

Employment and Efficiency Effects of Social Security (SS) and Social Protection (SP) Systems in the Context of an Informal Sector and Market Imperfections: A Conceptual Review.

Albert Berry

SERIE DOCUMENTOS DE TRABAJO

No. 198

Abril de 2017

Employment and Efficiency Effects of Social Security (SS) and Social Protection (SP) Systems in the Context of an Informal Sector and Market Imperfections: A Conceptual Review.

Albert Berry
Profesor Honorario
Facultad de Economía
Universidad del Rosario, Colombia

Profesor Emérito de Economía
Centro de Estudios Internacionales
Universidad de Toronto, Canadá

Abstract

Most developing countries suffer serious “imperfections “in their labour, capital and 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 (increasingly in recent years) through social protection systems (SP) designed to achieve greater overall coverage, e.g. to reach informal workers. Among the main policy issues that arise around the combined existence of these features are how to limit such negative effects as the market imperfections may have, how best to design SS and SP systems in such settings, and whether to pursue formalization be pursued as a policy and if so. The analysis of the impact of SP policies is very complicated and cannot be safely undertaken in a two sector (formal-informal) model, nor in one that assumes labour homogeneity or perfect competition in the product and capital markets. The analysis of any given case is likely to be complicated enough to pose serious barriers to the attainment of reliable conclusions, but is nonetheless worthwhile to narrow the range of uncertainty as to the impact of policy. For labour market functioning to have direct policy implications on the desirability of SP requires that labour allocation between the two (or among the three) sectors will respond to the presence of SP (as in non-queueing cases where the elasticity of formal demand for labour is not close to zero elasticity).

JEL 821, 822, 824, 911,914, 915

Keywords: social security; social policy, informality; formalization; registration; labour market; capital market; product markets; regulations; market imperfections; economic efficiency

Introduction: Social Security and Social Policy Systems in a Context of Informality and Market Imperfections

Most developing countries suffer serious “imperfections “in their labour, capital and 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 (increasingly in recent years) through social protection systems (SP) designed to achieve greater overall coverage, e.g. to reach informal workers. Three main policy issues arise around the combined existence of these features:

- i) what can be done to limit such negative effects as the market imperfections may have?
- ii) what is the best design for SS and SP systems in such settings?
- iii) should formalization be pursued as a policy and if so, how?

Many variables affect the impacts of the SS and SP systems and of informality, including labour demand elasticities, the effectiveness and *modus operandi* of the social security (SS) and social protection (SP) packages, and the presence of product and factor market imperfections, both those originating in policy—interventions like minimum wages (MW) and those occurring naturally in the economy—like the various forms of market power, externalities, etc. Accordingly, it can be very misleading to analyze the implications of a SS system, a SP system or any combination of the two, against too simple a set of background assumptions about how the economy functions. “Simple” usually means “in the absence of other market imperfections” such as monopoly power, monopolistic competition, and other market structures that are not perfectly competitive. And, although there is clear value to making an analytical distinction between a formal sector and an informal one in looking at such questions (e.g. because SS systems are limited to the formal sector), a two-sector framework also runs a serious risk of failing to pick up the most relevant aspects of reality and should, in general, be replaced by a three-sector one.

Various complementary methodological approaches can be employed in the attempt to better understand these issues, including theoretical attempts to ascertain how policies like SS and SP would be expected to affect outcomes under a range of specified conditions (the approach of this study), and econometric studies of how such policies have *de facto* affected outcomes. Both approaches face severe challenges to the achievement of convincing results, so it is important that each be complemented by the other to the degree possible.¹

The paper uses a step-by-step approach to the microeconomic analysis of these issues, beginning with an overly simplistic two-sector framework with only one “imperfection”, then gradually complicating the set of assumptions to more closely approximate reality and to understand how the policy conclusions change between less and more realistic assumptions. The first set of cases adopts a two-sector (formal and informal) model with perfect competition in product and factor markets to analyse the impacts of SS and minimum wage legislation on overall societal welfare and that of identifiable groups like formal sector workers, informal sector workers and owners of capital. Later sections introduce product market imperfections, capital market imperfections and, finally, a three sector model in which the formal sector is disaggregated into a large-scale modern-technology sector and a small and medium scale intermediate technology sector. The paper ends with some tentative conclusions on the probable separate and combined impacts of SS, SP, and MW systems. As is emphasized there, drawing any useful conclusions at all requires quantitative evidence on a number of structural features of an economy, so empirical research needs to be directed towards accumulating a better understanding of those structural features in order to better manage the set of policy issues under discussion.

