Prominent features of most developing countries are serious “imperfections” in many markets, including the labour, capital and various product markets; a high level of informality (lack of connection with the state, e.g. non-registration, non-compliance with tax and labour regulations, etc.); and attempts to improve the welfare of specified groups through social security (SS) systems in the formal sector and (sometimes) social protection systems (SP) that cover informal workers. Three main policy issues arise:

i) what can be done to limit such negative effects as the imperfections may have?

ii) what is the best design for SS and SP systems?

iii) should formalization be pursued as a policy?

Many variables affect the impacts of both the SS and SP systems and that of informality, including labour demand elasticities, the effectiveness and \textit{modus operandi} of the social security (SS) and social protection (SP) packages, and the presence of the market imperfections cited.

To better understand the way these features of a system interact to determine the outcomes, it is convenient to begin with the simplest possible case, and then approach reality step by step. Although it provides some useful insights, we conclude that the formal-informal sector dichotomy is not the appropriate conceptual framework within which to analyze these issues and needs to be supplemented and/or replaced by a three-sector model. It is clear that no general theoretical conclusion can be reached about the impact of SS and/or SP systems on employment or economic efficiency. The analysis identifies a number of the variables whose empirical values can be predicted to determine those effects.

\textbf{Case 1: Formal and informal sectors; pure and perfect competition in all product and factor markets; no SS or SP systems; workers are identical in all respects$^1$; no taxes except on consumption}

In this case, though the distinction may not yet carry much if any policy meaning, it is convenient to differentiate the (typically small) firms that will be conveniently labelled “informal” after taxes and other factors that can contribute to market imperfections are introduced. Workers move freely between the two “sectors” so worker decisions determine the amount of employment in each. As long as there is no non-economic

\footnote{1 The assumption of homogenous labour is not necessary to the analysis. Each type of labour which is divided between a formal and an informal sector can be handled separately in the same way as illustrated here.}
preference on their part for one or the other sector, the wage must be the same in both for there to be a stable equilibrium and no further movement of workers. The demand curve for labour by the formal sector \((D_f)\) is drawn from the left side axis of Figure 1 while that for the informal sector \((D_i)\) is drawn right to left from the rightward axis. Under the assumed condition of perfect competition among firms in each sector, the labour demand curves are also the value of the marginal product of labour \((VMPL)\) curves, \(MPL_f\) and \(MPL_i\) respectively. The distance between the two axes is the total supply of labour, assumed for simplicity to be independent of the wage rate and hence constant for purposes of this analysis. Where the two demand curves meet (point E) determines the equilibrium wage \(W_e\) and the allocation of labour between the two sectors \((0L_f^*\) to the formal sector and \(L_i^*0'\) to the informal sector). Total output in the economy, assuming nothing can be produced without labour, is the area under the two VMPL curves, \(0AEB0'\). The wage bill is the rectangle \(0W_eW_e0'\) and the rest of the area under the VMPL curves (area M plus area N) goes to other factors of production.

**Case 2a. Formal and informal sectors; pure and perfect competition in all product and factor markets; no SS or SP systems; workers are identical in all respects; a tax on formal sector workers**

The simplest case in which to illustrate the deadweight loss resulting from labour misallocation between two sectors of an economy is one in which the only imperfection is a simple (say a fixed percentage) tax on formal sector labour (Figure 2a). To illustrate the equilibrium process and its implications requires a distinction between labour costs or wages gross of taxes (here the VMPL\(_f\) curve), and workers’ benefits (the wages received) based on the labour demand curve facing the workers, \(D_f\) which by assumption lies below the labour cost curve VMPL\(_f\) by a fixed percentage. The new labour market equilibrium occurs at point C, since the formal sector labour demand curve \((D_f)\) confronting the workers is now below the VMPL\(_f\) curve by the ratio of the tax relative to gross wages. The triangle CGE gives the deadweight loss associated with labour misallocation between the two sectors \((0L_f^*\) to the formal sector and \(L_i^*0'\) to the informal sector). The tax has increased informal sector employment by \(L_i^*L_i\) workers. It has also lowered the equilibrium wage from \(W_e\) to \(W_c\), so that workers total earnings are lowered by the tax (to the tune of \(W_eW_e0'W_c0'\)). Formal sector capital (and other non-labour factors involved) loses by HGE\(_W_e\) while government collects HGC\(_W_c\) in taxes. Informal sector capital

2 Note that the portrayal of a two-sector economy in the simple way used here is most apt in a situation where only one good exists and the two sectors compete in its production. A multi-good economy is more complicated, among other things because the relative price of goods varies according to the relative output of each and this shifts the curves describing the value of the marginal product of labour.

3 The value of the marginal product of labour is the marginal product of labour (number of units added to total production by employing one more worker) times the price of the good or service being produced.

4 At this point it is necessary to provide some sort of definition of the formal sector to explain why the labour tax will only be collected there. The simplest is to assume that for some reason or reasons the tax can be collected from some workers but not from others, plausibly because it is easier to tax larger, more visible firms.
(together with any other non-labour factors) gains, so the overall impact on the earnings of those factors is theoretically indeterminate; it depends on the elasticities of the two VMPL curves. In any case the net (or post tax and expenditure) impacts on the welfare of these factors of production and of labour depend on how the government spends the taxes collected. More generally, the other (than those portrayed in Figure 2a) impacts of the tax depend on the relationship between the social benefits arising from the associated public spending on the one side and on the costs of the tax, taking the form of (a) direct loss of welfare to those who pay the taxes, (b) administrative costs of collecting the tax and spending the money and (c) any disincentive effects on economic activity arising from the imposition of the taxes. Category (c) could be thought of as including the labour misallocation effect, but since it is central to the present discussion we treat it explicitly. In a well-managed public sector, taxes will only be imposed up to the level where the associated marginal benefits outweigh the sum of the marginal increases in these costs.

**Appropriate Policy Response**

The imperfection giving rise to the deadweight loss is the labour tax so the most direct way of ending it is to impose instead a sector-neutral tax. If this is not feasible it becomes relevant to ask whether there are indirect ways to reduce the loss from this imperfection. For example, can the encouragement of formalization contribute to reducing the loss? In this initial static framework, the answer is likely to be no, but this could, nonetheless, depend on how formalization was encouraged. If a defining feature of informality is costliness of imposition of taxes, it is conceivable that the tax system could be made more efficient, thereby bringing more workers under the formal umbrella. It is possible that some resources collected through non-distortionary taxes could be used to provide an incentive to raise the relative attractiveness of formality to workers. If such non-distortionary taxes were available, they should have been used in the first place rather than the distorting wage tax. And where informality is correlated with poverty, there is in any case a justice argument against the taxation of those working in the sector. Note that this case is of interest mainly for heuristic purposes, since most taxes on labour income are designed with *quid pro quo* benefits to the workers, the case to which we now turn.

**Case 2b. A SS contribution with associated worker benefits is imposed on formal sector workers; workers are identical in all respects; the government does not subsidize the system**

When instead of a pure tax (collected and used to provide benefits that are not specifically directed to those who pay the tax) there is a SS contribution by the formal

---

5 Assuming mobility of capital between the two sectors, there will now be a shift from the formal to the informal sector, which will raise the VMPLc curve and lower the VMPLf curve, leading to a further shift of employment to the informal sector. This shift will further increase the deadweight loss from the tax, since at the same time that it brings the returns to capital into equality between the two sectors, it also pushes the economy even farther from the initial efficient allocation of both labour and capital between the two sectors. Thus the qualitative direction of the effects of the distortion are not reversed, and a first analysis, without taking account of the second round shift in capital between the two sectors, gives a reasonable feel for the impact of the labour tax. In practice, since the capital market functions less well than the labour one, it is an open question how much capital reallocation would occur between the two sectors in response to the imposition of a labour tax on the formal sector.
sector worker which basically buys him/her a set of services,\textsuperscript{6} there may be no labour misallocation (or any other loss) involved. This depends, among other things, on how much the worker values the services in question. In the simplest case, the worker considers the services provided from the SS contributions to be just equal to the value of the contributions (i.e. the welfare lost through having to pay them).\textsuperscript{7} The labour costs from the firm’s point of view are not all paid out as current wages to the workers but their current wage equivalent (CWE) value to the workers is the same as if they were paid out currently. The worker is by definition indifferent between a given wage in the informal sector (where wages constitute total labour costs to the firm) and the same CWE in the formal sector. The allocation of labour between the two sectors is the same as if neither the SS system nor a pure tax on formal sector labour existed. The SS system has no net positive or negative effect on the workers for whom it is designed and no effect on labour allocation between the two sectors (Figure 2b). The wage rate in the informal sector remains the same; that in the formal sector falls by the amount of the tax, but the CWE of the wage plus SS package is equal to the informal sector wage (and to the previous formal sector wage). Neither informal workers nor other factors of production are affected in terms of earnings; formal sector workers earn less nominally but retain the same welfare level as before. Market forces have pushed the nominal wage rate in the formal sector down to \( W_b \), as workers must be paid less in the formal sector to prevent more of them from shifting into it given the SS benefits.

\textit{Where the value of the benefits package differs from that of the foregone wages, with no one else affected positively or negatively.\textsuperscript{8}}

\textsuperscript{6} SS systems typically involve a payment by the worker and one by the employer, and sometimes a contribution by the state, either directly or in the form of the non-taxation of worker and employer income directed to the SS system. In the absence of other imperfections like the minimum wage, however, the employer contribution will be a nominal one only, since the wage rate will be bid down by an identical amount. Thus it is legitimate, at this level of abstraction, to simply assume that the worker is the only one who pays for the SS services. This assumption is modified later.\textsuperscript{7} This condition could occur under the combination of perfect and costless private insurance markets and perfect efficiency of the SS system. Since both of these assumptions would be wildly unrealistic, the equality would normally result from the same degree of falling short of the ideal in both the private management of the risks in question and their social management within the SS system. In any case, the bottom line is that the overall benefits from the formal sector labour contract are just equal to the value to the worker of the full labour costs paid to him/her as current wages.\textsuperscript{8} Note that the appropriate measurement of GDP becomes less than clear when some goods and services provided in non-market settings have a value to their users different from the cost of production. Traditionally, publicly provided goods and services are valued at cost in national accounting, whereas the value of privately produced goods and services is measured by their mark prices. For simplicity we adopt here the usual convention that GDP includes a value of publicly provided services equal to their cost. When that convention is followed, one must recognize that when the value of goods and services produced by the public sector or under SS-type arrangements is not equal to their actual value to the recipients, then GDP is a less valid measure of societal economic welfare than it would otherwise be.
The SS system is designed to be of benefit to the workers, and though of course in practice it may turn out not to be, it is at least possible that it may provide net benefits to them, as where it allows them to buy insurance against risks which cannot be as well handled in the private market (plausible), or facilitates savings which they would otherwise find it hard to do, perhaps for institutional reasons well described in the literature on the savings problems of lower income families. Defining a worker’s preference ratio (P) as the ratio of his/her valuation of the SS package benefits to the cost of same, P=1 describes the person who values those benefits just at their cost to him/her, P>1 refers to someone to whom the package provides net benefits above cost and P<1 someone for whom the costs exceed the benefits. Where the SS does provide net private benefits, such that the worker prefers receiving the services rather than the foregone labour earnings that pay for them (assuming, as noted above, that the system is not subsidized by the government and that, although nominally the firm pays part of the costs, this component is in fact paid by the workers in the form of lower wages), then worker preference leads to a larger formal sector than would be the case in the absence of the SS system. To illustrate the equilibrium process and its implications requires a distinction between marginal formal sector labour costs (given by the VMPLf curve) and the marginal workers’ perceived benefits, shown in Figure 2c as the CWEf curve.

