the level of prefectural city. Data on social insurance participation is available from 2012 on. We choose 2019 as the final year of the analysis, because the availability of other indicators reduced drastically afterwards, and because we want to avoid distortions connected to the Covid pandemic. In China, statistical data is transferred upwards through the administrative hierarchy by Chinese cities, but not all local indicators form consistent time series. Therefore, we conducted extensive data triangulation and consistency checks, excluding indicators where we perceived problems with validity,[1] reliability,[2] and objectivity. For triangulation, we used indicators at provincial level from the China Labour Statistical Yearbook (CLSY) and the China Statistical Yearbook (CSY).

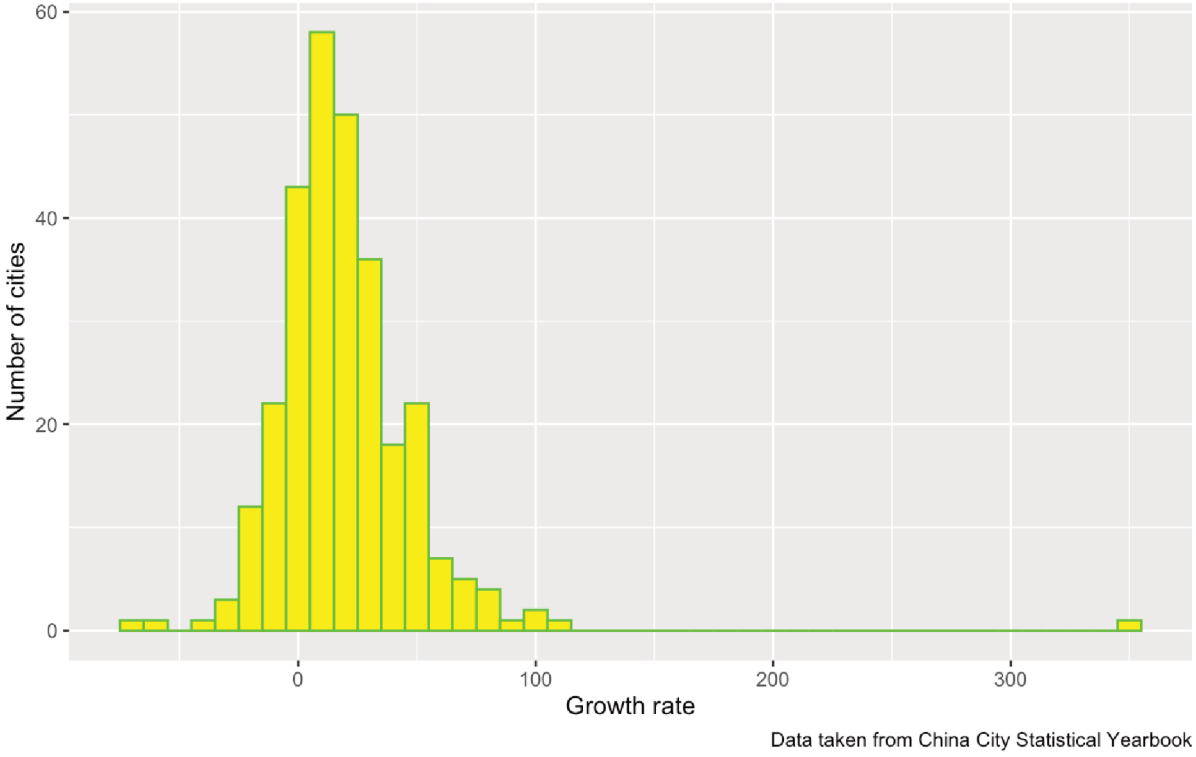
A2.1.1 Dependent variable

We measure coverage as the growth rate of the number of people insured in unemployment insurance in Chinese cities at prefectural level or above. Our dependent variable is calculated from the NBS indicator for the number of insured people in unemployment insurance, which is of good quality overall. Validity is less of a problem because there is only a single unemployment insurance program which covers people at working age, but not pensioners. Local social insurance bureaus have the number of enrolled people, and there is but one possible figure to report. This also helps with reliability because glitches and discontinuities in reporting can be easily avoided. Triangulating CCSY data with CLSY provincial indicators revealed a strong overlap of dynamics, with only a few outliers and errors. Objectivity also appears to be un-

1 Some indicators did not provide appropriate measurements for the underlying concepts. For example, fiscal revenue does not capture extra-budgetary revenues, which is a substantial share of local revenue, and thus provides but an incomplete indicator for local fiscal capacity. In a similar way, import-export data may not capture a substantial part of trade through special economic zones.

2 We also assessed the reproducibility of measurement (reliability) by examining the full time series from 2012 to 2019 for each city and province and triangulating it with data from different sources wherever possible. If time series displayed many and strong discontinuities or did not fit with provincial-level data from other yearbooks, we avoided using the indicator.

Plot A2.1a: Unemployment Insurance Coverage Growth Rate (2012 to 2019, 2-year averages)



problematic, as local governments have little or no incentive to misrepresent enrolment numbers. Using single-year values produces two extreme outliers: Lhasa (502%) and Shenzhen (243%). To avoid distortions, we thus used two-year averages. As the histogram shows, most cities experienced a growth in coverage. Lhasa still has the potential to distort the results of the regression analysis. We thus exclude the city of Lhasa from the analysis. Furthermore, we check the models for influential outliers.

A2.1.2 Population

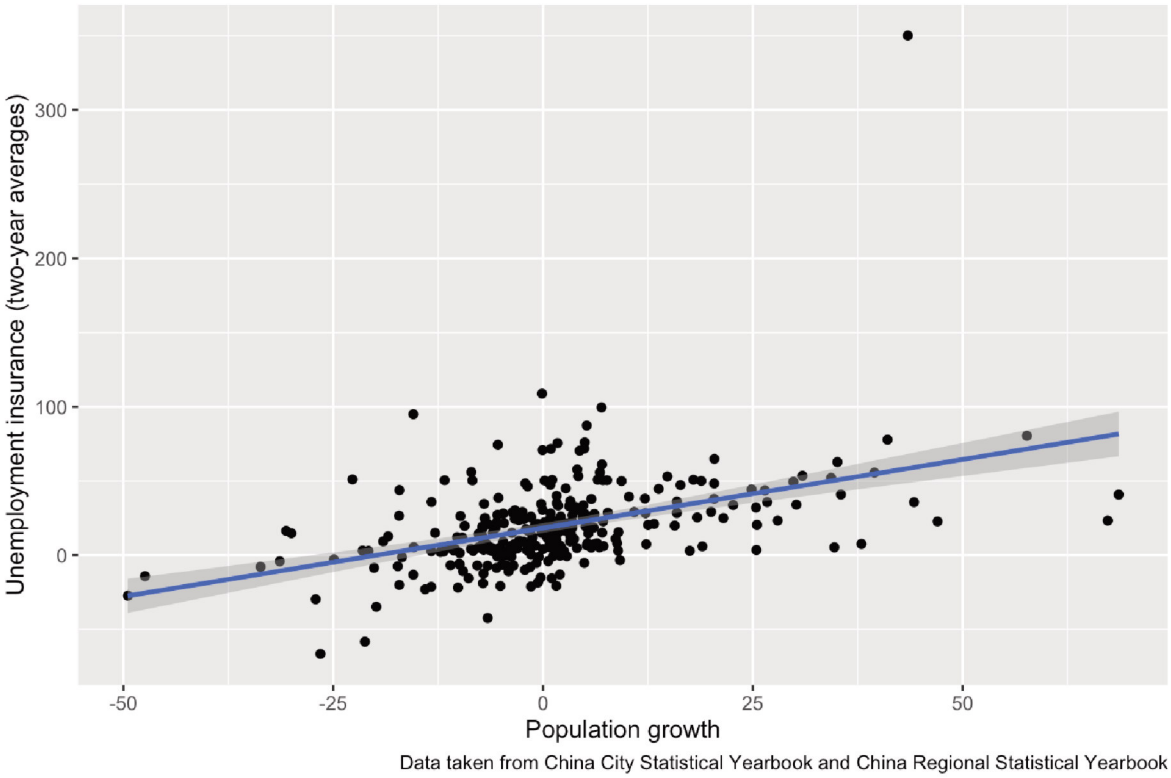
Population data presented a minor challenge, for the CCSY merely reports the *de jure* population, i.e. the population with a local household registration (*hukou*). Due to high levels of informal migration, *de facto* population may differ substantially in some places (Gibson & Li, 2017). We thus also used the more realistic indicator of permanent population (*changzhu renkou*). The 2012 indicator was reported in the Regional Economy Statistical yearbook and should be reasonably accurate given that a national census was carried out in 2010. No indicator was reported for 2019, but we expect the 2020 indicator to be accurate as a national census was carried out in the same year. Therefore, we calculated the 2019 indicator by subtracting the average annual difference from the 2020 indicator. The scatter plot illustrates how population growth and coverage growth relate to one another. Overall, there is a positive connection.