The large size of the informal sector in many economies has led many observers to speculate that this outcome is inefficient and/or in some sense abnormal, perhaps due to policy failure. Misallocation of labour across the sectors of an economy, e.g. between a more formal and a less formal sector, can arise from many factors. One is a labour tax imposed on formal sector workers, e.g. in the form of required payments into a social security (SS) system, whose benefits are perceived by the worker to be of less value

¹ Perhaps the main existing study which has attempted to do this is Levy (2008) in the context of Mexico.

than the contribution he/she must make to secure them. In the absence of institutional constraints (such as a minimum wage), the full worker and employer contribution will in fact be borne by the worker.² This means that, whereas labour would otherwise have been optimally allocated³ with the same value of marginal product in both on sectors, in such a case there is a net (deadweight) loss to society as a whole, consisting of two distinct types. First, some of the resources that fund the SS system are wasted (lost) because of the weak design or implementation which leads to its cost exceeding the benefits to the workers. 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, the value of the marginal product of a worker is less there than in the formal sector. Plausible parameter numbers (on the elasticities of labour demand in the formal and informal sectors respectively) suggest that the SS inefficiency loss, when it occurs, will often be larger than the resource misallocation loss that arises from the SS system's pushing labour into the informal sector.

Where the SS system is a net plus to workers (i.e. the benefits exceed the cost borne by the worker) then its presence creates a direct societal gain, but also leads to a misallocation of workers, this time in favour of the formal sector and an accompanying decrease in GDP, with the latter loss at least partially offsetting the gain from the effective SS system. Outcomes differ again when the benefits of the SS system are not limited to worker beneficiaries for whom they are intended but also accrue in part to others (see below).

The gains or losses attributable to a well or badly functioning SS system and the losses that may result from labour misallocation between the formal and informal sectors are associated with a variety of effects on the incomes of labour and of other factors of production, again depending on the parameters involved, in particular the labour demand elasticities in the two sectors. For example, where the labour demand curve of the informal sector is very elastic, the equilibrium wage cannot be pushed down very far by anything that happens in the formal sector and a decrease in formal sector employment is more likely to lower the income of the other factors of production. However, where the opposite is the case, the wage rate can be significantly lowered and labour will be the main loser among the factors of production.

It is normal that a given SS system fit the needs of some workers better than of others, and in that case, it matters for overall efficiency of the system that market forces determine which workers wind up in the formal sector with its SS system. When workers are chosen in some way other than according to their preference to participate in the SS system (e.g. at random) there is typically a double loss vis a vis a system where the market sorts out who goes where: the net social benefits from the SS system are less (since the group of people who work in the formal sector are not the same as those who could most benefit from it), and there is also likely to be a resource misallocation loss because too few or too many people work in the formal sector (depending on whether most workers gain from participation in the SS system or most lose from that participation).

To better understand how 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

² Usually both the worker and the employer make a contribution to this system, but, in the absence of market interventions, it can be assumed that the wage received by the worker will fall by the amount of the "employer's contribution", hence that in fact the worker is paying the whole cost of the SS system.

³ Except as noted below, we adopt the standard measure of economic efficiency, i.e. the maximization of total output in the economy. When this definition is clearly inappropriate this is taken into account.

three-sector model. It is clear that no general theoretical conclusion can be reached about the effects of SS and/or SP systems on employment or economic efficiency. The analysis identifies a number of the variables whose empirical values are among the determinants of those effects.

Case 1 (Benchmark): Two (formal and informal) sectors; pure and perfect competition in all product and factor markets; no SS or SP systems; workers identical in all respects⁴; no taxes except on consumption