Workers are now indifferent between employment in the two sectors when the informal sector wage (Wi) is equal to the formal sector CWEf (point F) so the equilibrium allocation of labour involves more of them than before in the formal sector (0Lf→). Although total output (GDP) in the economy has fallen (by area FEJ) and the current wage bill may have fallen (the wage rate falls in the formal sector and rises in the informal one), total societal welfare has risen by area MFJR minus area FEJ. Since the distance FJ represents the net gap between total employment-related benefits accruing to each formal sector worker from the SS package of services and the wages foregone to get access to that package when Lf = Lf→, area MFJR would be the net benefits to the economy if there were no loss of output brought on by the shift of workers to the formal sector; FEJ (that loss) must be subtracted to give the overall net impact on society.9 Labour is optimally allocated from the perspective of maximization of worker welfare as well as societal welfare, even though GDP falls. The income of capital falls by FNWcE in the informal sector and rises by WeEJR in the formal sector; the total rises as long as the slope of VMPLf in the relevant range EJ is greater than that of VMPLi in the relevant range EF.

Worker welfare, labour costs and non-labour incomes all depend on the elasticities of the

9 Another way to see who is affected in what way is to note that, overall, workers are better off by rectangle MNWcWc; each worker’s welfare is raised by amount Wc, regardless of sector, though the increase takes the form of current wages in the informal sector and SS benefits in the formal one. All of this area except KFE was originally part of capital income in one or the other sector but is no longer, but area WcEJR has been added to the capital income of the formal sector. For society as a whole, this outcome thus boils down to the gain of MFJR minus FJE, as mentioned in the text.
VMPL_f and VMPL_i curves in the relevant range EF. Worker welfare is greater the lower is the elasticity of VMPL_i and the higher is that of VMPL_f. At the limit, each worker (in both sectors) gains by the benefits gap FJ when VMPL_i is perfectly inelastic; at the other limit, they gain nothing when that curve is perfectly elastic.

Per worker labour costs now diverge between the two sectors by the size of the net benefits gap created, FJ. Whether total labour costs rise or not depends on the relative elasticities of the two MPL curves. If the elasticity is higher in the informal sector, the gap needed between the two VMPLs to bring equilibrium will come disproportionately from an increase in the informal sector labour cost (equal to the wage) rather than a decrease in the formal sector, so total wages paid out will increase. If, as often assumed, the elasticity is lower in the formal sector, then there will be a net fall in labour costs due to the reallocation. This creates a likelihood that other factors will benefit, which will be the case as long as labour costs fall more than does GDP. In the extreme case of perfectly elastic demand for labour in the informal sector, labour costs in the formal sector will fall by the amount of the net benefits gap, the wage will remain constant in the informal sector and will now be equal to the CWE in the formal sector, so all of the benefits of the SS system will accrue to other factors of production than labour, and worker welfare will be the same as before. This situation corresponds to one of the prominent variants in the labour surplus economy literature.

Finally, the amount of labour reallocation is positively related to the elasticity of both MPL curves.

In summary, the provision of a good SS package cannot make workers worse off under the assumptions used here, but will make them better off as long as the demand for labour in the informal sector is less than perfectly elastic and, that condition met, it will benefit them more the lower the relative labour demand elasticity in the formal sector vis a vis that in the informal sector. This case also introduces a situation in which GDP is not the correct indicator of societal welfare, even under the standard assumptions of much economic analysis which do make it a relevant indicator.10

It is also possible, however, that the bundle of services not be well designed and thus be worth less to the typical worker than the CWE, a situation discussed in a significant literature which focuses on inefficiencies in the design and provision of SS benefits. In this case, the above analysis is essentially reversed in all of it details. The SS package leads to a shifting of workers to the informal sector (Figure 2d); the new equilibrium

10 Apart from perfect markets, no externalities and no public goods (all explicitly or implicitly assumed here), that standard set includes the assumption that the marginal utility of income is the same for all people. The interpretations of results from that standard analysis must therefore be thought of in a ceteris paribus sense, i.e. if there were no grounds for expecting that the states of the economy being compared would differ in any of these ways, then a higher measured GDP would on average imply higher societal welfare than a lower measured GDP.
point F implies a formal sector employment level of $0L_f$. There is a net loss to society as a whole, consisting of two distinct types. First, some of the resources that fund the SS are wasted (lost) because of its weak design or implementation. The area HGFJ measures this loss—the per person loss (GF) times the number of workers in the formal sector (HG or JF). Second, because this problem leads to a shift of workers into the informal sector, GDP declines because over the range of workers shifting into the informal sector, $VMPL_i$ is below $VMPL_f$. This loss is measured by the triangle GEF. In the case portrayed the great majority of the loss comes from the bad design of the SS system, rather than the misallocation of labour between the two sectors. This is likely to be the normal outcome, since it appears improbable that the elasticities of demand in the two sectors would be such as to reverse it. It is consistent with the tendency for deadweight losses due to resource misallocation of the sort discussed here to be quite small as a share of GDP, a conclusion reached early by Harberger (1954). In any case the size of this loss in the present context depends on the elasticities of the two labour demand curves, which also determine the extent of labour misallocation. (The two elasticities affect loss in different ways, reflecting the direction of the labour misallocation. The extent of labour reallocation due to an increase in formal sector labour costs (like that brought about by an inefficient SS or by a MW) is determined solely by the elasticity of labour demand in the formal sector. The elasticity of $DL_i$ can be thought of as helping to determine the component of GEF that falls below $W_i$ i.e. area ???, while the elasticity of formal sector labour demand both determines how much labour reallocation will occur and affects the loss per worker who shifts. Except for special cases, as where all formal sector employment has been blocked by the labour costs, deadweight loss is a monotonically increasing function of the elasticity of the informal sector labour demand curve but a monotonically decreasing function of that of the formal sector. The elasticity of labour demand is usually assumed to be greater in the informal sector; this tendency would thus tend to limit the deadweight loss. This elasticity is usually assumed to be low in the formal sector; a pattern that would also tend to limit the deadweight loss.

The likelihood that the inefficient SS loss will swamp the labour misallocation loss can be deduced by the fact that the distance GF, measuring that inefficiency gap, figures in both losses. The inefficient SS loss is that gap multiplied by all formal sector workers; the labour misallocation loss is that gap multiplied by the number of workers misallocated between the sectors, all divided by about two (exactly two if the two labour demand curves are linear). Thus if the labour misallocation were to involve 5% of the labour force (an implausibly high number) and formal sector employment were 20% of all employment (a relatively low figure) the ratio of inefficient SS loss to labour misallocation loss would be 8:1. Given that the SS loss per person is a determinant of the labour misallocation loss, the two types of loss have a co-determinant, which makes it less likely that the labour misallocation loss could be very large in comparison with the inefficient SS loss.

How the deadweight loss due to labour misallocation is divided between labour and other factors depends on the same considerations mentioned above. In the extreme case of a perfectly elastic $VMPL_i$, curve, worker welfare will remain constant; that of other factors
must therefore fall. Except for that case, workers will lose, with their loss an increasing function of the elasticity of each of the two labour demand curves. Once again, one factor (here non-labour) can gain even though society as a whole has lost.

Where the bundle is not optimally constructed, its weaknesses may transform what could have been net benefits into net losses, implying that reforms to the structure of the SS system could be very beneficial.

*Alternative interpretations in cases where workers’ valuation of SS differs from its cost, especially when it is below the amount contributed: people other than the worker who pays for the SS may be affected.*

The SS service bundle costs the worker a certain amount, and when he/she values it at less, several different factors may be at work. The simplest case is that in which all of the resources collected to provide the bundle are in fact used up to provide that bundle to the worker (groups of workers), the worker completely “uses up” these services made available but those services are simply not worth as much to him/her as what he/she paid in. Another possibility is that some of the available services are, for whatever reason, not accessed by the worker and either those services or the funds available to provide them (e.g. funds to provide pensions) can then be redirected to someone else. In this case, the worker’s evaluation of the services, while accurate for himself/herself does not correspond to their societal value. Once again, the worker’s perception determines the allocation of labour, but societal loss is now less than before. At the limit there might be no loss through bad service design (in the specific sense that benefits are not fully channelled to those who paid for them), or even a gain, but there would be some output foregone because of the labour reallocation. Once again, GDP would not be the appropriate criterion of overall performance.

A distinction is thus needed according to whether components of the bundle that are not accessed by the workers in question, whether through their own preference or for other reasons) then go to waste (deadweight loss) or somehow become available to someone else. The other person may be another worker paying for SS or someone who does not. In the former case, for example, where available child care is not drawn on by some of those who are entitled to it, it might then become available in greater quantity to someone else (benefits thus being greater than expected), or the financial cost of SS contributions to others could be lowered. In such cases, a P< 1 for one worker might simply imply P>1 for another if the extra availability for the second worker were predictable by him/her, a situation to which we turn below. Alternatively, these benefits could accrue to others not covered by SS at all. In either case a low worker regard for it will be a natural result of there being services he/she does not value, but taking into account worker preferences does not give us the whole story, since those preferences do not distinguish cases where the services are not of value to the worker and those where they are not of value to anyone, or the resources for them are somehow wasted.

---

11 In the extreme, they may be stolen by operatives of the SS system or by others, an unfortunate form of redirection, but still creating value for someone.
The fact that any difference between the wages foregone to pay for SS services and the value of those services to the worker in question may or may not constitute deadweight loss to the society means that four values need to be borne in mind in thinking about the implications of adding one more worker to the formal sector:

i) the VMPL of that worker,
ii) the wages paid to the worker
iii) the current wage equivalent (CWE) of the worker’s wages together with the available SS package, and
iv) the societal value of the worker’s wages together with the SS package, equal to the sum of (iii) plus any net positive effects on anyone besides the worker, as just discussed. We refer to the curve that measures this sum as the MSBf curve.

As long as no SS-related benefits accrue to others than the workers, total societal welfare will rise or fall according to whether worker evaluation of benefits accruing to himself/herself (assumed to be accurate) is above or below the resources extracted from them. In either case, conventionally measured GDP will fall since worker allocation is moved away from the one that maximizes GDP but, as noted, this loss will be outweighed by direct welfare gains from SS in the case where workers value the package more than the wages foregone.

A couple of simple benchmark cases are worthy of mention. Where worker valuation of services received is equal to the wages foregone but benefits also accrue to others, the allocation of labour between the two sectors will not be affected by those additional benefits, but total societal welfare will be increased, the gains going to whomever receives the benefits, presumably either (i) SS affiliated workers of some category but perhaps not the lower income ones who are often considered to be those for whom some benefits of the SS package are of least value (see Perry et al, 2007, 194) or (ii) SS employees (perhaps in the form of above equilibrium wages).

Another case of special interest is that in which, although the SS bundle is worth less to the worker than the corresponding value of wages would be, none of the differential between the two values constitutes a social loss. In this case socially optimal allocation of labour would be the one where VMPL<sub>i</sub> equals VMPL<sub>f</sub>, e.g. the no-intervention equilibrium illustrated in Figure 1. But since the SS package is not worth the foregone wages to the workers, a labour shift occurs towards the informal sector, leading to a deadweight loss, occurring because the private gains of the decision-making workers are not fully aligned with social gains.