Compared to the permanent population, the registered population displayed a stronger concentration of values between 25% and -25%, but also more extreme positive values. Some cities are thus formalizing large numbers of in-migrants, but they are exceptions. For the analysis, we limit the impact of outliers by winsorizing the indicator with a threshold of 0.05.

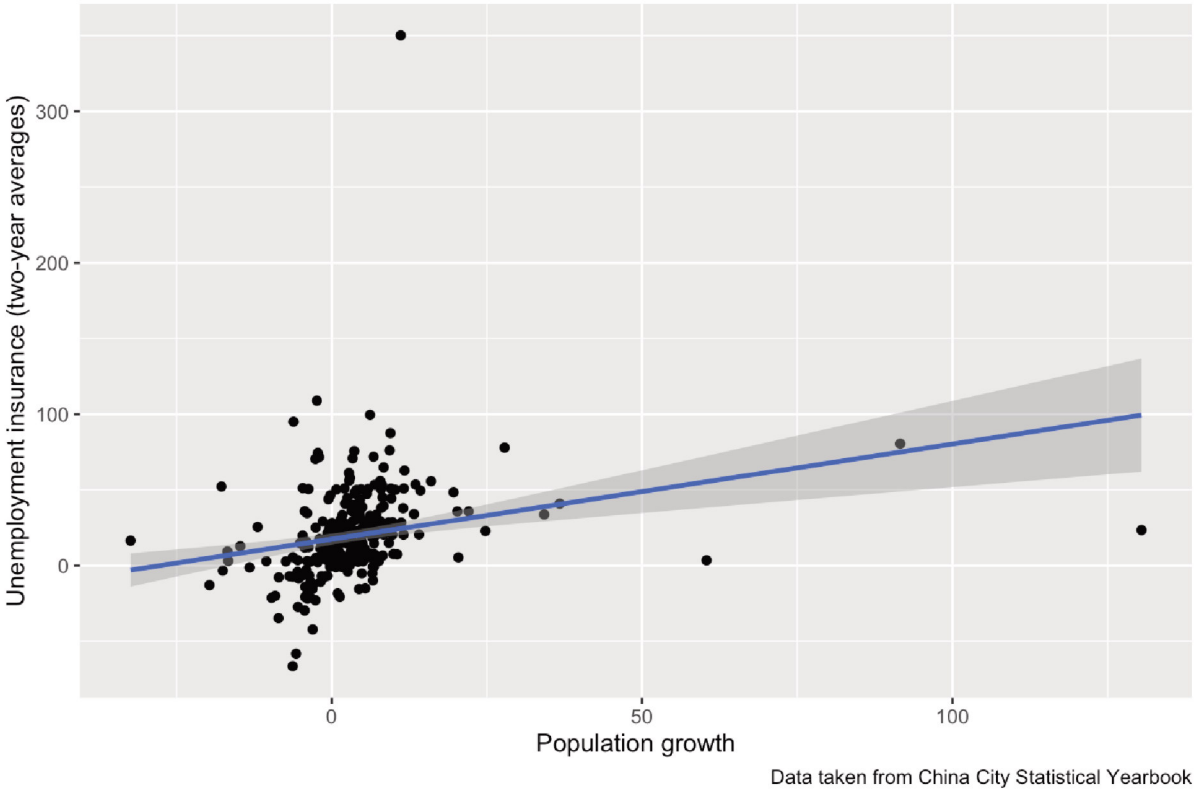
A2.1.3 Employment

There are overall three indicators for employment in prefectural cities, which largely match the indicators at national level presented in Figure 1 above. Formal employment is largely

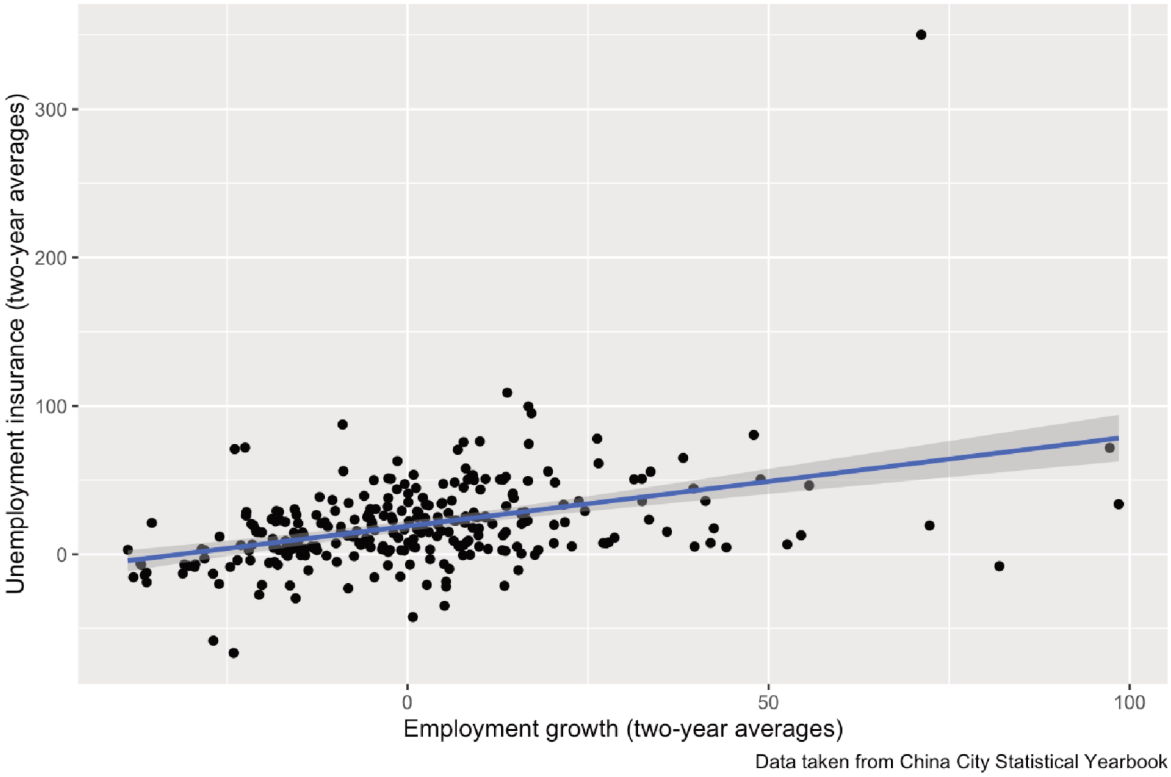
Plot A2.1b: Unemployment Insurance and Permanent Population



Plot A2.1c: Unemployment Insurance and Registered Population



Plot A2.1d: Unemployment Insurance and On-the-Job Employment



captured by the average number of on-the-job employees (*zaigang zhigong pingjun renshu*) in the CCSY, which is reported as a year-end indicator in the CLSY for the provinces. Furthermore, a broader year-end indicator of formal employment in the CCSY captures employees in urban units (*chengzhen danwei congye renyuan*). This second indicator is numerically equivalent to the broader indicator for employment in urban units in the CLSY, and the first indicator displays very similar trends. Numerical equivalence exists despite minor inconsistencies in indicator definitions between the different sources,[3] so we expect a certain level of noise in the data and only consider results with strong significance. Therefore, we use all four indicators for triangulation and use a two-year average of on-the-job employees as independent variable. The scatter plot points to a positive connection.

Figures for self-employment, small household companies and private companies in CCSY are, unfortunately, characterized by strong discontinuities and divergence from provincial figures reported in the CSY. We thus refrain from their usage due to doubts regarding reliability. Contract-based employment and insurance coverage are not exclusive to the sectors covered by the first two indicators, as survey data illustrated. However, this data stems from a set of enterprises directly reporting to the government, which typically have either a dominant position in their sector or are representative of companies in it (Holz, 2013). We thus expect the indicators used to also capture the dynamics of private enterprises to some extent, but not of self-employment and small household companies.

3 CCSY definitions are the same for 2013 till 2019 but differ slightly for 2012. CLSY definitions, conversely, display a minor shift between 2017 and 2018. Connection to the ownership status of employers is less obvious in the CCSY.

