

Employment Structure and the Rise of the Modern Tax System*

Anders Jensen [†]

January 2016

[Job Market Paper - most recent version [here](#)]

Abstract

This paper studies how the transition from self-employment to employee-jobs over the long run of development can explain growth in income tax capacity. I construct a new database which covers 90 household surveys across countries at different income levels and 140 years of historical data within the US (1870-2010). Using these data, I first establish three new stylized facts: 1) within country, the share of employees increases over the income distribution, and increases at all levels of income as a country develops; 2) the income tax exemption threshold moves down the income distribution as a country develops tracking employee growth; 3) the employee share above the income tax threshold remains high and constant at 80-85 percent. These findings are consistent with a model where a high employee share is a necessary condition for taxation and where the rise in income covered by information trails through increases in employee shares drives expansion of the income tax base. To provide a causal estimate of the impact of employee share on the exemption threshold, I study a state-led US development program implemented in the 1950s-60s which shifted up the level of employee share. The identification strategy exploits within-state changes in court-litigation status which generates quasi-experimental variation in the effective implementation date of the program. I find that the exogenous increase in employee share is associated with an expansion of the state income tax base and an increase in state income tax revenue.

*I thank Oriana Bandiera, Tim Besley, Henrik Kleven and Emmanuel Saez for relentless support and continued counsel on this project, and Torsten Persson for detailed comments. I thank Philippe Aghion, Miguel Almunia, Alan Auerbach, Pierre Bachas, Michael Best, Anne Brockmeyer, Robin Burgess, Benjamin Faber, Francois Gerard, Hilary Hoynes, Ethan Ilzetski, Michael Keen, Camille Landais, Peter Lindert, Guy Michaels, Gerard Padro-i-Miquel, Joana Naritomi, Dina Pomeranz, Joel Slemrod, Johannes Spinnewijn, Silvana Tenreyro, John Van Reenen, Danny Yagan, Gabriel Zucman, and seminar audience and conference participants at UC Berkeley, London School of Economics, IIES, ZEW, the 71st Annual Congress of IIPF, the 30th Annual Congress of EEA, for helpful comments. I am grateful to university librarians Paul Horsler (LSE) and Jesse Silva (UC Berkeley) for their help in locating historical US financial and employment data; Peter Lindert (UC Davis) for kindly providing me with income and occupation estimates for the US in 1870; and Jon Bakija (Williams College) for kindly providing me with access to the US historical state income tax calculator program.

[†]Department of Economics, London School of Economics. Contact: a.jensen1@lse.ac.uk.

1 Introduction

Tax capacity grows as economies develop. This is true across today's developing countries, and historically within today's advanced countries. In this paper, I show how the transition from self-employment to employee-jobs explains growth in tax capacity. This increase in employee share is a defining characteristic of changes to employment structure over the long run of development. Micro evidence shows that transitions into employee-jobs are associated with improved compliance at the individual level by creating third-party information trails (Kleven et al., 2011), but this evidence has little to say about state tax capacity over the long run. Macro evidence provide correlations between employee shares and tax take (Besley & Persson, 2014; Kleven, 2014; Kleven et al., 2015), but this evidence lacks clearly identified empirical channels. To build a bridge between the micro and macro contributions, I propose a research design which combines descriptive evidence and quasi-experimental evidence. In this design, I empirically identify a new channel through which employee share impacts tax capacity along the development path. To implement the design, I construct a new dataset with microdata for 90 countries at all levels of development and 140 years within the US (1870-2010).

The novel channel explains decreases in the income tax exemption threshold through increases in employee share that occur further and further down the income distribution. To motivate the channel, Panel A of Figure 1 shows four countries at increasing levels of development [India, Indonesia, Mexico, US]. Within each country, it plots employee share across deciles of the income distribution and the location of the exemption threshold above which earned income becomes liable for taxes. In India, the exemption threshold is located in the top decile of the income distribution, the only one where employee share is high. As countries reach higher levels of development, the threshold gradually moves down as the employee-share goes up in deciles further down the income distribution. This close co-movement is also observed within the US over time (Panel B, Fig 1). This paper explains the co-movements as the impact of increases in employee share on the exemption threshold.

Figure 1 shows that the statutory income tax base in US is 30 times larger than in India (Panel A). This large variation suggests the statutory threshold may be an important determinant of tax capacity. This observation complements previous papers that document a weak or zero correlation between statutory instruments and income tax capacity. While those studies have focused on variation in statutory tax *rates*, I study variation in the statutory tax *threshold*. This directly addresses the extensive margin of compliance of moving workers into the tax base. The focus on the threshold is

consistent with development tax policy which emphasizes the extensive margin of compliance as an important driver of tax capacity (Keen, 2012)

I construct a new micro database containing information on type of work and income from nationally representative household surveys across 90 countries at all income levels and within the US between 1870 and 2010. This collection effort offers two main advantages. First, it allows me to shed new light on employment structures in countries at medium and lower levels of development (below \$4000 per capita), where I source 83 percent of surveys directly from national statistics offices or government ministries. These surveys have a larger sample size and contain income data for a wider range of work-types than pre-existing surveys from public access licensed databases or external repositories. Second, it allows me to provide new evidence on long run changes to employment structure in the US pre 1900, where I draw on previously unexploited microdata produced by a joint effort between economic historians and IPUMS USA (Lindert & Williamson, forthcoming). My micro survey data offer the key advantage, relative to administrative records, that they permit studying employment structure for all types of work both above and below the exemption threshold.

In a first part of the paper, I provide three new stylized facts on employee shares and taxation. Stylized fact #1 shows that within country the employee-share is increasing through the income distribution, and over development the employee share is increasing in deciles locally further down the income distribution. I characterize stylized fact #1 in a quantitative exercise. I show that for each decile of the income distribution, there exists a level of per capita income beyond which employee growth in the decile stops and employee share reaches a steady state level. I estimate these per capita income levels for all deciles to find that: the income levels are decreasing in the deciles; and, the steady state employee share is constant across deciles at 80-85 percent.

The second stylized fact shows that the exemption threshold moves gradually down the income distribution in close co-movement with increase in employee share in the deciles locally to its left. The third stylized fact shows that the employee-share above the threshold remains constant and high at 82-85 percent at all levels of development. All three stylized facts are remarkably similar *in levels* when I compare a given US historical profile to a synthetic profile constructed from countries at similar levels of development. This robustness suggests a causal impact of employee share on the tax threshold.

In a second part of the paper I estimate the causal impact of employee share on the exemption threshold. Identification is based on exogenous timing in implementation dates of a US state-led development program. Through the Industrial Development Bonds program (IDB), states facilitated

the transition into manufacturing employee-jobs by constructing debt-financed leasable industrial facilities in rural areas. Implementing IDB required the state House to vote in a legal amendment to exempt IDB from the constitutional 'public purpose' provision whereby the state may only enter the debt market for a public purpose. But the lack of historical precedent to the IDB amendment meant the program was considered legally uncertain until the highest state court would litigate to uphold it.

The estimation strategy uses within-state changes in court litigation status as identifying variation. It assesses impact by comparing changes in outcomes before and after the upholding event, relative to counterfactual changes before and after the vote in event within the same state. I provide two pieces of evidence to support this identification strategy. First, I confirm graphically that the effects I find are driven by sharp "on impact" changes around the upholding event with clear break from the vote in pre-trend. Pretrends are stable regardless of the length between vote in and upholding (≥ 10 years in 40 percent of cases). Second, I show that, conditional on vote-in, the only significant predictor of faster timing to upholding is a time-invariant dummy for civil law origins. As an improvement upon cross sectional specifications, my empirical strategy remains identified even under time varying unobservable shocks *within state* to willingness or capacity to tax that coincide with the potentially endogenous policy period of vote in. Ultimately however, the causal interpretation of the results is driven by the exogenous timing of the upholding event, which I support by finding identical estimates in cross-sectional specifications (including paired synthetic matching) between IDB and non-IDB states.

I use changes in IDB court litigation status as an instrument to estimate the impact of employee share on the exemption threshold. I provide evidence to support the exclusion restriction that changes in IDB court litigation status impact the exemption threshold only through changing employee share. I first show changes in litigation status had no impact on proxies for other threshold determinants, including earnings structure, tax rate structure, enforcement capacity, demand for redistribution. Second, I find no impact of the upholding event on any other source of state tax revenue apart from income tax. Third, the program had no impact on labor force participation, migration, and sectoral spillovers. Fourth, I show that the impact on exemption threshold is sharp and occurs with a precise lag to the impact on employee share. Variation which violates the exclusion restriction would have to produce an immediate break from trend; occur with precise time lag to upholding event; not be captured by proxies; not impact any other tax instrument nor any other tax revenue.

In the first stage, I find that the program led to a large transition from self-employed to employee-jobs but had no impact on earnings structure, overall employment, rural migration, industry spillovers.

I find that the IDB program led to a large reduced-form decrease in the tax exemption threshold and increase in (income tax/GDP). The instrumented estimates suggest that a one standard deviation increase in length of IDB program through exogenous changes in IDB court litigation status led to increase in employee share which account for 24 percent of the expansion of the income tax base and 10 percent of the rise in (income tax/GDP).

The US instrumented estimate is a weighted average of the causal impact of employee share on the tax policy over complier states at different initial levels of employee-share. It has predictive value in the cross-development setting for three reasons. First, the mechanisms underlying the instrumented estimate closely match the cross-development stylized facts: the IDB impact on employee-share led to a shift leftward of the employee-share profile in the income distribution (consistent with fact #1); the IDB decrease in threshold was driven by employee-share increases occurring locally to its left (fact #2); the employee-share above the threshold remained constant at 80-85 percent before and after IDB (fact #3). Second, I find that complying states' labor markets were characterized by large underemployment amongst self-employed in rural areas - thus matching a key characteristic of rural labor markets in developing countries (ILO, 2009). Third, growth in US states' tax take over time has been driven by the rise of personal income tax - thus matching the key cross-development tax fact. State-led industrial financing programs similar to IDB have been implemented in India, Mexico and South Africa, and are considered elsewhere in developing countries (UN Financing for Development, 2009)

The results can be rationalized in a model where government maximizes revenue from setting the income tax exemption threshold. Employees and self-employed differ in evasion cost due to differences in information trail coverage. An exogenous increase in employee-share around the threshold lowers the fiscal costs of a local threshold decrease. I predict the threshold location in the 90 countries of the cross-section in the database, allowing only the employee shares over the income distribution to vary across countries, and holding constant values of enforcement and administrative capacity, earnings, demand for redistribution and public goods. The predicted tax base can account for 62 percent of the observed growth in tax base size across the 90 countries.

The following section discusses related literatures. Section III describes the micro database and provides new stylized facts on employment structure and tax structure. Section IV identifies the impact of employee share on the exemption threshold. Section V provides a model to rationalize the results and quantify the importance of the employee-share channel. Section VI concludes.

2 Related literature

This paper is related to the micro and macro studies on information trails as a determinant of individual compliance and state tax capacity. Kleven et al. (2015) show theoretically that collusive behavior between employees and the employer is hard to sustain when there exist business records, making third-party information reporting by firms a powerful tool of tax enforcement. Gordon & Li (2009) show how information reporting by financial institutions can also improve tax enforcement. These models are supported by a large set of empirical studies in both developed and developing countries that show tax enforcement is affected by information reporting. Kleven et al. (2011) use Danish random audits to show that increases in information coverage associated with employee jobs dramatically improves income tax enforcement. Best (2014) shows, using matched employer-employee data from Pakistan, that third-party reporting also limits income tax evasion in a developing country. But third-party reporting coverage of transactions remains much less prevalent in developing countries. Carillo, Pomeranz and Singhal (2014) use a natural experiment from Ecuador to show limits to third-party information effectiveness when taxpayers can adjust on non verifiable margins. Kumler, Verhoogen and Frias (2015) show that third-party enforcement of Mexican payroll taxes works better in larger firms and when employees have stronger incentives to monitor employer wage reports. Naritomi (2015) uses a Brazilian reform to show that increased incentives and improved technology of firm monitoring by consumers leads to a large increase in firms' reported revenues. In Chile, Pomeranz (2015) shows that randomized audit threats have less impact on transactions that are subject to double reporting from both buyers and sellers, indicating increased deterrence of evasion through double reporting. Finally, Bachas (2015) and Best et al. (2015) show, respectively in Costa Rica and Pakistan, that taxes based on turnover can be a useful alternative to corporate profit taxation because sales are easier to observe than profits. This paper provides micro evidence on changes in information trails related to employment-structure along the full development path.

The paper also relates to the literature on the determinants of government growth over development (see Cage and Gadenne (2015) for a comprehensive study of tax revenues in developing and developed countries). Demand side determinants of this growth include 'Wagner's law' whereby public goods have a income elasticity above one (see e.g. Musgrave, 1966); and, democratization and increased political power of the poor (Acemoglu and Robinson, 2000). Besley and Persson (2011, 2014) model investments in fiscal capacity as a response to demand for public goods and increased cohesive-

ness of institutions. This paper is more closely related to supply side studies that show how changes in economic structure impact the capacity to supply tax revenue (Bird and Oldman, 1964; Hinrichs, 1966; Kleven et al., 2015). I contribute by providing descriptive and identified evidence on a new tax policy channel through which economic structure affects the capacity to raise taxes.

The paper is related to the literature on changes in employment structure over development. Current evidence focuses on the cross-country stylized fact that self-employment declines over increasing levels of per capita income (Banerjee and Duflo, 2007; Gollin, 2008; 'Jobs' World Development Report, 2013; La Porta and Shleifer, 2014) Studies of structural transformation exploit both cross country and within country patterns, but focus on sectoral changes (review in Herrendorf et al., 2014). The exception is McNaig and Pavcnik (2014, 2015), who show that structural transformation in Vietnam 1990-2008 was accompanied by transitioning out of household businesses. This paper provides new evidence on the decline of self employment over development in larger cross-country and longer time series samples; and, at disaggregated levels over a country's income distribution. Using a previously unexploited US development program, I also contribute with identified evidence on an industrial policy's impact on sectoral and employment changes. I define self-employment versus employee-job based on whether the job generates an information trail relevant for tax enforcement. My self-employment category captures the informality category used by ILO (2009). Thus, this paper complements the literature on informality and development (La Porta and Shleifer, 2014), with micro based evidence on changes to informality over the income distribution along the development path. Finally, this paper's methodology relates to studies of macro economic changes using newly constructed micro evidence (Gollin, Waugh and Lagakos, 2013; Bick, Fuchs-Schundeln and Lagakos, 2015).

The paper also relates to studies of US states as laboratories in development and taxation. (Dincecco & Troiano, 2015; Gillitzer, 2013) While these papers study introductions of new tax instruments, I analyze determinants of state income tax exemption thresholds. I provide evidence on a previously unexploited state led development program, which differs from previous studies of US development programs, including Moretti & Kline (2013) in two dimensions. First, the program was narrow in scope as a stand-alone policy of industrial financing. Second, it was unique in its revenue-neutral funding scheme through issuance of revenue bonds. This type of state-led industrial financing is considered potentially important in developing countries, but there is little empirical evidence on its impacts (Platz, UN Financing for Development, 2009).

3 Descriptive evidence on employment structure and tax structure along the development path

I provide new stylized facts on changes to employee share and tax structure along the development path. I first describe the microdata and methodology. I then present descriptive and quantitative results.

3.1 Data and methodology

Data I construct the micro database from nationally representative household surveys. This represents two key advantages over alternative data sources. First, using household surveys allows a systematic study of changes to employee share over the income distribution which does not depend on the location of tax exemption threshold. This is in contrast to administrative records which typically measure earnings and type of work only for those above the exemption threshold. Second, using household surveys as opposed to firm surveys allows measures of types of work which are not restricted to activities for which individuals receive a wage, but cover all forms of self-employment and unpaid family work. This is especially important in less-developed countries where I document that the share of the workforce below the exemption threshold engaged in non-employee jobs is substantial.

I collected recent household surveys from 90 countries around the world at all levels of per capita income. I searched for surveys based on two main criteria: information on type of work and continuous measures of total gross earnings (as opposed to expenditure proxies) at the individual level; nationally representative coverage of all types of work. The first criteria allows construction of an income-distribution that is defined consistently across all countries. The second criteria ensures that all types of work are covered.

Many household surveys which are available as licensed data or through public access external repositories do not measure non-wage income.¹ This is the case, per example, for a significant number of Living Standards Measurement Surveys (LSMS). The absence of non-wage income measures in public access surveys is also more pronounced in least developed countries. Consequently, in these countries I sourced the surveys directly with the country's national statistics office, or in some cases, the Department for Planning. I collected 83 percent of surveys in countries below \$4000 per capita

¹Some datasets do contain step-function measures of non-wage income. I do not include such surveys in the database.

from such direct sources. These surveys have an average sample size 6.5 times larger on average than the country's LSMS (and are on average much more recent). I chose a living conditions survey over a labor force survey whenever possible. This is because the latter sometimes exclude certain types of work such as casual daily wage labor or family businesses.

The result is 90 surveys in countries ranging from \$125 per capita to \$80250 per capita. In 2 cases, I rely on expenditure to measure income. In another 2 cases, the sample is only representative of the urban population. In all other cases, the survey is nationally representative and contains continuous measures of earned income from all types of jobs. Full description of all surveys is in appendix B2.

I construct the within country dimension of the database by combining new and previously used micro sources from the US. I collect data between 1950 and 2010 from Census microdata extracted via IPUMS USA. Before 1950, the Census did not record work type and continuous measures of income at the individual level.² I use the 1935-36 Study of Consumer Purchases, which was jointly conducted by Bureau of Labor Statistics, the Bureau of Home Economics and the Department of Agriculture. Considered the precursor to the Census methodology of data on income at the individual level, the survey was meant to "ascertain for the first time in a single national survey the earning and spending habits of inhabitants of large and small cities, villages, farms." (ICPSR study 8908, 2009) 300,000 households were interviewed based on sampling units chosen to represent the "demographic, regional, and economic characteristics of the United States." (ICPSR, 2009) Both the work type and income categories in the 1935 survey are consistent with the later Census-based definitions. In particular, the gross income variable contains continuous measures of wage-earnings, business income, and farm income. I provide further details on the ICPSR dataset in the appendix.

