Employer power and employment in developing countries

Nancy H. Chau,¹ Ravi Kanbur,² and Vidhya Soundararajan³

September 2022
Abstract: The issue of employer power is underemphasized in the development literature. The default model is usually one of competitive labour markets. This assumption matters for analysis and policy prescription. There is growing evidence that the competitive labour markets assumption is not valid for employment in developing countries. Our objective in this paper is to review this evidence, to present theoretical and policy perspectives that follow from it, and to highlight areas for further research.

Key words: employer power, employment, developing countries

JEL classification: J42, O15

Acknowledgements: This paper has been prepared for the JICA-IPD Task Force on Employment, directed by Joseph Stiglitz, Arjun Jayadev, and Akbar Noman. We thank them and participants in Task Force meetings for their helpful comments. This paper was completed while Chau was visiting scholar at the UNU-WIDER. The hospitality of the institute and the comments and suggestions received are gratefully acknowledged.
1 Introduction

The issue of market power is underemphasized in the development economics literature. In particular, as concerns employment, the default model is one of competitive labour markets, but this assumption matters for analysis and policy prescription as there is growing evidence that this competitive assumption is not valid. Our objective in this paper is to review this evidence, to present theoretical and policy perspectives which follow from it, and to highlight areas for further research.

The idea that employers can have market power over workers is not new in economics. In Book 1, Chapter 8 of the Wealth of Nations, Adam Smith (1776) highlights such power and the effect it has on wages:

What are the common wages of labour, depends everywhere upon the contract usually made between those two parties, whose interests are by no means the same. The workmen desire to get as much, the masters to give as little as possible . . . It is not, however, difficult to foresee which of the two parties must, upon all ordinary occasions, have the advantage in the dispute, and force the other into a compliance with their terms. The masters, being fewer in number, can combine much more easily . . . In all such disputes the masters can hold out much longer . . . Masters are always and everywhere in a sort of tacit, but constant and uniform combination, not to raise the wages of labour above their actual rate.

Economic theorizing about employer power took a giant leap forward with Joan Robinson’s classic The Economics of Imperfect Competition (Robinson 1932). She famously coined the term ‘monopsony’ for the case of a labour market with a single employer, and developed the formula linking the elasticity of labour supply to the wage markdown when compared to the competitive equilibrium outcome. We will return to this formula many times in this paper.

Adam Smith also notes that ‘the law, besides, authorizes, or at least does not prohibit their combinations, while it prohibits those of the workmen’. In the history of developing countries, the legal backing from colonial powers to employers strengthening their bargaining power is well recorded. Indentured labour from India was effectively captive labour on plantations throughout the British empire, in Mauritius, South Africa, Fiji, the West Indies, Ceylon, etc. The draconian laws that increased employer power in these colonies are well recorded by Hugh Tinker (1974). Within India, the state notoriously supported European indigo plantation owners against the poor workers and sharecroppers, leading to Mahtama Gandhi’s first major agitation, the Champaran Satyagraha in April 1917 (Remesh 2006). Even without state backing, the burden of debt reduced the bargaining power of peasants against landlords, who in many cases were also their creditors, as documented by Sir Malcolm Darling (1928). These concerns on asymmetry in market power between landlords and peasant workers are also present in the history of Latin America and the role of large latifundia estates and landed elites in the political economy of development.

With this background, the literature on labour markets and employer power in developing countries seems to have developed in two directions. One direction continued the focus on the power of landlords as employers, through mainly micro-level studies in rural areas. Among the best known studies in this vein are those for India, by authors such as Beteille (1974) and Srinivas (1987) in sociology and anthropology, and Bardhan (1984), Datt (1996), and Rudra (1982) in economics. The empirical economic investigations were accompanied by theoretical understandings of the nature of power relations and the role of interlinkages across credit and labour markets in sustaining asymmetric market power (Bardhan 1980; Basu 1990; Bell and Srinivasan 1989). These types of studies in development economics went hand in hand with, and sometimes underpinned, policy interventions in credit markets and in land markets.
The second strand of literature in development economics is remarkable for being somewhat separate from, and seemingly uninformed by, the above literature on the nature of labour markets in developing countries. Two classic papers illustrate this strand. Arthur Lewis’ (1954) ‘Economic Development with Unlimited Supplies of Labour’ and Harris and Todaro’s (1970) paper, ‘Migration, Unemployment and Development: A Two-Sector Analysis’. The first of these papers was cited as a key item in Lewis’ Nobel Prize; the second has garnered more than 10,000 citations on Google Scholar, and was chosen as one of the top 20 articles published in the American Economic Review for its centenary edition in 2011.

Both of these papers present different sector models, and it is instructive to consider the labour market specifications in each. In the Lewis model, the rural or agricultural sector is famously the source of unlimited supplies of labour. Either because the marginal productivity of labour is zero and there is simply a socially determined wage, or as in subsequent literature because of household adjustments (Sen 1966), withdrawal of a unit of labour from the rural sector does not change the supply price of labour to the urban/industrial sector. There is thus no modelling of, and certainly no emphasis on, the rural labour market and market power asymmetries in this market. The modelling of industry is in terms of firms facing an infinitely elastic supply curve of labour. There is thus no market power modelled in this sector either. Profits are reinvested and demand for labour increases without an increase in the wage. The process continues till the surplus labour is exhausted and this phase of ‘dualistic development’ comes to an end.