Figure 2e illustrates a perhaps more likely case, in which the shortfall between the costs to the worker of SS and the benefits she receives is only partly counterbalanced by net gains to other members of the society. The CWE<sub>f</sub> curve falls well below the VMPL<sub>f</sub> curve, reflecting her much lower valuation of the SS package than of the wages foregone to acquire it, but the gap between CWE<sub>f</sub> and the marginal social benefits curve (MSB<sub>f</sub>) measures the benefits to other members of society from her involvement in the SS system. The gap between that curve and VMPL<sub>f</sub> then measures the deadweight loss to
society from the existence of the SS system. Since, as always, it is the benefits received directly by formal sector workers that drives the allocation of workers between the two sectors, the equilibrium in this case occurs at point M, with $L_{fm}$ workers in the formal sector. GDP has fallen by MNE vis a vis Case 1 because of this shift into the informal sector, and the direct deadweight loss from the badly designed SS system is area RNTU. The optimal allocation would have corresponded to point V, where the MSB$_f$ curve cuts the VMPL$_i$ curve and societal loss would then have been lower by area MTV.

In the contrasting case where CWE$_f$ falls below VMPL$_f$ but MSB$_f$ falls above it, less than $L_f*$ workers will once again be employed in the formal sector, although the optimal allocation would have been above that level. Conventionally measured GDP will fall but societal welfare will rise as long as the net gains from the SS system outweigh the loss of output (GDP); this will depend on the parameters involved.

Workers may suffer myopia or some other informational gap relative to the designers of the SS system, such that the true value of the SS benefits to them differs from their own perception/evaluation of those benefits. Workers might, for example, benefit more from savings than they realize or might not in the absence of SS be able to save as much as they wanted to. In terms of final societal benefits and any deadweight loss due to misallocation of workers between the two sectors, this case is essentially the same as the one just discussed, since it does not matter whether the worker’s own evaluation of SS benefits differs from total societal benefits because he/she understates the benefits to himself/herself or because he/she does not take account of benefits to third parties. In the present case, and still assuming that all workers are the same in these relevant respects, one must distinguish labour costs, actual benefits to workers (including wages and SS benefits) and perceived benefits to workers. The equilibrium allocation of workers between the two sectors reflects the position of the CWE$_f$ curve (perceived benefits) rather than the MAPB$_f$ curve (actual benefits to the worker). This latter curve plays the same role in the analysis as did MSB$_f$ in the previous case.

**Case 2c. A SS contribution with associated worker benefits is placed on formal sector workers; workers are identical in all respects; the government subsidizes the system**

Many SS systems receive public sector subsidies in the financing of these bundles. Other things equal, this will increase the likelihood that P will be high, and hence increase the attractiveness of formal sector employment to the workers. It will also constitute a source of misallocation of labour between the two sectors and will often encourage inefficiency (or corruption) within the SS system. Absent the latter effects, a subsidy will lower social welfare when the worker (correctly) values the SS less than the foregone wages (i.e., CWE > current wages + SS benefits) since it will lead to an increase in formal

---

12 It has been noted that access to severance payments has often provided the funds to allow a start-up of a small firm by a former paid worker. In some cases, no doubt, the worker would have been able to save up on his/her own, but presumably not in all.
employment even though the payoff to formal sector employment is \((CWE_f)\) is less than that to informal employment \((VMPL_i)\). When workers underestimate the SS benefits, either to them or to others, the subsidy can offset the otherwise existing inefficiency from too few workers opting for formal sector employment; when they overestimate the benefits, the subsidy will increase the deadweight loss.

**Case 2d Valuation of the bundle of SS benefits varies across formal sector workers**

Outcomes are altered when the valuation of the bundle of SS benefits varies across people, even though they are identical in terms of productivity. The effects of an SS system now depend, *inter alia*, on the way people are selected into the formal sector as well as on the average valuation of SS benefits and its distribution across workers.

Two extreme ways of conceptualizing the selection of people into the formal sector are (i) at random, unless and until some workers opt out of that sector and (ii) (for all workers), strictly according to the worker’s valuation of the SS package. As discussed above, when \(P=1\) for all workers (i.e. each person values the SS package as equal to the wages foregone in acquiring it), the presence of the SS system has no impact on labour allocation or on total societal welfare. The same outcome would on average hold when the mean value of \(P\) across workers is one and they are chosen at random for entry into the formal sector. If the distribution of \(P\) across all possible entrants is such that for \(L_f = L_{f^*}\) (as in Figure 1) there are \(L_{f^*}\) workers for whom \(P \geq 1\) and there is self-selection into formal sector employment, labour allocation between the two sectors will again be unaltered from the no-SS case, but societal welfare will have increased since all workers in the formal sector except the marginal one have benefited from that system.

In cases where \(P \leq 1\) for all workers, loss occurs due to the badly designed SS system but is less if workers are self-selected by \(P\) into the formal sector than if any other selection mechanism is used. Assume for simplicity that \(L_{f^*}\) is half of total employment and that, for the \(L_{f^*}\) workers with the highest \(P\), the value of that variable declines linearly from \(P=1\) down to \(P=0.5\) and for the rest ranges on down to \(P=0\); thus on average \(P=0.5\).

Assume also that the SS contribution is equal to the base wage.

Consider first the case where workers are self-selected into formal employment:

The equilibrium condition determining the allocation of labour between the two sectors of this economy is \(W_{Df} = W_{Sf}\), where \(W_{Df}\) is the wage at which a given quantity of labour \(D_f\) is demanded in the formal sector and \(W_{Sf}\) is the wage at which a given quantity \(S_f\) is supplied. The demand curve is \(D_f\) while the supply curve is the arc \(TB\), coinciding with the VMPL\(_i\) curve at its two end points but lying below it between those two points. Its position can be deduced as follows. The lowest wage at which any worker will be supplied to the formal sector (neglecting the condition that they must receive subsistence for the wage to be feasible over the medium or long run) is either the vertical intercept of the VMPL\(_i\) curve on the left-hand axis or zero if VMPL\(_i\) does not reach this axis (as in the example portrayed in Figure 2f). For any given worker the supply price is what he/she could earn in the informal sector (the height of the VMPL\(_i\) curve) minus the net loss to that worker from his/her involvement in the SS system. In the present case that net loss
gradually falls from zero for the first worker hired in the formal sector to the full SS contribution (equal to the base wage) for the last one hired away from the informal sector.

The equilibrium condition is

$$W_{Df} (L_j) = W_{Sf} (L_j)$$

where $$W_{Sf} (L_j) = \frac{VMPL_i (L_j)}{L_j / L_t} \left(1 - \frac{SSR}{VMPL_i (L_j)}\right)$$

where SSR is the ratio of the social security contribution to the basic formal sector wage, 1.0 in the case shown in Figure 2f) and

$$\frac{L_j / L_t}{1}$$ is the share of workers in the formal sector when worker j joins that sector.

In words, the supply price to the formal sector for any quantity of labour is the VMPL in the informal sector were that marginal worker to work there minus a share of that informal sector productivity equal to the SS tax rate (SSR) when the share of workers in the formal sector is zero, falling to zero by the time all workers are in the formal sector, reflecting the fact that the SS benefits are worth the full cost of them for the first worker the formal sector would hire and worth nothing to the last worker hired by that sector.13

(Under the present assumptions the arc TB would be one quarter below point E directly under that point, since for the worker found there P=0.5.)

The equilibrium, defined by the intersection of Df and the arc TB occurs at point R. Under the specific assumptions made here point R will involve less employment in the formal sector than Lf*, the optimal level in the absence of the SS system. The deadweight output loss is EFG and the SS inefficiency loss is the area FGJK. For no misallocation loss to occur in this specific case it is necessary that half or more of workers have P \(\geq\) 1. There is then no SS inefficiency loss either since only those with P \(\geq\) 1 are found in the formal sector. As the distribution of P becomes less favourable (shifts down), the SS inefficiency loss naturally rises but so also does the misallocation of labour loss.

Now if selection to enter the formal sector is at random, the average directly SS-related loss will be half of the worker’s payment into the system as long as all workers offered a job in the formal sector opt to work there, and the CWEf curve will be as shown in Figure 2f. However, for formal sector employment levels of above Lf the formal sector wage will be less than Wh, with the result that those workers with P=0 will now opt out of formal sector employment. As more workers enter the formal sector and the wage goes down farther, a higher and higher share of workers will opt out.14 But total formal sector

---

13 Put another way, workers moving to the formal sector get a formal sector base wage plus SS ebnenfits. Inducing more workers into the formal sector requires the sum of these two components to rise. By the time all workers have been hired, with the last one having P=0, this supply curve intersects the VMPLi curve (on the rightward axis).

14 ****repeats Suppose now that workers are selected according to their P level, plausible since those with high P levels will be most desirous of getting employment in the formal sector. The CWE curve in this case, CWEf*, coincides with the MPLf curve on the vertical axis, then falls inearaingly below it and intersects the line FfE half way
employment will not be as high as in the previous case of worker self-selection. This can be seen as follows. In the self-selection case, were formal sector employment equal to $L_{f^*}$ (counterfactual) then all workers with $P>0.5$ would be in the formal sector, with the last worker hired having $P=0.5$. In the case of random selection, no one opts out of formal sector employment until $L_{H}$ workers have been hired there, but by the time employment had reached $L_{f^*}$ all those with $P\geq 0.5$ would have opted out. Thus the allocation of the workers between the two sectors would be the same. But for lower levels of formal sector employment the self-selection is by definition less perfect (efficient) in the random case, implying that the equilibrium formal sector employment will be smaller. Further, there will be greater SS-related loss because of that imperfect selection. Thus both types of loss will be greater than in the full self-selection case.

Another way to see this difference is that, in the self-selection case as $L_f$ rises from zero to $L_{f^*}$ that selection linearly lowers $P$ from 1 to 0.5. Thus, by $L_f = 0.9 L_{f^*}$ the last worker hired has $P=0.55$ and all those with $P$ under 0.55 are in the informal sector. In the random selection case, those with $P$ less than one only start opting out when by $L_f / L_{f^*}$ is equal by $L_f = 0.9 L_{f^*}$

And only when $L_f$ has covered 90% of the distance between $L_H$ and $L_{f^*}$ will all those with $P \leq 0.55$ have opted out to the informal sector.

To summarize, self-selection of workers for the formal sector by their supply price minimizes the sum of the direct deadweight loss from SS inefficiency and the loss from labour misallocation (or maximizes the gain, depending on the situation); in other words,
it maximizes social welfare given the VMPL curves and the pattern of loss within the SS system. If enough workers have $P \geq 1$ there will be no loss and can be a net gain.

An alternative way of portraying the socially optimal labour allocation in this case is as the equilibrium defined by the intersection of the VMPL\textsubscript{i} curve and the MSB\textsubscript{i} curve, defined as the VMPL\textsubscript{f} curve plus or minus the net welfare effects of the SS system for the worker in question. With workers ranked by their $P$ ratios, then the first worker to appear on the VMPL\textsubscript{f} curve is also the one with the highest $P$ ratio; each marginal worker on this curve has a lower VMPL than the preceding worker and also a lower $P$, such that the MSB\textsubscript{f} curve has a greater slope at each level of $L_f$ than does the VMPL\textsubscript{f} curve. If the MSB\textsubscript{f} curve lies above the VMPL\textsubscript{f} curve to the right of point $E$, then the socially optimal allocation of labour involves more formal sector employment than $L_f^*$.