All national surveys carried out in the late 19th century and the end of the 1920s focused on sampling the work and living conditions of employed wage-earners.³ To construct a historical pre-1900 profile of employment structure, I rely on previously unexploited data resulting from a collective effort between Williamson & Lindert (forthcoming) and IPUMS USA. Unlike previous pre-1900 estimates of US wealth which are built up from production or expenditure approaches (Berry, 1968; Gallman, 2000), Williamson & Lindert build their data from local personal income and work type records. To build the dataset, the authors used local tax assessments and occupational directories for

²Several US historical cross-sections of individual-level data contain binary measures indicating whether an individual earns non-wage income in excess of a given amount. These variables appear hard to map into a continuous measure of individual gross income, and for that reason I choose not to use such datasets.

³Such as the "Cost of living in the United States, 1917-1919" (ICPSR 7711, 1986) and the "Cost of Living in the United States and Europe, 1888-1890" (Haines, 2006).

'registered occupations' and local censuses for 'unregistered occupations'. Labor force counts using the 1 percent US Census sample were provided specifically for the data-project by IPUMS USA. I use the Williamson & Lindert computations of gross earned income, which include wage income, farm income and non-farm business income. However, unlike the surveys from 1935 onwards which contain harmonized employee and self-employed variables, the 1870 data required building types of employment categories. I use a text search algorithm which exploits the highly detailed work titles from the enumerator instructions to the 1870 Census in order to construct self-employed and employee categories. Per example, the 1870 enumerators were explicitly instructed to "not call a man a 'shoemaker', 'bootmaker', unless he makes the entire boot or shoe in a small shop. If he works in a boot and shoe factory, say so (...) Cooks, waiters, etc., in hotels and restaurants will be reported separately from domestic servants." I discuss the US historical data in the appendix.

Methodology For each survey, I construct a nationally representative distribution of earned gross income over the subsample of respondents who declare being active in the labor force (ILO definition). I partition the distribution into ten deciles $s = 1, \dots, 10$. Within each decile, I compute the agricultural share of total employment, ι_s , based on ISIC industry classification. Within non agriculture $1 - \iota_s$, I compute the self-employed and employee shares of non-agriculture employment, denoted respectively $(\varphi | 1 - \iota)_s$ and $(1 - \varphi | 1 - \iota)_s$. I define self-employed and employee shares such that φ and $1 - \varphi$ are mutually exclusive of $1 - \iota_s$. Together with ι_s , the three categories are jointly exhaustive and mutually exclusive in employment in decile s .

I focus on self-employed and employee shares outside the agricultural sector in order to study changes in employment structure amongst workers whose earnings are subject to income tax. The predominant practice in lower and middle income countries is to exclude agricultural earnings from the liable income tax base. I maintain the non-agriculture categories in high income countries where agriculture earnings are subject to income tax. I do this to ensure that variables studied remain the same across countries throughout the analysis of this section. Although in principle this is inconsistent with the definition of the relevant employment base for income tax in high income countries, in practice it has no impact. That is because agricultural employment is very small in high income countries. But importantly, as I document in the appendix, it is also because agricultural employment-share in

high income countries is spread uniformly across all deciles.⁴

I code employment as self-employed versus employee on the basis of whether the work generates derivative information trails that can be used for income tax enforcement. In advanced countries, such information trails are mainly generated through third party reporting of employee wages by employers. In less developed countries where third-party coverage is limited, such information includes paper trails generated by contractual arrangements such as labor contract. This information trails definition of employees is conceptually consistent with the contractual definition of employees in Banerjee & Newman (1993).

Following this classification, I code an agent as self-employed if she responds working in a business without a registered employer. This category is mainly composed of small family businesses and domestic workers. I also code as self-employed any respondent who reports working as a casual daily-wage laborer. Naturally, I also include in the self-employment category all respondents who report working 'on their own account' or in a firm of size 1. This information trails classification produces a self-employment category which encompasses all work categories in the ILO (2002) classification of informal employment: "self-employment in informal enterprises such as unregistered enterprises, employers and unpaid family workers; and paid employment from informal jobs such as casual or day labor and unregistered help for informal and formal enterprises, households or temporary employers."

A compelling feature of the methodology is that industry and type of employment categories are defined in a fully consistent way across all surveys in the cross-section and the within-section. In all surveys, I only code the type of work reported in the first job. This is usually defined as the the job defined towards which the largest number of working hours is used. In doing so, I omit variation in type of work status especially among respondents who may be best considered as having a portfolio of jobs (Banerjee & Duflo, 2007).

The interpretation is that employee work generates an information trail that can be used for income tax enforcement. To the extent that increases in enforceability is driven by movements into large firms, my employee category represents a consistent but 'fuzzy treatment' proxy for enforceable income.⁵ If

⁴It is interesting to consider why agricultural earnings are exempt from personal income tax at lower levels of development. One possibility would be that most agricultural production in less developed countries is carried out by self employed and very small units of production. According to the logic of this paper, the absence of information trails on the earnings in these units would make it worthwhile to exempt agricultural sources of income from taxation.

⁵Kleven et al. (2015) provide audit evidence from Denmark, which shows dramatic decrease in evasion rates on the 'extensive margin' between self-employed and employee groups, and on the 'intensive margin' between large and small firms.

transition into employee jobs is associated with increases in gross income, then my employee share measure represents a lower bound on actual increase in enforceable income. On the other hand, if there is systematic under-reporting of earnings in the survey by self-employed, then the employee-share is an upper bound measure of enforceable mass in a given decile.⁶ In the household surveys, I find no salient evidence of excess earnings mass locally below the threshold. I regard this as evidence that large misreporting of income the survey in response to location of the threshold is not a first order concern. This observation comes with the caveat that the sample sizes in the surveys do not have the power to fully rule out the existence of such underreporting behavior.

Finally, note that unobservable increases over development in the capacity to detect under-reporting of self-employment earnings would appear as a decrease in the employee share. Underlying growth in enforcement capacity thus works against my finding of gradual increases in employee share along the development path.

3.2 Results

I first document on changes to employee share using the cross-section of 90 countries. To build profiles of employee share representative at incremental development levels, I partition the sample of surveys into ten equal sized bins. The bins are based on the constant per capita income of the country-year in which the survey was collected. At each 'development level', I construct the profile of self employed share $(\varphi | 1 - \iota)_s$, and employee share $(1 - \varphi | 1 - \iota)_s$, in each decile s of the income distribution as the weighted average over shares in decile s of countries that belong the the development bin. The country-weights are constructed as the country-survey representative sample share in the total sum of representative country samples in the development bin. These weights only vary across countries, not across country-decile. The resulting profile represents employee and self employed shares across the income distribution of the representative country at given development level.

Stylized fact #1: Within country employee share increases over the income distribution, and at all levels of income as a country develops

The results are reported in Figure ???. At the initially lowest level of development (\$277 per capita), employee-share is concentrated in the top decile. Profiles for individual countries, such as India, suggest that even within the top percentiles, there is an extremely steep increase in the employee-

⁶Or if there is systematic under-reporting in both categories, but self-employed under-report a larger fraction of true income.

share. As the first stylized fact, the panels show how transitions into employee jobs occurs in stages of development. Indeed, at each successive development level, the increase in employee share is concentrated in deciles gradually further down the income distribution (stylized fact #1). At low levels of development (\$730 to \$3286), increase in employee share is concentrated in the top third deciles. At middle levels of development, (\$4638 to \$13512), the increase in employee share is concentrated in the middle third deciles. Finally, at higher levels of development (\$27596 to \$53234), the increase in employee share is concentrated in the bottom third deciles.

The distributional patterns of increase in employee share are strikingly similar when comparing the long run historical evolution of the US to 'synthetic' profiles from the cross-section at similar levels of development. In Figure 3, I pool the historical US profiles and the cross-country profiles using the Maddison dataset. This results in a loss of 33 surveys from the cross-section for which the year is more recent than the latest Maddison year. For each historical US profile, I construct a paired synthetic country. This synthetic country is the average over countries for which the real per capita income in the survey-year lies within 10 percent of the per capita income of the historical US survey-year. Per example, India is included in the synthetic country that is paired with the US 1870 survey year. Remarkably, the US profile of employee share in 1870 on the eve of its second industrial revolution, matches closely the synthetic Indian-based profile in levels. Over the long run from 1870 to 2010, the US profiles systematically match the synthetic cross-country profiles in levels of employee share, and in trends of growth in employee share gradually further down the income distribution.

Stylized fact#2: Tax exemption threshold moves down the income distribution as a country develops in co-movement with increases in employee share

For all countries in the micro database, I extract the nominal value of the personal income tax exemption threshold from the tax code in the year of the survey. I locate it in the gross income distribution, as illustrated in Figure 1 using four different countries (panel A), and four points over time within the US (panel B). In all countries, I compute the size of the income tax base as 1 minus the percentile location of the exemption threshold. The size of the income tax base thus describes what fraction of the income distribution is liable in statutory terms to pay taxes. I calculate the employee share above the threshold as the employee-share over all percentiles that lie above the threshold. The second stylized fact describes the tight co-movement over development between increase in employee share occurring gradually further down the income-distribution and decreases in the exemption threshold (stylized fact #2). At lowest levels of development, the exemption threshold is systematically located

in the top percentiles of the income distribution. At increasing levels of development, the increase in employee share to the left of the tenth decile is associated with decreases in the threshold. At highest levels of development, growth in employee share is concentrated in the bottom deciles where the threshold moves down. Through this co-movement with employee share, the threshold gradually decreases such that the size of the income tax base progressively increases over development. This is shown across 90 countries in the LHS of Panel B. Remarkably, the RHS of panel B shows that for a given level of development, the size of the income base is very similar when comparing US historical variation to cross country variation.

Stylized fact #3: Employee share above the tax exemption threshold remains constant and high at 80-85 percent as a country develops

The third stylized fact shows how despite large increases in the size of the income tax base, the employee-share on the income tax remains constant and high around 85 percent. In the LHS of Panel C, I cannot reject with statistical confidence that the employee share above the exemption threshold are the same in India and the US. This constant employee share on the tax base occurs despite the tax base being 30 times larger in the US than in India. The RHS of Panel C shows that, for a given level of development, the employee share on the tax base is the same in historical US and across countries.

These stylized facts are suggestive of the impact of employee share on the exemption threshold along the development path. In the appendix, I show that the three stylized facts also hold within Brazil between 1970 and 2010. They are consistent with a mechanism where employee-share proxies for enforceable income, and growth in enforceable income through increases in employee share drive expansions of the income tax base.

The tight match in the stylized facts provides a bridge of external validity between the suggestive cross country patterns and a within country identified estimate. In Section 4, I provide a within country identified estimate of the impact of the employee share on the exemption threshold.

Quantitative exercise To complement the descriptive facts, I quantify the relationship between employee share and levels of development, separately for each decile of the income distribution. I adapt a regression method used to quantify ‘development-stages’ of structural processes (Imbs & Wazciarg, 2003)

In order to quantify a fully flexible relationship, I attempt to impose as little structure on the functional form as possible. This motivates the use of a non-parametric regression method, which is locally

robust⁷ and allows me to recover the estimated coefficients that I use to for later estimation.⁸

The method is adapted from robust locally weighted scatterplot smoothing ('lowess'). In the cross country sample, I partition the data into S subsamples according to overlapping constant per capita income intervals of size $J = \$1,000$, with an overlap of size $\Delta = \$250$. Each interval thus has a midpoint which is \$250 away from the following midpoint. For each decile, I run a regression on each subsample of the decile-specific measure of employee-share on per capita income. I use the estimated coefficient on per capita income and constant values to plot the fitted value against the income midpoint of each estimation subsample.

The resulting curve for a given decile j yields a robust non-identified shape of the evolution of employee-share in every decile j over development. I plot the curves for all deciles in the Appendix. I can statistically confirm stylized fact #1. At low levels of development, growth in employee share is stronger in the top decile and is significantly different from growth in other deciles. Over increasing levels of per capita income, the statistically significant growth in employee share occurs in deciles gradually further down the income distribution.

More interestingly, the curves suggest that beyond a real per capita income level which is specific to each decile, growth of employee share in a given decile has come to a halt and employee-share in the decile has reached a steady-state level. The regression technique allows me to robustly estimate these decile-specific per capita income points. I calculate the decile-specific per capita income point by setting to 0 the derivative of the predicted decile-specific employee-share with respect to per capita income. Per example, at real \$7500 per capita, employee growth is estimated to have come to a halt in the 6th decile. At this per capita income level, the employee share predicted by the statistical curves has reached 85 percent. At higher capita income levels, the predicted employee share remains constant at this steady state level. I estimate the 'steady state' per capita income specific to each decile. I find that the per capita income points are smoothly but steeply decreasing in all deciles (appendix Figure). Remarkably, the fitted curves further suggest that the steady state employee share is constant across deciles, ranging between 80 and 85 percent.

In the appendix, I repeat the exercise for the sample of US states over time. The panel structure allows me to include state fixed effects into the subsample regressions. I can therefore calculate the

⁷This is unlike polynomial/semi-parametric functions where the shape of the curve at one point can be determined by shape of the curve at other extreme points.

⁸Other smoothing methods simply compute the mean of y for subsamples of data centered around x_s . This distinction is important because I am not only interested in the shape of the general relationship between employee share and per capita income (which is delivered by these other smoothing methods), but also in the sign and the statistical significance of the coefficients.

'steady state' real per capita incomes within the average state over time. I find that the estimated real per capita incomes are also smoothly decreasing in the deciles, but they exhibit less curvature than in the cross country setting. This difference in curvature is similar to the findings in Imbs & Wacziarg (2003) who finds that patterns of sectoral concentration are more pronounced across than within countries.

I use the fitted curves and estimated 'steady state' per capita levels in the model section.

4 Direct estimate of impact of employee share on exemption threshold and tax collection

In this subsection, I provide a causal estimate of the impact of employee share on the exemption threshold and income tax revenue. I first provide background information and details of the development program. I then discuss data and identification strategy. Finally, I present graphical and regression results.

4.1 Background and program details

Background US states are a compelling setting to study development of tax systems. Each state defines and enforces a wide range of tax bases. Historically, the large growth in state tax-to-GDP ratios (Wallis, 2000) has been entirely driven by an increase in personal income taxes (Panel A, Figure 5). The rise of the modern income tax system in individual states matches the key tax capacity stylized fact (Besley & Persson, 2014; Kleven et al., 2015).

In parallel to the rise of the modern income tax, individual states also witnessed large changes in employment structure over time. Panel B of Figure 5 shows the non agricultural employee-share of total employment along the income distribution of the average state over time between 1950 and 1980. The employee-share is increasing in the income distribution, and over time the employee share increases in all income deciles. This is consistent with stylized fact #1. Panel B shows the state exemption threshold gradually moves down the income distribution in co-movement with increases in employee share locally to its left, consistent with stylized fact #2. Finally, the employee-share above the threshold remains constant at 85 percent over time, consistent with stylized fact #3.

Throughout this section, employee and self-employed are calculated as shares of total employment

that includes agriculture. I do this in order to follow the same criteria for calculating employee shares as in Section III: in the US, agriculture has never been exempt from state income tax base. I focus, however, on changes in non-agricultural employee share to maximize comparability with employee share in Section III. This choice also highlights that results are not confounded by movements out of agriculture. Indeed, I show that the development program led to no change in volume of agricultural employees, but a large movement out of self-employed farming. Including agricultural employees into the employee share variable does not affect results.

Program details To establish a direction of causality from employee share to the exemption threshold, I exploit exogenous variation in implementation date of the Industrial Development Bonds (IDB) program. Through the IDB program, the state built leasable manufacturing facilities in rural counties characterized by underemployed self-employment. (Area Redevelopment Administration Commission, 1963; Cobb, 1993) The IDB program thus acted as a level-shifter in employee-share.

Financing of the IDB program was directly incompatible with the state 'public purpose' Constitutional provision, whereby government debt may only be issued for public purpose. Implementation therefore required the state House to vote in a legal act which exempted IDB from the public purpose provision. But there was no legal historical precedent to such development program. The voted act and by extension any program funding would therefore remain legally uncertain until the highest state court would litigate to uphold the legality of the IDB program through a specific court case (Pinsky, 1963; Abbey, 1966; Rollinson, 1976; Cobb, 1993).⁹

The time-lag between the vote-in and the upholding events was substantial. The lag has mean of 6.67 years and standard deviation of 6.77. In 40 percent of cases, the time-lag exceeded 10 years. I digitize archived Moody's state financial records on issuance of IDB debt. I use this data in Figure 6 to show that upholding was a necessary condition for any issuance of IDB debt. This observation holds amongst states with the longest time lag. This suggests that timing of IDB implementation cannot be explained by systematic implementation delay since vote in due to constrained administrative capacity.

⁹A large Federal reform in 1968 made large changes to IDB regulations. To maintain comparability, I therefore limit the treatment definition to the set of states that implemented the IDB program prior to the reform. Because my estimation strategy relies only on within-state variation, this sample choice has no bearing on the construction of a counterfactual and hence on the estimates. It does, however, condition results on the sample properties of the pre-reform IDB states.

4.2 Data

I construct employee shares of employment from decennial Census between 1910-2010. I construct counts of the working labor force. I construct counts of non-agriculture and agriculture class of workers using the Census class of worker categories 'self-employed' and 'works for wages' and the Census 1950-based industry classification. I construct the employee share as the ratio of non-agriculture employees to agriculture and non-agriculture workers. The self-employed share is constructed as the ratio of all self-employed workers to agriculture and non-agriculture workers. I interpolate the numerator and denominator between Census years using a natural cubic spline (Herriot & Reinsch, 1973). The main results are robust to collapsing all data to a simple pre-post sample.