A2.2 Satellite data

We use satellite data to improve on the official indicators regarding the growth of urban land use and for GDP growth. For the former, the CCSY did not provide a valid indicator for the period of observation. The objectivity of Chinese GDP data is subject to controversial debate. On the one hand, local governments have strong political incentives to report biased figures (Edin, 2003), for example to overstate economic growth, or to understate local unemployment.[4] Furthermore, the NBS itself has been reported to misrepresent economic growth (W. Chen et al., 2019; Martínez, 2022).[5] On the other hand, in many cases there is no alternative to NBS data, and its overall quality improved considerably in recent decades. By contrast, night-lights emissions captured by satellites are frequently used as proxies for economic development, especially in contexts where other data is not available. We use a combination of official GDP data and night-light emissions to get a better approximation of true economic growth.

A2.2.1 Lights data

We use satellite-detected night-time lights (NTL) data to supplement reported GDP data as a measure of city economic growth. These NTL data are popular in research where GDP data are either unavailable or manipulation is suspected (Martínez, 2022) because they are outside of political control and are comparable over space irrespective of varying statistical capacity and economic informality (which can undermine administrative data). Despite this popularity there are concerns about using NTL data for changes in economic activity at local levels because they only weakly relate to traditional measures, such as changes in commune-level employment and household expenditures (Goldblatt et al., 2020) and changes in county-level GDP (X. Zhang & Gibson, 2022).

In light of these concerns, we chose a NTL data source that should be best for our setting – NASA's Black Marble near-nadir annual composites (Román et al., 2018). Night-time lights mainly reflect urban activity (Gibson, 2021) and in a 3-D landscape such as cities there is less distortion from angular viewing effects if satellites are directly overhead; only Black Marble data let users restrict attention to near-nadir images. These data also have no issues with saturation or top-coding, unlike the widely used DMSP data where top-coding causes central cities to seem no brighter than dimly-lit suburbs (Bluhm & Krause, 2022) or smaller towns (Gibson, 2021). Black Marble data have onboard calibration for temporal consistency whereas DMSP data reflect unrecorded changes in amplification over the lunar cycle and suffer from large inter-satellite differences. The ground footprint for Black Marble is at least 45 times more spatially precise than for DMSP (Elvidge et al., 2013).

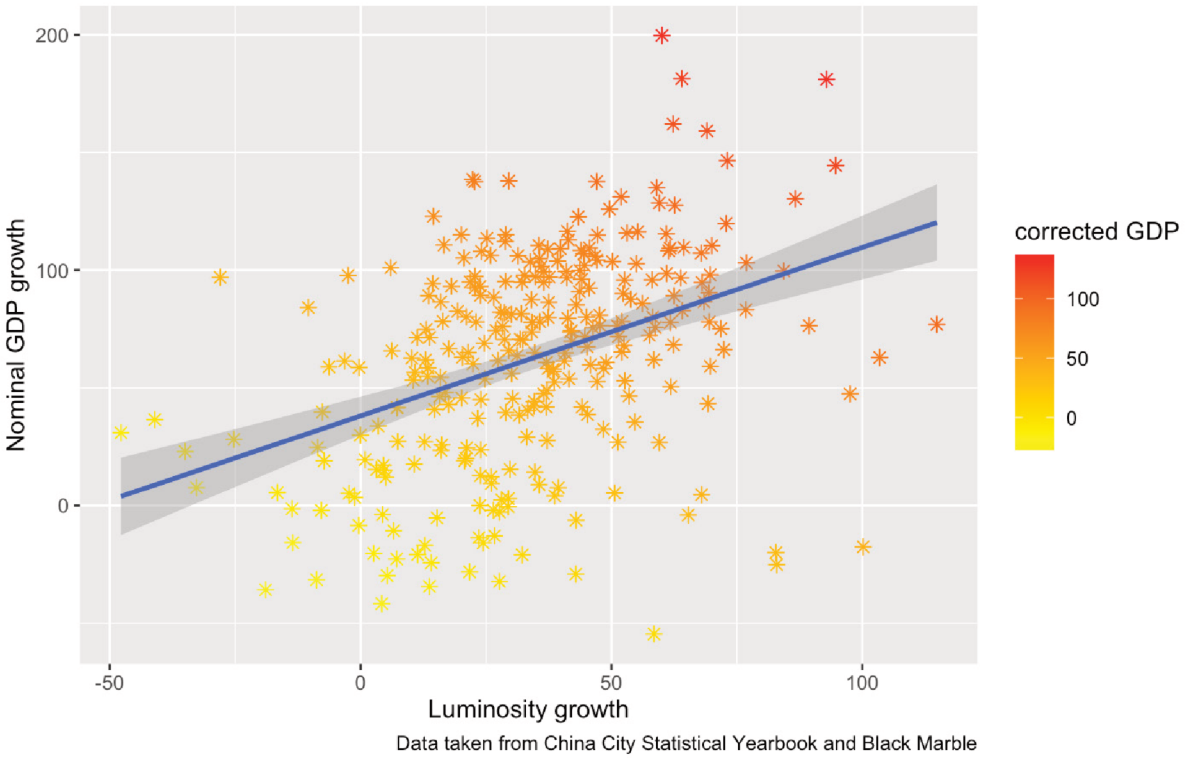
Another advantage of Black Marble data is that corrections for stray light and snow allow more nights to be used for annual composites. The 2012 annual composites covering China rely on images from 130 nights, on average, yet related products rely on just 90 nights and DMSP composites use only 60 nights; by 2019 the gap favouring Black Marble is even greater (X. Zhang & Gibson, 2022). A measure of annual economic activity based on lights seen on almost one-half of the nights of the year should better capture seasonal and other intra-year fluctuations than data using less than one-sixth of the nights of the year.

The final relevant aspect of our NTL data is that any errors (in measuring luminosity, or in luminosity as a proxy) should be independent of errors in reported GDP. The combination of these two noisy measures will thus be a better

4 The official unemployment rate is usually truncated around 5% for political reasons, thus not being a valid indicator for actual unemployment. Furthermore, time series between 2012 and 2019 are implausibly discontinuous, indicating that it is also not a reliable indicator for the number of registered unemployed potentially entitled to unemployment benefits. We thus chose to exclude the indicator.

5 Some economists even treat the GDP growth rate as an input to the economy, rather than an outcome, due to the politicized nature of growth targets (Klein & Pettis, 2021).

Plot A2.2a: Luminosity Growth and GDP (Considering Light-Corrected GDP)

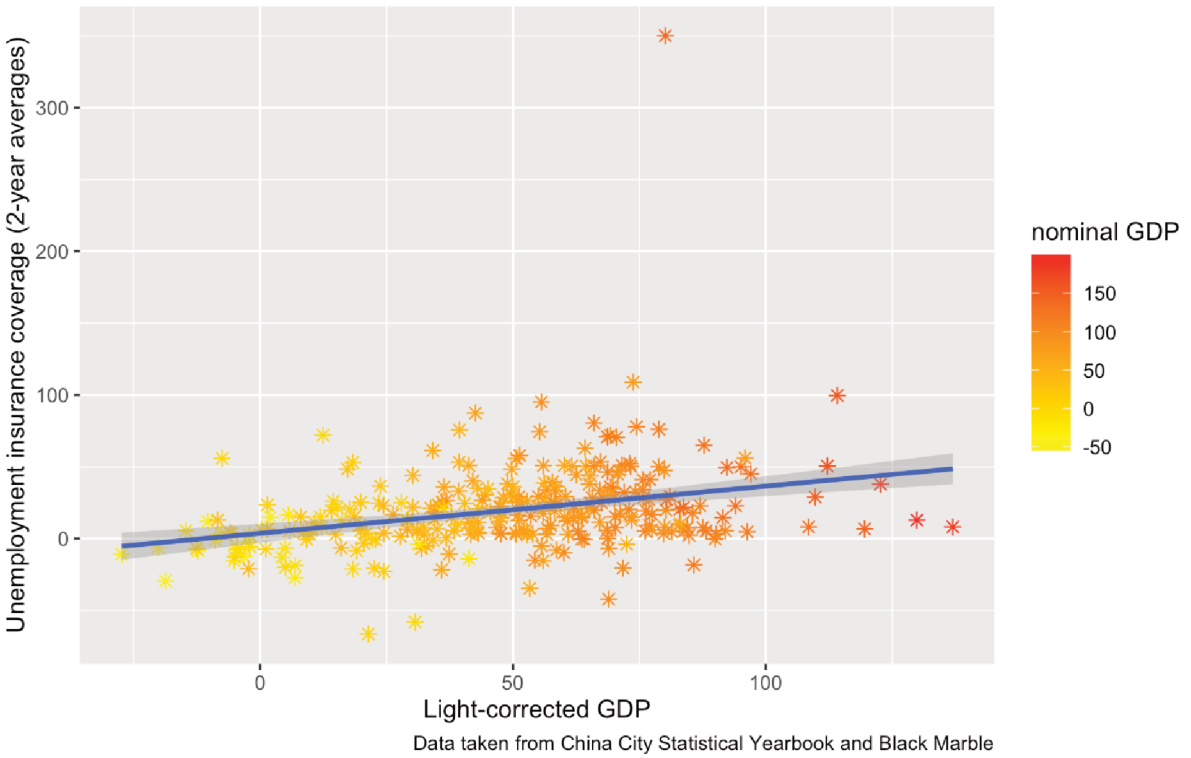


proxy for true unobserved growth than using just one measure (Henderson, Storeygard, and Weil 2012). Therefore, we create a light-corrected GDP indicator by averaging over the growth rates of GDP and lights. The outcome has a correlation with the original GDP growth rates of 0.96, whereas the correlation between original GDP and luminosity growth is 0.69. The independence of the errors in night lights and official GDP data cancels out errors in each of the original indicators, and thus avoids potential bias in the calculations.