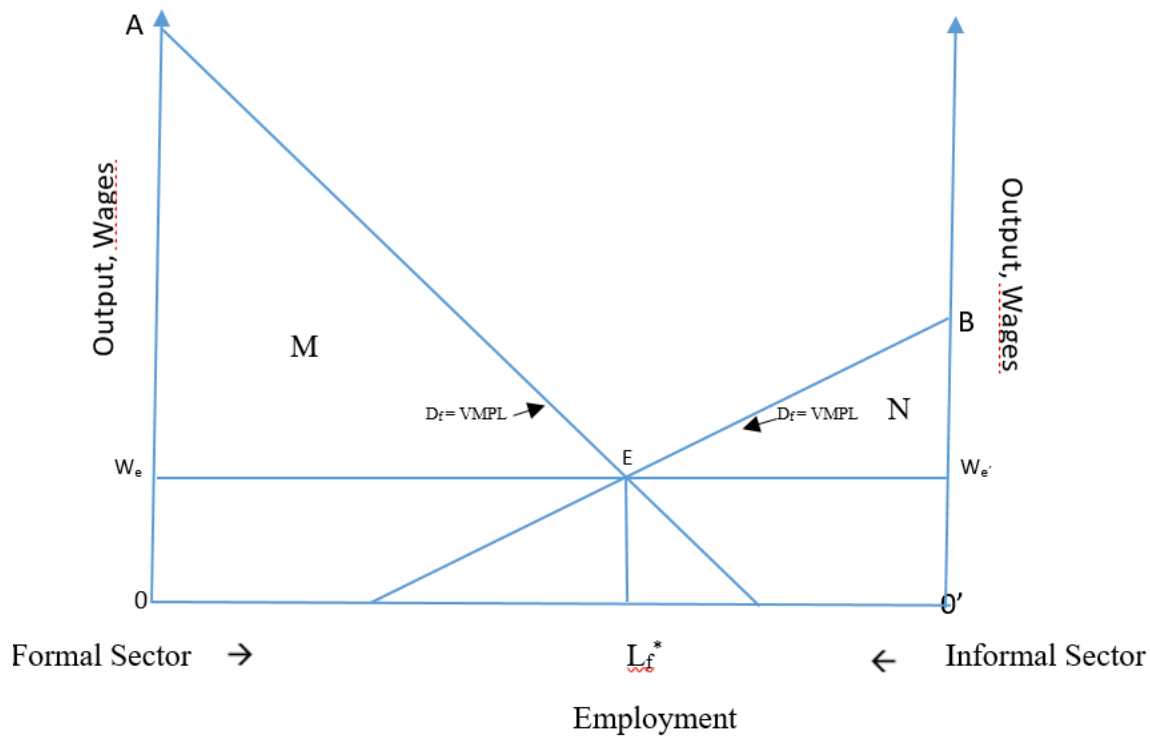
In this case, though it does not yet carry any policy meaning, we make a distinction between larger higher-technology firms and the typically small ones that will more meaningfully be labelled “informal” when taxes and other factors that contribute to market imperfections are introduced into the analysis. In this unrealistically simple economy, 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 preference on their part for either sector, the wage must be the same in both for there to be a stable equilibrium with no further movement of workers.⁵ 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 marginal product of labour (VMPL) curves, $VMPL_f$ and $VMPL_i$ respectively.⁶

⁴ The assumption of homogenous labour is not necessary to the analysis. Each type of labour which is found in both a formal and an informal sector can be handled separately in the same way as illustrated here.

⁵ 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 reasons 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.

⁶ The value of the marginal product of labour is the marginal product of labour (number of units added to total production by the last worker hired) times the price of the good or service being produced.

Figure 1



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 (OL_{f^*} to the formal sector and $L_{f^*}O'$ to the informal sector). Total output of the economy, assuming nothing can be produced without labour, is the area under the two VMPL curves, $OAEBO'$. The wage bill is the rectangle OW_eW_eO' ; the rest of the area under the VMPL curves (area M plus area N) therefore accrues 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 identical in all respects⁷; a tax is imposed on formal sector workers

The simplest way 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 to the firms, or wages gross of taxes (here the $VMPL_f$ curve), and workers' benefits (the wages received) based on the formal sector labour demand curve facing the workers, D_f , which lies below the labour cost curve $VMPL_f$ by the fixed (for simplicity) percentage defined by the tax. The new labour market equilibrium occurs at point C. The triangle CGE gives the deadweight loss associated with the non-optimal labour allocation between the two sectors (OL'_f to the formal sector and L'_fO' to the informal sector). The tax has increased informal sector employment by $L'_fL_{f^*}$ workers. It has also lowered the equilibrium wage from W_e to W_c , so that workers' total earnings are lowered by the presence

⁷ 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(s) the tax can be collected from some workers but not from others, plausibly because it is easier to tax larger, more visible firms.