Continuous measures of employee counts are retrieved from 'State and Area Employment, Hours and Earnings' collected by the Bureau of Labor Statistics. The series provides number of employees in non-agriculture by industry categories in state-years 1939-2002. Volumes of earnings by industry and type of work (employee, self-employed non farming, and self-employed farming) is from the historical series "SA5H: Personal income by major component and earnings by industry" produced by BEA in all state-years 1929-2005. I use the BEA series to construct continuous measures of employee and self-employed shares of gross individual earnings. Gross earnings excludes all government transfers and taxes, and non-work income (dividends, interest). I combine the historical BEA and BLS series to construct continuous measures of average employee earnings, and average employee earnings by major industries. I combine historical BEA series with interpolated Census data to construct average workforce earnings, and average self-employed earnings.

Tax-to-GDP measures between 1929-2010 are based on state government finances published by US Census. GDP is proxied for by using state personal income. I use US historical state tax calculator 1900-2007, constructed by Bakija (2009), to construct measures of the state PIT-base and the state PIT-rate structure. I am grateful to Bakija for providing me input from the calculator. I calculate the state income tax exemption-threshold, K , for an individual earner who files as being single, reports having one dependent, and claims the standard deduction.

I measure dates of vote in and upholding of IDB-program independently across legal reviews and Federal administrative records. I define the year of IDB-upholding as the year where the highest court-instance in the state upholds IDB through a particular court case. Leading cases cited are consistent across legal reviews.

I digitize the full set of Book of the States 1935-2010 (Council of State Governments). I record the

number of agencies administering major taxes: this ranges between 1 and 6. I construct the ratio of annual administrative salary in revenue-taxation relative to annual salary as Treasurer; as Attorney General; and, as average state government administrator. I use these variables to proxy for investments in enforcement capacity. I use political outcome variables from Besley et al. (2010). I use the Democratic vote-share across all state elections, the Democratic seat share in the state Houses, the existence of a Democratic Governor as measures of increasing demand for redistribution. I use the Besley et al. (2010) measure of political competition.

I construct the distribution of average income across deciles of the state income-distribution by combining Census data 1950-2010 and the 1935-36 Consumer Survey. The income by decile is interpolated across missing years. I digitize the series of significant provisions of state unemployment insurance laws (Department of Labor, 1937-2009) to construct a measure of firm size coverage of state UI.

The sample used for the main set of regressions is a panel of the 48 continental states, between 1939 and 2005. A more detailed description of all variables and the data-sources can be found in the appendix.

4.3 Identification strategy

Identifying variation The estimation strategy exploits institutional features of IDB implementation. I use changes in court upholding litigation status as identifying variation, and assess program impact by comparing changes in outcomes before and after the upholding event, relative to counterfactual changes before and after the vote in event within the same state.

This specification estimates the causal impact of the IDB program on employee share under the identifying assumption that the vote-in period represents a valid counterfactual for the upholding period. That is, absent the IDB program, the outcome of interest would have been on parallel trends throughout the vote in and upholding periods within the same state. I provide two main pieces of evidence to support this identifying assumption.

First, I find that the only significant predictor of the time lag is a state specific time invariant dummy for civil law origins. Table 1 reports the results from non-parametric Cox proportional hazards models. These models use state time-varying and time-invariant regressors to predict the conditional probability of the upholding event occurring, conditional on the vote in event having occurred. The

civil law dummy significantly predicts a higher conditional probability of upholding. None of the economic variables (manufacturing share of labor force, employee share of labor force, 'redevelopment'-share of the labor force, log per capita income, log population), political variables (political competition measure, existence of a poll tax and or literacy tests as voting restrictions), taxation variables (size of exemption threshold, share of income tax to GDP, share of total tax to GDP), geographical variables (dummy for Southern region) predict significant changes in the conditional probability, once the civil law origins dummy has been included. This variable is drawn from Berkowitz & Clay (2005), and codes a state with civil law origins if, by the time of American acquisition, its colonizers had a civil law legal system (as opposed to a common law system).¹⁰ The faster time to uphold associated with civil law states is consistent with studies across US states (Berkowitz & Clay, 2005) and across countries (La Porta et al., 2008).¹¹ In the appendix, I show that the civil law residual time lag, and the civil law dummy, are both uncorrelated with outcomes in a pre-IDB cross-section.

Second, I plot changes in outcomes before and after the upholding event and confirm that the effects I find are driven by sharp on-impact changes. I show that the outcome of interest is trending in a stable manner over the full pre-event interval regardless of the length of the time-lag between vote in. I show there is no discernible change in the years immediately preceding the upholding event. Any secular trends are largely eliminated by the inclusion of a state-specific linear trend in the empirical specification.

Using within state identifying variation alleviates identification issues related to cross-sectional estimates in the current setting where vote in is likely endogenous to the economic and political environment. Centering treatment around the upholding event also overcomes issues of fuzzy treatment due to legal uncertainty during vote in. The within-state specification in 1 controls for any state-time varying unobservable shocks to political and economic environments which occur at time of vote in and which are common to vote in and upholding periods. Estimates are lower bound on the causal impact if there was any program impact during vote in.

¹⁰Ten of the continental American states were settled by France, Mexico or Spain and had civil law legal systems by the time of the American Revolution. These ten states are: Alabama, Arizona, Arkansas, California, Florida, Louisiana, Mississippi, Missouri, New Mexico, Texas. The 38 other had a common law system or were unsettled. Note that an additional five states - Illinois, Indiana, Michigan, Ohio, Wisconsin - were also originally settled by a civil law country, but were acquired by Great Britain prior to the American Revolution.

¹¹One interpretation of this result, drawing on Berkowitz & Clay, is that civil law produces a Constitution with more statutory components, rather than framework provisions, and that the existence of statutory laws created more frequent demand for constitutional change among affected groups as the political and economic climates change over time. This explanation is consistent with the difference in IDB-litigation procedures observed across states: civil law origins states were more likely to vote in statutes, as opposed to Acts, which was likely to be revised more quickly.

Exclusion restriction The objective of this section is to obtain an estimate of the effect of employee share on the exemption threshold which is not affected by simultaneity or omitted variables. To achieve that goal, I instrument for employee share using changes in court IDB litigation status.

I show in the following subsection that court IDB litigation status has a large impact on employee share, with an break from trend immediately 'on impact'. The previous subsection showed that the only significant predictor of change in litigation status was a dummy for civil law origins. This evidence suggests changes in employee share caused by exogenous change in IDB litigation constitute a valid first stage.

In order for changes court IDB litigation status to be a valid instrument, it must also satisfy the exclusion restriction. That is, changes in court IDB litigation must only affect location of the exemption threshold through changes in the employee share. I provide several pieces of evidence to support this claim. First, I show that change in litigation status is not correlated with proxies for enforcement capacity, tax rate structure, earnings structure, demand for redistribution, political competition. These variables are chosen based on the exemption threshold model derived in the following section. I show that these proxies are meaningful confounders, in the sense that they strongly correlate with location threshold and tax revenue collections. Second, I show the program has no impact on revenue from any other tax base apart from the income tax. This suggests court litigation status did not lead to any change in revenue requirements from the issuing of IDB debt, which could also have caused a decrease in threshold. This is consistent with the nature of the IDB debt, which was issued as revenue bonds that only pledged repayment of principal and interest against the future income derived from the leasing of the IDB facilities, not against the 'full faith and credit' of the state. Third, the program had no impact on economic outcomes which are closely associated with changes in employment structure and correlate with tax capacity, including overall labor force attachment, migration rates, sectoral spillovers. Fourth, I show in the following subsection that the change in litigation status led to a sharp change in exemption threshold, with a consistent lag in timing relative to the impact on employee share.

Together, these pieces of evidence narrow the possible variation that could violate the exclusion restriction. Such variation would have to produce an immediate and sustained break from trend; impact the exemption threshold and income tax revenue with precise sequential time lags to the upholding event; not be captured by any of the proxies; have no impact on any other tax instrument nor any other tax revenue.

Baseline specification The baseline empirical specification is

$$y_{st} = \beta + \alpha \mathbf{1}(\text{Vote in})_{st} + \theta \mathbf{1}(\text{Upheld})_{st} + \lambda \mathbf{X}_{st} + \mu_s + \gamma_t + \phi_s \cdot t + \varepsilon_{st} \quad (1)$$

where s denotes state, t denotes time, $\mathbf{1}(\text{Vote in})_{st}$ is a dummy indicating whether the state has voted in the constitutional amendment, and $\mathbf{1}(\text{Upheld})_{st}$ is a dummy indicating whether the state-court has upheld the legality of the constitutional amendment. All main results reported include a state-specific linear trend, $\phi_s \cdot t$, and a set of political controls (dummies for the existence of voting rights restrictions and for election years), policy environment (dummy for the existence of right-to-work laws and continuous measure of the firm-size coverage of state unemployment insurance schemes) and log per capita income. Results are fully robust to replacing the state linear trend with interactions between three cross-sections of structural determinants calculated in 1930 and a linear trend.

All standard errors are clustered at the state level to allow for correlation over time within a state. I confirm that a block-bootstrap at the state level yields similar significance levels. I also implement non parametric tests of significance levels using permutation tests, and find similar results. Bertrand et al. (2004) Significance is also robust to ignoring all time series information by collapsing data to a two-period pre-post DiD specification.

The causal interpretation of the results is based on the exogenous timing of the upholding event. I confirm this interpretation by showing that results are identical in cross-sectional specifications, both standard and paired synthetic matching, which estimate program impacts off a non-IDB counterfactual.

In the main results, I use the full sample of years and states to allow precise estimation of covariates in 1. Results are robust in a sample of IDB states in a small interval around vote in and upholding.

4.4 Employment results

In this subsection, I show that court upholding of the IDB program led to an increase in the non-agriculture employee share, but had no impact on earnings structure. This provides the first stage of the IV.

Graphical evidence I begin with some graphical evidence. Figure 7 provides evidence on an immediate and large increase in employee-share upon upholding. The full set of IDB states have been grouped according to the time-lag between vote in and upholding: 0 to 5 years; 5 to 10 years; 10 to 15 years; in excess of 15 years. Relative to a normalized outcome value of 1 in the year of vote in, I calculate the average within state change in trend before and after the event, across the four groups. Centering the trend graphically around the year of vote in both provides a test for any endogenous impact around the vote in event where my data suggests the IDB program was not active. It also provides a simple but real test of any impact around the upholding event which is free from visual bias. In these graphs I use the employee-share of earned income.¹² I do this because it is available on a continuous basis before and after vote in and upholding. In the regressions, I find no impact of upholding on continuous measure of average employee earnings. This suggests an increase in employee share of earnings can be interpreted as increase in employee share of employment.¹³

Within each group, there is a systematic break in trend and increase of employee-share immediately following the upholding event. In contrast, within each group the employee-share is smoothly trending through the vote in event - with the exception of the the group for whom vote in and upholding coincide. There is no visible change in pretrend in the years immediately preceding the upholding event. This is neither the case when comparing the same group to itself over successive intervals of the vote in period, nor when comparing across groups in given years since vote in. Finally, the trend since vote in is small and almost identical across groups up to ten years after the vote, prior to upholding. This suggests confounding differences between early and late upholding groups in initial employee-share at time of upholding is not a first order concern.

In Figure ?? I provide a direct graphical equivalent of the estimation strategy in 1. I assess changes in outcome before and after the upholding event and compare them to changes before and after the vote in event within the same state. I focus on the subset of states in which the time-lag is in excess of 15 years, and define a post vote period of 10 years and a (non overlapping) pre upholding period of 5 years.¹⁴ I make these choices in order to construct visually long periods around each event. I show in the appendix that the results hold when constructed using a subsample with smaller time lag (and

¹²This variable excluded all transfers received from Federal and state government. The denominator contains all sources of earned income: employee farming, employee non farming, self employed farming, self employed non farming. The denominator excludes all sources of non earned income, such as dividends and interest payments.

¹³In the regressions, I also find that the increase in employee share of employment and the increase in employee share of earnings are almost identical. This further suggests that the impact on employee share uncovered in the graphs is driven by a change in employment rather than average earnings.

¹⁴Which represents around 20 percent of the sample of IDB states according to appendix Figure.

hence a larger set of states). Panel A shows an immediate increase in employee share upon upholding. The pre-trend is stable up until the upholding event.

Regression results The first 9 columns of Panel A of Table 2 present results from studying changes in log volume of total employment in specific employment-industry categories, using 1. The program led to no change in overall employment levels, but a large transition from self employment into manufacturing. I find no impact of the program on volume of employee jobs in construction, trade, government, and employee agriculture. I uncover a marginally significant decrease in volume of services, suggesting a small switch between industries within the employee share. These results suggest the specification is precisely picking up the program impacts with a movement out of self-employed into manufacturing. The results are inconsistent with secular sectoral changes due to structural transformation, which would also predict movement into construction, retail and services. In the appendix I show that the the program did not lead to any changes in net migration rates, also consistent with the characteristics of the program. The last two columns of Panel A show that these changes in employment volumes map into a 3.6 percentage point increase in the non-agricultural employee share of total employment. The increase is entirely accounted for by a corresponding decrease in self-employment share. Given the specification, note that these are lower bounds on the true program impact.

Panel B of Table 2 shows that the program was not accompanied by any discernible change to the gross earnings structure. It did however lead to an increase in volume of employee employee-related amenities. The program led to no change in gross overall earnings, nor in total volume neither per member of the workforce. The program led to an increase in volume but not in average earnings of employee-jobs, and a mirrored decrease in volume but not in average self-employment earnings. The program increased employee-related amenities, but had no impact on other state transfers. The historically relevant employee-specific amenities included state unemployment insurance.¹⁵ Such employee-targeted amenities are consistent with current practices in developing countries (Gerard & Gonzaga, 2014). The last columns of Panel B show changes to volume and average earnings sources translate into a 4.29 percentage point increase in the employee-share of gross resident earnings.

¹⁵At its onset in 1938, UI extended only to private firms employing eight or more persons at least 20 weeks a year. By 1958, coverage was broadened to include firms employing one to three employees. Only in 1978 was coverage extended to agricultural firms employing a minimum of 10 workers in at least 20 weeks a year or having a \$20,000 quarterly payroll, and to employers paying a quarterly minimum of \$1,000 to domestic workers (Bureau of Labor Statistics Handbook of Methods, 1997).

Discussion The employment and earnings results are consistent with a classical Lewis model (1954) in a setting of a local rural labor market. In this setting, the IDB program can be interpreted as an exogenous positive shock to available capital which shifts out initially constrained demand for manufacturing employee jobs. The setting is characterized by existence of large cohorts of potential workers relative to constrained capital stock. This generates 'underemployed' individuals who are self-employed in order to satisfy subsistence needs. The workers have infinitely elastic supply to the manufacturing sector. These characteristics fits the description of the the typical IDB-targeted areas. In this model, self-employed make the transition into employee-jobs even if the newly offered wage is no higher than pre-existing earnings. This is because the transition is associated with monetary gains due to available employee-specific (non taxable) amenities. It can also be due to non-monetary utility gains from moving to a more secure income stream. Consistent with my results, the model predicts that the IDB program will lead to a large movement from self-employment to low-skilled employee-jobs; have no impact on average earnings in manufacturing nor overall earnings; and, increase employee specific amenities.¹⁶ Accounting for the equalization of average earnings, my employment and income results are fully consistent with other findings on development programs in the US, noticeably Kline & Moretti (2013) on the Tennessee Authority program.¹⁷

I find no change to overall volumes of total employment, suggesting the transition was not driven by movements into the labor force of initially unemployed. The finding of no significant spill over to other sectors, both in terms of employment and earnings, suggests limited importance of general equilibrium effects. This finding is consistent with Federal and academic reports which document that IDB spurred local plant expansion of pre-existing firms which operated in low-skill industries such as textile and food-processing. These low-skill firms mainly exported goods to out of state consumer markets. Finally, the simple model outlined above abstracts from effects due to income accruing to IDB plant owners. If owners consume outside local markets (perhaps because they physically reside elsewhere), one would expect no such impact. I find no impact on retail and services at the state

¹⁶One key test of the model is whether earnings in employee and self-employment jobs are equalized prior to the IDB program. I use county level data between 1940 and 1950 to confirm this to be the case. I discuss in the Appendix the details of this exercise, including the assumptions made to calculate average earnings of self-employed. To fit the test of equal earnings as closely as possible to the model, I compare earnings in counties likely to be treated under the IDB program. I rely on Federal administrative reports (Area Redevelopment Administration, 1963), which coded all counties in the US as 'redevelopment areas' if they satisfied the two IDB criteria: excess under-employment of self-employed, farmers; lack of private credit supply. Using the 'redevelopment counties' as proxy for IDB counties, I find that in IDB counties a t-test cannot reject equality of average self-employed and average employee earnings.

¹⁷In particular, they also uncover a large movement from agriculture into manufacturing. They find that aggregate income does increase, arguing that this occurs because manufacturing paid higher wages than agriculture. Differences in income results between their study and mine could arise from underlying differences in types of counties studied - indeed, I find that in non-redevelopment counties, the equalization of earnings across self-employment and employee-jobs does not hold.

level. This is consistent with anecdotal evidence that the IDB-program was dominantly taken up by firms whose headquarters were located outside the IDB state. The local labor market model seems to account for effects of IDB at the level of the IDB state, but not at the national level.