The first scatter plot illustrates how growth in night lights (on the x axis) relates to growth in official GDP figures (on the y axes), and the colouring reflects light-corrected GDP. There is clearly a positive connection, represented by the regression line, which strongly corresponds to the colouring. As noted above, the independence of errors in night lights and official GDP figures helps decrease bias in the light-corrected GDP indicator by adding to growth in cities with low reported growth rates but higher light emissions, and by decreasing growth rates in cities with high reported growth but low light emissions.

The second scatter plot illustrates the connection between our light-corrected GDP indicator and coverage growth. Cities with higher growth rates in light-corrected GDP also had higher growth rates of unemployment insurance. The colouring reflects official GDP figures, and aligns relatively well to the overall connection, indicating that official GDP figures pick up a substantial part of economic development in mainland China. But there are many visible corrections: some points are lighter than their immediate environment, indicating that official GDP figures understate economic activity. In such cases, luminosity data may for example capture informal economic activity that GDP figures do not capture. In other cases, points are darker than their immediate environment. Here, city governments may have over-reported growth to cover up local economic contraction, or to meet official growth targets without issuing loans for

Plot A2.2b: Unemployment Insurance and Light-Corrected GDP (Considering Nominal GDP)



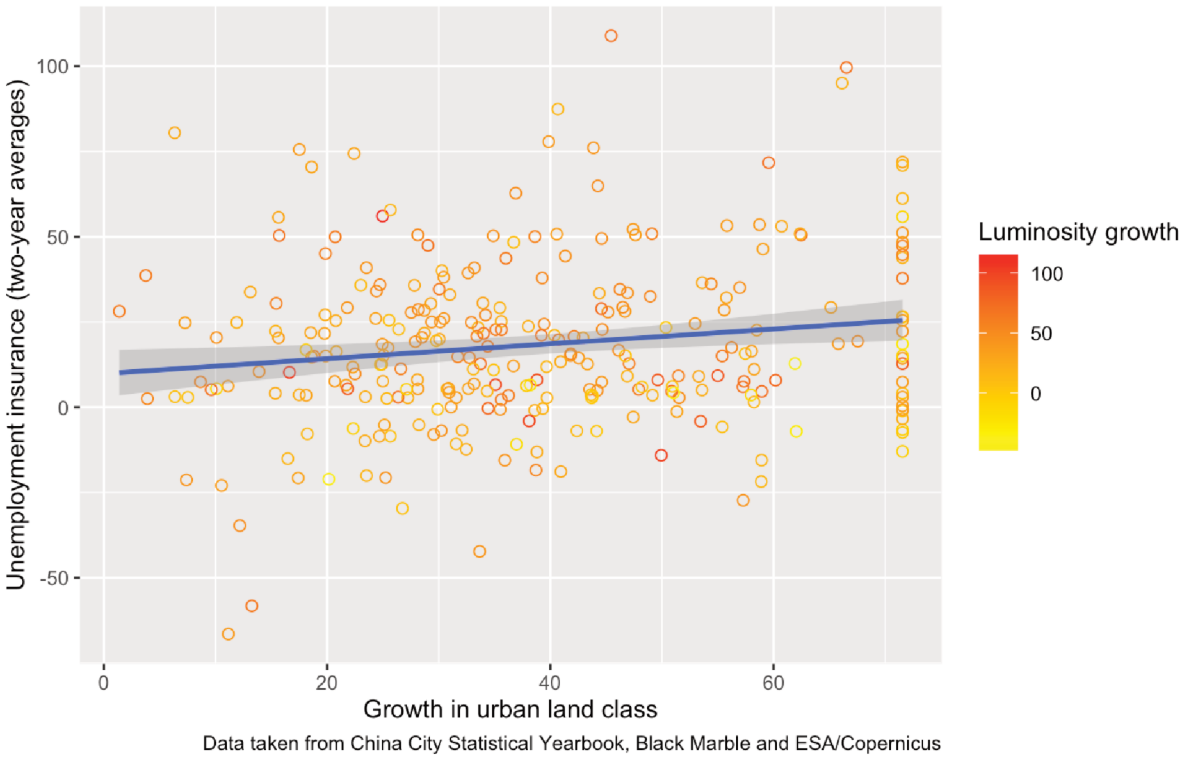
unproductive economic activity. Light-correction does not change the overall connection between the variables, but it reduces the bias of GDP figures.

A2.2.2 Urbanization data

We also use satellite-detected land cover data to provide an independent measure of the rate of urban area growth of each city between 2012 and 2019. While China's City Statistical Yearbooks report built-up area, only indicators with sub-optimal validity were reported throughout the period of observation. More generally, there are concerns about the accuracy of these figures. One concern is that higher levels of government set limits on converting rural land to urban use but local governments rely on land sales and taxes on urban construction for much of their revenue (Lichtenberg & Ding, 2009). Hence, a possible conflict of interest arises that may create incentives for local governments to under-report the amount of newly built urban area. The administrative data may also be incomplete because in the secondary market an entity who obtains land use rights from the state may rent or mortgage that land to others and this often involves a change in land use that is not registered with the authorities (S. Liu & Schmitt, 2014).

We use land cover data provided by the Copernicus Climate Change Service (C3S, 2019), which is the earth observation component of the EU Space Programme. These global land cover maps categorize the land surface into 22 classes at a spatial resolution of 10 arc-seconds (corresponding to about 250 meters at China's latitude). One of these 22 classes is urban area, based on artificial and impermeable surfaces. The estimates are updated yearly, with images obtained from the Sentinel 3 satellites, and are consistent with earlier global annual land cover data for 1992 to 2015 produced by the European Space Agency (ESA) Climate Change Initiative (CCI). We calculate the sum of built-up area by city in 2012 and in 2019 to derive the rates

Plot A2.2c: Unemployment Insurance and Urban Land Class Growth



(Considering luminosity growth, excluding Lhasa)

of change. There are a few extreme outliers for urban growth, especially in Ningxia. Shizuishan tripled in size, and Zhongwei and Yinchuan almost doubled. The median of growth is 36.1% and the mean is 44.1%. For the analysis, we limit the impact of outliers by winsorizing the indicator with a threshold of 0.05. The scatter plot illustrates the connection between winsorized growth of urban areas and unemployment insurance coverage. Overall, the connection is positive. The colouring reflects luminosity growth, which does not display a clear connection to the growth of urban space.