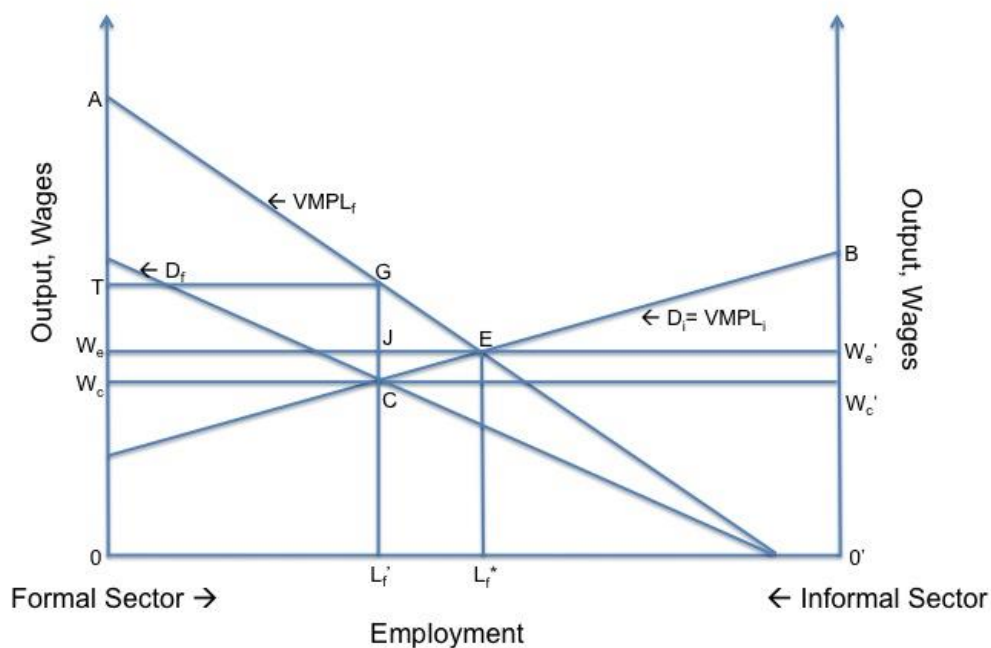
of the tax (by the amount $W_e W_e \cdot W_c \cdot W_c$). In the absence of any shift of capital between the two sectors, formal sector capital⁸ loses by $TGEW_e$ ⁹ while government collects $TGCW_c$ in taxes. Informal sector capital gains by $CEW_e \cdot W_c$; the overall impact on the earnings of capital is theoretically indeterminate; it depends on the elasticities of the two VMPL curves.¹⁰ In any case the net (or post tax and expenditure) impacts on the welfare of capital-owners 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 and the costs of the tax, which take the form of (a) direct loss of disposable income 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 just-identified labour misallocation effect, but since that effect 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 tax-related costs.

⁸ For simplicity “capital” is assumed here to represent all non-labour factors involved in production.

⁹ As its income falls from the previous level AEW_e to the new level AGT .

¹⁰ Allowing for mobility of capital between the two sectors, there will now be a shift of that factor from the formal to the informal sector, which will raise the $VMPL_i$ curve and lower the $VMPL_f$ curve, leading to a further shift of employment to the informal sector. The impact of this shift on the deadweight loss is not predictable theoretically, i.e. it requires the details of the two production functions in order to calculate it (as per the “Theory of the Second Best” [Lipsey and Lancaster, 1956-57], whereby the direction of the impact on deadweight loss from creating or putting an end to one distortion (here the different levels of the marginal product of capital between the two sectors) in the continuing presence of another distortion cannot be predicted without a detailed analysis based on the two production functions. Except for very exotic cases, however, it can be presumed that the termination of the second imperfection will not totally erase the deadweight loss from the first one. In practice, since the capital market functions less well than the labour market, 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. Thus, the qualitative direction of the effects of the labour market distortion will not be reversed, and a first analysis--without taking account of the second round shift in capital between the two sectors, is likely to give a reasonable feel for the impact of the labour tax.