4.5 Tax exemption threshold results

In this subsection, I study impacts on the exemption threshold. I first show a reduced form relationship between the IDB program and the threshold. I then estimate the impact of employee share on exemption threshold where I instrument for employee share using changes in IDB court litigation status (from Section 4.4)

Graphical evidence: reduced form The graphs in panel B of Figure ?? document on a reduced form impact of the program on the exemption threshold. In the LHS graph of Panel B I show that the program led to active tax policy change, through an increased likelihood of passing a legislative reform to the nominal value of the exemption threshold.¹⁸ The break from trend is salient, and occurs with a two year lag relative to upholding. In the RHS graph of Panel B, I find a sharp decrease in the ratio of the exemption threshold to average earnings. This ratio proxies for the relative position of the exemption threshold in the income distribution, where a decrease implies the threshold has effectively moved down the distribution. The decrease in location of threshold could come both from a rightward shift in earnings for a constant nominal value of the threshold,¹⁹ and from a real reform that lowered the nominal value of the threshold. The break from trend in threshold coincides exactly in timing with the year of break from trend increase in the likelihood of a legislative reform. This suggest the upholding event led state policy-makers to actively expand the income tax base through lowering of the exemption threshold. In the regressions, I find that upholding had no impact on earnings.

The break from trend impacts combined with stable pre-trends are simple evidence of a reduced form impact of IDB court upholding on a policy change to the location of the tax exemption threshold. The timing of sequential changes suggests the decrease in exemption threshold was driven in part through increase in employee share.

¹⁸This normalized likelihood is constructed as state specific empirical cumulative distribution of number of legislative reforms passed over the full sample period. I code a year as legislative reform whenever the nominal value of the exemption threshold breaks from the previous year. This coding is not confounded by inflation-indexing to the exemption threshold, which states started implementing in the late early 1980s, well beyond the average year of IDB upholding.

¹⁹Similar to the bracket creep phenomenon occurring in periods of inflation with non-indexed threshold values.

Regression results: reduced form I modify the baseline regression specification to include a full set of average incomes z by decile j in state s in year t . The new specification controls in a flexible way for movements in the earnings distribution which may have impacted changes to the threshold. Formally, I estimate the following model

$$[K/y]_{st} = \beta + \alpha \mathbf{1}(\text{Vote in})_{st} + \theta \mathbf{1}(\text{Upheld})_{st} + \sum_{j=1}^{10} \omega_j z_{jst} + \lambda \mathbf{X}_{st} + \mu_s + \gamma_t + \phi_s \cdot t + \varepsilon_{st} \quad (2)$$

where $[K/y]_{st}$ is the ratio of the nominal value of the exemption threshold to the average gross resident earnings. Note that a decrease in the ratio implies a lowering of the threshold location in the income distribution. In Table 3, Col. 1, I find a large, statistically significant decrease of $-.7218$ in $[K/y]$, which translates into a 18.23 percentile decrease of the location of the threshold in the average IDB state's pre-period income distribution.

Maintaining the reduced-form specification, I provide three pieces of evidence which directly match the cross-development stylized facts from Section 3. I estimate the program impact on employee-share separately for all deciles j in the income distribution using 1 and study the coefficients $\hat{\theta}_j$. Panel A of Figure 9 plots $\hat{\theta}_j$. Panel B plots the implied post-IDB employee-shares $\theta_j^{POST} = \hat{\theta}_j + \theta_j^{PRE}$ where θ_j^{PRE} is the average employee share in decile j calculated in the pre-IDB period in the IDB states. The panels compellingly show that the distributional impact of the program was to shift to the employee-share profile leftwards, resembling closely the stylized fact #1. The panels also locate the implied post IDB percentile location of the threshold calculated as $K^{POST} = K^{PRE} + \widehat{dK}$. $\widehat{dK} = -18.23$ corresponds to the reduced-form estimate. K^{PRE} is the average percentile location in the pre-period in the IDB states. Inspecting θ_j^{POST} relative to K^{PRE} makes clear that the program lead to increase in employee share exclusively *below* the pre-location of the threshold. Combining θ_j^{POST} and K^{POST} provides strong evidence that the threshold tracked growth in employee-share occurring locally to its left, which matches stylized fact #2. Finally, comparing K^{PRE} and $\theta_{j > K^{PRE}}^{PRE}$ with K^{POST} and $\theta_{j > K^{POST}}^{POST}$ suggests that before and after the large changes in employee share and location of the threshold, the employee share above the threshold remains constant, consistent with stylized fact #3. Col 4 of Table 3 confirms in a regression that employee-share above the threshold remained constant upon upholding.

Regression results: instrumental variables Having demonstrated in the previous subsections a relationship between IDB court litigation status and employee share, as well as a reduced form relation-

ship between such litigation changes and the exemption threshold, I now apply instrumental variables to estimate the elasticity of the exemption threshold with respect to employee share.

I assume that changes in exemption threshold and in employee share are determined according to

$$[K/y]_{st} = \beta + \varphi \text{Employee-share}_{st-1} + \sum_{j=1}^{10} \omega_j z_{jst} + \lambda \mathbf{X}_{st} + \mu_s + \gamma_t + \phi_s \cdot t + \varepsilon_{st} \quad (3)$$

The first stage estimates changes in employee share from changes in IDB court litigation status according to 1. If changes in employee share predicted by changes in IDB court litigation status provide a valid first stage, and if changes in court litigation status only affect the exemption threshold through employee share, then the IV-estimated φ^{IV} in 3 provides a causal estimate.

Table 3, Cols.2-3 show the instrumented value of employee-share is negative and significant. It is one-third larger in magnitude than non-instrumented employee-share, a sign of simultaneity.

I provide evidence for the exclusion restriction assumption that changes in IDB court litigation only impacted the exemption threshold through changes in employee share. I use the formula for the threshold determinants, derived in the following section, to construct proxies for confounding channels. I show in the appendix that the proxies meaningfully correlate with K and tax revenue. I proxy for changes in enforcement capacity in two ways. First, a decrease in the number of tax departments responsible for collecting distinct sources tax revenues reflects improved administrative capacity to centralize and cross-check information sources in order to enhance enforcement. Second, measures of wages of tax administrators relative to other high ranking administration officials reflects improved incentives for tax collectors. (Khan et al., 2015) I construct approximations to the earnings hazard ratio by using measures of average income by decile of the income distribution and the methodology outlined in Saez (2001). I proxy for demand for redistribution by using measures of democratic vote share at the state level across all types of elections, measures of Democratic seat share in the state House, a dummy for democratic governor, and the political competition measure in Besley et al. (2010). I show in the following subsection that there was no impact of upholding on any other source of tax revenue apart from income tax. This suggest the program did not led to a change in common marginal value of funds. Finally, I show that none of a range of statutory tax rate measures change with upholding.

Cols. (5)-(8) show that the program had no impact on the marginal bottom tax rate, the earnings hazard ratio proxy, the number of tax agencies proxy for enforcement, or the democratic vote share proxy for redistribution. Appendix tables show the change in litigation status did not either corre-

late with any of the other proxies for confounding determinants. I find in the following subsection that there was no impact of the program on any other source of tax revenue apart from income tax, suggesting no change to a marginal value of funds. This combines with findings from the previous section that the program had no non-employment impact on local economic structure as evidence in favor of the exclusion restriction.

Based on the instrumented employee-share value, I derive an elasticity of the exemption threshold with respect to employee share. I find a value of -7.548. This is close in magnitude to the elasticity implied by the ratio of reduced form impacts on employee share and the exemption threshold, which equals -6.678 (Column 1, Table 3).

The φ^{IV} estimate captures the local average treatment effect of employee share on exemption threshold for states that implemented IDB. I use the Federal administration's classification of counties as 'redevelopment areas' (Area Redevelopment Administration, 1962) to calculate the state's share of labor force in the 'redevelopment counties' characterized by underemployment of self-employed in rural areas. The mean 'redevelopment' share in IDB states, 16.5, is three times larger than in non IDB states. This suggests the IV estimates carry external validity in labor markets of less developed countries. Indeed, ILO (2009) considers underemployment as a defining characteristic of labor markets at lower levels of development. Furthermore, ILO documents that in developing countries, underemployment tends to be concentrated in rural areas with large proportions of self employed workers. The IV estimates may also be relevant in developing countries such as India, Mexico and South Africa, which have implemented industrial financing programs like IDB. (Platz, UN Financing for Development, 2009)

I use the instrumented elasticity to understand what extent of observed variation within the average IDB state over time in size of income tax base can be explained by (policy-led) increase in employee share. A one standard deviation increase in length of IDB program due to random changes in court litigation status leads to an increase in size of tax base via an increase in employee share which can account for 26% of the average IDB within-state standard deviation growth in tax base.

4.6 Income tax capacity results

In this subsection, I proxy for tax capacity by using the ratio of tax revenue from a given base to total state GDP. I show graphical and regression evidence of a reduced form impact on income tax capacity.

I then estimate an instrumented elasticity of income tax capacity with respect to employee share.

Panel C of Figure ?? shows a sharp increase in income tax to GDP, with a three year lag to the upholding event. The increase occurs in the year following the legislative reform to lower the exemption threshold (Panel B of Figure ??). This sequential timing suggests the rise in income tax capacity was in part channeled through the decrease of the exemption threshold.

Using specification 2, I find a large, positive reduced-form impact of the program on personal income tax. Column 1, Table 4) I proceed to estimate the elasticity of [income tax/GDP] with respect to employee-share. I assume the second stage follows 3. Employee-share is again instrumented using variation in court litigation status. I find an instrumented value of employee-share which is one-third larger than the non-instrumented one. The instrumented elasticity of (income tax/GDP) with respect to employee share is .90. The IV-elasticity is smaller than the elasticity implied by the ratio of IDB reduced form impacts on income tax and employee share, which equals 1.292 (Column 1, Table 4)

This elasticity is a consistent estimate under the exclusion restriction that changes in court IDB litigation status only affect (income tax/GDP) through its impact on employee share. To support this claim, Cols.4-7 show that changes in court litigation status had no impact on the four major other sources of state tax revenue: corporate income tax, general sales tax, selected sales tax, and license taxes. This suggests the program did not impact any general marginal value of funds, either per se or through increased revenue requirements following the issuance of IDB debt. As discussed earlier, this is consistent with the type of IDB debt that did not pledge repayment against the state's own revenue. I show in the appendix that for selective sales and license taxes, the null result holds when considering specific categories (motor vehicle, tobacco, public utilities, alcohol).²⁰ In the previous subsections, I found no impact of change in court litigation status on enforcement and administrative capacity, earnings structure, tax rate structure, demand for redistribution. I also found that the sharp increase in income taxes after the upholding event occurred with consistent time lag to the increase in employee share. Variation which violates the exclusion restriction would have to occur precisely around the time of upholding; produce an immediate and sustained break from trend; not be captured by any of the proxies; impact income tax revenue with precise time lag to employee share impact; have no impact on any other tax revenue.

²⁰In the appendix, I show there was a decrease in the relative reliance within state on license taxes levied specifically on occupations and small businesses, which are effectively taxes on self employed. This decrease in revenue is consistent with a behavioural response induced by the IDB program out of self-employment. Relative to GDP, these specific licence taxes represented a trivial amount, with a mean of .0002 at the onset of the IDB program. When the outcome variable is [occupation licence taxes/GDP], the revenue loss estimated from upholding comes out insignificant, and represents 0.00008 percent of the estimated increase in income tax revenue.

This IV-estimate is the local average treatment effect of employee share on (income tax/GD) for those states that implemented the IDB program. The magnitude of the instrumented elasticity suggests a standard deviation increase in exposure to state-led increase in employee share through random court litigation changes can account for just under 10 percent of the average within state change in income tax effort over time.

I argue that the IV estimated effect of employee share on income tax effort was in part channeled through changes in the exemption threshold. Recall that there was no impact on employee-share in deciles of the income distribution that lied above the pre-IDB location of the exemption threshold. (Figure 9). Growth in employee share below the exemption threshold can lead to increase in income tax revenue through two channels. First, the employee share growth can cause the exemption threshold to decrease. Second, the employee share growth can lead to increased ability to collect income tax, conditional on an exogenous decrease in the threshold. I do not rule out the second channel, but interpret the results in Figure 9 as strong evidence of the first channel.

4.7 Robustness

The causal interpretation of the results is driven by the exogenous variation in timing of the court upholding decision. In the appendix, I provide further evidence for this claim by showing that results are nearly identical in cross-sectional specifications which are based on very different counterfactual variation. For the key set of outcomes (employee share, exemption threshold, employee share above threshold, income tax rate, income tax to GDP ratio) and confounders (average earnings, average employee earnings, bottom and top marginal tax rates, political competition, democratic vote share), I find very similar magnitudes and significance in a simple cross-state specification between upholding and non upholding states. The results are also robust to using synthetic matching. In this specification, each IDB state is paired to a synthetic control state which is the weighted average over all non-IDB states that maximizes pre-upholding trends on employee-share. I use the the full set of economic and political covariates to predict the pretrend.

Results are robust to fully ignoring the time dimension of the panel. I do this in a specification which collapses all time series information into pre and post upholding periods. Standard errors block-bootstrapped by state deliver similar patterns of significance. To alleviate concerns in DiD specifications over bias in standard errors and over-rejection of the null of $\theta = 0$ (Bertrand et al., 2004), I

implement a non-parametric permutation test for $\theta = 0$. I construct constructing placebo triplets of [IDB state]-[year of vote in]-[time lag to upholding]. I then re-estimate the main specification 1 using 500 random number generator seeds, for employee share and exemption threshold outcomes. The significance levels are obtained as the cumulative distribution of placebo effects below (above) the main specification employee (threshold) estimate, and remain very similar.²¹

The estimated magnitudes are not driven by events which occur long after the upholding event. Magnitudes remain very similar when I narrow the sample to the IDB states and a time window of 15 years before the vote in and 15 years after after upholding. Results are robust to interacting pre-IDB cross-sections of determinants of structural transformation (illiteracy rate; urbanization rate; population density) with time.

Results are robust to allowing non-parametrically for civil law origin states to be on a different time path.

5 Model

5.1 Setup and empirical prediction

I consider a fixed distribution of income z across workers with pdf $h(z)$ and cdf $H(z)$. I assume exogenous employment shares of self-employed and employees at each income level, denoted respectively φ_z and $1 - \varphi_z$. I do not model the development process that leads to changes in employment shares, but assume it follows the patterns documented in stylized fact #1.

If the agent reports income $z \geq K$, then she is liable to pay $\tau(z - K)$. Otherwise, she is not liable for income tax. K is the exemption threshold and τ is the marginal tax rate. I assume linear utility to abstract from income effects. I assume agents have access to an evasion technology which allows them to pay c in order to report income at K and fully evade taxes. This evasion technology generates ‘bunching’ of reported income at K , in line with large set of evidence on evasion behavior. The cost is assumed to be infinite for employees: $c^E(z) = \infty$. For self-employed, the cost depends flexibly on total income z (due perhaps to a ‘visibility’ effect) and on the distance between income and the threshold $z - K$ such that $c^{SE} = c(z, z - K) > 0$. The cost is assumed to be increasing and convex in z . In this setting, there will exist a ‘marginal buncher’ at income \bar{z} who is indifferent between bunching

²¹The non-parametric significance levels are nearly identical, in a permutation exercise which construct placebo [year of vote in]-[time lag to upholding] but which maintains the observed set of IDB states.

and full compliance: $\bar{z} - c(\bar{z}, \bar{z} - K) = K + (1 - \tau)(\bar{z} - K)$. All self-employed with income z : $K \leq z \leq \bar{z}$ will under-report and bunch at the threshold. An increase in τ unambiguously leads to more evaders: $\frac{\partial \bar{z}}{\partial \tau} > 0$. An increase in the threshold will lead to less evaders if the marginal gain from compliance is larger than the marginal gain from under-reporting after the threshold decrease, that is

$$\frac{\partial \bar{z}}{\partial K} < 0 \text{ if } \tau > c_K(\bar{z}, \bar{z} - K) \quad (4)$$

I will assume the condition in 4 holds. The revenue base reflects evading self-employed between K and \bar{z} :

$$R = \int_{z \geq \bar{z}} \tau(z - K) \varphi_z dH(z) + \int_{z \geq K} \tau(z - K) (1 - \varphi_z) dH(z)$$

Consider a reform which locally decreases the threshold: $dK < 0$. This reform will have two effects on revenue: a mechanical gain and a behavioral loss. The mechanical gain, dM , reflects the marginal increase in revenue collected due to the reform on the inframarginal agents, assuming no behavioral responses

$$\begin{aligned} dM &= -dK \tau \left[\int_{z \geq \bar{z}} \tau(z - K) \varphi_z dH(z) + \int_{z \geq K} \tau(z - K) (1 - \varphi_z) dH(z) \right] \\ &\geq 0 \text{ if } dK < 0 \end{aligned} \quad (5)$$

The behavioral loss, dB , reflects loss in revenue due to behavioral responses of the marginal agents

$$\begin{aligned} dB &= -\frac{\partial \bar{z}}{\partial K} dK \tau (\bar{z} - K) \varphi_K \\ &\leq 0 \text{ if } dK < 0 \end{aligned} \quad (6)$$

where I have used the local approximation that $\varphi_K \approx \varphi_{\bar{z}}$, which is plausible if the last buncher is not located too far above the threshold. At the revenue maximizing optimum, K^{Rev} , it must be that $dB + dM = 0$. This yields the characterization for the location of the threshold

$$\frac{K^{Rev}}{\bar{z}} = \frac{1}{\left[1 + \left[\frac{\text{Mech gain}}{\text{Beh loss}} \right] \cdot [\varepsilon_{\bar{z}, K} \varphi_K]^{-1} \right]} \quad (7)$$

where Mech gain = $\int_{z \geq \bar{z}} \tau(z - K) \varphi_z dH(z) + \int_{z \geq K} \tau(z - K) (1 - \varphi_z) dH(z)$, Beh loss = $h(\bar{z}) \bar{z}$, and where $\varepsilon_{\bar{z}, K}$ denotes the elasticity of the marginal buncher with respect to the threshold. By chang-

ing the mass of agents who respond to the local reform, the model predicts the main empirical result of this paper

Empirical prediction: An increase in employee share (self-employed share) locally around the threshold leads to optimally lower (larger) size of threshold

$$\frac{\partial K^{Rev}}{\partial \varphi_K} > 0 \quad (8)$$

Extension: administrative costs A policy literature in developing and developed countries (Tanzi, 1987; OECD, 2015; IMF, 2015) discusses differences in administrative costs between reconstructing information trails for self-employment earnings and aggregation of employee information trails by employers. I model the administrative cost of taxing an income segment z as an increasing function of the self-employed share on the income segment, $c(z) = c(\varphi_z)$. Revenue net of administrative costs equals

$$R = \int_{z \geq \bar{z}} \tau(z - K) \varphi_z dH(z) + \int_{z \geq K} \tau(z - K) (1 - \varphi_z) dH(z) - \int_{z \geq K} c(\varphi_z) dH(z)$$

The local threshold decrease $dK < 0$ will lead to an additional administrative marginal cost $dC = dK \cdot c(\varphi_K) < 0$ if $dK < 0$. The revenue maximizing threshold now equals

$$\frac{K^{Admin.Rev}}{\bar{z}} = \frac{1}{\left[1 + \left[\frac{\text{Mech gain} - dC(\varphi_K)}{\text{Beh loss}} \right] \cdot [\varepsilon_{\bar{z}, K} \varphi_K]^{-1} \right]} \quad (9)$$

where the threshold is now predicted to increase due both to behavioral distortions and administrative costs that increase as the self-employed share goes up.