In the Harris–Todaro (1970) model, the labour market is a standard competitive market with wages equal to the marginal product of labour with a standard production function. The action is in the urban market, where there is a formal sector with a minimum wage above the marginal product of labour so that there is urban unemployment. Thus the market power in this model is in the hands of urban formal sector workers. Firms are wage-takers with no market power. The focus of the model, rather, is famously on the migration equilibrium between the rural and urban sectors, where migrants compare the rural wage to the expected wage in the urban sector, allowing for the probability of actually securing a high-paying formal sector job.

The informal sector, where the urban unemployed congregate, is not modelled in detail by Harris and Todaro (1970), but subsequent literature essentially treats the urban informal sector in a competitive frame with no market power on either side, even when it models the process of transition from unemployment to employment in a more sophisticated fashion (Basu et al. 2019; Fields 1975). The idea of the informal sector being competitive has become deeply ingrained, and the wage being above the marginal product in the formal sector is then seen as a distortion that leads to inefficiency. But this notion of perfect competition in the informal labour market does not sit well with those who work at the ground level with these workers, especially women. The experience of organizations like SEW A (www.sewa.org) is that these workers need to be protected from employer power in their labour markets. Basu et al. (2015) provide a theoretical analysis where contracting difficulties in the informal sector labour market end up delivering market power into the hands of employers, in the very sector in which it is thought that competition reigns supreme.

Thus, in contrast to the tradition of microeconomic analysis of rural sector labour markets, in which market power of employers is emphasized, such asymmetry in market power is largely missing in the broad macro-level multisector analysis literature in development economics, following the lead of its classic progenitors. Indeed, one of these departures from competition is identified as being in the direction of worker power. This tradition leads naturally to a policy stance of deregulation of labour markets, starting from the removal of minimum wage laws, as the basis of a vibrant employment and development strategy. A competitive view of developing country labour markets also leads to a corresponding stance on the distributional consequences of trade liberalization. A standard ‘Stolper–Samuelson’ perspective says that trade liberalization will benefit the abundant factor of production, which in developing countries would be labour, so that the distributional consequences would be egalitarian on standard
assumptions. But the Stolper–Samuelson theorem depends on competitive factor markets, and in particular competitive labour markets. Thus it becomes important to establish whether labour markets in developing countries are indeed competitive.

Recent years have seen an increase in research on questions of employer power in developed country labour markets, and in economics in general (see, for example, the review by Manning 2021). In the wake of this, there has been a reawakening of research interest in the topic in developing countries, where policy issues such as the minimum wage have been and are being hotly debated (see, for example, Bhorat et al. 2016, 2021). Section 2 of this paper provides a review of the growing but still limited empirical literature on employer power in developing countries. Section 3 builds on this evidence to present some theoretical perspectives on labour markets in developing countries. Section 4 considers the policy implications and Section 5 concludes.

2 Empirical evidence on employer power in developing countries

The basic theory as developed by Robinson (1932) says that the wage markdown in monopsony is inversely related to the elasticity of labour supply. The empirical studies investigating monopsony power follow two broad methods. First, studies directly measure the labour supply elasticity. The lower (higher) the elasticity, the higher (lower) the monopsony power. The main challenge in estimating elasticity is in overcoming the problem of wages being endogenously determined through both supply and demand changes. An exogenous variation in wages is required to elicit supply elasticities, separately from demand elasticities. Second, without necessarily estimating the supply elasticity, studies indirectly illustrate the presence of market power—for example, using rent pass-through to wages, measuring differences between wages and marginal revenue product, and understanding changes in wages as a result of mergers and acquisitions. Below, we present an overview of both types of studies, with a focus on developing markets, but also covering key insights from developed countries.

2.1 Direct measures

In developed country settings, monopsony has been traditionally measured by exploiting changes in wage and employment regulation. For example, studies exploit exogenous wage changes among school teachers (Falch 2010; Ransom and Sims 2010) or nurses (Staiger et al. 2010) and exploit minimum employment requirement restrictions for nurses to find elasticity estimates close to zero. Other studies have utilized randomized controlled trial designs in online labour markets (Dube et al. 2020) or instrumental variable regressions for faculty hiring (Goolsbee and Syverson 2019) to estimate low supply elasticities. The estimated elasticities in these studies range from 1.4 to 5. Studies show that supply elasticities for women are smaller than those for men, indicating that employers exert higher market power on female workers (Hirsch et al. 2010; Ransom and Oaxaca 2010).¹

For developing countries a variety of methodologies have been used to estimate labour supply elasticities. Amodio and De Roux (2021) exploit predetermined variation across plants in sales export destinations combined with exchange rate variation to generate plant-specific shocks in Colombian manufacturing. Due to a positive marginal revenue shock, while employment would increase under all types of labour markets, wages will increase under perfect competition but not under monopsony. Following this approach, they demonstrate the presence of monopsony and estimate a firm-level labour supply elasticity of around 2.5. Dal Bó et al. (2013) randomize wages through a public programme in Mexico called the Regional Development Program. Different salaries were announced randomly across recruitment sites, with the following job offers also subsequently randomized. Experimental estimates of labour supply

¹ See Naidu and Posner (2022) for an overview of the literature on monopsony in developed country labour markets.
elasticity were measured to be around 2.15. Brummund (2011) uses the production function approach to compute labour supply elasticities in Indonesian manufacturing plants and finds that over half of the manufacturing establishments have a significant amount of market power, with the median firm facing a labour supply elasticity of 0.52. Notably, the interval of elasticities estimated for developing countries are lower than those for developed countries.