Figure 2a



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, though 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. As for Case 2a but a SS contribution with associated worker benefits is imposed on formal sector workers; workers identical in all respects; the government does not subsidize the SS system

Value of the benefits package to workers is equal to that of the foregone wages¹¹

¹¹ For simplicity we assume the equality in question refers to the total cost of the SS payments and the total benefits to the worker, as would follow, for example, if the benefits package came in natural units and the amount involved were one unit. More generally, for the SS payments and associated benefits to be optimal, the criterion would be that the

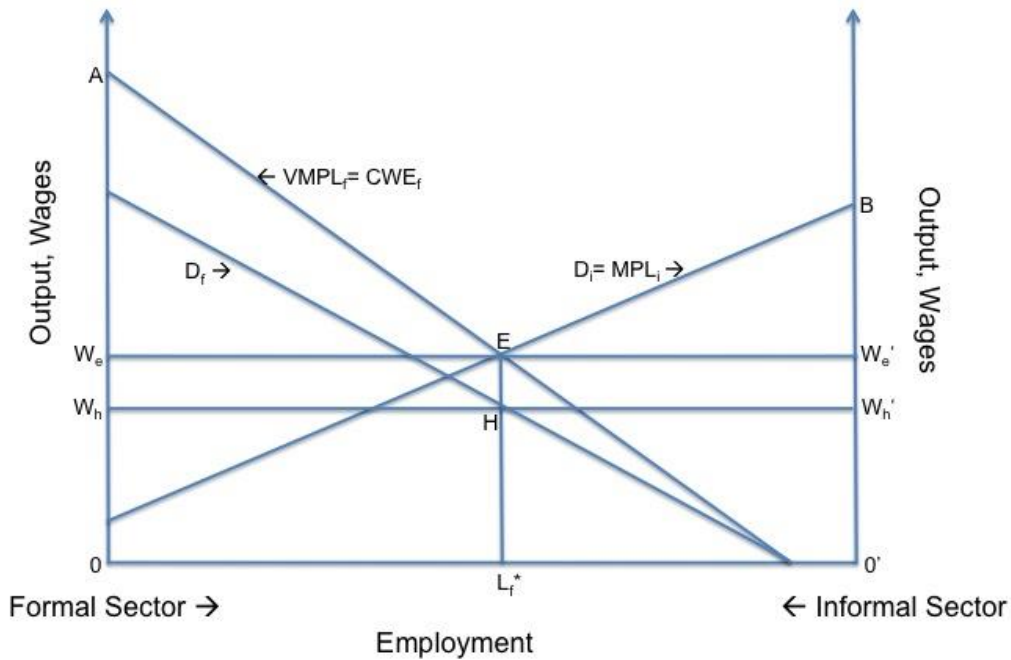
When instead of a pure tax (collected and used to provide benefits that are not specifically directed to those who pay the tax) there is a SS contribution by the formal sector worker which basically buys him/her a set of services,¹² there may be no labour misallocation (or any other loss) involved. This 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 he/she makes (i.e. the welfare lost through having to pay them).¹³ Not all of the labour costs borne by the firm are paid out as current wages to the workers but their current wage equivalent (CWE) to the workers is the same as if they were all paid out currently. (In other words, the worker places the same value on the combination of wages now and deferred benefits in the future as he/she would place on the full labour cost to the firm if it were all paid out as current wages.) Under these conditions, 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; 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_h , as workers must be paid less in the formal sector to prevent more of them from shifting into it given the SS benefits.

marginal cost to the worker of an additional unit of SS payments was equal to the marginal benefits from the additional wages given up to get that additional amount of SS benefits.

¹² 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.

¹³ 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 inefficiency (or 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 key point is that the overall benefits from the formal sector labour contract are just equal to the value to the worker of the full costs the firm bears in hiring him/her.