In the appendix I show that the formula 9 is robust to any general social welfare function.

Discussion: objective functions The full set of empirical results on the impact of employee share on the threshold and income tax revenue are consistent with an objective function of revenue maximization over the exemption threshold. An extension to the objective function which is also consistent with the full set of results is to include a social preference for a 'fair tax base'. Discussed especially

in a setting of low enforcement capacity countries, fairness relates to the idea that the tax base should not discriminate against particular groups in terms of compliance. On the income tax base, such fairness would imply that a group's share in effective contribution to tax revenue should be equal to its share in statutory contribution. I formalize this channel in the online appendix by modeling a 'misrepresentation' index given by the ratio of employee-share of income on the statutory income tax base to the ratio of employee-share of income on the compliant income tax base. Society faces social loss with parameter μ from any deviation of this index from a situation of perfect representation (with index value 1)

$$\text{Horizontal inequity} = \mu \left(1 - \left| \frac{\text{Employee-share on statutory income tax base}}{\text{Employee-share on compliant income tax base}} \right| \right) \quad (10)$$

So long as self-employed evade more than employees, the inequity cost associated with a lowering of the income tax threshold, dE , will always be smaller when the employee-share at the local threshold is larger. In the appendix I extend the formula for the optimal threshold that solves $dM + dB + dC + dE = 0$. This horizontal equity channel delivers a non-trivial prediction for movements in exemption threshold driven by gradual increases in employee share (stylized fact #2) and for constant employee share above the threshold (stylized facts #2 & #3) in the simplest possible setting of costless full evasion by self-employed (unlike the behavioral distortions channel and the administrative cost channel).²²

The empirical prediction 8 is also consistent with an objective function that minimizes social distortions across tax bases subject to a minimum revenue requirement.²³ But such a model predicts decreases in revenue collected from distortionary tax bases as employee share increases, which is not consistent with the US regression results.

5.2 Model fit and new proxy for fiscal capacity

I use the K^* derived under 9 to perform two simple quantitative exercises. First, I predict the location of the threshold in all 90 countries in the cross-country section of the micro database. I use the calculated distribution φ_j of employee shares across deciles j of the country's income distribution as the only source of cross-country variation, and assume constant values for ε , ω , λ , and the earnings-

²²In a setting where agents differ in skill level and in cost of avoidance, Kopczuk (2001) derives conditions under which high marginal tax rates can exacerbate horizontal inequity.

²³In which case individual tax instruments will be set such that their marginal social costs are equalized.

distribution which characterizes the mechanical gain and behavioral loss expressions.²⁴ I calculate the predicted size of the income tax base as the sum of percentiles that lie above the predicted percentile-location of the exemption threshold. I detail the choice of values for the different parameters in the appendix. I compare the predicted size of tax base to observed size of tax base, calculated as the sum of percentiles above the actual location of the exemption threshold. Panel A of Figure 10 shows results from this exercise. Variation across development in the employee-share predicted exemption threshold can account for 62 percent of the observed cross-country variation in exemption threshold (based on a simple R-squared). The predicted threshold lies within 3-5 percentiles of the actual threshold in countries at levels of development ranging between China, Indonesia and Mexico. At higher levels of income, the predicted threshold is on average ten percentiles lower than the actual threshold, due to the large increase in self-employed in the bottom decile. It is as large as 25 percentiles in countries like Italy and Spain where the self-employed share is prevalent in the the bottom third deciles. At lowest levels of development, the predicted threshold is on average 10 percentiles below the observed threshold. This masks heterogeneity between countries where the model performs very well (Rwanda: 1 percentile gap), fairly well (India: 6 percentile gap) and poorly (Kenya: 25 percentile gap). The average over-prediction in less-developed countries could be closed down by allowing enforcement capacity to exogenously increase over development. This highlights interactions between economic growth processes and investments in state capacity to explain the full variation in tax take along the development path. The gap could also be closed down by varying demand for redistribution or public goods.

In a second exercise, I construct the predicted distribution of employee shares at each of ten estimated steady state per capita income points from Section III. I predict the location of K^* from 9 in these ten constructed income distributions using the formula from 9 . I plot the ten combined minimum income points and associated size of income tax base in Panel B of Figure 10. I interpret these changes as proxies for 'stages' of fiscal capacity to expand the tax base. At per capita income point associated with steady state employee share in decile j , a country has reached a level of development which is associated with an enforceable tax base of size $10 - j$.

²⁴The distributions of φ_j in each country were used to construct the profiles for the stylized facts in Section III.

6 Conclusion

This paper has provided evidence and supporting theory to show that transition into employee jobs over development is an important driver of tax capacity. The paper introduces a novel channel, in which increases in employee share occurring gradually further down the income distribution causes broadening of the base through lowering the exemption threshold.

By focusing on the exemption threshold, this paper has provided evidence on the interaction between the extensive margin of income tax compliance and a statutory tax instrument. While this margin of moving agents into the income tax net is potentially an important compliance margin to explain tax capacity, there currently exists no well identified micro evidence on it.²⁵ Providing an estimate of the extensive margin of income tax compliance is an area for future research (Agostini & Jensen, 2016). The large variation in the location of the tax threshold across development suggests statutory tax instruments may be important determinants of tax capacity. On the other hand, current literature has focused on enforcement technologies and political economy as the main determinants of tax capacity. On-going research (Abramovsky, Bachas & Jensen, 2015) attempts to build new measures of statutory tax instruments to investigate further their relative importance in explaining tax capacity.

The evidence in this paper suggests the importance of studying jointly the drivers of development and their impacts on taxation. A simple but robust finding has been the close match both between less developed countries and currently advanced countries at similar levels of development. This suggests that a small income tax base in a less developed country reflects the same factors which lead to a low employee share. Future research could study factors which explain the patterns of gradual increases in employee share over the income distribution.

The research design of this paper has highlighted the usefulness of building micro evidence to answer macro questions. Such design could be applied to study other questions in taxation and development. There exists compelling micro evidence on the enforcement gains from exploiting sales connections between firms (Pomeranz, 2015). This evidence could be combined with the development-macro evidence on the growth in complexity and interconnectedness of firms (Poschke, 2011) to explain patterns of sales tax structure and revenue collection over development.

²⁵Field experiments in developing countries including Sri Lanka, Brazil, and Malawi study the impacts on business registration from varying the costs of formalization, but do not address any impact on labor income taxes (review of evidence in Bruhn and McKenzie, 2014).

References

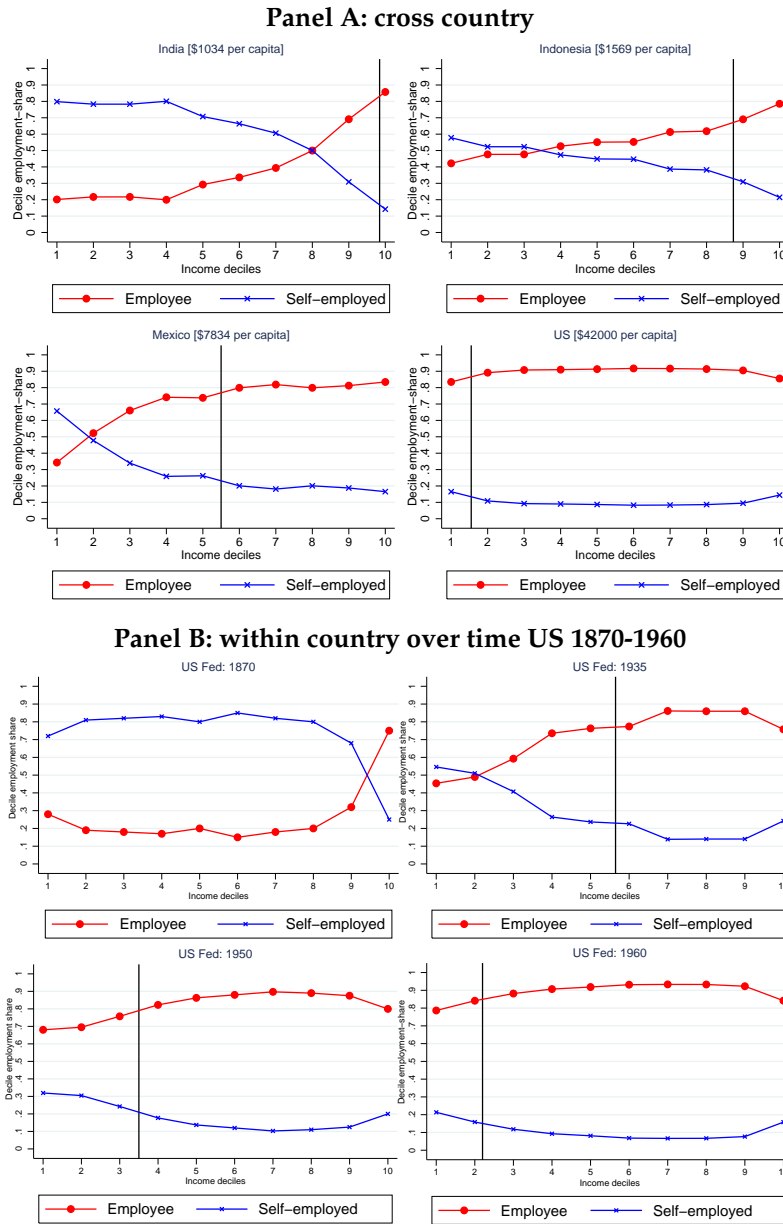
- [1] Abbey, A. [1966]. "Municipal Industrial Development Bonds", *Vanderbilt Law Review*, 19.
- [2] Abramovsky, L., P. Bachas, A. Jensen. [2015] "Tax Structures and Size Based Tax Policies: A Cross Country Database and Analysis", IGC working paper.
- [3] Acemoglu, D., J. Robinson [2000]. "Why did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective", *Quarterly Journal of Economics*, 115(4).
- [4] Agostini, C., A. Jensen [2016]. "Employment Choice and the Extensive Margin of Compliance: Evidence from Chile", unpublished.
- [5] Area Redevelopment Administration [1961]. *Records of the Office of Planning and Research. Several volumes*. Washington D.C: Department of Commerce and National Archives
- [6] Area Redevelopment Administration [1963]. *Statistical Profiles for Rural Redevelopment Areas. All states*. Washington D.C: Department of Commerce and National Archives
- [7] Bachas, P. [2015]. "Not(ch) Your Average Tax System: Corporate Taxation Under Weak Enforcement", Unpublished manuscript.
- [8] Bakija, J. [2009]. "Documentation for a Comprehensive Historical U.S. Federal and State Income Tax Calculator Program", Williams College working paper.
- [9] Banerjee, A., E. Duflo [2007]. "The Economic Lives of the Poor", *Journal of Economic Perspectives*, 21(1).
- [10] Banerjee, A., A. Newman [1993]. "Occupational Choice and the Process of Development", *Journal of Political Economy*, 101(2).
- [11] Best, M. [2014]. "The Role of Firms in Workers' Earnings Response to Taxes: Evidence from Pakistan", Stanford Institute for Economic Policy Research working paper.
- [12] Best, M., A. Brockmeyer, H. Kleven, J. Spinnewijn, M. Waseem [2014]. "Productiob Versus
- [13] Bertrand, M., E. Duflo, S. Mullainathan [2004]. "How Much Should We Trust Differences-In-Differences Estimates?", *Quarterly Journal of Economics*, 199(1).
- [14] Besley, T., T. Persson [2011]. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton University Press.
- [15] Besley, T., T. Persson [2014]. "Why Do Developing Countries Tax So Little?", *Journal of Economic Perspectives*, 28(4).
- [16] Besley, T., T. Persson, D. Sturm [2010]. "Political Competition, Policy and Growth: Theory and Evidence from the US", *Review of Economic Studies*, 77(4).
- [17] Berkowitz, D., K. Clay [2005]. "American Civil Law Origins: Implications for State Constitutions", *American Law and Economic Review*, 7(1).
- [18] Bick, A., N. Fuchs-Schundeln, D. Lagakos [2015]. "How do Average Hours Worked Vary with Development? Cross-Country Evidence and Implications", Unpublished manuscript.
- [19] Bird, R., O. Oldman [1964]. *Readings on Taxation in Developing Countries*. Johns Hopkins University Press.

- [20] Brautigam, D., O. Fjeldstad, M. Moore [2008]. *Taxation and State-Building in Developing Countries: Capacity and Consent*. Cambridge University Press
- [21] Bruhn, M., D. McKenzie [2014]. "Entry Regulation and the Formalization of Microenterprises in Developing Countries", *World Bank Research Observer*, 29(2).
- [22] Bustos, P., B. Caprettini, J. Ponticelli [2015]. "Agricultural Productivity and Structural Transformation: Evidence from Brazil", forthcoming *American Economic Review*.
- [23] Cage, J., L. Gadenne [2014]. "Tax Revenues, Development, and the Fiscal Cost of Trade Liberalization, 1792-2006", Unpublished manuscript.
- [24] Carillo, P., D. Pomeranz, M. Singhal [2014]. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement", *NBER Working Paper* No.20624.
- [25] Cobb, J. [1993]. *The Selling of the South: the Southern Crusade for Industrial Development 1936-1990*. University of Illinois Press
- [26] Council of State Governments [years between 1939 and 2009]. *Book of the States, all Parts*. Washington DC: The Council of State Governments.
- [27] Dincecco, M., U. Troiano [2015]. "Broadening the State: Policy Responses to the Introduction of the Income Tax", *NBER Working Paper* No. 21373.
- [28] Frias, J., T. Kumler, E. Verhoogen [2015]. "Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico", Unpublished manuscript.
- [29] Gallman, R. [2000]. "Economic Growth and Structural Change in the Long Nineteenth Century", in S. Engerman & R. Gallman (eds.), *The Cambridge Economic History of the United States: Volume II: The Long Nineteenth Century*. Cambridge University Press.
- [30] Gerard, F., G. Gonzaga [2014]. "Informal Labor and the Efficiency Cost of Social Programs: Evidence from 15 years of Unemployment Insurance in Brazil", unpublished.
- [31] Gillitzer, C. [2013]. *Essays in Macroeconomics and Public Finance*. Doctoral dissertation: University of Michigan.
- [32] Gollin, D. [2008]. "Nobody's Business But My Own: Self-employment and Small Enterprise in Economic Development", *Journal of Monetary Economics*, 55(2).
- [33] Gollin, D., D. Lagakos, M. Waugh [2013]. "The Agricultural Productivity Gap", *Quarterly Journal of Economics*, 129(2).
- [34] Gordon, R., W. Li [2009]. "Tax Structure in Developing Countries: Many Puzzles and a Possible Explanation", *Journal of Public Economics*, 93(7-8).
- [35] Haines, M. [2006]. *Cost of Living of Industrial Workers in the United States and in Europe*. ICPSR study 7711. Ann Arbor, Michigan: Inter-university Consortium for Political and Social Research.
- [36] Hinrichs, H. [1966]. *A General Theory of Tax Structure Change During Economic Development*. The Law School of Harvard University.
- [37] Herrendorf, B., R. Rogerson, A. Valentinyi [2014]. "Growth and Structural Transformation", Chapter 6 in (eds.) Aghion, P., S. Durlauf, *Handbook of Economic Growth*, Vol.2.