2.2 Indirect measures

Studies have also tested the presence and extent of monopsony power using indirect means without necessarily measuring labour supply elasticity. In developed country markets, for example, studies show the deviation between wages and marginal revenue product (Isen 2013), find the effect of patent rents on additional earnings (Kline et al. 2019), and use fiscal stimulus shocks from American Recovery and Reinvestment Act (ARRA) contracts to examine wage changes (Cho 2019). Others explore the relationship between employer concentration (Herfindahl—Hirschman index, or HHI) and wages and show a negative relationship. Higher concentration is associated with lower wages (Azar et al. 2022), and the relationship between productivity and wage levels is weaker when labour markets are more concentrated (Abel et al. 2018; Benmelech et al. 2022). Further, if mergers increase local labour market concentration, earnings and wage growth fall for M&A workers (Arnold 2019; Prager and Schmitt 2021).

In developing country contexts, Naidu et al. (2016) exploited relaxation of visa restrictions of migrant worker movements across employers in the United Arab Emirates. Before the reform, workers needed to obtain permission from employers to switch to new contracts when their current contract expired. The reform allowed workers to transition to new employers without approval. The results indicate that real earnings and labour retention improved for incumbent workers after the reform, and that firms paid workers 51 per cent of the marginal product before the reform, but 72 per cent after. Brooks et al. (2021) find that building transportation infrastructure, specifically the Golden Quadrilateral highway in India, disrupted local monopsony power. Labour markdowns reduced among firms near newly constructed highways relative to firms that remain far from highways. Highway construction increased the labour share of income by 1.8–2.3 percentage points. Using firm-level data and variations in tariff reductions across industries in China, Kondo et al. (2022) find that input tariff liberalization is associated with lower labour markdowns across firms. The decline in labour markdown after trade liberalization is more pronounced for skill-intensive firms compared to non-skill-intensive firms.

2.3 Monopsony and the minimum wage

Textbook theories indicate that the introduction of a minimum wage or an increase in the minimum wage has a negative effect on employment. These results presuppose that labour markets are competitive. However, if the labour markets are monopsonistic, the minimum wage does not adversely affect employment, and in some conditions also improves employment outcomes (Stigler 1946). In the presence of imperfect enforcement and imperfect competition, Basu et al. (2010) predict that employment responses to introducing a minimum wage can be negative, positive, or muted, depending on the level of enforcement and the minimum wage.

Owing to these clear predictions and the popularity of the minimum wage laws, plenty of studies empirically investigate the effect of the minimum wage on employment. In developed countries there is emerging evidence in favour of monopsony (Card and Krueger 1994; Dube et al. 2010), but studies also continue to show negative employment effects, illustrating that labour markets are competitive (Burkhauser et al. 2000; Machin et al. 2003), especially among very low-skilled workers—in particular, for teens (Neumark et al. 2014).
In developing countries null employment effects were observed by Hohberg and Lay (2015) in the case of Indonesia and El-Hamidi and Terrell (2002) in the case of Costa Rica, across both formal and informal workers. No adverse employment effects were observed in South Africa, except in the case of agriculture (see Dinkelman and Ranchhod (2012) for domestic workers; Bhorat et al. (2013) for a variety of sectors). The effect of a minimum wage on construction workers in India suggests evidence in favour of monopsonistic labour markets (Soundararajan 2019). The study in India exploits variation in enforcement, and hence the probability of prevailing minimum wage being enforced. The results indicate that higher enforcement of minimum wage laws increases both market wages and employment. Further, null or positive employment effects are observed in districts with higher employer concentration.

Bhorat et al. (2017) provide an overview of the results in the minimum wage literature on developing countries. The study examines 98 papers reviewed by Neumark and Wascher (2007), along with 17 recent studies focused on lower-middle-income countries. They find that 55 per cent of employment elasticities are not significantly different from zero, and that the median elasticity is $-0.08$. This metas- tudy suggests that the minimum wage has either benign or slightly negative employment effects taken together across many settings. Broecke et al. (2017) review evidence on minimum wage effects in 14 major emerging economies (Argentina, Brazil, Chile, China, Colombia, India, Indonesia, Mexico, Poland, the Philippines, the Russian Federation, South Africa, Thailand, and Turkey), and find minimal impact on employment. Vulnerable groups, such as youth and the low-skilled, are more negatively affected.

3 Theory from a developing country perspective

Monopsony power and wage markdowns are inextricably interlinked via a simple elasticity formula in this fast-growing literature on employer power. The canonical model features a profit-maximizing producer faced with an upward-sloping labour supply schedule. Profit maximization prescribes a markdown on wages— a proportionate wage shortfall relative to the marginal value of the product of labour—to equal to the elasticity of the inverse labour supply schedule as the employer internalizes the wage consequences of hiring at the margin. The competitive case, at the limit, optimally sets the markdown at zero as the elasticity goes to zero. The intermediate case of Cournot labour market competition yields predictions that are consistent with developed country evidence as well—the higher the HHI of labour market concentration, the steeper the wage markdown (e.g. Azar et al. 2020, 2022).