Under the assumptions of this case, the market outcome is always the optimal one, taking as given the SS arrangements. And it is always as good as or better than the no-SS outcome if $P \geq 1$ for at least $L_f^*$ workers. When less than $L_f^*$ workers have $P \geq 1$ then the socially optimal level of formal sector employment is less than $L_f^*$ and once again loss is minimized by the market solution. Absent some policy to prevent this outcome (such as an imposed minimum wage—see below), however, it is the natural outcome.

Is this para now unnecessary???

When for whatever reason worker allocation does not reflect $P$ (as when the workers are chosen at random for formal sector employment), there is an additional loss because too many people are in the informal sector. Figure 2f portrays this case. The CWE\textsubscript{f} curve intersects MPL\textsubscript{i} at the equilibrium point $Z$, which is also the socially optimal point given the existence of the inefficient SS system and the random selection of workers into the formal sector. Although there are more than $L_f^*$ people in the informal sector, a fact that lowers GDP (as normally measured) by $YEZ$, it does not in this case lower societal welfare; if an additional worker moved into the formal sector (with $P = 0.5$) this would raise total output by $ZY$, it would also lead to a loss of that amount as the worker pays that much more into the SS system than he/she will get back. Further labour shifts would lower total societal welfare as the GDP gain would be less than the additional loss from the weak SS system.

If $P$ ranges up to 1.0 or above so that enough workers do not lose at all, then the fact that the package is not optimal for everyone has no impact and the equilibrium will involve no loss of either type. But as soon as there is imperfect bundle loss there will also be misallocation loss.

**Case 2e. Application of SS varies across formal sector firms**

As noted above, the implications of an SS system can vary greatly according to whether there is or is not a degree of self-selection involved in who winds up having it, through the mechanisms of worker choice to work in the formal or the informal sector. A parallel element of choice may involve the possibility that not all otherwise formal firms will
adopt a SS system. As discussed below in Case 7?, a more realistic portrayal of the economy in which the issues under discussion play themselves out is one of three sectors, with an in-between sector made up largely of intermediate technology SMEs. Application of such social instruments as SS tend to be partial within this sector (whereas they are virtually complete in the fully formal modern technology sector and nearly absent in the informal sector). This makes the determination of when they are applied and when they are not crucial to their overall impacts on the allocation of labour among the three sectors, on economic efficiency, on the wage rate(s), etc. As with worker allocation between sectors, the application of SS regulations can be decided at random, can reflect the preferences of workers and firms or can be affected by other factors, such as corrupt officials who wield this tool against some firms when bribed to do so by other firms.

In broad terms, flexibility in the application of SS regulations can act like worker self-selection into the formal sector in diminishing the likelihood of deadweight loss resulting from them. In this case, however, the outcome depends not just on workers but possibly also on firm and regulatory agency behaviour. Firm interests and overall worker interests are likely to be consistent with each other, but the behaviour pattern of the regulators is not so predictable. It could vary from what might be called efficient flexibility—where the SS requirement was not imposed whenever its presence would be damaging (whether because of internal inefficiencies such that P<1 or because it would raise labour costs and through that mechanism lead to deadweight loss), to inefficient flexibility, where SS was imposed precisely where it would do the most damage, whether a result of pressure from interested groups, bureaucratic incompetence, corruption or whatever other factor that might be at play.

Case 3 Minimum Wage (MW) above the Equilibrium Wage, Applied to the Formal Sector
Another frequent real world constraint is the existence of a minimum wage (MW), applied to the formal sector. Assuming no loss due to imperfect SS design (because P=1 for all workers or for enough of them to fill the available formal sector jobs) then the pre-MW equilibrium at E (Figure 3) will involve no loss. The real world norm is that a MW applied in the formal sector (e.g. \( W_m \)) be above the informal sector wage. This shrinks formal employment (to \( L_{fm} \)) and in the case shown makes it superior to informal employment for all of those who can be absorbed in the sector, even if their P were =0 (in which case their total benefits from employment would simply be the wage received (\( W_m \)). Thus there is an excess supply of workers available to the sector and a rationing device different from self-selection has to be part of the labour allocation process. As long as that device does not rank workers by the net benefits accruing to them from

---

15 Note that if the presence of SS is the only thing that defines formality, then this outcome is tautologically impossible. In fact, however, formality is as often linked to the firm’s being registered and being on the tax rolls, so the present point is not irrelevant.

16 A similar mechanism may arise on the labour side when worker selection is “supposed” to be random or when an excess supply of labour creates a queue of workers wanting to enter the formal sector.
formal sector employment (wage plus worker’s evaluation of SS benefits) there will be
deadweight loss through mis-selection of which workers go to each sector. This is
additional to the deadweight loss from there being too many people in the informal sector
due to the MW and any loss from the fact that less than Lfm workers have P≥ 1 (so that
even if those with the highest Ps were selected to work in the formal sector there would
be loss associated with the inefficiency of the SS system). Suppose the workers are
chosen randomly with respect to their P ratio and the average for those chosen is P=0.5.
Then the former loss will be PFLQ where HL (value of SS services to the worker) equals
half of HF (cost of those services). The deadweight output loss resulting from the
combination of a MW and the inefficient SS system is area FEG. The former area could
in theory vary from zero if P=1 for all workers to the whole area PFHWm if P=0 for all. It
could be very low (high) if P varies among workers and their selection into the formal
sector is according to how high (low) that P level is.

The loss from a minimum wage that applies only to part of the economy (which has to be
the case if its level is above the equilibrium wage and it is not allowed to generate large
open unemployment) has been widely analyzed (e.g. Heckman and Pages, 2004). For
present purposes, the interesting aspect of the MW is the way losses from its presence
(and any possibly positive redistribution effects) are related to the presence of SS, which
can also independently create loss and misallocation of labour between the two sectors. In
the simplest of cases, an inefficient SS (with Ps<1) simply constitutes an increase in
labour costs so whatever damage is done by a MW will be accentuated by the SS. The
presence of a binding MW will normally assure that an imperfectly designed SS
magnifies the original losses.

Another way of dissecting the factors contributing to loss here is to note that, in the
absence of SS the imposition of a minimum wage at level Wm would lead to a deadweight
loss of KEJ, a fall in informal sector wages (by RJ) and a rise in formal sector ones (by
RK). In the context of an SS system operating as assumed in Figure 3 the deadweight
output loss is the much greater FEG plus PFLQ. The big increase in loss results partly
from the fact that the deadweight loss per additional worker misallocated to the informal
sector is an exponential function of the number misallocated17 and partly from the
addition of the SS inefficiency loss. In short, the details of the SS system (both its costs
and its efficiency) are likely to matter much more in the presence of a distorting
minimum wage. This result follows from the fact that, in terms of the impact on formal
sector employment and associated deadweight loss, the SS contribution is equivalent to a
higher MW by that same amount, since the sum of these two items determines labour
costs, which determine the level of employment. A really efficient SS system, in which
P≥ 1 for most or all workers will naturally cut losses vis a vis an inefficient one like that
portrayed in Figure 3 (by reducing, eliminating or evenising the loss area PFLQ. But the
benefits of this efficiency will not be manifested in a greater level of employment in the
formal sector since, in contrast to cases with no distorting MW, workers are not allowed
to accept lower wages in return for the SS benefits. With P=1 for everyone, the
deadweight loss will remain at FEG; when P exceeds one the overall loss will be FEG

\[17\] For the first worker misallocated the loss is asymptotically close to zero, but for the last worker (in
Figure 3) it is GF.
minus the difference between SS benefits and the payments required to get them. Those benefits would have been greater had more people been allowed into the formal sector, but the minimum wage prevents that outcome.

A Minimum wage when labour is heterogeneous in productivity

The impact of a MW is affected by the degree of heterogeneity of productivity across workers, assumed in the above example to be zero. In fact, such productivity may vary rather widely. Under perfect markets and no interventions, firm-worker bargaining will lead to the same payment per unit of work to each worker, with the total wage (per unit of time) thus varying across workers with their productivity. The wage system is thus equivalent to a piece rate system. Where a given worker’s relative productivity vis a vis others is the same whether engaged in the formal or the informal sector, then prior to the imposition of the MW the allocation of low and high productivity workers between the formal and informal sectors would be unrelated to productivity, unless some additional feature of the situation connected the two.

Consider first the impact of the imposition of MW where productivity (call it “absolute” productivity) differs across workers in the same way for both sectors. A MW which sets a floor under the wage per unit of time may or may not diminish the capacity of the firms to equalize wages per unit of work. As long as they control who will be hired, they will now select the more productive workers first. As long as there are enough such workers so that, following this procedure, the firms can hire as many units of work as before (albeit less workers) without raising the payment per unit of work, then their profitability will be unaffected as will their total employment measured in work units. More workers will now remain in the informal sector but since they are less productive than the average ones, the economy-wide payment per unit of work will remain the same as before, as will the distribution of output between the formal and informal sectors. Any employment reducing impacts of an inefficient SS system will be similar to those discussed above for Case 2d, where the loss depended on how well the market could allocate different people between the two sectors. When, through some combination of a relatively high MW and a relative scarcity of workers with well above average productivity, the firms start to have to pay more per unit of work, then the equilibrium changes and formal sector output and employment (of work units) falls; the MW has the same real effects as occur in the standard homogeneous labour case. Any additional employment reducing impacts of an inefficient SS system will now be magnified as in that case. The main point here, though, is that worker heterogeneity can help to avert both initial MW deadweight loss and the additional inefficient SS-related loss that can accompany it.

Turn now to the case where relative productivity as between the formal and informal sectors (call it Rf/Ri) is not constant across workers. This means that both absolute productivity and relative productivity between the sectors can vary across workers. We assume that these features of worker productivity are known to the workers and to potential employers. In the absence of any imperfections in the market, those with the highest relative productivity in the formal sector would be found there and those with lower values would be in the informal sector. Absolute productivity would not affect where a worker found a job since the payment per unit of work would be constant across workers—payment per day would vary according to absolute productivity in the sector in
which the person was employed. The relative size of the two sectors and the distribution of relative productivities across workers would determine how far down the $R_f/R_i$ scale the formal sector workers ranges in the initial equilibrium. Once again, the imposition of a MW might or might not disturb this prior equilibrium (except in the sense that some people will trade places as between the two sectors). The answer will turn on whether there is a large enough supply of people available who, earning the pre-MW payment per unit of work, would also be earning the MW per day. If not, then the MW will raise the labour cost per unit of work and curtail employment in the formal sector.

Note that the MW can be set well above the daily earnings of many people in both the formal and informal sectors prior to its imposition without affecting the equilibrium, since they can now be shuffled around in such a way that the only ones earning less than MW per day after that imposition are found in the informal sector. This outcome is especially likely when a high $R_f/R_i$ is correlated across workers with a high absolute productivity.

Assuming for simplicity that absolute productivity in the formal sector varies within the range 0.8 to 1.25 with a rectangular distribution. And that relative productivity takes on three values, 1.25, 1.0 and 0.80. Then those at the bottom of the absolute productivity distribution (0.8) will have productivity in the informal sector of either 0.64 (those with $R_f/R_i = 0.8$), 0.8 (those with $R_f/R_i = 1.0$) or 1.0 (those with $R_f/R_i = 1.25$) while those at the top of the absolute productivity distribution in the formal sector will have informal sector productivity of either 1.0 or 1.25 or 1.71 in the informal sector. Others will be in between the two extremes of 0.64 and 1.71. The supply curve of workers to the formal sector, based on their productivities in the informal sector will thus start at 0.64, have a linear segment up to 0.8, then a steeper linear segment up to 1.25 and a flatter linear segment up to 1.71 (as shown in Figure ??).