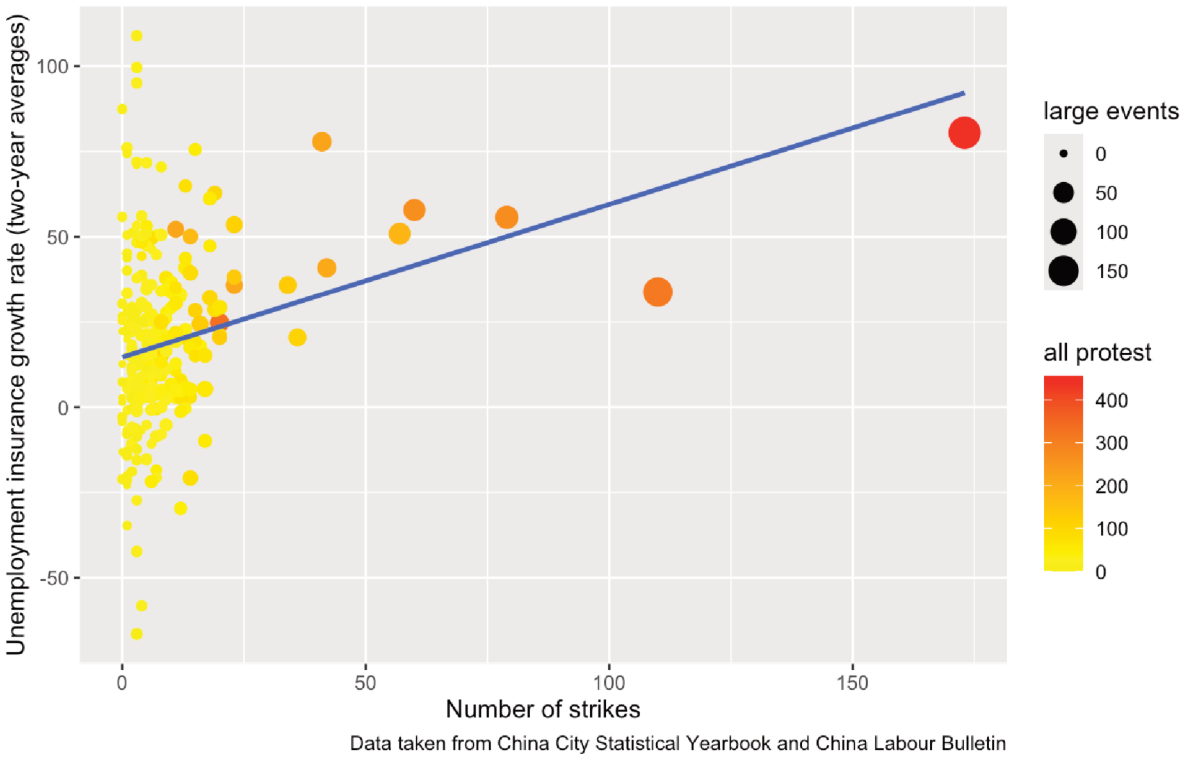
A2.3 Labour Unrest data

The China Strikes data collected by the China Labour Bulletin in Hong Kong is the most comprehensive collection of labour unrest data in the 2010s. This data is very rich and provides descriptions of many of the events, including standardized variables regarding the nature of demands, the type of event, the scope of the event, and the reaction by management and local authorities.

We choose a dual strategy to deal with the particularities of strike data in an authoritarian context: limiting our analysis to readily interpretable indicators covering at least 10% of total events and integrating these indicators at the very end of the analysis. Doing so helps us minimize distortions due to the potentially uneven spatial spread of the networks of the China Labour Bulletin, and the resulting possibility that strike events in some areas have a higher probability of being recorded than in other areas.[6] Other limitations stem from the time of observation: China Labour

6 Given the often-illegal nature of strikes and the opaqueness of the political regime, it is prudent to expect even the best data available not to provide an exhaustive picture of all labour protest in China. We expect those networks to be particularly strong in Guangdong, and in industrial regions like the Yangzi Delta. Conversely, we expect them to be weaker in the hinterlands and in smaller cities and rural areas.

Plot A2.3: Unemployment Insurance and Labour Mobilization



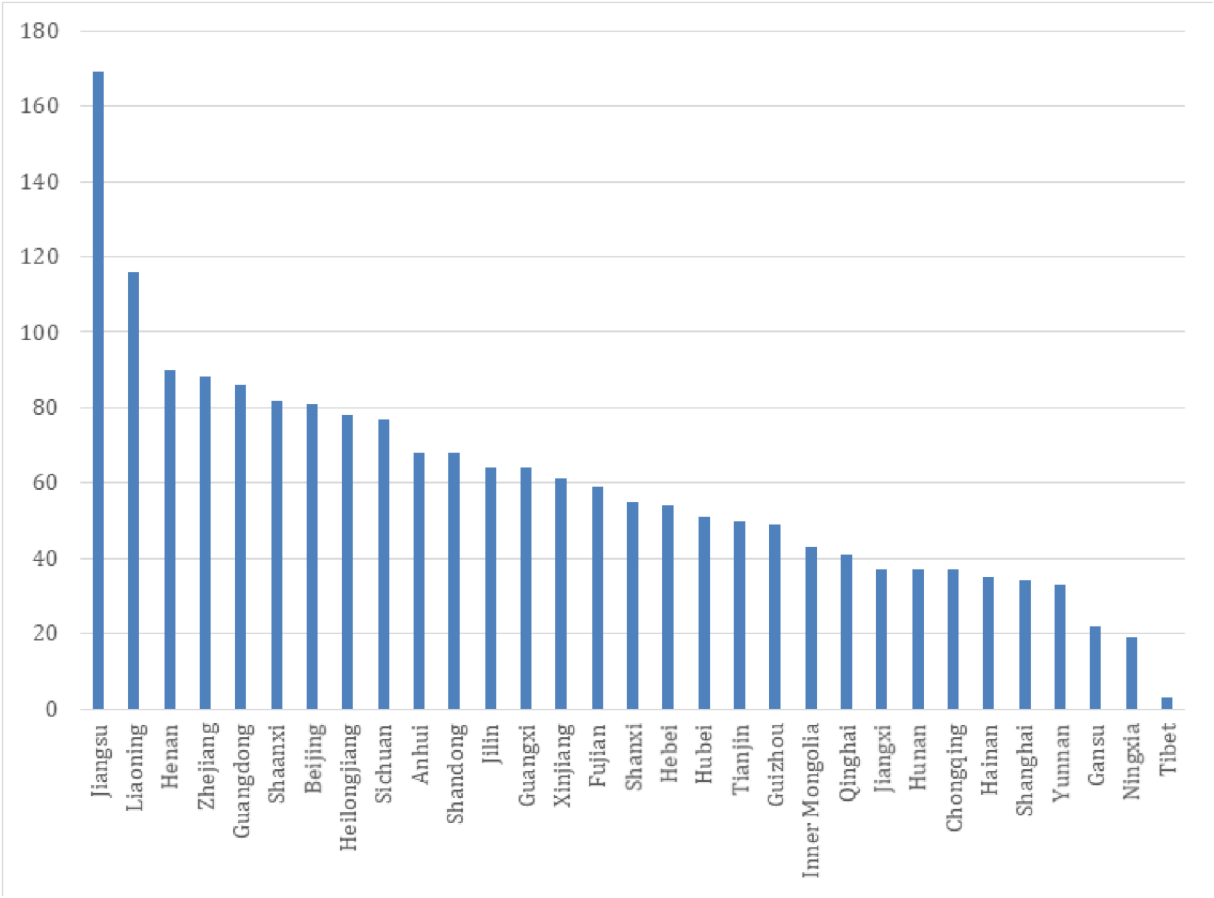
Bulletin data is available starting from 2011. We cannot calculate a proper growth rate in the incidence of strikes between 2012 and 2019, because we lack full data for the reference years 2004 to 2011. For example, a large strike wave in Guangdong in 2010 is not covered by the data (Butollo & ten Brink, 2012), but may have facilitated rapid expansion of unemployment insurance coverage in 2012 and 2013. Instead of growth rates, we use three event counts: the total number of recorded protest events, the number of strikes, and the number of protest events with more than 100 participants.

The scatter plot illustrates the standardized count of strikes in each city in relation to the growth rate of unemployment insurance. The size of the points is adjusted to the standardized count of large-scale events with more than 100 participants, and the colouring is connected to the standardized total count of protest events. The cities with the largest standardized number of strikes are Shenzhen, Dongguan and Guangzhou in the Pearl River Delta, followed by Shanghai and Suzhou in the Yangzi Delta. Those are followed by Chongqing and Chengdu, which are in turn followed by Huizhou and Foshan in the Pearl River Delta again. On the one hand, the chance of recording a strike should be higher in regions close to Hong Kong and where the networks of China Labour Bulletin are strong. On the other, we expect Shenzhen and Dongguan to be meaningful outliers, because the observability of strikes in neighbouring Huizhou and Foshan should be like Shenzhen and Dongguan, but their numbers are much lower.[7] The case

7 It is somewhat difficult to disentangle meaningful information and bias in the strike data. On the one hand, we would expect more labour mobilization in the coastal areas. Industry is concentrated here and went through contraction and conflict due to automatization and the relocation of low-value-added and polluting industries further inland. Furthermore, the stagnating supply of cheap migrant labour and the previous experiments in Guangdong with collective bargaining support the idea that labour mobilization is concentrated in those areas. On the other hand, the networks of China Labour Bulletin may also be concentrated in those areas. The

Global Dynamics of Social Policy CRC 1342
socium Research Center on Inequality and Social Policy

Plot A2.4: Number of Documents by Province



of Shenzhen is most likely distorting the picture. So, some tests for influential outliers in the final models will be in order (see below).