- [38] ICPSR based on US Department of Labor [2009]. *Study of Consumer Purchases in the United States, 1935-1936*. ICPSR study 8908. Ann Arbor, Michigan: Inter-university Consortium for Political and Social Research.
- [39] ICPSR based on US Department of Labor [1986]. *Cost of Living in the United States, 1917-1919*. ICPSR study 8299. Ann Arbor, Michigan: Inter-university Consortium for Political and Social Research.
- [40] Imbs, J., R. Wacziarg [2003]. "Stages of Diversification", *American Economic Review*, 63-86.
- [41] International Labour Organization [2002]. "Report VI: Decent Work and the Informal Economy", in *International Labour Conference 90th session*. Geneva: International Labour Office.
- [42] International Labour Organization [2009]. "International Definitions and Prospects of Underemployment Statistics", ILO Bureau of Statistics, unpublished.
- [43] International Monetary Fund [2015]. "Current Challenges in Revenue Mobilization - Improving Tax Compliance", *IMF Policy Paper*.
- [44] Jensen, A. [2015]. "Tax-mix over development: historical evidence from US states", LSE working paper.
- [45] Keen, M. [2012]. "Taxation and Development - Again" IMF Working Paper 12/220.
- [46] Khan, A., A. Khwaja, B. Olken [2015]. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors", *Quarterly Journal of Economics*, forthcoming.
- [47] Kleven, H. [2014]. "How Can Scandinavians Tax So Much?", *Journal of Economic Perspectives*, 28(4).
- [48] Kleven, H., M. Knudsen, C. Kreiner, S. Pedersen, E. Saez [2011]. "Unwilling or Unable to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark", *Econometrica*, 79(3).
- [49] Kleven, H., C. Kreiner, E. Saez [2015]. "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries", forthcoming *Economica*
- [50] Kline, P., E. Moretti [2013]. "Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority", *Quarterly Journal of Economics*, 129(1).
- [51] Kopczuk, Wojciech [2011]. "Redistribution when Avoidance Behaviour is Heterogeneous", *Journal of Public Economics*, 81(1).
- [52] La Porta, R., A. Shleifer [2014]. "Informality and Development", *Journal of Economic Perspectives*, 28(3).
- [53] Lewis, A. [1954] "Economic Development with Unlimited Supplies of Labor". *The Manchester School* 22(2).
- [54] Lindert, P., J. Williamson [2016]. *Unequal Gains: A New History of American Growth and Inequality*. Forthcoming, Princeton University Press.
- [55] McCaig, B., N. Pavcnik [2013]. "Moving Out of Agriculture: Structural Change in Vietnam", *NBER Working Paper Series*, No.19616.
- [56] McNaig, B., N. Pavcnik [2015]. "Informal Employment in a Growing and Globalizing Low-Income Country", *American Economic Review: P&P*, 105(5).

- [57] Moody's Investor Service [1974]. *Municipal and Government Manual of Issuers*. New York: Moody's Investor Service Inc.
- [58] Musgrave, R. [1966]. *Fiscal Systems*. New Haven: Yale University Press.
- [59] Naritomi, J. [2015]. "Consumers as Tax Auditors", Unpublished manuscript.
- [60] Pinsky, D. [1963]. "State Constitutional Limitations on Public Industrial Financing: an Historical and Economic Approach", *University of Pennsylvania Law Review*, 111(3).
- [61] Platz, D. [2009]. "Infrastructure Finance in Developing Countries - the Potential of Sub-Sovereign Bonds", *DESA Working Paper*, No.76.
- [62] Pomeranz, D. [2015]. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax", *American Economic Review*, forthcoming.
- [63] Porta, R., F. Lopez-de-Silanes, A. Shleifer [2008]. "The Economic Consequences of Legal Origins", *Journal of Economic Literature*, 46(2).
- [64] Rollinson, M. [1976]. *Small Issue Industrial Development Bonds*. Capital Publishing Corporation.
- [65] Saez, E. [2001]. "Using Elasticities to Derive Optimal Income Tax Rates", *Review of Economic Studies*, 68.
- [66] US Advisory Commission on Intergovernmental Relations [1963]. *Industrial Development Bond Financing: A Commission Report*. Washington D.C., US: Advisory Commission on Intergovernmental Relations.
- [67] US Internal Revenue Service [years between 1945 and 2000]. *Statistics of Income: Individual Income Tax Returns*. Washington DC: Department of the Treasury.
- [68] World Bank [2-13]. *World Development Report 2013: Jobs*. Washington DC: World Development Bank.
- [69] Wallis, J. [2000]. "American Government Finance in the Long-Run: 1790 to 1990", *Journal of Economic Perspectives*, 61-82.
- [70] Yitzhaki, S. [1979]. "A Note on Optimal Taxation and Administrative Costs", *American Economic Review*, 475-480.

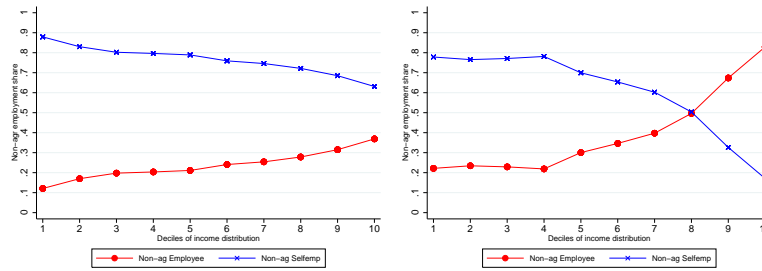
FIGURE 1: EMPLOYEE SHARE OVER INCOME DISTRIBUTION AND DECREASE IN INCOME TAX EXEMPTION THRESHOLD



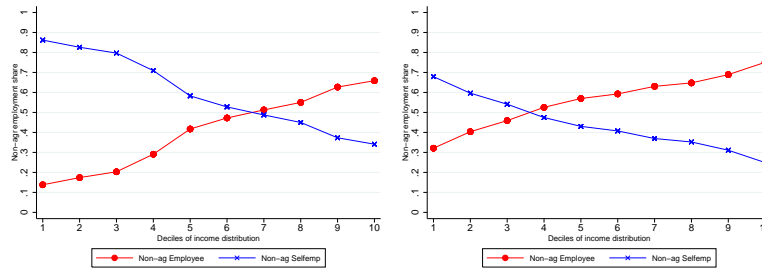
These figures plot the employment-shares of employees and self-employed over deciles of the income-distribution, for different countries (Panel A) and within-country over time (Panel B). The share of each work-type is defined as the share of total non-agricultural employment in the decile of the income-distribution. Employees are defined as individuals working in a firm with size > 1 ; self-employed are defined as individuals who report working as own-account workers, or as employees in a firm of size 1, or in a family-business with no employer. In each graph, the black solid denotes the location of the personal income tax (PIT) exemption threshold, taken from the tax code of the relevant country-year. The PIT threshold is defined as the level of gross income above which an individual earner becomes liable to pay personal income tax. The source for each graph is a household micro-dataset containing a nationally representative sample; in all underlying household surveys, the work-type status is mutually exclusive at the level of the individual. Source: Appendix and Section 3.

FIGURE 2: EMPLOYEE SHARE: REPRESENTATIVE DEVELOPMENT PROFILES

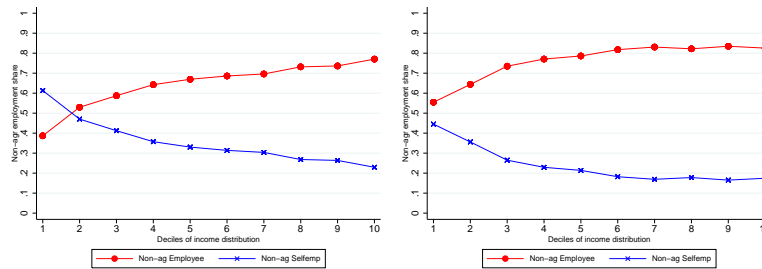
Profile for average country at \$277 pc [LHS] and \$730 pc [RHS]



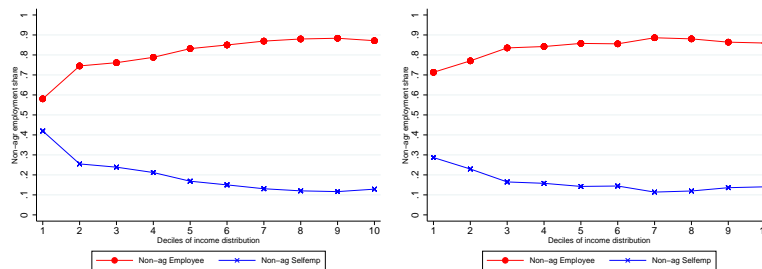
Profile for average country at \$1422 pc [LHS] and \$3286 pc [RHS]



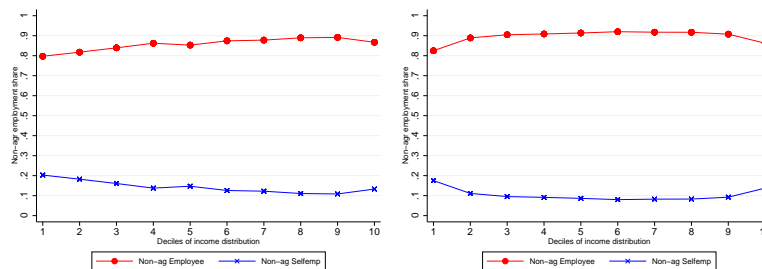
Profile for average country at \$4638 pc [LHS] and \$6945 pc [RHS]



Profile for average country at \$13512 pc [LHS] and \$27596 pc [RHS]



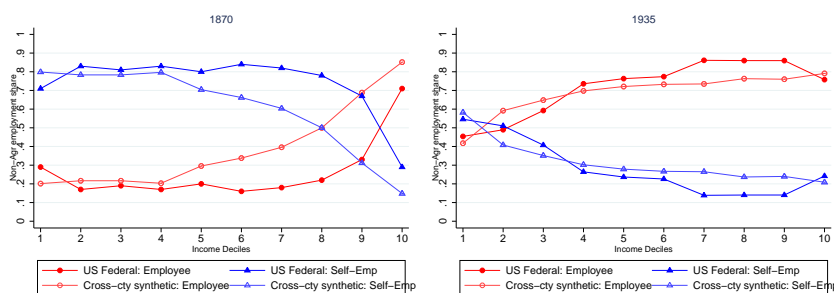
Profile for average country at \$37369 pc [LHS] and \$53234 pc [RHS]



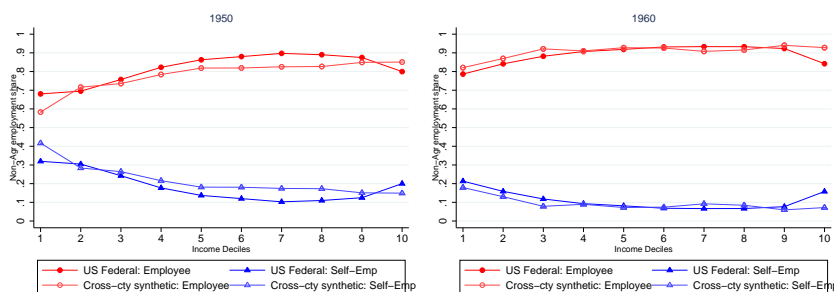
These panels depict the average employment structure profiles over deciles of the income-distribution. Red dotted (blue cross) observations indicate the employee (self-employed) share of non-agricultural employment in an income decile. A profile of employment shares is first constructed for the 90 individual countries in the cross-section of the micro database. Then an average profile is constructed over the profiles of countries that lie in a bin with indicated average real per capita income. The bins correspond to deciles of the real per capita income distribution across the 90 countries. Source: Appendix and Section 3.

FIGURE 3: EMPLOYEE SHARE: WITHIN US AND PAIRED SYNTHETIC CROSS-COUNTRY PROFILES

Federal US 1870 profile [\$2445 per cap] and Federal US 1935 profile [\$6212 per cap]

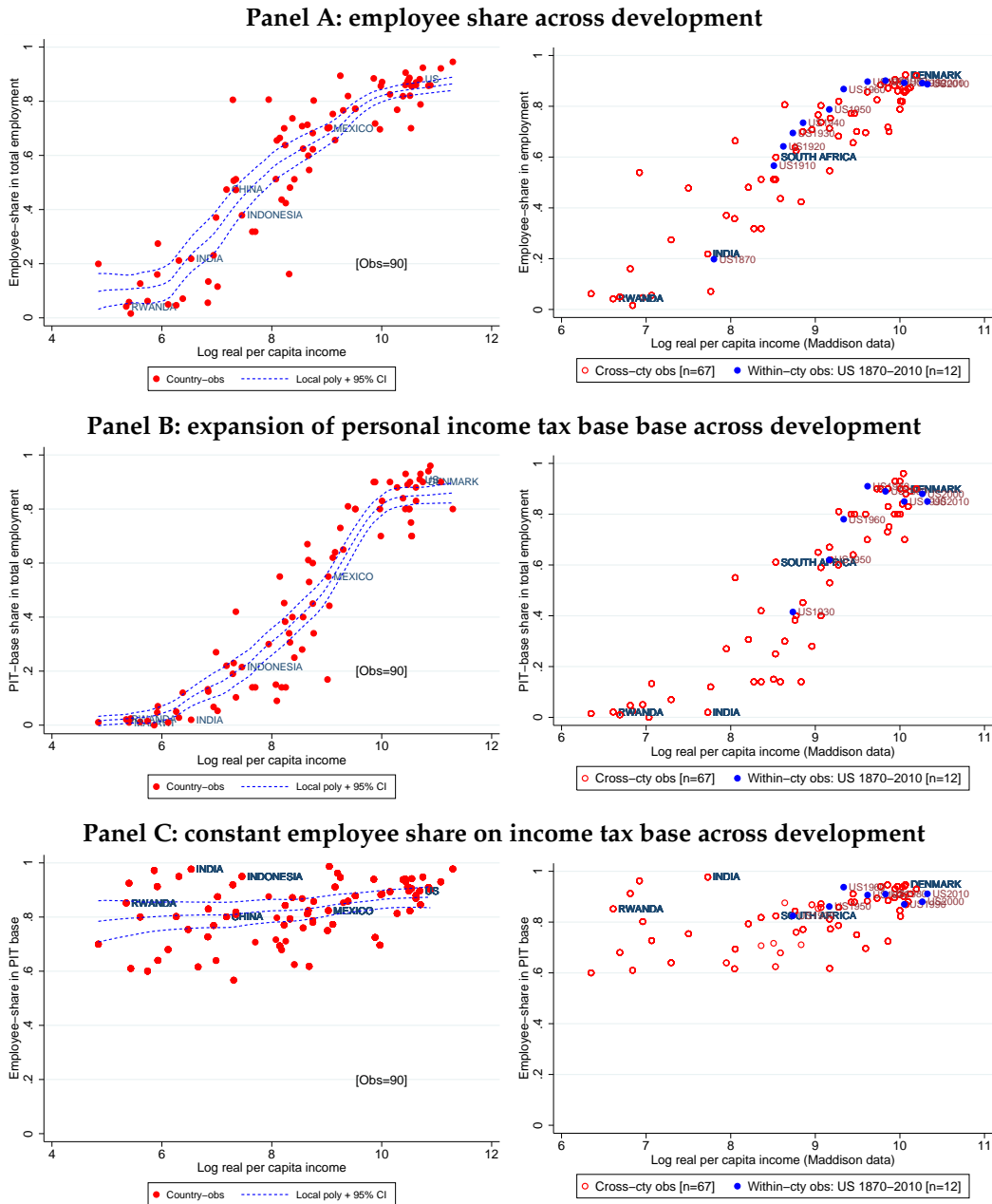


Federal US 1950 profile [\$9561 per cap] and Federal US profile 1960 [\$11338 per cap]



These panels plot employment structure shares over the income distribution in the US between 1870 and 1960. Each graph are constructed in exactly the same way as the profiles in Figure X, but using a nationally representative sample of the U.S. employed population in 1870, 1935, 1950 and 1960. In each panel, the solid circle and triangle denote the employment shares in the US historical year. The hollow observations denote the employment-shares in the synthetic country based on the cross-country section of the micro database. For each US profile, the paired synthetic cross-country profile is created by taking the average over profiles of countries whose real per capita income lies within ± 10 percent of the US real per capita income, using the Maddison dataset. The per capita income reported in brackets corresponds to the average income in the US in the year of the survey in the Maddison data. Source: Appendix and Section 3.