Before imposition of the MW expressed in terms of daily earnings, workers would be allocated to the formal and informal sectors exclusively on the basis of their $R_f/R_i$. Depending on its size, the formal sector might have just some of the workers with $R_f/R_i = 1.25$, all of the former group and some of those with $R_f/R_i = 1.0$, or all of both of these two groups plus some of those with $R_f/R_i = 0.8$. Then, depending on where the daily MW is set, it may or may not affect economic efficiency (and hence GDP), the per unit of work wage rate, and other relevant variables. Figure ?? exemplifies cases where there will or will not be an effect. In the pre-MW equilibrium occurs within the range where some but onto all workers with $R_f/R_i = 1.0$ are in the formal sector. This means that all workers with $R_f/R_i = 1.25$ are located there. For this latter group the minimum absolute productivity (which can be thought of as daily productivity) is 1.0. For the $R_f/R_i = 1.0$ group, minimum absolute productivity is 0.8. If the MW were now set at 0.8 it would have no impact of any sort. If set at 1.0 it would not force any of those formal sector workers of the $R_f/R_i = 1.25$ group out of that sector, but it would imply that the half of the

---

18 Recall that this range of productivities reflects differences among workers; the actual labour productivities observed in the formal (or informal) sector will also reflect the amount of accompanying capital and the technologies available.
$R_f/R_i = 1.0$ group with absolute productivity less than one could not be hired there. If less than half of this group were in the formal sector in the pre-MW situation, the equilibrium would not be affected, although all of the workers with absolute productivity below 1.0 would have to move to the informal sector, being replaced by others with absolute productivity equal to or greater than one. But if so many of the $R_f/R_i = 1.0$ workers were already in the formal sector so that the above reshuffling could not be achieved without diminishing employment there, then the equilibrium would be affected and economic inefficiency would result. When the level of $R_f/R_i$

is a continuous variable rather than a discrete one as above, the equilibrium wage and accompanying allocation of labour between the two sectors will correspond to a specific level of that variable and there will not be a large number of workers with exactly that level. Thus, as was that a dual MW will fail to affect the equilibrium only if it is below the daily earnings of everyone currently employed in the formal sector, which means it must be well below the average earnings in that sector. In that case there will be no reshuffling of workers between sectors as a result of the MW. But as soon as the MW pushes someone out of the formal sector (because his/her daily wages were below the new MW), that person’s least cost replacement from the informal sector must have a $R_f/R_i$ below that of the currently marginal formal sector workers, so the exchange of workers much lower overall worker productivity and hence disturb the previous equilibrium. In either of the two scenarios just discussed, it is true, as always, that the higher the MW is set, the greater the likelihood that its negative impact will be large. What is special about this case, however, is that a daily MW above the previous level of some formal sector workers will not necessarily lower formal sector employment.

In the real world case there is a (perhaps normal) distribution of both absolute and relative productivities. Workers would, as in the simple case outlined above, sort themselves out according to where their absolute productivity was higher. Once again, the impact of a MW in the formal sector would depend on whether anyone currently working in that sector had a daily wage below the new MW, which presumably they would unless that MW were well below the average daily earnings in both sectors, since some workers near the level of $R_f/R_i$ that defines the border between the two sectors would have low absolute productivity. Increases in that wage would now have the impact of producing a reshuffling of workers between sectors. In the initial equilibrium, all workers at the margin between the two sectors would have the same $R_f/R_i$, some with high absolute productivities and others with lower ones. All workers with higher $R_f/R_i$s than this would be in the formal sector and all those with lower ratios would be in the informal sector. The variance in daily earnings of those at this margin will reflect the variance of absolute productivity of those in the two sectors. When the MW is imposed

19 Note that, since the basic measure of labour input here is the work unit, it is no longer the case that an unchanged distribution of work units between the formal and informal sectors implies an unchanged distribution of people. But the former is the more relevant one in economic terms.
the cost of those towards the bottom of the previous daily wage distribution (and the absolute productivity distribution on which it is based) become too expensive for formal sector employers (at least more expensive than some other workers with already higher daily wages.

Workers with sufficiently high absolute productivity such that even those of them with the lowest $R_f/R_i$s still earn the new daily minimum will not be part of this shuffle. But workers with lower absolute productivity and the previously marginal $R_f/R_i$, or a somewhat higher one will now be shifted to the informal sector and replaced by workers with $R_f/R_i$ only a little below that marginal level but with higher absolute productivities. This shift will lower total output in the economy whenever some workers in the formal sector have lower $R_f/R_i$s than some in the informal sector, and the criterion of efficient allocation of labour between the two sectors is that all those in the formal sector should have $R_f/R_i$s equal to or higher than all workers in the informal sector. If the variance of $R_f/R_i$ across workers is small relative to that of absolute productivity, then the reallocation of workers in response to the MW will produce only a low deadweight loss, whereas if the opposite is true the loss will be high.20

The MW and resulting labour reallocation will raise the average wage per unit of work in the formal sector, because in effect it brings about a shrinkage of the supply of labour to that sector, so defined), and it will lower the wage it in the informal sector since more workers are pushed into it. The average daily wage earnings will rise even more (than average wages per unit of work) in the formal sector and fall even more in the informal sector. The impact on total labour earnings is ambiguous, since as always with minimum wages, it depends on the labour demand elasticities in the two sectors. A basic impact is to push workers with low absolute productivities out of the formal sector. If such low productivities are due to a shortage of learning experience, and if a better learning experience can be had in the formal sector, an additional loss in longer run labour productivity will be incurred. Where low current wages are a commonly bargained trade-off with the benefits of learning (as in apprenticeship arrangements) this useful option will be cut off.

Under what circumstances would there be an excess supply of labour to the formal sector, i.e., a non-equilibrium wage? There is excess supply as soon as the MW is created since everyone currently earning less than that amount wants to enter the formal sector, and the firms select out those with lower formal sector productivity. Because of the variance in worker productivity, the loss is less than it would be otherwise, or is this true, given that it is hard to compare the cases. If the MW is set the same distance above the previous

20 Note that, assuming capital is fixed in each sector and matters to productivity, the demand curve for labour, expressed in terms of productivity units of labour, slopes down for that reason. Assuming those workers with the highest $R_f/R_i$ are selected first into the formal sector one could also note that average per worker marginal productivity would thus fall faster than would per unit of work productivity. With workers chosen only by $R_f/R_i$, there would at any point on the curve be a mixture of workers according to absolute productivity levels, making it necessary to speak in terms of their average marginal productivity rather than in terms of a single number.
average earnings of workers as between cases of homogeneous and heterogeneous workers, then it is not clear in which case the loss is greater.

Suppose a SS tax is imposed on one sector but not the other, in the presence of workers who are heterogeneous in productivity and in the absence of a MW. Once again, it will have no ill effects unless it is “inefficient” in which case the implications are as in the simple Case 2b above. The additional complexities related to a heterogeneous productivity labour force have no general effects on the impact of a SS system. When a MW is present, however, the addition of an SS system again raises the cost of labour to the formal sector and contracts employment further below its optimal level.

**Other artificial barriers to entry into a formal sector**

Any “artificial barrier” to worker entry into the formal sector is likely to have some of the same loss-causing effects as a MW. The misallocation inefficiency created by such barriers typically assures that any inefficiencies in the SS system will have additional negative impacts, as discussed noted in connection with minimum wages. The absence of a self-selection system, now more probable since worker supply prices can no longer be that selection mechanism, is likely to increase the total loss since to the other sources of loss is now added the deadweight loss associated with allocation of the wrong workers to each sector. Assuming, for simplicity, that half of the workers had $P=1$ and the other half had $P = 0$, the minimum inefficiency SS loss would now be zero (if only workers with $P = 1$ were chosen to work there) and the maximum loss would be $PFHW_{ms}$ (of Figure ??) twice the loss occurring under random choice. Where $P$ varies widely among potential workers, the selection criterion can be the main determinant of the amount of societal loss from the combination of a MW and a SS system.

The extent to which self-selection can still play a role thus becomes an important aspect of any scenario of the sort under discussion here. Another posited explanation for inter-sectoral wage and other differentials is a system of “efficiency wages.”21 (Mazumdar, 1959; Basu and Felkey, 2009). An efficiency wage mechanism does not by itself prevent a trade-off between wages and SS benefits and is also more consistent with worker self-selection into the formal sector than are some other scenarios.

Note first that the idea of efficiency wages is based on the proposition that workers will be more productive when paid more and in particular that a wage differential in their favour vis a vis other workers can keep their productivity high enough to make it profitable for the firm to pay them an above going wage. Such a view may have a plausible base simply in the idea that better fed workers are more productive or that a wage advantage provides a good deal of motivation to workers (e.g. a strong desire not to lose their job); it may also have a foundation in the idea that some workers are better learners than others and that since the firm does not want to lose them after they have acquired firm–specific human capital, it makes double sense to pay them more. Alternatively, any combination of these factors and others cited in the literature could be at play. In any case, one implication of the presence of an efficiency wage logic in a

---

21 Though the net effects of an efficiency wage system also depend on its impact on firm-level efficiency, the raison d’être of its existence in the first place.
given scenario is that the imposition of a binding MW may cause a smaller deadweight loss than it would in the absence that mechanism.

The most obvious implication of the presence of an efficiency wage mechanism is that since a substantial wage differential between a higher productivity sector and the rest of the economy is that the higher wages are paid by choice, at least up to a point, the MW may at the limit be simply irrelevant to wage-setting in the formal sector and hence to the allocation of workers between sectors. If it is high enough, of course, the MW will eventually start to bite. Also worth noting is the fact that, even if the efficiency wage process was not previously very important, the imposition of a MW creates a strong incentive for firms to get the most they can out of highly paid workers. They may therefore undertake conscious search for such the benefits, which may be of the type highlighted in the efficiency wage theories, or more related to the creation of firm-specific human capital, which now has more logic than before.

How does SS fit into this picture of an efficiency wage scenario? The implications are no different from previous cases in the sense that to the extent that it pushes wages above their equilibrium level for formal sector firms it can discourage employment. But they differ in other respects. First since, as noted, a MW well above the going informal sector wage may not constitute a distortion, it follows that the addition of a SS system will not add further deadweight output loss. Second, some components of SS may be substitutes for wages, so this case is likely to approximate one in which worker wages are traded off for SS benefits, even if the MW is well above the informal sector level.

A simple portrayal of how an efficiency wage system might interface with SS and a MW is given in Figure 4.

Policy Response
A policy implication of this last point is that, if and when SS systems are improved, it becomes additionally important that minimum wages systems not constitute an impediment to entry into the formal sector.

An important empirical question bearing on the size of the misallocation is the degree of queuing for formal sector jobs.

Introducing queues and non-queuers here???

In this case, it is generally true that removal of either “imperfection” will lower loss. Which is more important depends on the details of the case.

Addition of Incomplete SS coverage within the formal sector
Partial coverage of SS within the “formal” sector provides another way for the system to avoid deadweight loss, somewhat parallel to that of variability of worker preference with respect to SS. However, since in this case there is the possibility of perverse application, the analysis is less straightforward and the possible implications vary more widely.
Elaborate?? It becomes especially important to have an empirical feel for what does determine when SS is applied and when it is not.