A2.4 Administrative documents

We compiled a corpus of relevant regulatory documents to extract relevant information on sub-national coverage and pooling arrangements via machine learning text classification, and other text mining techniques. Relying on Beijing University's PKUlaw database (*falü fagui shujuku*) we collected 1,935 relevant documents. To restrict the search to meaningful documents, we selected only those the title of which featured the words "unemployment insurance" (*shiye baoxian*). The documents were downloaded as txt files and subsequently loaded in an R data frame. We used conventional text-mining techniques such as regular expressions to identify core information such as the title, the year of enactment, the jurisdiction and the administrative code and number of the documents. Then, we created an Access database with the documents and used forms and SQL queries to check the results for consistency.

A2.4.1 Jurisdictions covered

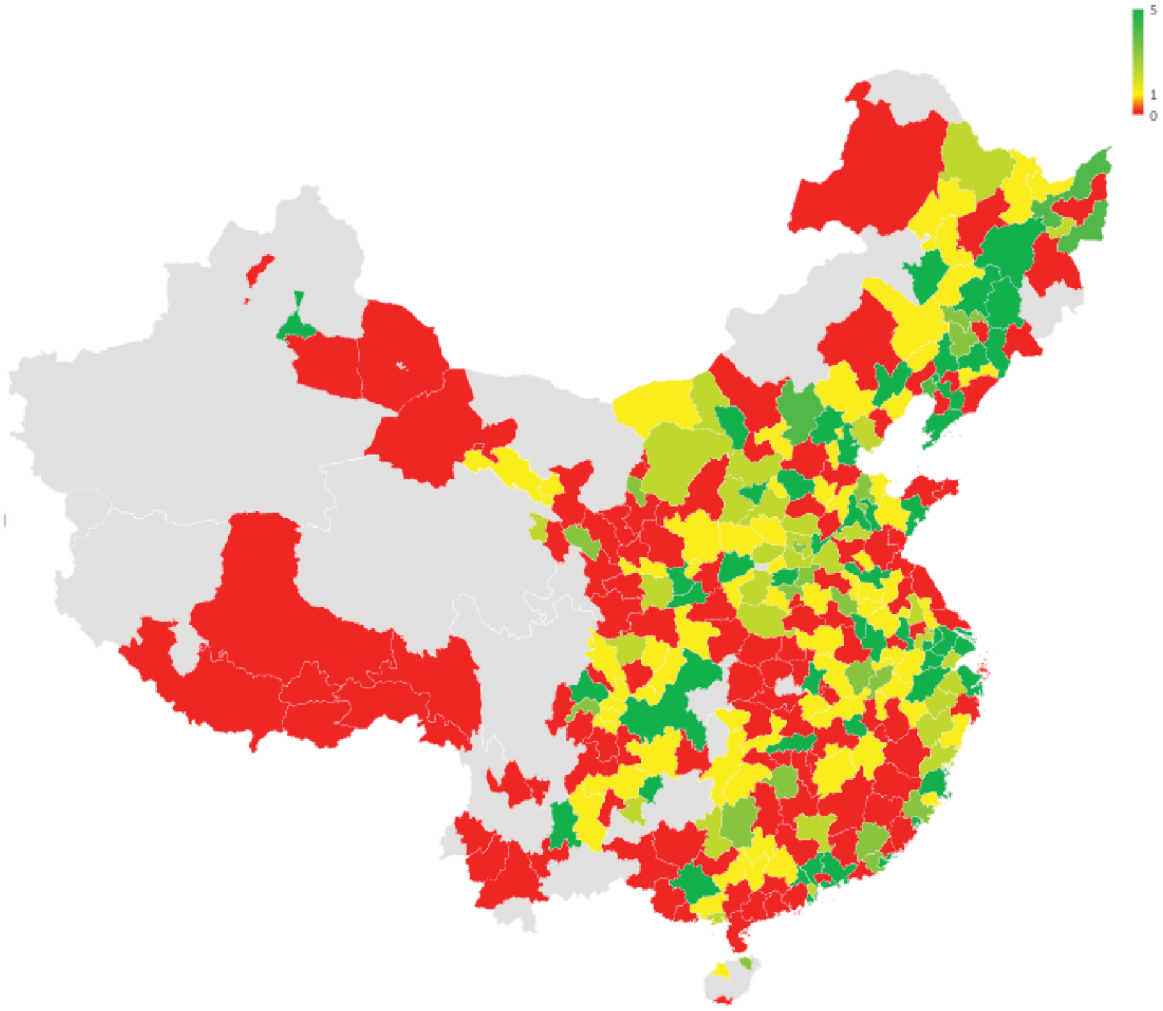
Overall, we identified 84 documents by the central level; 802 documents of territorial provinces;[8] and 1016 documents by provincial and prefectural cities,[9] the units of analysis in our

likelihood of them overlooking strike events in smaller cities and inland areas is arguably higher, which facilitates a concentration of values at the lower end. Due to the current political climate, it was not possible to contact China Labour Bulletin to talk about this.

8 Additionally, 5 documents were issued jointly by central and provincial organs.

9 Additionally, there were 12 documents by their subordinate units at county level, and 16 documents by prefectures who had not yet attained city status.

Map A2.4 Distribution of Documents at Prefectural City Level



data set. All 31 provincial jurisdictions are represented, but only 179 prefectural and provincial cities. The figure and map below illustrate the distribution of the documents across different jurisdictions. At provincial level, the number of documents arguably reflects population size as well as political priorities and the local history of unemployment insurance.

The map is coloured as follows: red jurisdictions are those with no documents available; yellow ones have at least one document available; and dark green represents 5 or more documents. The map includes various jurisdictions that were not in our dataset for calculation, especially prefectural cities formed from prefectures during the 2010s in Tibet and Xinjiang, for which no documents are available. Apart from that, the cities without documents are spread evenly across the provinces. However, there is some concentration on a line from the Southeast to the Northwest, and in the mountainous Southwest, which partly overlaps with regions of low coverage illustrated in Map 1 in the manuscript.

A2.4.2 Text classification for coverage regulations

We conducted text classification to quickly identify documents and sentences regulating coverage. To this end, we split the documents into a total of 48,136 sentences, from which 7,472 sentences were selected randomizing across documents and sentences. Three research assistants coded the sentences in line with instructions from the first author, who cross-checked all the positive examples and conducted consistency

Global Dynamics of Social Policy CRC 1342
socium Research Center on Inequality and Social Policy

Table A4.2: Classifiers

	Core specifications	Accuracy	Recall	Vectorization
Logistic Regression	L2 penalty, C = 10, Thresholds: 0.5 | 0.3	87%|76%	79%|90%	FastText embeddings
Bagging	N_estimators = 500, Rpart-like defaults	86%	87%	Multihot
Boosting (Adaboost)	N_estimators = 500, Learning rate 0.8	95%	79%	Multihot
Chinese MacBERT-large	Token length: 300	96%	81%	BERT embeddings

Software: Python 3.10, sklearn version 1.6.1

Note: Code files are available at this repository: https://github.com/arminmueller81/Working_Paper_Unemployment_2025.git

checks across the data set. Overall, only 498 sentences conveyed information directly regulating coverage or indirectly pointing to the inclusion of groups for whom participation was not mandatory. This sums up to only 6.67% positive observations in the sample. We trained machine learning models focusing on two types of vectorizations of the textual data: FastText embeddings, and multihot encoding.

We tried a broad range of classifiers, and got comparatively good – though not ideal – performance from logistic regression, bagging, boosting, and the Chinese version of MacBERT. We created an index from the classifications of the latter three, while consulting probabilities from logistic regression separately. For analysis, we summed up the index and the probabilities for each document, and then grouped the documents in descending order. We subsequently analysed the documents one by one, starting with central documents and proceeding towards the lower levels of government. For each document, we first checked the paragraphs that included the sentences with the highest index values, and then checked for relevant keywords and in typical places (especially the beginning and the end of a document) to detect false negatives. The extracted information was organized in an excel file, which includes central regulations and summaries of provincial and prefectural level regulation. This allowed for cross-checking whether regulations at prefectural level were more inclusive than the respective provincial regulations.

Finally, we used targeted SQL queries to extract additional regulations using non-standardized language that the algorithm did not detect on a sentence level, but which turned up during the analysis of the coded regulatory documents. These largely pertain to self-employed persons, for whom there are different denominations, each of which features rarely in the corpus. We extracted the relevant paragraphs and organized them in an Excel file.

A2.4.3 Text mining for the pooling level

Identifying the level of social pooling was somewhat simpler because the relevant formulations always involve the word "pooling" (*tongchou*) and the respective level of government. The challenge here is rather that the word *tongchou* may also mean coordination or harmonization. We therefore used simple SQL searches to identify relevant documents at different administrative levels and read the relevant sections in context to determine which meaning of *tongchou* applied. We extracted the relevant paragraphs and organized them in an Excel file.