Naidu and Posner (2022) and Manning (2021) expand on the definition of the wage elasticity of labour supply and observe that high labour supply elasticities and steep wage markdowns are associated with concentration and non-concentration drivers. Drawing an analogue from a product market power model driven by random utility (Gabaix et al. 2016), Naidu and Posner (2022) point out that wage markdowns can persist even in labour markets with many employers, such as in multinomial choice models with heterogeneous worker preferences (e.g. Card et al. 2018). Alternatively, search friction can also give rise to wage markdown in the presence of many competing firms, particularly when job offer arrival rates to job destruction rates are low (e.g. Burdett and Mortensen 1998; Manning 2003).

This elasticity-centric story of the wage markdown glosses over many developing country labour market characteristics, leaving important questions unanswered. For example, how appropriate are standard prescriptions of antitrust actions as a cure for steep wage markdowns in developing country labour markets? Beyond the perceived labour supply schedule, are there worker-level, institution-specific, and

2 A host of other policies studied in the developed country context continue to be open questions in developing country contexts (Naidu and Posner 2022). These include means-tested transfers (Posner and Weyl 2018), tacit agreements on no-poaching rules and non-compete contracts (Ashenfelter et al. 2021), and mergers and acquisition policies (Benmelech et al. 2018).
policy-related drivers of the wage markdown? In this section we present insights from three models that respectively showcase three sets of oft-noted features of developing country labour markets:

- Formal sector wages are often regulated via minimum wage laws, although full enforcement of the minimum wage is the exception rather than the rule (Basu et al. 2015).
- Even when a formal wage anchor (e.g. a minimum wage) is absent, workers may nonetheless harbour fair wage concerns and penalize employers by reducing effort when fair wage demands are not met (Basu et al. 2019).
- When wage contracts are not fully enforced by third parties to different extents, depending, for example, on the formality of the job, self-enforcement of open-ended employment relationships are subject to non-reneging constraints (Basu et al. 2015).

In what follows we briefly describe the three models. We then show the corresponding revision in wage markdown formulae in relation to the canonical prescription, discuss potential causes that drive compression or magnification in the markdown, and point out the role of labour market policies in the determination of the markdown.

3.1 Wage markdown, imperfect enforcement, and the minimum wage

We present a simple setup in which wage markdowns are outcomes of (1) monopsony power, (2) minimum wage, and (3) imperfect enforcement of the minimum wage. \textit{aR}(\ell)\text{ denotes a downward sloping and continuously differentiable revenue schedule of an monopsonistic employer hiring \ell workers, and \textit{aR}_f(\ell)\text{ is the associated marginal revenue product of \ell. a denotes a Hicks neutral parameter of monopsony productivity. Let \textit{w}(\ell)\text{ denote an upward-sloping and continuously differentiable inverse labour supply schedule with elasticity} d\log \textit{w}(\ell) / d\log \ell \equiv \varepsilon.}

Let a minimum wage policy be a wage and enforcement intensity combination \{\bar{w}, \lambda\} where \lambda \in [0, 1] denotes the probability of inspection. In response, the employer has three options: over-compliance, exact compliance, and non-compliance. An over-complying employer sets the wage at \textit{w} > \bar{w}. The total cost of employing \ell workers is just

\[ W(\ell) = \textit{w}(\ell)\ell, \quad \textit{w}(\ell) > \bar{w} \]

For an employer in exact compliance, the total cost is

\[ W(\ell) = \bar{w}\ell \]

Finally, for a non-complying employer, the total wage cost is:

\[
W(\ell) = (1 - \lambda)\textit{w}(\ell)\ell + \lambda(\bar{w} + \delta(\bar{w} - \textit{w}(\ell)))\ell, \quad \textit{w}(\ell) < \bar{w}
\]

where \delta is the transaction cost of transferring back wages amounting to \((\bar{w} - \textit{w}(\ell))\ell\) to workers when non-complying employers are discovered. Taken together, the marginal labour cost schedule is piece-wise continuously differentiable, and given by:

\[
W_\ell(\ell) = \begin{cases}
\textit{w}(\ell)(1 + \varepsilon) & \text{at } \ell > \ell \\
\textit{w}(\ell)(1 + \varepsilon) + \lambda(1 + \delta)(\bar{w} - \textit{w}(\ell))(1 + \varepsilon) & \text{at } \ell < \ell
\end{cases}
\]

where \ell = \{\ell|\textit{w}(\ell) = \bar{w}\} is the level of labour supply consistent with full payment of the minimum wage.