**Open Unemployment**

Open unemployment is part of any economy, and quite important in some labour markets. Since it is viewed as reflecting some form of market imperfection, it is usually not introduced in the first steps of an analysis like this one. It can take on particular importance as a form of deadweight loss when the unemployment is the result of queuing for better jobs in a segmented labour market. When, on the other hand, it reflects efficiency-improving job search it does not constitute a sign of inefficiency. How open unemployment is linked to issues of formalization is thus likely to depend on the causes of that unemployment.

The conceptual framework within which open unemployment would probably figure most prominently in the impacts of a forma-informal segmentation is the Harris-Todaro model (Harris and Todaro, 1972), in which relatively high levels of urban unemployment result from the large wage or earnings gap between urban formal sector jobs and rural activities. In such a case, one must add to any losses considered in the cases analyzed above the fact that people who could otherwise have been producing and earning in the informal sector opt out of work entirely in order to be in a position to take the highly desirable formal sector jobs. Taken simply and literally, this modifies that earlier analysis in that, instead of the suboptimal level of formal sector employment causing a deadweight loss associated with the gap between the VMPLs in the formal and informal sectors it is now associated with the formal sector wage, since the openly unemployed in the HT model produce nothing. Given that this deadweight loss is larger, on a per person basis, than it would otherwise be, the extent to which the HT phenomenon characterizes developing countries becomes a matter of importance. It has not been easy to settle this issue empirically. Later versions of the original model posited that many informal sector workers were in effect queuing for formal sector jobs (Reynolds, 1969). In that case there might be no additional deadweight loss from the queuing, although it was argued by some that workers moved from higher productivity rural activities to lower productivity urban informal ones in order to be better positioned for formal sector jobs. This latter possibility can only be analyzed in a come complete model than the one used here.

**Case 4. Product Market Imperfections in one or Both Sectors**

Thus far perfect competition has been assumed in both sectors. In real life, perfectly competitive markets are an exotic anomaly, probably characterizing under 5% of any economy. It is thus necessary to consider how much the present analysis will change when the assumptions about how markets function are shifted in the direction of reality.

When the assumption of perfect competition in both sectors is dropped the coincidence of the firm demand curves for labour and their VMPL curves disappear. As a result, no clear results on the impacts of additional imperfections like the MW can be deduced from theory except in special cases. The outcomes depend, inter alia, on the nature of the

---

22 The “theory of the second best” applies here; its general message is that whereas microeconomic theory can say useful things about the impacts of a market imperfection when it is the only one in the economy,
product market imperfections, which on the supply side may take the form of monopoly, monopolistic competition, oligopoly or various combinations of these and on the demand side their monopsonistic et al counterparts.

While product market structure has impacts on all outcome variables, its effects on the allocation of labour between the formal and informal sectors are likely to depend greatly on how market structure differs as between the two. Analysis is simplest when the only form of imperfect competition is monopoly. The more interesting cases are thus where the degree of monopoly power differs markedly between the formal and informal sectors. An extreme case would involve a set of monopolies constituting the formal sector and a set of competitive firms and industries in the informal sector. Too few workers and all other factors as well will be found in the formal sector, leading to deadweight loss, whose size depends on the elasticity of demand for the monopoly good; the lower it is the greater the loss.23

An illustrative feel for what sort of outcome may emerge under these conditions (illustrative since in fact there will be interactions among the monopolists and between them and the competitive informal sector through the demand side, etc) can be obtained by treating the group of monopolists as if they were just one firm, but a firm without monopsony power in their labour and other inputs markets. Under these conditions, the socially optimal allocation of labour between the two sectors is at point E of Figure 4 where the two VMPL curves intersect. But since the monopolist’s demand curve for labour is MRPL_f it occurs instead at the intersection of MRPL_f and VMPL_i (point F) with only L_m workers in the formal (in this case monopoly) sector. The dead weight loss is EFG, wages in both sectors are at W_m instead of W_e. 24 25

The imposition of a minimum wage on the monopoly sector will further curtail output and employment and lead to a further decline in wages in the informal sector.26 The same

---

23 This is portrayed in Figure ?? (see p. 123 diagram)
24 In fact, since a single monopoly firm would also have monopsony power in the labour market, formal sector employment and output would be even lower than shown here, where the implicit assumption is that labour supply to the firm is perfectly elastic.
25 In the case where the formal sector consists of a set of monopolies, each with a VMPL curve and a marginal revenue product (MRPL) curve, the latter constituting its demand curve for labour, under the assumption of independence among these monopolies in their product markets (including inputs) the aggregate labour demand and VMPL curves would be the horizontal summation of the corresponding curves of the individual firms. In this case it would be realistic to assume away the possibility of monopsonistic power in the labour market by the formal sector employers.
26 Note that when some of the monopolies are big enough to also exercise monopsony power, the imposition of a minimum wage can actually increase employment, since it makes the labour supply curve perfectly elastic over the range of the underlying labour supply curve that lies below the level of MW.
goes for an (inefficient) SS system which leaves CWE less than the labour costs borne by the firm.

Where the labour productivity gap between the two sectors is particularly large, as it is likely to be in this case, and it suggests a high level of labour misallocation, it is often tempting to lay the blame (or search for the explanation) on something about the informal sector firms—low productivity, small size or whatever. But in fact either the combination of minimum wages and a SS system that raises labour costs or the exercise of monopoly power in the formal sector can make that sector smaller than it should be. In the presence of market power in product or other input markets, one must always countenance the possibility that the true cause of a too-small formal sector is the exercise of that power by the large formal firms, which succeed in preventing entry into their spheres of activity.

Where monopoly is the dominant structure, MRPL is routinely below MVPL, in which case the deadweight loss can be greater if the degree of monopoly is greater in the one sector than in the other, which is likely between the formal and informal sectors.

For a single monopoly, the demand for labour (the MRPL curve) is by definition below the VMPL curve, the ratio gives a commonly used index of monopoly power. When all firms in a sector are monopolies, the total demand for labour curve will be correspondingly below the horizontal summation of the VMPL curves. When some are monopolies and some are not, the aggregate labour demand curve will still fall below the horizontal summation of the VMPL curves. If monopoly presence is comparable in the two sectors, the main impact will not be to cause a misallocation of labour between them but rather to keep the wage rate low and the returns to monopoly capital high (Berry, 1969).

The elasticity of demand for the monopolist’s product (which determines the degree of monopoly power) is all-important in predicting the effects of this market imperfection. The longer run elasticity of product demand is naturally higher than the short-run one, so short-run costs and losses will not necessarily be translated into longer run ones. The question of interest here is whether a monopolist’s activity in the labour market tends more to reflect its short-run product demand curve (and the corresponding MRP curve) or the longer run one. Where a monopolist has a long lasting market niche (like Coca Cola) the presumption is that it responds to its long run demand curve, but in such a case even that curve is likely to be quite inelastic since substitutes do not emerge even over a long period of time. Other firms accept that their market power in a specific product will not last too long (though it may even be replaced by another product giving comparable market power over the following period) and thus calibrate their labour demand accordingly. Though Schumpeter’s creative destruction will always be at work creating and destroying monopolies, the fact that in the long run most former monopolies do lose much of their market power does not mean that they will not be replaced by new ones who replicate that power, and hence that at a given point of time the system involves much monopoly power. Much monopoly power is based on the guild practices of groups like lawyers and some other professions, and is by definition long-lasting.
Monopoly power can be based on bureaucratic regulations, on economies of scale, on superior technology (e.g. one that has not yet been dispersed because the patent is held), on brand recognition and related repetitive buyer behaviour, or on crime/illegal use of power. Most large fortunes are amassed on the basis either of speculation in assets markets or of monopoly power with roots in one or more of the above sources. It has often been argued that freer international trade and globalization should lower the degree of monopoly power by enlarging the scope of competition. Often, however, this has led to the replacement of a national monopoly with larger multinational oligopolies which allocate many resources to developing and maintaining market power in each country where they operate. Therefore, it is hard to judge whether overall market power has risen or fallen during this period.

It is also hard to judge how much monopoly and other forms of market power affect labour demand. One of the reasons is related to the close association between extremely high executive salaries and a firm’s market power. It might otherwise appear that a good overall indicator of the extent of monopoly power in an economy would be the capital share of GDP (or the other side of the coin, the labour share). However, when those high salaries are a result of monopoly power and are recorded as part of labour income in the national accounts, the meaning of these factor shares becomes blurry.

**Appropriate Policy Response**

When market power rather than things like labour regulations explains the smallish size of the formal sector, entry of additional firms to that sector will often not occur unless some policy steps are directed at that power. Where bureaucratic barriers to entry are the problem, it simply requires getting rid of them. However, when the monopoly power rests on economies of scale, the theoretically appropriate step is to regulate the firm, for example through price controls. In any case, it is pivotal to recognize that increasing the size of the formal sector may be impossible if it’s reduced size is a natural result of the exercise of market power and if policy is powerless to confront that exercise of power. At the extreme, this monopoly sector would take control of as many product areas as it wished to; its limited labour absorption would keep labour cheap and monopoly profits high.

Raising informal sector labour productivity might seem a good approach under these conditions. When the informal sector firms involved are independent of the monopoly sector, this may be the case. But monopolists are usually adept at controlling the main components of demand and production. If informal firms raise productivity in a group of product niches that they do control, relative price may fall so that they gain little benefit from the increased productivity. If they supply inputs to the monopolies, an increase in their productivity is even more likely to redound to the monopolist/monopsonist’s benefit.

If the predominant market structure is monopolistic competition, in which case the marginal revenue product (MRPL) of one more worker in a typical firm is again below
the VMPL of that firm (reflecting the monopoly aspect of the firm), but where in equilibrium the representative firm has zero abnormal profits and it is hard to sort out the social efficiency implications of the fact that all firms are operating on the downward sloping segment of their average cost curves (Dixit and Stiglitz, 1977), there may or may not be much resulting social inefficiency/deadweight loss. In addition, wages are not being bid down to the benefit of monopoly profits as in the pure monopoly case. Therefore, there is nothing identifiably wrong with this scenario. Again, the main interest arises when the market structure differs as between the two sectors. When monopoly dominates in the formal sector and monopolistic competition in the informal one (a reasonable first approximation), then too few of all resources will be found in the formal sector and the results will probably be very similar to those where monopoly confronts perfect competition, discussed above.

One implication of the monopolistically competitive markets, like much of retail commerce, is that they may provide a more elastic demand for labour than do other scenarios. This is an implication of the relative freedom of entry that characterizes this market structure, through a combination of modest economies of scale and other features. This structure can guarantee at least a small income to anyone in a position to share a given market, such as street salespeople. One person’s income comes in part at the expense of others, but this can constitute a sort of last resort wage-sharing scheme.

The implications of monopoly or market power in an economy depend on the details of where it is located within the system and how it is exercised. In most economies there is considerable economic interaction between firms with market power and others without it, as where the former buy inputs from or supply them to small competitive firms. Value chains often involve a large firm with market power at one stage of the chain and some competitive small firms at other stages. The former, in acting as monopolist and/or monopsonist extracts the available rents in the trade between the two but also, more relevant here, manages the chain in such a way that it behaves as an integrated monopolist. Thus, even in a system where most production is undertaken by competitive firms, if they are all found in market chains dominated by firms with monopoly power over a segment of the total process the whole system will behave on monopoly principles, i.e. identically to one composed only of vertically integrated monopolists. The demand for all factors of production will be kept artificially low and monopoly rents will make up a significant share of national income.