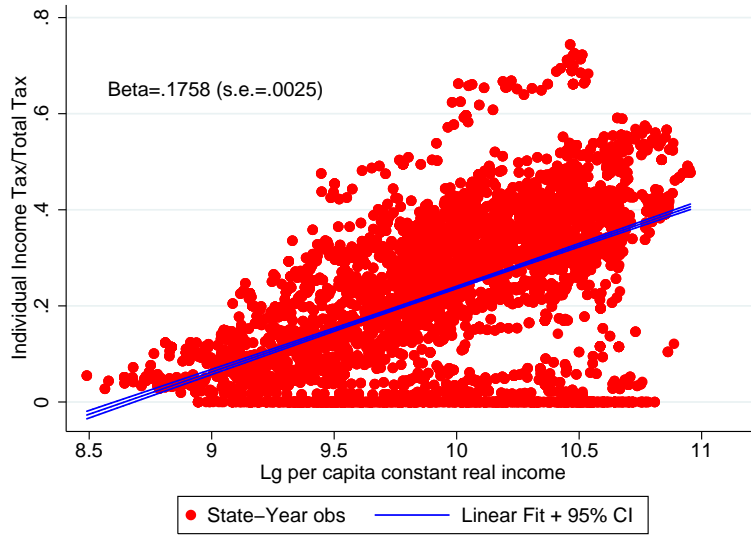
FIGURE 4: EMPLOYEE SHARE, SIZE AND EMPLOYMENT-COMPOSITION OF INCOME TAX BASE



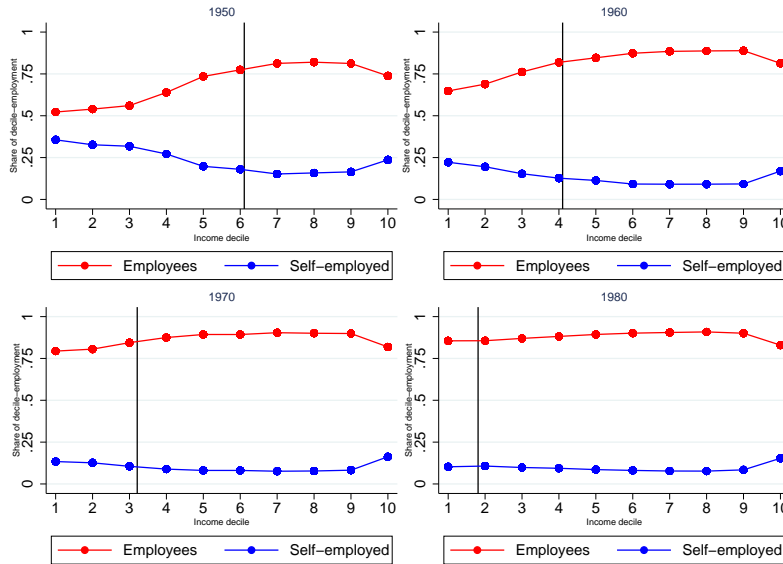
These figures plot correlations between log per capita income and: share of non-agriculture employees in total employment (Panel A); share of employment above the personal income tax threshold (PIT) in aggregate employment (Panel B); share of non-agriculture employees in employment above the PIT (Panel C). Each country-observation is calculated using a household micro-dataset containing a nationally representative sample; in all underlying household surveys, the work-type status is mutually exclusive at the level of the individual. Within each panel, the LHS graph plots the correlation using the full sample of surveys in the cross-country section of the micro database; the RHS graph uses the full set of surveys from the within-country section of the database together with the subset of surveys from the cross-country section which could be appended using the Maddison real per capita income database. Dashed lines denote the local polynomial fit on the underlying observations together with a 95% confidence interval. In Panel A, the employee-share of employment is defined as the share of workers who report working in non-agriculture industries, and in firms of size > 1 and which are not family units or casual daily-wage laborers. In Panel B, the PIT-base share in total employment is defined as the number of percentiles of the gross earnings income distribution which lies above the income tax exemption threshold. In Panel C, the employee-share in the PIT base is defined as the share of non-agriculture employees in total employment counted over the percentiles of the country's income distribution which lie above the income tax exemption threshold. Source: appendix and section 3

FIGURE 5: RISE OF INCOME TAX, TAX BASE EXPANSION AND EMPLOYEE SHARE: US STATES

Panel A: State income tax share of total taxes: all states 1939-2010

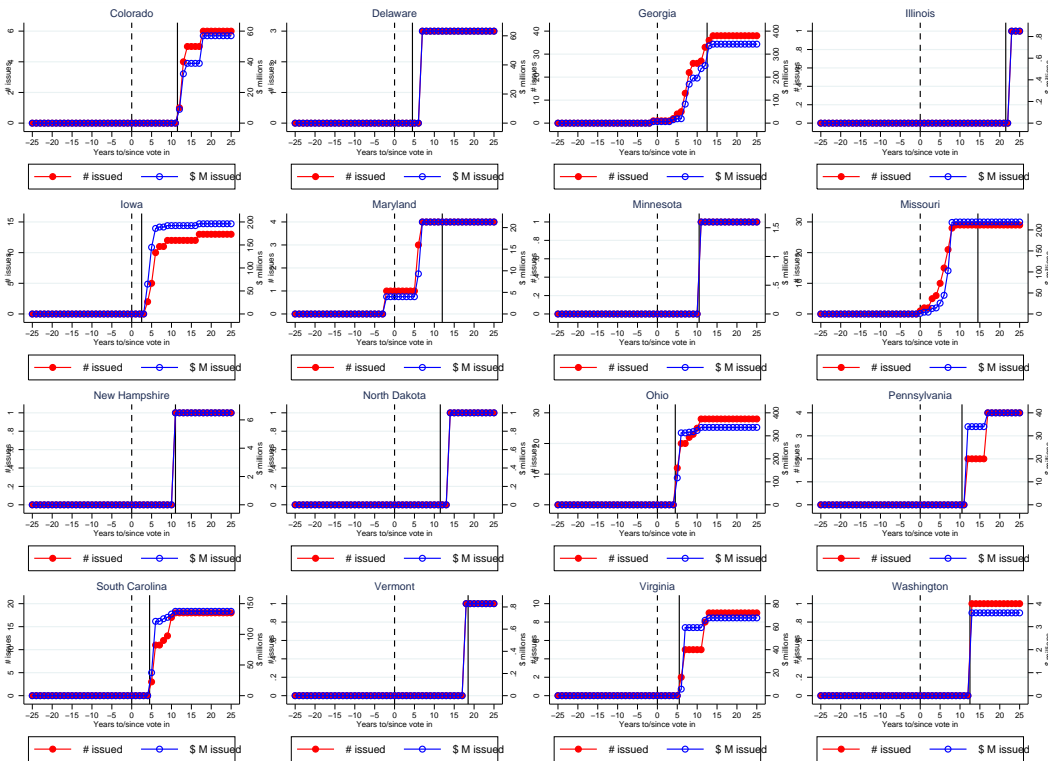


Panel B: employee share and state exemption threshold: average state 1950-1980



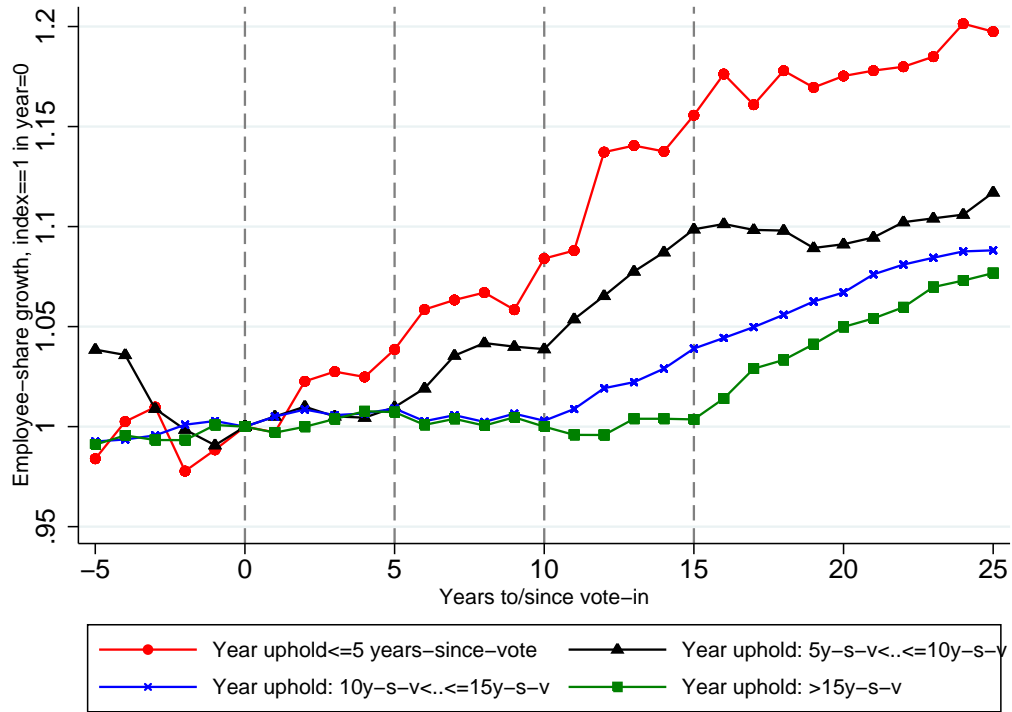
Panel A plots the state income tax share of total state taxes, using all state-year between 1939 and 2010, against real per capita income. The tax-mix observations are from Besley & Jensen (2015), originally sourced from historical Census records. The real per capita income is constructed as the per capita income in a state-year from the historical BEA series, deflated by the historical CPI. The solid lines denote the linear fit with a 95% confidence interval from the regression on the full underlying state-year observations. Panel B plots the employment-shares of employees and self-employed over deciles of the income-distribution, for the average state in the US between 1950 and 1980. The employee-share is defined as the share of employed agents who report being employed in a non-agricultural industry; the self-employed share is defined as the share of agents who report working on their own account, or as employee in a firm of size 1. The PIT threshold corresponds to the state PIT threshold (which in most cases differs from the Federal PIT threshold), and is calculated at the state-year level using the Bakija (2015) historical state-tax calculator. The income-decile distribution of employment-shares is first calculated for each state-year, then the average is taken over all the continental states ($N=48$) in a given year; the value of the exemption threshold is calculated in every state-year, then the average is taken over all threshold values. Source: appendix and section 4.

FIGURE 6: INDUSTRIAL DEVELOPMENT BOND PROGRAM: VOTE IN, COURT UPHOLDING AND ISSUANCE



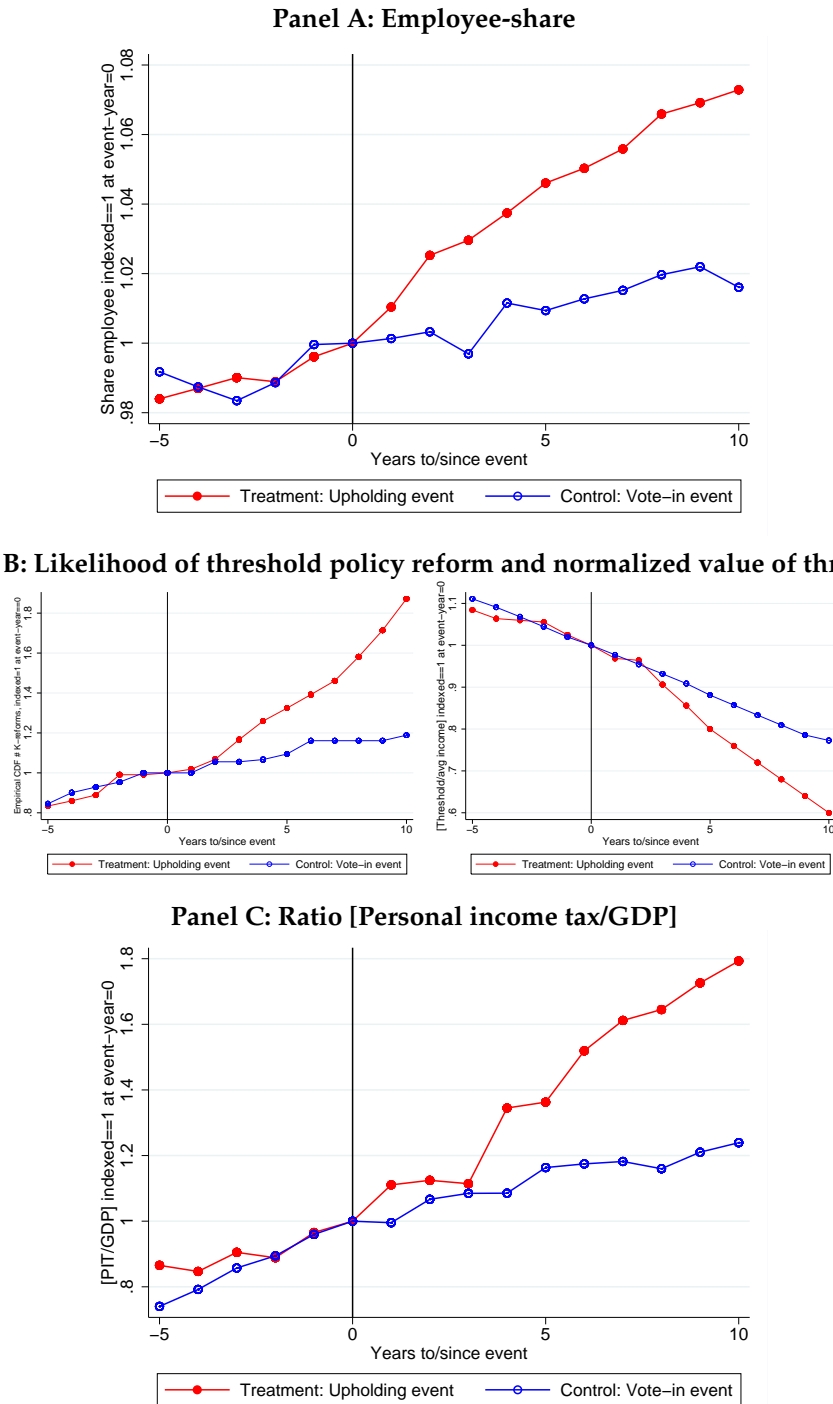
This panel plots time-series of cumulative IDB-debt in number of issues and in millions of \$ of principal, for the selected 16 states with the largest time lag between upholding and vote in. In each state, time is indicated as years to/since the vote in event, which is year 0 (black dashed line). In each state, the solid vertical line denotes the year of the court upholding of IDB. Source: appendix and Section 4

FIGURE 7: GRAPHICAL EVIDENCE ON IMPACT OF UPHOLDING EVENT ON EMPLOYEE SHARE



This figure shows the evolution of employee-share of earned income for groups of states for which the time-lag between the vote-in and the upholding event differed. For the circle-group, $\text{timelag} \in [0, 5]$; triangle group, $\text{timelag} \in (5, 10]$; cross group, $\text{timelag} \in (10, 15]$; square group, $\text{timelag} \in (15, 20]$. Each series shows the evolution of employee-share for the average state in the group, where employee-share is indexed to 1 in the year of the vote-in for all groups. Vertical dashed lines denote years since vote-in where a first upholding-decision occurs within a group. Source: appendix and Section 4.

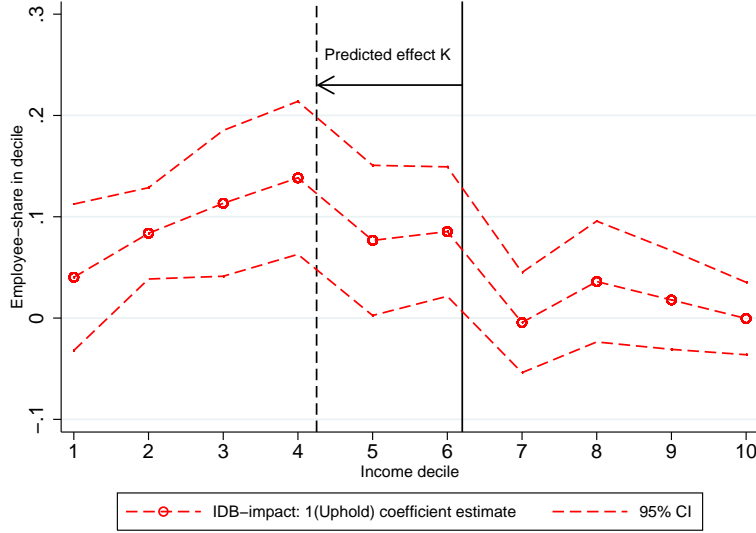
FIGURE 8: EVENT-STUDY WITHIN STATE AROUND VOTE IN AND UPHOLDING



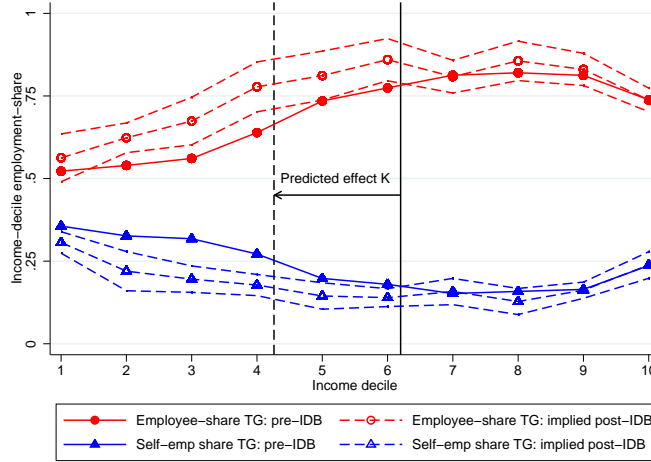
This figure shows the evolution of employee-share of earned income (panel A), size of PIT base and likelihood of PIT reform (panel B), and PIT to GDP ratio (panel C), within-state over time for the average treatment state around two events: vote-in, upholding. In the hollow-circle series (filled-circle series), the treatment-control is based on years before and after the event of state House vote-in of the Constitutional amendment to issue IDB (event of state-court upholding IDB). For each series, the evolution is normalized to 1 in the year of the event. In order to show the two events within-state for ten years after each event, these graph use only the subset of states for which the time-lag between vote-in and upholding exceeded 15 years, and the pre-event period is set to 5 years. These are arbitrary choices, and in the Appendix, I report the same exercise but for a shorter time-window to/since the event. In Panel B, the PIT-base is proxied for by the ratio of the state personal income tax exemption threshold to the state average earnings (right-hand side panel). The left-graph of Panel B shows the evolution of the state-specific empirical cumulative distribution of number of legislative reforms to the state exemption threshold: this measure controls for potential cross-state differences in frequency of tax-reform, and isolates any changes in likelihood that a reform to the state threshold will be passed around the event. Panel C shows the evolution of the personal income taxes relative to total resident earnings. Note the appealing time-lag across outcomes in materializing of the effect of the upholding-event: employee-share increases after 1 year, reform to PIT base and size of PIT base change after 2 years, and PIT-take increases after 3 years. Source: appendix and Section 4.