Figure 1 illustrates this upward-sloping marginal labour cost schedule in bold as \textit{W}_\ell(\ell), showing the standard marginal labour cost equalling \textit{w}(\ell)(1 + \varepsilon) for over-compliers. For non-compliers, the marginal
labour cost curve is a weighted average of the minimum wage \( \bar{w} \) and the original marginal labour cost \( w(\ell)(1+\varepsilon) \) in the absence of the minimum wage policy. Being a weighted average, this revised marginal labour cost pivots at \( \hat{\ell} \)—the level of employment at which the original marginal labour cost \( w(\ell)(1+\varepsilon) \) intersects the minimum wage. In particular, \( W(\ell) \) lies below the original marginal labour cost curve whenever \( \bar{w} \) is less than \( w(\ell)(1+\varepsilon) \) or when \( \ell > \hat{\ell} \), and above the original marginal labour cost curve whenever \( \bar{w} \) is greater than \( w(\ell)(1+\varepsilon) \), or when \( \ell < \hat{\ell} \). Finally, at \( \bar{\ell} \)—the level of employment consistent with the minimum wage—there is a sharp drop in the marginal labour cost schedule to reflect a sharp drop in the marginal labour cost in the transition from over-compliance to non-compliance, as

\[
\bar{w}(\bar{\ell})(1+\varepsilon) + \lambda(1+\delta)(\bar{w} - w(\bar{\ell})(1+\varepsilon)) < w(\bar{\ell})(1+\varepsilon)
\]

because \( \bar{\ell} > \hat{\ell} \).

Figure 1: The analytics of a minimum wage with monopsony power and imperfect enforcement

Consider therefore a profit-maximizing monopoly that sets employment \( \ell^* \) at a level that implicitly solves \( aR_\ell(\ell^*) = W(\ell^*) \). We make the following observations.

Changes in the wage markdown formula

Since the marginal labour cost schedules contains three distinctive segments, the markdown formula in fact changes depending on the productivity of the firm. For sufficiently high-productivity monopsonies (e.g. \( A_1 \)), the minimum wage is non-binding and as such the markdown is independent of the minimum wage policy. The corresponding wage markdown consistent with profit maximization is given by:

\[
\frac{aR_\ell(\ell) - w(\ell)}{w(\ell)} = \varepsilon
\]

for employers that operate in this range.

For the next range of productivities (e.g. \( A_2 \)), there is a clustering of employers who exactly comply with the minimum wage policy. These are employers that find it profit-maximizing to pay the minimum wage, hiring exactly \( \bar{\ell} \). The wage markdown consistent with profit maximization is given by:

\[
\frac{aR_\ell(\ell) - \bar{w}}{\bar{w}} = \frac{aR_\ell(\bar{\ell}) - \bar{w}}{\bar{w}}
\]
for the cluster of employer productivity types with exact compliance and $\ell = \bar{\ell}$ is labour supply consistent with the minimum wage. This markdown is higher the higher the firm-level productivity. Furthermore, the markdown is also strictly decreasing in the minimum wage $\bar{w}$, but completely independent of the enforcement intensity $\lambda$.

Now, for productivity levels that are lower still, the employer is a non-complier (e.g. $A_3$ or $A_4$). The equilibrium wage markdown can be written as

$$\frac{aR_e(\ell^*) - \bar{w}(\ell^*)}{\bar{w}(\ell^*)} = \varepsilon - \lambda (1 + \delta)(1 + \varepsilon)$$

Evidently, here the wage markdown once again depends on the elasticity of labour supply, but with an adjustment that depends on firm-level productivity, the minimum wage, the enforcement intensity $\lambda$, and the transaction cost associated with being fined for incurring back pay.

**Markdown compression or magnification**

A natural question at this point is whether an imperfectly enforced minimum wage can compress the wage markdown, in an effort to achieving a more equitable division of surplus between employer and workers. As it turns out, the answer to this question is nuanced, and highly non-uniform depending specifically on the productivity of the monopsony.

To start, note that the wage markdown at each level of employment $\ell$ in the end depends on the marginal labour cost, for by profit maximization:

$$\frac{aR_e(\ell) - \bar{w}}{\bar{w}} = \frac{W_e(\ell) - \bar{w}}{\bar{w}}$$

By inspection of Figure 1, which displays the marginal labour cost curve before and after the minimum wage policy, the minimum wage policy has no effect on the wage markdown, clearly, on over-compliers, as both employment and wages are independent of the policy for this group of employers. The minimum wage policy does indeed compress the wage markdown for intermediate productivity employers, however, who hire $\ell^* > \hat{\ell}$ but less than $\bar{\ell}$ for the minimum wage policy $W_e(\ell)$ around $\hat{\ell}$. Finally, the minimum wage policy has the opposite effect of magnifying the wage markdown among the lowest productivity producers, for who $\ell^* < \hat{\ell}$.

To sum up, the effect of the minimum wage policy on wage markdown runs the gamut from unchanging, to a compression, and then further to a magnification effect across firm types.

**Interplay between wage markdown and labour market policies**

The case of an imperfectly enforced minimum wage offers a particularly nuanced view on how labour market policies can impact the markdown. The lesson is simple—any policy combinations that give rise to a reduction in the marginal labour cost will compress the equilibrium wage markdown. However, what remains underappreciated in the literature is that observed wage markdowns should be understood case-by-case as the joint outcome of the interaction between labour market policy and firm-level characteristics. To give two examples based on our simple model, a gradual magnification of the wage markdown can be a symptom of rising monopsony wage-setting power, or a change in labour market policies consistent with an increase in the marginal labour cost at a constant labour supply schedule $w(\ell)$. Meanwhile, an increase in enforcement intensity may reduce markdown among some employers (higher productivity), but increase it among others (lower productivity).