Whatever the details of the use of monopoly power in the formal sector, there will be a problem if the exercise of that power severely curtails employment in the formal sector itself (helping to generate an inelastic and low demand for labour), and simultaneously makes it difficult for informal sector producers to move into the same market, through a variety of defensive techniques that monopolists employ, or curtails informal sector production of inputs to the formal sector through the use of monopsony power. Under these circumstances it may be very difficult to achieve increases in productive employment in the informal sector (whether by formalizing it or in other ways) since the firms there may be essentially competing against each other, in which case any increases in productivity that they achieve may simply lead to lower prices.
Case 5. Capital Market Imperfections but no Product Market Imperfections

Capital market imperfections are serious in all economies, making the flow of this factor of production between some segments of the economy difficult and partial, and gives certain types of firms better access to capital and savings opportunities than others have. A general presumption is that the informal and usually relatively small-scale sector is relatively deprived of capital and that a large-scale sector has good access.

Analysis of the implications of this imperfection is conceptually simplest under the assumption of fixed capital stock and technology in each of the two sectors, and perfect competition in each, as carried out above in Cases 1-3 (???). In fact, however, when some workers move between sectors they take at least some of their capital with them, as where someone previously employed in the formal sector and putting savings in the bank (in which case his/her savings would be channelled to formal sector firms) now becomes informal and takes those savings with him/her to invest in a business. In this case the deadweight loss associated with there being too few workers in the modern sector will be less sensitive to the degree of misallocation of labour than it would otherwise be.

The normal situation in terms of factor allocation between formal and informal sectors is that there is too much capital in the formal sector and too much labour in the informal one.27 This leaves it theoretically ambiguous whether in some general sense there are too many factors in one or the other sector and too much output from one or the other. Clearly, the optimal outcome is one that corrects both factor market imperfections. But if it is not possible to do so in one (say the capital market) then it becomes unclear what impact the other imperfection has. On the one hand it may reduce an over-allocation of output to the capital rich sector; on the other it increases the factor choice inefficiency within each of the two sectors.28 When factor substitutability is low in both sectors, the latter source of inefficiency is small so a labour tax in the formal sector will raise overall efficiency by offsetting the lower price of capital in that sector. But if substitutability is high the opposite result may emerge. As always, the elasticities of labour demand in the two sectors are definitive for the size of deadweight loss. It requires only that elasticity be low in the formal sector to keep this loss low.

When the degree of this capital market imperfection is a given (i.e. is not subject to policy) its presence does not directly affect the above analysis in any qualitative sense. It does contribute to a VMPL curve which begins high and is steeply downward sloping, and thus affects the overall equilibrium reached in the economy. Since the elasticity of labour demand in the formal sector is a determinant of deadweight loss for all but the marginal worker shifting between sectors, this capital market imperfection does increase

27 Note that when the movement of both labour and capital between a formal sector and an informal one are taken into account, formality itself becomes endogenous in the sense that firms can move in either direction when a specific factor market becomes more competitive.

28 This inefficiency results from the underpricing of capital relative to labour in the formal sector and the opposite factor price distortion in the informal sector.
the potential gain from a shift of workers into the formal sector. The imperfection could also be causally linked to other variables of interest, such as the MW that is set, the character of any SS system, and so on. Analysis of such links would call for a more complex model.

Where does the addition of product and factor market imperfections leave us in terms of the likely implications of a SS system in the formal sector? A first, and undoubtedly oversimplified cut at this would suggest the following:

i) labour misallocation loss (additional to any design inefficiency loss from the SS depends on there being some actual or potential segmentation (wording???) leading to queuing for entry into the formal sector, such as that resulting from a MW;

ii) given such actual or potential queuing the loss depends negatively on the degree of worker self-selection into the formal sector by preference for the SS benefits, on the extent of efficiently differential application of the SS regulations, on the extent to which efficiency wage phenomena are present and play a role, and on the variance in absolute and relative productivity among workers.29

iii) product and capital market imperfections, apart from the deadweight loss to shish they give rise directly, are likely to decrease the labour misallocation loss from distortions like those under discussion, because both of them are likely to contribute to an inelastic demand for labour. At the same time, it must be noted that these features tend to amplify the loss from the imperfections that give rise to them and thus highlight the fact that under some circumstances it is of little use to worry about a labour market imperfection when the ones that really count are elsewhere.

**Case 6. Social Protection Policy**

The arrival of SP changes the net benefits implicit in having a formal sector job and the associated SS package and can thus have a variety of impacts, many of which have been identified in the literature. Consider first its implications in the simple case where there are no product or factor market imperfections, where all formal workers have \( P=0.5 \), where there is no MW, and where access to SP is contingent on working in the informal sector, then the effect on labour allocation between the sectors can be seen as the result of an upward shift of total benefits accruing to workers from their employment in the informal sector. Thus in Figure T (???) the curve VBLi shows the value of the total benefits from informal sector employment, the sum of wages and the SP benefits. More workers now shift to the informal sector. Disregarding the impact of SP per se on the benefitting workers as a separate matter, there is a welfare loss due to overproduction in the informal sector equal to area EFG. If SP benefits have the same weight in the social utility function as the benefits given up to fund them, then that area measures the net loss from the policy. If the benefits exceed those given up by more than this amount, then the policy will raise total social welfare.

The impact on labour allocation is equivalent to that of a lowering of the \( P \) of the formal sector workers. With \( VMPL \) higher in the formal sector, this increases the labour

\[ \text{29 The impact of this last factor may, however, be more complex than discussed here.} \]
misallocation deadweight loss. However, it also cuts the loss from the imperfectly designed SS bundle. In Figure 5, the previous equilibrium at point E implied output deadweight loss of EFG, plus the inefficient SS loss of ??? Now the misallocation loss area expands by JLF. 

When the SP payment is not contingent on employment in the informal sector but remains either unavailable or uninteresting to those working in the formal sector, the main likely effect in the labour market is that the supply of labour will decrease. The welfare impacts of such a shift cannot easily be shown in the simple figures used above (where a perfectly inelastic supply of labour has been assumed) but one way to portray what happens is to assume that the distance OO’ now shrinks by a fixed amount (the number of people who drop out of the labour market, regardless of the wage). With no distinctions yet made among workers and no changes in complementary factors available in the two sectors, the VMPLi curve will now be closer to the leftward axis (Figure R??), the equilibrium point will also shift to the left and employment will fall in both sectors, with wages moving higher in the formal sector and falling in the informal one. The shrinkage in total labour supply (the high labour supply underlies the large size of the informal sector) is likely to reduce non-labour incomes and hence improve income distribution.

A more likely scenario is one in which those who drop out of the labour supply as a result of the SP system are workers who would otherwise be working due to poverty. This behaviour pattern may or may not change the supply very much from its previous level. In any case, if it does so, the impact will be qualitatively the same as noted above. (Figure V). Whether the informal or the formal sector shrinks more will depend on the relative elasticities of demand for labour in the two sectors; with the formal sector demand less elastic, a reasonable guess is that employment will fall mainly in the informal sector. In short, SP may help to shrink that sector rather than, as it may under other circumstances, expand it.

When the SP is the only “imperfection” in the system (because SS does not introduce one and there is no MW), the likely deadweight loss will be small unless the SP payments are implausibly large. If the VMPLf and VMPLi curves are inelastic, the shift of workers to the informal sector will be small, limiting the deadweight loss. When both these curves are relatively elastic, it will be greater.

These surmises lead us back to an earlier proposition: the sort of quantitative shift that is likely to occur toward the informal sector is likely to bring really significant welfare loss only when that sector is already much too big, such that the marginal loss per person added to it from the formal sector is large. It is thus important to focus on the conditions which make this the case. A case of particular interest is that in which the SS system, although not ideal for all workers, only becomes really damaging in the presence of a MW which makes self-section of workers into the formal sector infeasible. Empirically speaking this makes it important to know to what extent such self-selection does operate.
When the SP is a more efficient use of public funds than SS is of private (and sometimes some) public funds, and through one mechanism or another its presence shifts labour into the informal sector then the shift towards the informal sector has both the effect of lowering GDP and of raising the overall efficiency of social spending. The GDP loss from such shifts will be either partly, wholly or more than wholly offset by the gain from that greater efficiency. Presumably the best outcome of all, of course, is where the SS system is repaired as it enters competition with the SP system.

If existing capital market imperfections cannot be diminished by policy, they may under some circumstances be irrelevant to policy on labour regulations and on SS and SP. But their presence often implies that economic efficiency can be raised significantly in more informal firms, it is possible that a SP implemented together with, say, as good microfinance program will be a good combination from an efficiency point of view. The improved access to capital can at least imply that a shift of labour into the informal sector will not imply falling marginal product of labour and wages in that sector. Alternatively, if a strong SP program can have some impact on inducing capital into the informal sector (taken by former formal sector workers) this makes the SP policy more beneficial than it would otherwise have been.

Another case of special interest is that in which the formal sector SS system is subsidized by the state. In that case, in that absence of other distortions, that subsidy will have led to an above optimal size formal sector, and the introduction of a SP system that benefits informal sector workers may offset the bias towards formal employment. Where, in spite of some subsidy to the SS system, the formal sector is smaller than optimal (due to other distortions such as MW), then the introduction of SP will increase rather than cut the deadweight loss due to misallocation of labour, though it may also decrease the inefficiency loss from the combined SS-PS system, so that once again it is possible that the overall impact on social welfare will be positive.

****** A SS wedge causes some loss if P is under 1 and/or if there is a state subsidy to the SS in the formal sector. If only the latter, and P =1, then some SP (equal to the subsidy to SS) will presumably rectify the deadweight loss. If P< 1 and the difference is 100% waste and no subsidy, the initial shift of workers into the informal sector is socially efficient, since GDP is the wrong indicator of societal welfare, it should be GDP minus waste associated with the h SS policy. One might call it NDP, net of that waste. When SP arrives on this scene, it encourages workers into the informal sector, lowers GDP and lowers NDP by the added deadweight loss plus the costs of administration.

The impact of SP can clearly depend on each of the aspects of the situation discussed above. The main issue is whether, prior to the institution of a SP system, there was a major misallocation of labour between the formal and informal sectors, one which would normally require a distorting MW as a building block, together with a further distorting SS system, and without the alleviating mechanisms of self-selection into the formal sector or efficiently partial application of SS within that sector. See below, however, for a

---

30 Either in absolute terms or, if total labour supply falls, in relative terms.
discussion of a three sector model, a more plausible conceptual framework within which to analyze these issues.

**Case 6.1 SP without queuing/rationing of entry to formal sector**

Where the relevant aggregate social indicator falls but poverty is reduced, a trade-off may be inevitable. Then the dynamics of the links between poverty and growth kick in. Can the SP raise investment in education, etc.

Need to distinguish those queued up and those not, with the latter including small entrepreneurs who can hire others. How much their costs are lowered is a genuine issue, though in many cases it seems that not too many workers paid are in the informal sector. See the figures though.

**Case 6.2 SP with queuing/rationing of entry to formal sector**

Note that a “wage efficiency” model will yield some of the same results as the MW case. What about the implications of MW affecting informal sector wages. Of course, they would be expected to if the workers in question are actually choosing.

In this case (see diagram) at the limit the addition of SP will have no impact on the allocation of labour between the two sectors. Therefore, no additional misallocation loss will occur.