Appendix 3: Regressions with robust standard errors

Table A3: Regression analysis with long differences and robust standard errors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Intercept)	4.396**	9.438***	0.487	1.160	-1.708	-0.546	-4.518	-3.128	-3.669
	(1.636)	(1.888)	(3.137)	(2.921)	(2.945)	(2.953)	(3.219)	(3.129)	(3.164)
GDP (light-corrected)	0.289***	0.181***	0.193***	0.146***	0.107**	0.110**	0.125***	0.123***	0.129***
	(0.032)	(0.037)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.037)
On-the-job employees									
0.308***									
					0.244***	0.193***	0.160**	0.160**	0.154**
0.289***									
0.173**									
		(0.060)	(0.058)	(0.056)	(0.054)	(0.054)	(0.058)	(0.059)	(0.059)
Urban space growth			0.216***	0.243***	0.268***	0.266***	0.294***	0.286***	0.293***
			(0.058)	(0.055)	(0.055)	(0.055)	(0.056)	(0.055)	(0.055)
Population 1: permanent				0.557***		0.322***	0.221**	0.264**	0.254**
				(0.070)		(0.081)	(0.079)	(0.081)	(0.081)
Population 2: registered					1.856***	1.202***	1.196***	1.204***	1.205***
					(0.243)	(0.292)	(0.288)	(0.294)	(0.292)
Protest events							0.059***		
							(0.018)		
Protest events (strikes)								0.156*	
								(0.063)	
Large protest events									0.177**
									(0.066)
Num.Obs.	287	287	287	287	287	287	287	287	287
R2	0.124	0.175	0.198	0.292	0.298	0.317	0.328	0.323	0.325
R2 Adj.	0.121	0.169	0.190	0.282	0.288	0.305	0.314	0.308	0.311

Note: Significance levels: +=0.1, *=0.05, **=0.01, ***=0.001 Note: Significance levels: +=0.1, *=0.05, **=0.01, ***=0.001

Appendix 4: Model diagnostics

A4.1 Multicollinearity

The correlation plot illustrates the degree to which the different variables are correlated. Correlations are highest between the labour mobilization variables, which were only used in separate models. Furthermore, there is high correlation between the permanent and hukou population, and between light-corrected GDP and urban growth. However, the model table illustrates that the combination of those predictors has not led to an escalation in explained variance.

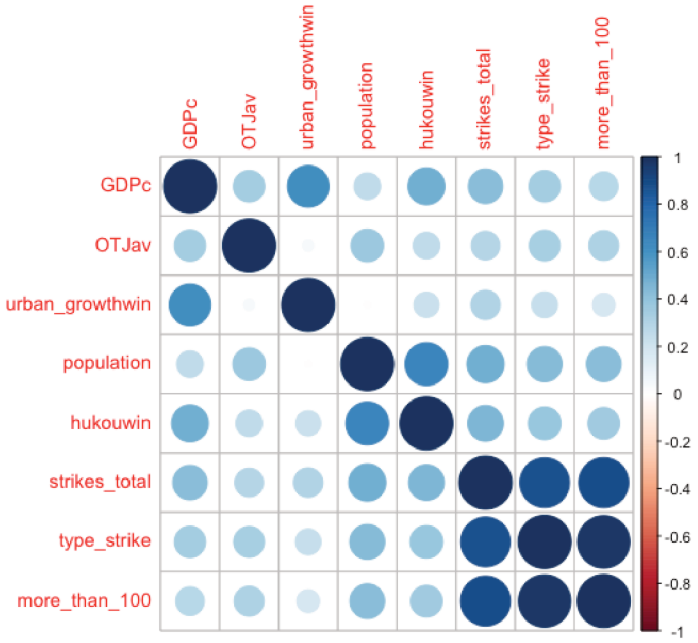
A4.2 Diagnostic plots for Model 6

The models were calculated with two-wave panel data using the plm package for panel data analysis in R. In the panel data, we first calculated the 2019 and 2012 values relative to the 2012 values, rendering 100% for 2012 and the growth rate plus 100% for 2019. Then, we subtracted 100% from both. Calculating OLS on the 2019 cross-section thus renders numerically equivalent coefficients, but less reliable standard errors. For the plots below, the models were re-created with the lm() function for linear models, which only features a cross-section of the growth rates. This calculation renders invalid standard errors, but the estimates are the same, which allows for some standard OLS diagnostics.

A4.2.1 Unbiasedness of estimates

As plot 1 illustrates, the model is roughly linear in parameters, and the conditional mean of the errors is near zero. The model can thus be considered unbiased. Plot 2 illustrates that none of the observations features a Cook's Distance value above 0.5 (with the city of Lhasa having been removed from the data set, and the variables for registered population and urban growth winsorized). Hence, there are no outliers inducing bias.

Plot A3: Correlation Plot



A4.2.2 Validity of variance estimates

The unbiasedness of variance estimates and the BLUE[10] character of the OLS estimator depend

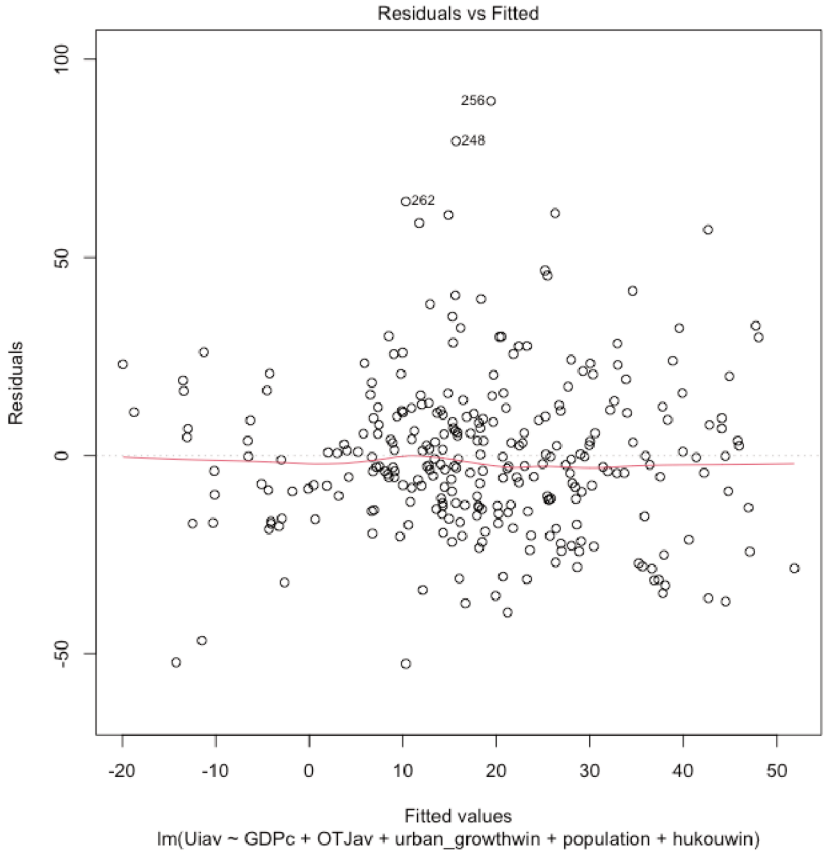
10 Best Linear Unbiased Estimator

```
Uiav_M6 <- plm::plm(Uiav ~ GDPc + OTJav + urban_growthwin + population + hukouwin,
                    model = "fd", index = c("City_ID", "year"), data = data)
summary(Uiav_M6)
```