---

3 Note that we have taken the $w(\ell^*)$ at equilibrium employment $\ell^*$ as the baseline wage in the markdown formula to capture wage markdown among workers receiving sub-minimum wages.
3.2 Wage markdown and fair wage considerations

When enforcement of wage contracts cannot be assured, what options do workers have to guarantee a living wage? In a seminal paper, Akerlof and Yellen (1990) posit the fair wage–effort hypothesis, in which worker reprisal in the form of a reduction in work effort takes place in the event wages are below a level that workers deem fair—the fair wage. There is emerging empirical evidence supporting this—see Breza et al. (2018) for India, and Verhoogen et al. (2014) and Dube et al. (2019) for the United States. The proportional reduction in work effort is taken to be equal to the wage shortfall relative to the fair wage in such a way that the employer is naturally incentivized to pay the fair wage to reap the full effort from the worker. Worker reprisal implies an upward-sloping labour supply at the worker level, where the effort of each worker $e$ is a function of the wage received $w$ relative to the fair wage $w^f$:

$$e(w) = \min\left\{ \frac{w}{w^f}, 1 \right\}$$

As shown, effort is bounded above at the maximal effort level per worker at unity.

Given this incentive structure associated with the fair wage–effort relationship where employers are incentivized to pay exactly $w^f$, the associated wage markdown naturally departs from the canonical formula, depending instead only on the determinants of the fair wage. In this regard, Akerlof and Yellen (1990) furthermore posit a fair wage to be given by the weighted average of the marginal value product of a worker at full effort, $p$, and the wage the worker can expect if she opts out of the fair wage contract, $w_o$, respectively:

$$w^f = \beta p + (1 - \beta)w_o$$

(2)

where $\beta \in [0, 1]$ is a fairness preference parameter indicating workers’ desire for pay commensurate with productivity $p$. The wage markdown is thus simply:

$$\frac{p - w^f}{w^f} = (1 - \beta) \left( \beta + \frac{w_o}{p - w_o} \right)^{-1}$$

Naturally, the wage markdown is strictly decreasing with respect to the fairness preference parameter $\beta$. Furthermore, the wage markdown reflects the gap between the marginal value product of labour $p$ and the opt-out wage $w_o$.

To that end, Basu et al. (2019) introduce more structure to the model to consider the profit-maximization problem of the employer. Specifically, they assume simply that the relationship between labour input and output is 1:1. The employer in question selects a price $p$, equivalently the marginal value product of labour, to maximize profits subject to a weakly downward sloping demand curve $x(p)$ with elasticity of demand $-\frac{d \log x(p)}{d \log p} = \rho$:

$$\pi(w, p) = \max_{p, w} (p - w/e(w))x(p)$$

(3)

subject to the fair wage–effort relationship:

$$e(w) = \min\left\{ \frac{w}{w^f}, 1 \right\}$$

4 Wage disparity is important in both developed and developing country labour markets. See Duman (2019) for a review of the literature and estimates from various studies. This disparity presents a departure from the canonical labour market model of monopsony power on a homogeneous workforce.

5 There are alternative specifications, including specifications that put weight on the wages of others in the same establishment in the definition of fairness, for example.
Changes in the wage markdown formula

The solution to eq. (3) reiterates the familiar output pricing solution to the firm’s maximization problem as

\[ p^* = \frac{w_o}{\rho} > w_o \]

which implies that the fair wage in equilibrium is ultimately determined by the opt-out wage \( w_o \), the preference parameter \( \beta \), in addition to the product market power of the employer parameterized by the demand elasticity \( \rho \). Substituting this solution to \( p^* \) into the fair wage markdown, we have:

\[ w^* = \beta p^* + (1 - \beta)w_o = w_o(1 + \beta(1 - \rho)/\rho) \]

The corresponding wage markdown can simply be expressed as a function of two parameters of the model, \( \beta \) and \( \rho \):

\[ \frac{p^* - w^*}{w^*} = (1 - \beta) \left( \beta + \frac{\rho}{1 - \rho} \right)^{-1} \]

**Markdown compression or magnification**

Contrary to the canonical prescription that the wage markdown is a function only of labour market power, the current setup presents an alternative view when wage contracts are enforced by threat of worker reprisal. The drivers of the corresponding wage markdown are the fairness preference parameter \( \beta \) and the monopoly power of the employer as given by \( \rho \). Naturally, markdown compression occurs when the fairness preference parameter swings in favour of giving a higher share of the workers’ marginal product as wages. Meanwhile, the more potent the monopoly power (lower \( \rho \)), the more magnified the wage markdown will be.

**Interplay between wage markdown and labour market institutions**

An important takeaway is that in the absence of third-party enforcement of wage contracts, the wage markdown is held in check by workers’ fairness preference, conditional on employers’ ability to manipulate the value of workers’ efforts. This form of tacit agreement is not universal, and indeed there are multiple ways to further extend this path of reasoning. Basu et al. (2019) make the argument that employers are incentivized to pay more if they are sure that the workers are direct beneficiaries of the higher wage. This direct employment relationship can be severed when workers are hired indirectly through subcontractors. Basu et al. (2019) ascertain the determinant of the wage markdown for subcontracted workers, who are paid by contractors, and not employers. It is shown that the payment of the fair wage to subcontracted workers is no longer guaranteed. This gives rise to effort and wage cost implications that employers must then take into account when making subcontracting decisions. What is important to note here is that labour market institutions enter into the markdown debate through a composition effect—in this case the share of regular versus contract workers—that can potentially shed new light in the research on the drivers of the economy-wide labour share.