**Case 6.3 SP with product and factor market imperfections**

As noted above, both product and non-labour factor market imperfections alter the impact of any other specific imperfection, such as a minimum wage or a formal sector SS system that raises labour costs above the CWE. Although the implications of these imperfections vary from case to case, two main ones are:

i) that capital market imperfection implies that firms currently in (or the sort of firms currently in) the informal sector are producing too little and the formal sector firms too much. The first-best remedy is to improve the functioning of the capital market. Where this is not possible, no general conclusions as to whether more or less labour should be in the informal sector are possible; this depends on a set of parameters.

ii) The product market imperfections are likely to imply too few of all resources used in the monopolistic sector (MC??) if it competes with the competitive sector, but that the…..

iii) Value chain systems where monopolistic and competing firms are part of the same chain will tend to operate like all-monopoly economies, with a negative impact on labour demand and wages.

**Case 7. A Three Sector Model**

To provide a conceptual framework for the discussion of the implications of SS and SP in economies with relatively high levels of informality we have thus far followed the
common tradition--a two-sector model (formal and informal) static model. In fact, everyone recognizes that the features used to define these two sectors do not in fact correspond to a real world dichotomy of firms; there is, rather, a spectrum from a high level of informality at one end to a high level of formality at the other. Distinguishing a small number of categories in terms of which to frame the analysis can be useful; the question remains, though, how many to distinguish and what should be the dividing lines between them? The answer depends on the issue under discussion; if it is the impact of minimum wage legislation, it may be three categories, if some other issue, four, and so on. The optimal cuts will also depend on the country. Meanwhile, static analysis implies that risk, uncertainty and firm dynamics have not been taken into account. Since the empirical evidence points to rapid rates of turnover among small firms (both those usually classified as informal and the smaller formal firms) and high levels of risk and uncertainty, it is clear that a static model cannot be assumed to be an adequate one. Consider first the number of sectors to be distinguished.

As soon as one’s framework shifts from a sectoral dichotomy to a larger number of groups, the analysis changes in a qualitative sense, since it no longer automatically the case, for example, that if one sector is too big for the overall efficiency of the economy the other one must be too small. It now requires us to know both which sector is too big and at the expense of which other sector. As a result, one needs to know a good deal more to judge which sectors should change size and whether and how this may be done.

A major reason for opting for a three-sector way of looking at some of the issues related to SS and SP is that a common and useful classification of firms and the workers they hire involves three groups. (See the large literature on the “missing middle”, e.g., ???.) Conceptually speaking, the nature of product competition and other forms of interaction between each pair of groups will differ. In the above two-sector analysis, a labour tax that raises wage costs in the formal sector increases the size of the informal sector with the efficiency loss as described. But in a three sector analysis a labour tax that raises labour costs for, say an SME group of firms, may then redound to the competitive benefit of either of the other two groups (or to both). If SME firms mainly compete with large very formal firms, curtailing their performance through such a tax will raise the output share of the very formal sector at the expense of the somewhat formal sector.31 Because this will likely reduce total labour absorption in these two sectors taken together, more workers will shift to the informal sector as an indirect result. The overall impact on GDP may go either way. If SME firms mainly compete with the informal sector, a labour tax on them will again cut into their competitive position and will, in this case directly, augment the informal sector but with the likely result that GDP will fall assuming the SME firms are relatively more productive in TFP terms.

Whatever the precise fallout from a tax on SMEs, what is clear is that to conceptualize the process as one in which the impact will be to discourage activity in the fully formal

31 It was recently noted that large formal firms in South Africa supported union pressure for the blanket application of high minimum wages, recognizing that this would improve their competitive position vis a vis the SME sector.
sector is likely to be wrong. Most of that sector already pays above MW and has SS, so these aspects of the situation can be taken as given. The question is the extent to which such regulations will be extended to the intermediate or SME sector as well. Applying or removing these impositions will seldom affect much of the fully formal sector directly, since it is too distant from the SME sector in terms of technology and institutional modus operandi. Estimates of potential productivity gains from transferring labour and other resources out of the informal sector should not include this high labour productivity sector since that is not likely to be the shift that happens.

A large literature, dating back at least 3-4 decades, has argued that the intermediate technology firms (typically small/medium in size) sometimes achieve higher total factor productivity than either large ones or small ones when all inputs are measured at the economically correct values (their social opportunity costs). Though this view is certainly contestable and the truth presumably varies considerably from country to country, a stronger argument can be made that the larger this group the better the labour market outcome, since the modern technology firms create too few jobs and the low technology firms achieve too low labour productivity.

Given the arguable importance of disaggregating what is often called the formal sector into a usually large usually modern technology component and a usually small-medium size and usually intermediate technology component it is clear that this is the way any analysis of this sort should start. Then, if it turns out that for purposes of the questions being asked, the formal sector is not an excessively heterogeneous one, the analysis can be folded back into that simpler framework. But it should not start there.

It seems clear that much of the labour movement into and out of the informal sector involves the SME component of the formal or fairly formal sector. This means that any comparisons of factor productivity used to draw conclusions with respect to the relative productivity of the two interacting sectors should be the intermediate-technology SME sector. The SS features borne in mind and the possible impacts of an MW should also be analysed for this sector.

Need to keep incentives up so that large firms will be induced to do lots of sub-contracting. See the discussion in main paper.

The upshot of a recognition that a three sector conceptualization of the economy is more helpful than a two-sector one complicates the analysis of the impacts of any given policy intervention, from MW, to SS to SP. Often, as where the focus is on the impact of

In addition, partial application is more likely—we know it to be an important phenomena.

Case 8. Allowing for risk, uncertainty, macroeconomic instability, and firm dynamics

34
As noted above, risk and uncertainty play an especially large part in the lives of many informal and small formal sector firms, and they are anything but “static” over time. How do these facts of life affect the likely implications of SS and SP systems on employment and on economic efficiency? How do they need to be taken into account in the analysis? Three features especially characteristic of many informal and small partially formal firms are: the learning process in which they find themselves over the first years after creation or after significant change of some sort; their need to build up capital in order to guarantee their continuing existence and success; and their need to be risk averse in order to minimize the likelihood of failure/collapse. These characteristics have a variety of implications, among them the fact that getting locked into various aspects of formality is risky for any firm which faces a significant chance of needed to lower costs in future in order to survive. Both taxes and labour regulations usually constitute such a “locking” in. The risk is less if the tax-regulatory apparatus is flexible, allowing a reduction of responsibilities when the firm’s condition worsens. When that is the case, the risk to the firm of becoming more formal is reduced.

Some features of the labour institutions that are not of particular importance under static full-information assumptions become significant in a world of risk and uncertainly. One is the rules on worker dismissal. Under certainty (with respect to worker performance, macroeconomic evolution and the firm’s own future), these are just one more determinant of average labour costs. However, under uncertainty on those counts, rigid and/or costly dismissal rules raise risk; they lower expected future profits of a firm and also bias its technology choice away from the use of labour. In the context to the static cases analyzed above the impact of this type of regulation is the same as that of a MW that raises labour costs. In a dynamic analysis, however, its implications are different and it needs to be taken into account in its own right.

Integrating the Complexities

Although for heuristic purposes most of the above analysis has been carried out in the context of a two-sector static model, this does not appear to be an appropriate one for most developing countries, so most policy thinking should be framed in a dynamic three-sector model. The main empirics that need to be understood in order to get a reasonable judgement as to the impacts of SS and SP systems on economic efficiency are the structure of interaction among the three sectors: what are the areas of greatest competition and of greatest complementation, how do these interventions affect the firms in these relations of interaction with one or both of the other two sectors? Another major need is a good understanding of capital market functioning, since to the extent that the informal and/or the SME sectors are disadvantaged by the way markers work this is likely to involve this factor market. As for the implications of dynamics.....??????

An Illustrative Possible Scenario

At a theoretical level, the number of permutations of assumptions that can be considered is excessively large and leads to quite different results according to which is chosen.
Accordingly, it is important to consider what the theory implies under apparently realistic sets of assumptions. For many countries, this set would include:

i) the three sectors distinguished by the positively correlated features of formality, technology and size;

ii) a SS system and MW that are applied systematically in the large/modern sector, not in the very small/informal sector and partially in the intermediate/small-medium sector

iii) institutional disincentives to formal sector labour hiring, through some combination of a high MW, rigid firing rules and possibly other elements of the labour regulatory package, possibly leading to a significant degree of queuing for formal sector jobs (as analyzed by Perry et al, 2007, ??), but with other possible factors also at play including efficiency wage behaviour by formal sector firms;

iv) serious imperfections in the capital market, such that access improves as a firm rises in the size-technology scale;

v) a considerable presence of monopoly and oligopoly in the large-modern sector, less market power in the intermediate/SME sector and least but still not necessarily a small amount in the informal sector, with the form in these last two typically being monopolistic competition;

vi) variable levels of purchase and sale across these size-technology-formality categories leading to a degree of complementarity between some firms of different categories, coexisting with some direct and indirect competition across the categories;

vii) varying degrees of competition from imports in the domestic markets of these goods and services;

viii) a high level of heterogeneity in the economic features of the smaller firms, including activity, market setting, quality of entrepreneurship, and access to capital.

A plausible setting is one in which the large more modern formal sector has a quite inelastic demand for labour (because of its modern technology and it market power, both of which work in this direction), with this elasticity greater in the SME formal sector and probably greatest in the informal sector. Labour regulations create a degree of disincentive to formal sector employment, but this effect may be small in the larger, more modern sector but important in the smaller formal sector. These disincentives both discourage the shifting of workers from informal to formal through hiring by existing formal sector firms and discourage the “graduation” of informal sector firms into formality.

It is natural to presume that the labour demand elasticities are considerably higher in the longer than in the shorter run for any given group of firms, hence that the impact of a change in labour regulations or other relevant policies may be considerably greater over a period of years than, say, over a short period like a year.

Numerous econometric studies have estimated the elasticity of labour demand and/or the impact of changes in labour regulations. Unfortunately, few have attempted to isolate
different groups of firms, e.g. distinguish between larger more technologically modern formal firms and smaller less modern ones.

At this point we can safely say that the SME intermediate technology sector creates more jobs in relation to capital or to output than does the larger modern technology sector. This means that if the country suffers from a “missing middle” (a relative lack of middles size intermediate-technology firms) then formal sector demand for labour is likely o be small and the elasticity of demand is also likely to be small. In that case, labour regulatory reforms may have very little impact, especially in the short run and a modest one even in the long run. The only way to achieve a significant increase in reasonably productive jobs in this sort of economy will be rapid growth in the intermediate sector. With respect to this goal, it may be very important whether certain labour regulations constrain the growth of this set of firms.

For countries not suffering the “missing middle” problem a big empirical question is how great the elasticity of labour in that sector is and how responsive employment is to possible reforms. For all countries, it is important to know how much the elasticities vary between the short and the long run and to understand the mechanisms that explain the difference, since some of them may be policy-induced and others may not.

Summary

For there to be direct policy implications on the desirability of SP, it has to be the case that labour allocation between the two sectors will respond to the presence of SP—true in non queuing cases where the elasticity of formal demand for labour is not close to zero elasticity (and increasingly true the more elastic it is??)

It is essential to remember that a lot of movement between the two sectors does not directly imply that the relative sizes could change much without greatly changing the relative VMPLs. In addition, essential to remember the GDP ceases to be the right aggregate welfare indicator under some circumstances.

RE indirect policy implications, this seems to depend on the case. If there is large-scale queuing, SP does not necessarily increase at all the loss from the reasons leading to that queuing.

Presumably, the rate of movement back and forth and the relative balance in direction of movement (among other things) will reflect the shape of the two MPL curves?

References