FIGURE 9: DISTRIBUTIONAL IDB IMPACTS ON EMPLOYEE SHARES AND THRESHOLD LOCATION

Panel A: IDB impact on employee shares and threshold location in income distribution



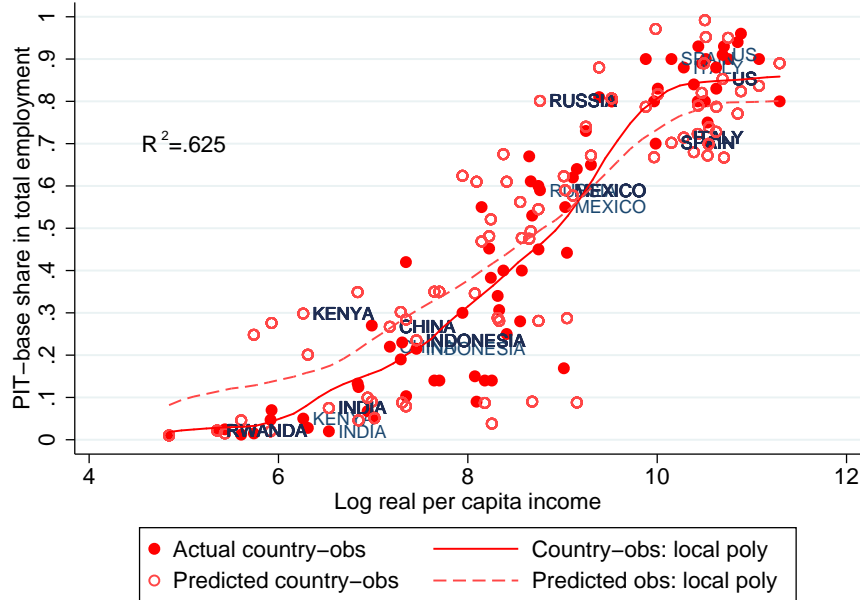
Panel B: implied IDB impact on employee shares and threshold location in income distribution



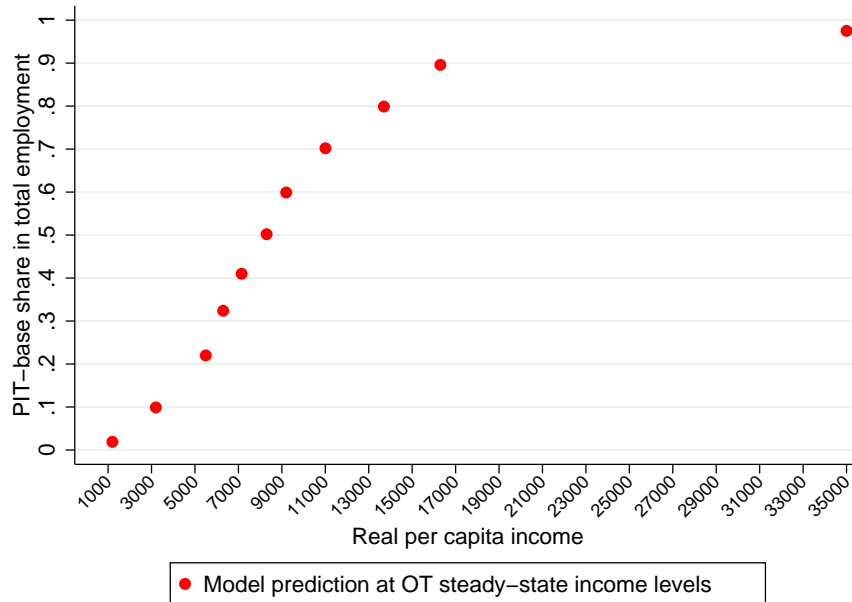
Panel A reports the coefficients $\hat{\theta}_j$ on the 1(Uphold) dummy in a regression on employee-share in decile $j = 1, \dots, 10$, using specification 1. Each hollow-circle denotes the decile- j point estimate $\hat{\theta}_j$ and the dashed lines denote the 95% confidence interval of the point-estimate (robust standard errors clustered at the state level). The black solid line denotes the location of the average PIT exemption threshold in the IDB-treatment state in the pre-IDB period, K^{PRE} . The dashed line shows the predicted post IDB percentile location of the threshold calculated as $K^{POST} = K^{PRE} + \widehat{dK}$, where $\widehat{dK} = -18.43$ corresponds to the reduced-form estimate using 2. Panel B plots the implied post-IDB employee-shares and self-employed shares. The implied employee-share in a decile j , θ_j^{POST} , is calculated as $\theta_j^{POST} = \hat{\theta}_j + \theta_j^{PRE}$ where θ_j^{PRE} is the average employee share in decile j calculated in the pre-IDB period. The construction of the implied self-employed shares is similar. The solid circle series denotes the pre-IDB distribution of employee shares, θ_j^{PRE} , while the hollow circle series denotes the post-IDB implied distribution θ_j^{POST} . Similarly for self-employed shares denotes with triangle symbols. Source: appendix and Section 4.

FIGURE 10: MODEL FIT

Panel A: Model predicted versus actual size of income tax base in 90 countries



Panel B: Model predicted size of income tax base at steady state employee share stages of development



Panel A plots the cross country actual (solid circle) and predicted (hollow circle) size of income tax base, for the 90 countries in the cross section of the micro database. The predicted size of income tax base is calculated based on the model derived in Section V, and predicts the location of the exemption threshold in all countries using the measured employment shares in the deciles of the country’s income distribution as the only source of cross-country variation. That is, values of enforcement capacity, administrative costs, earnings structure, demand for redistribution, marginal value of public goods are all assumed to be constant across countries. The highlighted countries are meant to illustrate cases of good and poor model fit. The cross country model fit is reported as the R-square from an OLS regression of the actual PIT base share in employment on the predicted PIT base share in employment. Panel B plots the predicted size of income tax base, using the model derived in Section V and the predicted decile distributions of employment shares based estimated at each of the ten development stages discussed in Section V. Similarly to Panel A, the model predicts the location of the threshold assuming constant values for all other threshold determinants. Source: Section 3 and Section 5

TABLE 1: DETERMINANTS OF CHANGE TO IDB LITIGATION-STATUS

	LHS=1(Upheld)										
	Non-parametric Cox proportional hazard model, hazard rate reported										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1(Civil law origins)	5.4351 (3.2017)***	4.6122 (3.4472)**	4.4784 (3.3070)**	6.7249 (4.7508)***	6.8493 (5.0120)***	7.0552 (5.3059)***	6.8491 (4.8204)***	5.4314 (4.6463)**	5.9501 (4.7442)**	9.2531 (9.3976)**	5.3948 (4.4142)**
Redev-share lab force		9.0393 (31.4626)	9.2889 (34.6653)	2.6910 (12.0902)	6.9380 (22.0226)	1.4096 (6.5460)	1.7451 (8.0408)	11.9789 (74.1761)	254.7608 (948.501)	.6733 (4.1576)	16.5004 (129.3488)
1(Poll tax)			1.0184 (.7572)	.9983 (.6685)	1.2002 (1.0741)	1.3098 (1.1022)	1.3129 (1.1176)	1.3968 (1.2716)	1.5858 (1.7699)	1.5628 (1.5572)	1.329 (1.5791)
1(Literacy test)			1.3063 (.7206)	1.2828 (.6659)	1.2042 (.5564)	.9580 (.4748)	.9745 (.5073)	1.2009 (1.1384)	.0372 (.0856)	.8025 (.7165)	1.2751 (1.6888)
Log(population)				.6759 (.2843)*	.5863 (.2467)	.6045 (.1880)	.6118 (.2023)	.6416 (.2100)	.5925 (.2283)	.5516 (.2010)	.6417 (.2123)
Manuf share lab force					7.4211 (33.6242)						
Employee share lab force						16.8598 (82.1034)	11.9849 (81.8919)	10.0700 (70.7272)	1794.923 (16589.73)	95.3021 (702.837)	9.6009 (70.2367)
Lg(per cap inc)							1.2221 (3.3499)	.9257 (2.3310)	.1472 (.5211)	.2766 (.7398)	.9216 (2.3287)
1(Southern state)								.6464 (.6681)	.1562 (.2125)	.6113 (.5602)	.6367 (.6839)
(K/y)									5.6413 (7.0495)		
(Pers income tax/GDP)										2.1414 (4.7415)	
(Total income tax/GDP)											.0626 (2.2671)
Obs	134	134	134	134	130	130	130	130	123	123	123

*, **, *** denote significance at the 10 percent, 5 percent, 1 percent level. Standard errors robust to clustering at the state level. This table reports the results of estimating non-parametric Cox proportional hazard models, where hazard rates are reported. Hence tests for significance relate to the null that the coefficient is equal to one. The unit of observation is state-year. A state enters the sample in the year where the Constitutional amendment allowing issuance of IDBs is voted in. The state drops the sample once the highest instance of the state court system has upheld the legality of the IDB-program. In Column 1, the baseline model includes a dummy for civil law origins. In the columns onwards, the baseline model is augmented with additional controls: col. (2) includes the state-share of labor force in redevelopment counties (time-invariant: TI) ; col.(3) includes indicators for whether the state has a poll tax and/or a literacy test for voting (time-varying: TV); col.(4) includes the log of state-population (TV); col.(5) includes manufacturing share of employment (TV); col.(6) includes the employee-share of total employment (TV); col.(7) includes the log of per capita income (TV); col.(8) includes a dummy for Southern states according to US Census definition (TI); col.(9) includes the ratio of state income tax threshold to average earnings (TV); col.(10) includes the ratio of personal income tax to state GDP (TV); col.(11) includes the ratio of total tax to GDP (TV). Sources: appendix and Section 4.

TABLE 2: IDB IMPACT ON EMPLOYMENT AND EARNINGS STRUCTURE

Panel A: Employment											
	Log(volume of employment)									Share of employment	
	Total (1)	Non agric employee (2)	Manufacturing (3)	Construction (4)	Trade (5)	Services (6)	Government (7)	Self employed (8)	Agric employee (9)	Non agric employee (10)	Self employed (11)
1(Vote in)	-0.0047 (0.0123)	0.0089 (0.0111)	-0.004 (0.0248)	0.0246 (0.0375)	-0.0145 (0.0131)	0.0023 (0.0185)	-0.0054 (0.0201)	.0084 (.0203)	-.0322 (.0309)	.0029 (.0035)	-.0042 (.0029)
1(Uphold)	-0.0243 (0.0241)	0.0394 (.0191)**	0.0763 (.0238)***	-0.036 (0.0544)	0.0252 (0.0215)	-0.0478 (.0277)*	.0054 (.0338)	-.0756 (.0240)***	-.0085 (.0496)	.0360 (.0073)***	-.0368 (.0052)***
R squared	0.9808	0.9878	0.9150	0.9067	0.9871	0.9939	0.9867	0.9422	0.9569	0.9810	0.9559
Number of states	48	48	48	48	48	48	48	48	48	48	48
Number of state-years	2890	2890	2890	2890	2890	2455	2890	2890	2890	2890	2890

Panel B: Earnings												
	Log(volume of earnings)					Log(average earnings)					Share of earnings	
	Total (1)	Non agric employee (2)	Self employed (3)	Employee-related transfers (4)	Other transfers (5)	Total (6)	Non agric employee (7)	Self employed (8)	Employee-related transfers (9)	Other transfers (10)	Non agric employee (11)	Self employed (12)
1(Vote in)	-.0089 (.0139)	-.0040 (.0133)	.0192 (.0216)	-.0101 (.0185)	-.0113 (.0236)	-.0004 (.0117)	.0022 (.0062)	.0215 (.0309)	.0000 (.0170)	.0049 (.0240)	.0048 (.0043)	.0000 (.0044)
1(Uphold)	-.0031 (.0209)	.0508 (.0215)**	-.0891 (.0420)**	.0880 (.0412)**	.0514 (.0429)	.0251 (.0200)	-.0004 (.0157)	-.0025 (.0361)	.0843 (.0472)*	.0435 (.0477)	.0429 (.0137)***	-.0407 (.0115)***
R squared	0.9988	0.9984	0.9896	0.9975	0.9982	0.9960	0.9989	0.9694	0.9959	0.9961	0.8913	0.8138
Number of states	48	48	48	48	48	48	48	48	48	48	48	48
Number of state-years	2855	2855	2855	2855	2855	2855	2855	2855	2855	2855	2855	2855

*, **, *** denote significance at the 10%, 5%, 1% level. Robust standard errors clustered at the state level in parentheses. Time period is 1939-2005. This table reports OLS estimates $\hat{\alpha}$ and $\hat{\theta}$

$$y_{st} = \beta + \alpha \mathbf{1}(\text{Vote in})_{st} + \theta \mathbf{1}(\text{Upheld})_{st} + \lambda \mathbf{X}_{st} + \mu_s + \gamma_t + \phi_s \cdot t + \varepsilon_{st}$$

where s denotes state, t denotes time, $\mathbf{1}(\text{Vote in})_{st}$ indicates whether a vote has occurred in the state-House to allow issuance of IDB but the IDB has not yet been upheld, $\mathbf{1}(\text{Upheld})_{st}$ indicates whether the court-system has upheld the legality of IDB. The vote-in and upholding events are mutually exclusive events. The set of state-time varying controls \mathbf{X}_{st} includes: log average resident earnings; dummies for the existence of a poll tax and a literacy test, both used for voting restrictions; dummies for state election years; dummies for the existence of a state corporate income tax and of right-to-work laws; a continuous measure of the firm-size coverage of state unemployment insurance laws. All regressions include a state-specific linear trend: $\phi_s \cdot t$. Source: appendix and Section 4.

TABLE 3: IDB REDUCED FORM IMPACT AND EMPLOYEE SHARE IMPACT ON INCOME TAX BASE

	Employee share impact on tax base:				Confounding channels:			
	Size of tax base		Composition of tax base		Tax rates	Earnings distribution	Enforcement	Demand for redistribution
	(PIT exemption threshold/average earnings)		[Employee-share above threshold]		Bottom MTR	Hazard ratio	Number tax agencies	Democratic vote share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1(Vote in)	-0.1400 (.1140)			-0.0052 (.0119)	.0018 (.0065)	.0484 (.0870)	.0401 (.0577)	.0072 (.0101)
1(Upheld)	-0.7218 (.3296)**			-0.0054 (.0185)	.0035 (.0075)	.0263 (.1473)	.0326 (.0965)	.0108 (.0178)
Employee-share		-20.318 (5.832)***	-29.588 (12.157)**					
1st stage F-test p-value			7.79 (.0012)					
Implied elasticity tax base - emp share	-6.678	-5.200	-7.548					
Number of states	48	48	48	48	48	48	48	48
Number of state-years	2931	2931	2931	2931	2931	2901	2815	2931
R-squared	0.7869	0.8111		0.7737	0.6762	0.3035	0.6719	0.6619
Method	OLS	OLS	IV	OLS	OLS	OLS	OLS	OLS

*, **, *** denote significance at the 10%, 5%, 1% level. Robust standard errors clustered at the state level in parentheses. Time period is 1939-2005. This table reports OLS estimates $\hat{\alpha}$ and $\hat{\theta}$

$$y_{st} = \beta + \alpha \mathbf{1}(\text{Vote in})_{st} + \theta \mathbf{1}(\text{Upheld})_{st} + \sum_{j=1}^{10} \omega_j z_{jst} + \lambda \mathbf{X}_{st} + \mu_s + \gamma_t + \phi_s \cdot t + \varepsilon_{st}$$

where s denotes state, t denotes time, $\mathbf{1}(\text{Vote in})_{st}$ indicates whether a vote has occurred in the state-House to allow issuance of IDB but the IDB has not yet been upheld, $\mathbf{1}(\text{Upheld})_{st}$ indicates whether the court-system has upheld the legality of IDB. The regression includes average income in all ten deciles of the state-year income-distribution, z_{jst} . The set of state-time varying controls \mathbf{X}_{st} include dummies for the existence of a poll tax and a literacy test, both used for voting restrictions; dummies for state election years; dummies for the existence of a state corporate income tax and of right-to-work laws; a continuous measure of the firm-size coverage of state unemployment insurance laws. All regressions include a state-specific linear trend: $\phi_s \cdot t$. In Col. (1), the implied elasticity of [exemption threshold/average earnings] is calculated based on the ratio of reduced-form estimates of $\mathbf{1}(\text{Upheld})$; in Col.(3), the implied elasticity of [exemption threshold/average earnings] is calculated based on the IV-estimated impact of employee-share on $[K/y]$. Source: appendix and Section 4.

TABLE 4: IDB IMPACT ON TAX TAKES

	Personal Income Tax/GDP			CorpIncT/GDP	GenSalesT/GDP	SelectSalesT/GDP	LicenceT/GDP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1(Vote in)	.0007 (.0009)			-.0001 (.0002)	-.0007 (.0008)	-.0005 (.0004)	.0003 (.0002)
1(Upheld)	.0017 (.0007)**			.0001 (.0004)	-.0002 (.0010)	.0001 (.0010)	-.0004 (.0003)
Employee-share		.0166 (.007)**	.0248 (.0109)**				
1st stage F-test p-value			7.96 (.001)				
Implied elasticity	1.292	.6225	.900				
Number of states	48	48	48	48	48	48	48
Number of state-years	2931	2931	2931	2931	2931	2931	2931
R-squared	0.9195	0.9280		0.7081	0.8653	0.8805	0.8414
Method	OLS	OLS	IV	OLS	OLS	OLS	OLS

*, **, *** denote significance at the 10%, 5%, 1% level. Robust standard errors clustered at the state level in parentheses. Time period is 1939-2005. This table reports OLS estimates $\hat{\alpha}$ and $\hat{\theta}$

$$y_{st} = \beta + \alpha \mathbf{1}(\text{Vote in})_{st} + \theta \mathbf{1}(\text{Upheld})_{st} + \sum_{j=1}^{10} \omega_j z_{jst} + \lambda \mathbf{X}_{st} + \mu_s + \gamma_t + \phi_s \cdot t + \varepsilon_{st}$$

where s denotes state, t denotes time, $\mathbf{1}(\text{Vote in})_{st}$ indicates whether a vote has occurred in the state-House to allow issuance of IDB but the IDB has not yet been upheld, $\mathbf{1}(\text{Upheld})_{st}$ indicates whether the court-system has upheld the legality of IDB. The regression includes average income in all ten deciles of the state-year income-distribution, z_{jst} . State-time varying controls \mathbf{X}_{st} include dummies for the existence of a poll tax and a literacy test, both used for voting restrictions; dummies for state election years; dummies for the existence of a state corporate income tax and of right-to-work laws; a continuous measure of the firm-size coverage of state unemployment insurance laws. All regressions include a state-specific linear trend: $\phi_s \cdot t$. In Col. (1), the implied elasticity of [PIT/GDP] is calculated based on the ratio of reduced-form estimates of $\mathbf{1}(\text{Upheld})_{st}$; in Col.(3), the implied elasticity is calculated based on the IV-estimated impact of employee-share on [PIT/GDP]. Source: appendix and Section 4