An alternative extension can involve endogenizing the opt-out wage \( w_o \) as a function of the labour demand of the employer. This of course takes us back to the canonical territory of wage markdown in the presence of monopsony power. The precise ways in which the elasticity of labour supply through \( w_o \) interacts with the fairness preference in the determination of a wage markdown is an interesting question for future research.

### 3.3 Wage markdown and the self-enforcement constraint

Beyond the threat of worker reprisal, credibility in labour contracting can also be achieved by ensuring that any wage promise made will be kept via a self-enforcing constraint. Basu et al. (2015) examine the
nature of self-enforcing wage contracts in the formal and the informal sectors, where minimum wage enforcement differs in intensity. A no-reneging constraint in this context gives the maximal wage that workers can demand such that employers are incentivized to pay as promised in order to maintain a long-term employment relationship with the same worker. The alternative is to terminate employment after reneging on the contract. The extent to which employers can expect to get away with the practice is taken to depend on the extent to which minimum wages are enforced. Assume henceforth that workers can find employment either in the formal ($f$) or the informal ($i$) sector where minimum wage enforcement intensities are $q_f$ and $q_i$, respectively, where $1 \geq q_f > q_i \geq 0$.

A self-enforcing wage contract in this context is the maximal wage satisfying a no-reneging constraint. Let $V_k(w^*, a)$ denote the value function of an employer who hires a worker with marginal value product $a$, and who promises to pay $w$ every period in sector $k = f, i$ in an open-ended employment relationship. Contracts are terminated at an exogenous rate $\delta$ every period. Thus, in a steady state at discount factor $\beta \in (0, 1)$

$$V_k(w^*, a) = a - w - b(1 - \delta)V_k(w^*, a) = \frac{a - w^*}{1 - b(1 - \delta)}, k = i, f$$

Alternatively, the expected profits of a reneging employer in sector $k$ is:

$$V^o_k(w, a) = a - q_k\bar{w}$$

The non-reneging constraint in $k$ requires that paying the promised wage $w^*$ is the better option:

$$V_k(w^*, a) = \frac{a - w^*}{1 - b(1 - \delta)} \geq a - q_k\bar{w}$$

Equivalently, the highest possible wage consistent with the no-reneging constraints implies strictly positive wage markdowns:

$$w^N_k(a, \bar{w}) = \max\{w | V_i(w, a) \geq a - q_k\bar{w}\} = b(1 - \delta)a + (1 - b(1 - \delta))q_k\bar{w} \quad (4)$$

Interestingly, the application of a minimum wage raises $w^N_k$ disproportionately more than $w^N_i$, since the lowest expected wage that a reneging employer will be enforced to pay by law is $q_f\bar{w}$ in the formal sector, and $q_i\bar{w}$ otherwise. The key observation here is thus that the potential wage savings from reneging, all else being equal, is strictly less severe in the formal sector thanks to government enforcement of the minimum wage for the same promised wage.

Changes in the wage markdown formulae

Using eq. (4), the wage markdown formulae are:

$$\frac{a - w^N_f}{w^N_f} = (1 - b(1 - \delta)) (b(1 - \delta) + q_f\bar{w} / (a - q_f\bar{w})^{-1} \quad (5)$$

$$\frac{a - w^N_i}{w^N_i} = (1 - b(1 - \delta)) (b(1 - \delta) + q_i\bar{w} / (a - q_i\bar{w})^{-1} \quad (6)$$

Thus, in the context of a self-enforcing constraint, the wage markdown offers the employer just enough surplus to reject reneging on the wage contract. These prescriptions bear close resemblance with that of the fair wage setting in our earlier discussion. Here, the contract termination risk-adjusted discount rate, $b(1 - \delta)$, takes the place of the fairness preference parameter in the fair wage setting. $q_k\bar{w}$ is of course just the expected level of third-party enforced wage payment, which now takes the place of the opt-out wage in the fair wage setting.
Markdown compression or magnification

The rationale behind a wage markdown in the current context is to set wages low enough so that employers choose voluntarily to maintain an open-ended employment relationship. As long as formal sector employers are faced with a better-enforced minimum wage, their wage savings from reneging is restricted relative to an informal employer promising the same wage. One would thus expect that the self-enforcing wage markdown in the formal sector is a compressed version of the wage markdown in the informal sector. And this is exactly what we find in eqs. (5) and (6), all else being equal.

The wage markdowns are driven by firm- and worker-level considerations as well. Specifically, uncertainty in operating a formal establishment, as parameterized by $\delta$, has the effect of magnifying the wage markdown by making open-ended employment relationships less attractive, while an increase in the productivity of the worker $a$ increases the wage markdown by raising the profitability of a fly-by-night employer. These considerations once again depart from the canonical elasticity-based markdown formula, bringing together policy-, firm-, and worker-level considerations.