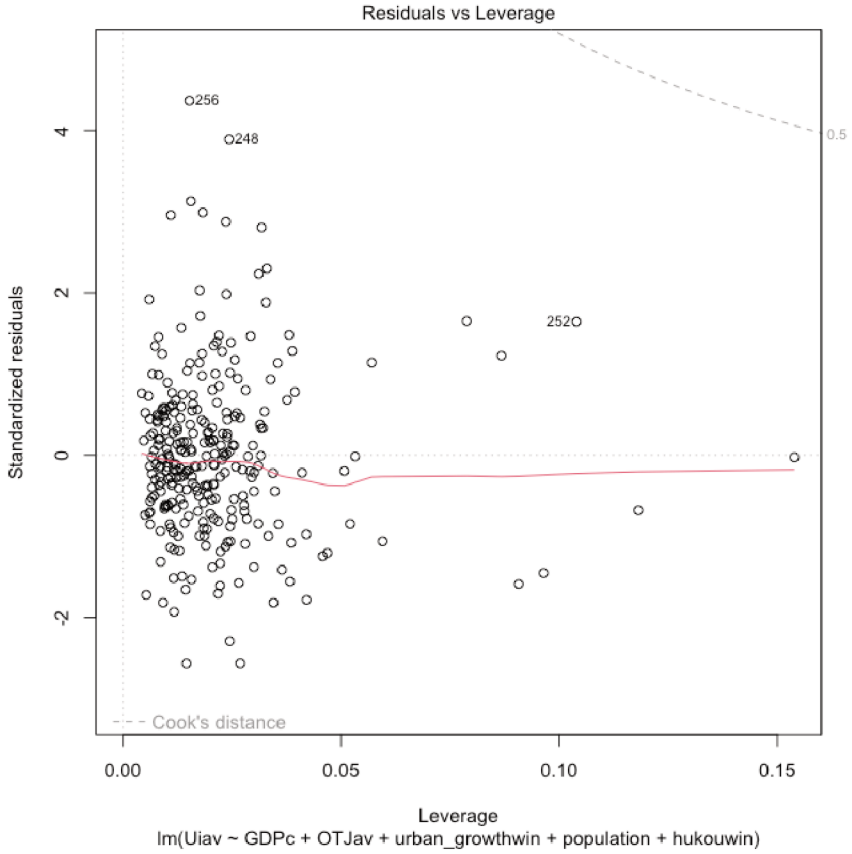
on the assumption of homoskedasticity. The standardized residuals in Plot 3 point to some heteroskedasticity problems. As noted above, robust standard errors decrease the standard errors, rather than increasing them, thus not improving on the problem.

Also, not all residuals are normally distributed, as plot 4 illustrates. In a survey-based study, a violation of the assumption of normally distributed errors would call into question the generalizability of the models to the underlying population. As this study is not based on sampling but on a complete survey of the population, the import of this violation is limited to robustness against measurement error.

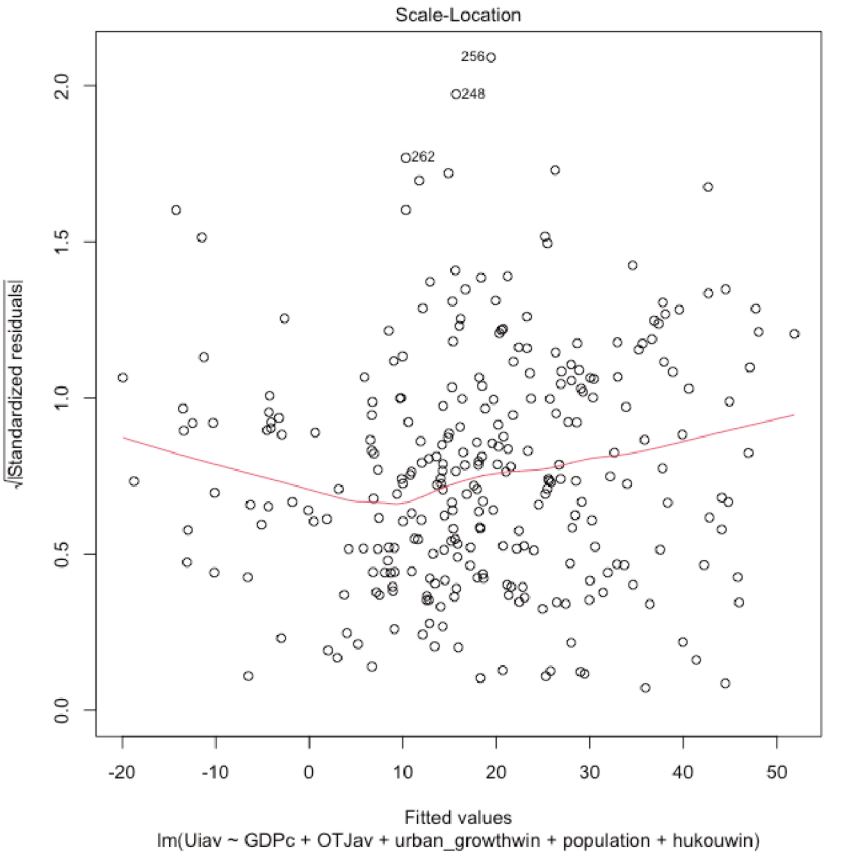
Regression plots: Plot 1



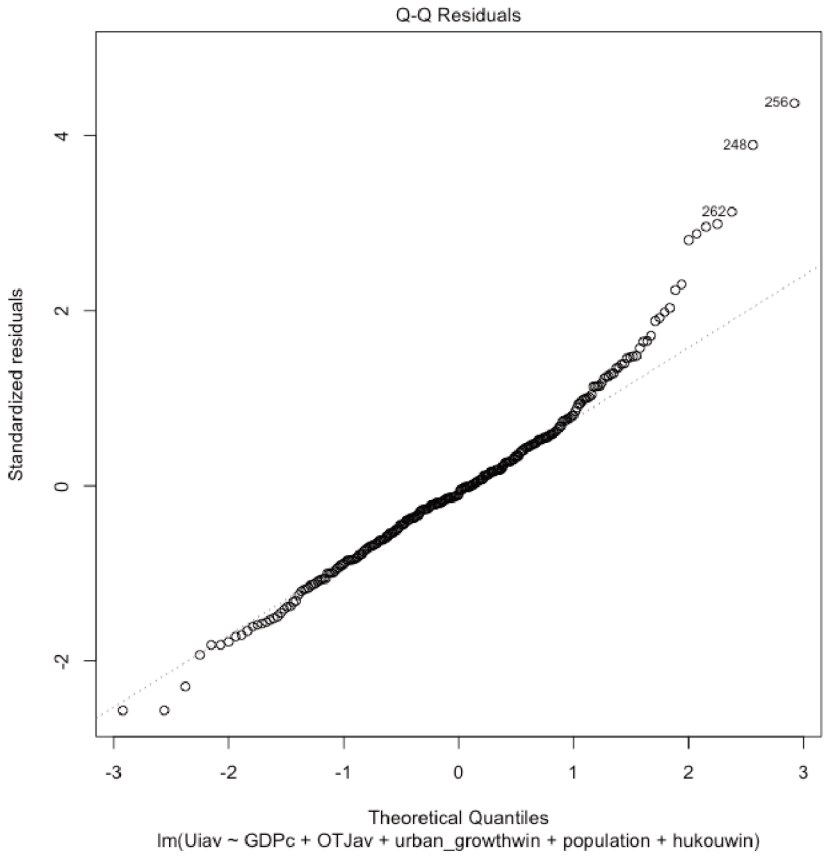
Regression plots: Plot 2



Regression plots: Plot 3



Regression plots: Plot 4

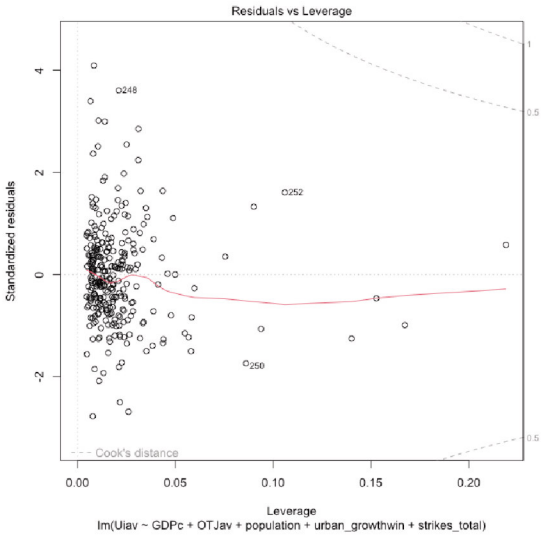


A4.3 Outlier diagnostics for strike models

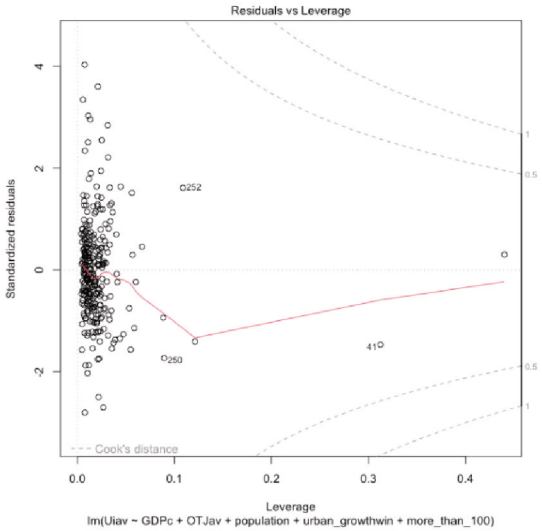
As the plots for model 7 to 9 illustrate, there are no influential outliers in the models with labour mobilization variables.

Outlier diagnostic plots

Model 7



Model 9



Model 8