Interplay between wage markdown and labour market institutions

Informal employment is a ubiquitous feature of developing country labour markets. Yet, our understanding of the determinants of wage markdown here is notably scant. The current setting is a first step in this direction, although useful extensions abound. Suppose that informal employment feeds into formal production by acting as an intermediate input supplier (e.g. Chen 2007); what is the nature of the interlinkage between wage markdowns in the two sectors when employers can cross-substitute between labour input providers? In addition, a worker self-selects into formal and informal sector employment (Basu et al. 2015) since their productivities naturally feed into wages in the two sectors as in eqs. (5) and (6). Worker productivity thus gives rise to another systematic driver of observed wage markdowns in the formal and the informal sectors. Specifically, the informal wage markdown falls with minimum wage enforcement, but the same markdown increases if relatively high-productivity workers simultaneously self-select to work in the informal sector because better enforcement of formal wages gives rise to formal job rationing. These two drivers have opposite effects on the informal sector wage markdown. Which one of these tendencies dominates in aggregate wage markdown observations? These and other related questions related to composition effects are promising areas for future research.

4 Conclusion: policy and analytical implications for developing countries

Conceptual and analytical perspectives matter for policy. A world view that is based on the assumption of competition everywhere will be inclined against intervention and regulation. When empirical evidence stacks up against such a position, it opens up questions on appropriate public policy to enhance social welfare. It also opens up analytical perspectives which are founded on features of the real world, including constraints on policy implementation.

We have already alluded to the profound influence of the Harris–Todaro perspective where the departure from competitive labour markets is assumed to occur in the formal sector through the power of labour unions—in other words, employee power. This reduces employment in the formal sector and displaces labour to the informal labour market, which is assumed to be competitive with no market power on either side. The policy conclusion follows straightforwardly. Since the ‘distortion’ away from perfect competition lies in the power of labour unions, this power needs to be dismantled—a policy which has indeed been followed in recent years. A similar trajectory from a competitive world view, to what is considered then to be an inappropriately instituted minimum wage in the formal labour market, to
reduction of employment and displacement of labour towards the informal sector, is to be found in critiques of minimum wage policies which are widespread in developing countries.

But what if labour markets are not competitive and, far from labour unions having power, it is employers who have such power? As reviewed in this paper, the evidence is certainly accumulating that this is the case. The case against regulation, and the case for intervention, then needs to be rethought. Specifically for the minimum wage, employer power justifies its use as a device to increase efficiency and equity. The argument now turns on context and detail of labour market structure, and on the level and structure of minimum wages. Nothing said so far precludes a critique of actual minimum wages as set—–they may be too high or not well structured—–but a blanket negation of this policy instrument is no longer warranted (Bhorat et al. 2021). Detailed economic analysis is indeed needed, and necessary.

Let us then accept that the need for state intervention cannot be blocked out a priori by the assumption that labour markets are competitive. But this still leaves open the question of implementation of interventions when state enforcement capacity is weak, as it is generally recognized to be in developing countries. The intensity of enforcement is a key factor in determining the efficacy of a minimum wage in delivering its efficiency and equity gains (Basu et al. 2010). At one extreme, zero enforcement means in effect there is no intervention. But intermediate levels of enforcement interact with different levels of the minimum wage in intricate ways. Apart from the empirical difficulties in making inferences from the variation in employment as minimum wage varies, when enforcement can vary as well, this raises the conceptual question of how is the enforcement intensity chosen and why? And how does the ability or inability to commit to enforcement (rather than to ‘turn a blind eye’) play into the choice of the minimum wage? Such issues are highlighted in the previous section.

Thus, while the competitive labour markets presumption is too simple and empirically invalid, so is the presumption of easily implemented policy correctives to employer power in labour markets. The difficulty of enforcing minimum wages is but one example. More generally, there is no easy transference of an ‘antitrust’ frame, very well developed for product markets in developed countries, to labour markets in developing countries. Defining the boundaries of a labour market with prevalence of informality, and then identifying measures and limits to concentration in the way it is done in conventional antitrust regulation for product markets in developed countries, will be extremely difficult in developing countries.

If directly addressing employer power in developing country labour markets can only be a partial answer, is there an alternative or a supplement? We get a clue from the basic wage markdown formula in the elasticity-centric framework. The lower the elasticity, the greater the markdown. In simple terms, the more workers are bound to supply labour no matter what the wage, the greater the power employers have over them. Thus raising the elasticity lowers the markdown. The attention then shifts analytically to the factors underlying the elasticity of labour supply, and from a policy perspective to interventions that can raise this elasticity. What might such interventions look like?

In the empirical section we have already alluded to the role of transport infrastructure. As developed by Basu et al. (2010), for example, transportation costs are a key determinant of the responsiveness of labour supply to changes in wages. And the empirical evidence reviewed in Section 2 suggests that development of transport infrastructure can lower the degree of employer power. But transportation is just a start. Other economic, social, cultural, and institutional barriers to labour force participation can have the same effect, and prompt us to investigate these features of developing countries more closely as underpinnings of employer power. The specific models of Section 3 are a start on such an exercise.

In sum, employer power is a feature of labour markets in developing countries. Taking this on board changes our policy perspective significantly on labour market interventions, and it opens up interesting and important areas of empirical and theoretical research.
References