Figure 2b

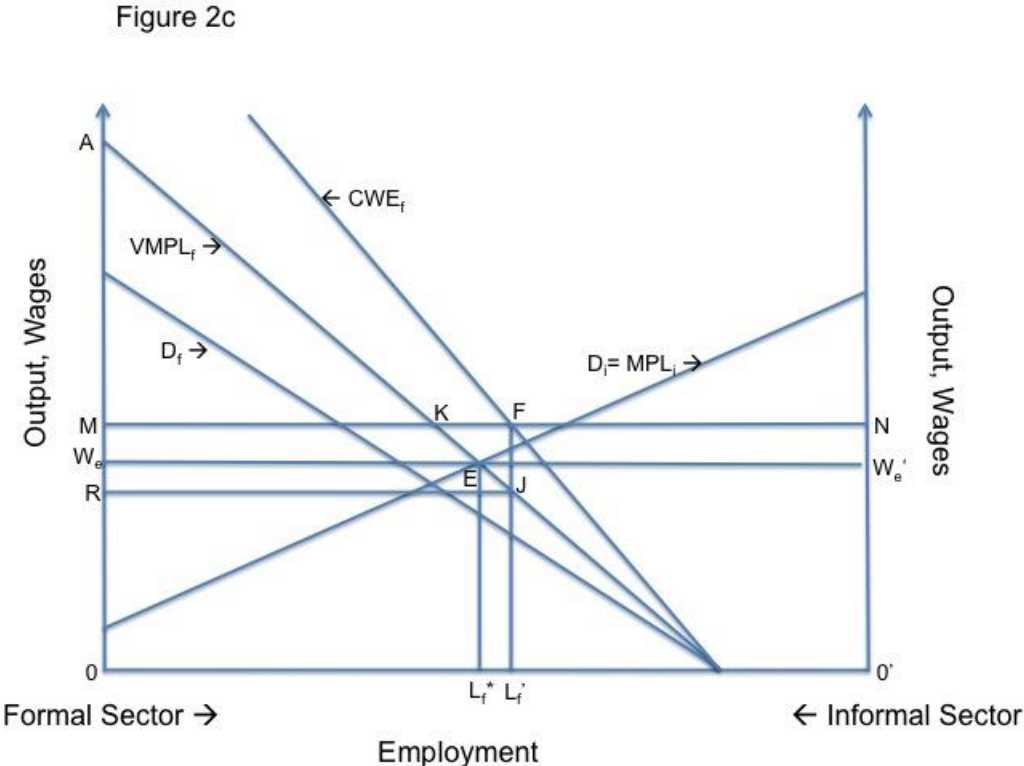


Value of the benefits package to the workers differs from that of the foregone wages, with no one else affected positively or negatively.

The SS system is designed to *provide net benefits to the workers* (though of course in practice it may turn out not to do so), 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 those benefits to him/her, $P=1$ describes the person who values those benefits just at their cost to him/her (as in the previous case), $P>1$ corresponds to someone to whom the package provides net benefits above cost and $P<1$ to 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 a positive 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—the firms' offer price for labour (given by the $VMPL_f$ curve) and the marginal worker's perceived benefits, shown in Figure 2c as the CWE_f curve, which in this case lies above the $VMPL_f$ curve.

Workers are now indifferent between employment in the two sectors when the informal sector wage (W_i) is equal to the formal sector wage equivalent CWE_f , a condition satisfied where curves CWE_f and $VMPL_i$ intersect (point F) so the equilibrium allocation of labour involves more workers than before in the formal

sector ($OL_{f'}$). 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.¹⁴ 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 $L_f = L_{f'}$, area MFJR would be the net benefits to the economy if there were no loss of output brought on by the shift of some workers to the formal sector; FEJ (that loss) must be subtracted to give the overall net impact on society.¹⁵



Labour is optimally allocated from the perspective of maximization of societal welfare (defined as the sum of income or GDP plus net SS benefits), even though GDP itself (as normally measured) falls. The income of capital falls by FNW_eE in the informal sector and rises by W_eEJR in the formal sector; the total rises as long as the elasticity of $VMPL_f$ in the relevant range EJ is less than that of $VMPL_i$ in the relevant range EF.

¹⁴ For simplicity, we adopt 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.

¹⁵ Another way to see who is affected in what way is to note that, overall, workers are better off by rectangle $MN W_e' W_e$; each worker's welfare is raised by amount MW_e 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 W_eEJR 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.

The amount of labour reallocation and the levels of worker welfare, labour costs and non-labour incomes all depend on the elasticities of the $VMPL_f$ and $VMPL_i$ curves in the relevant ranges EJ and EF. The extent of labour reallocation is positively related to the elasticity of both VMPL curves. 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 VMPL curves. If the elasticity is lower 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 rather than a decrease in that of the formal sector.¹⁶ If, as often assumed, the elasticity of labour demand 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--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 full amount of the net benefits gap, the wage will remain constant in the informal sector, now being 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 essentially corresponds in one way to a prominent variant of the labour surplus model.¹⁷

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. As noted, 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.¹⁸

It is also possible, however, that the bundle of services be badly designed and thus 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.¹⁹ In this case, the above analysis is essentially reversed in all of its details. The existence of the SS package leads to a shifting of workers to the informal sector (Figure 2d);

¹⁶ Total labour costs in the economy are made up of the sum of per worker labour cost in each sector times the number of workers in that sector. Total wage costs may also rise or fall, depending on the relative wage movements in each sector and the relative amount of employment found in each.

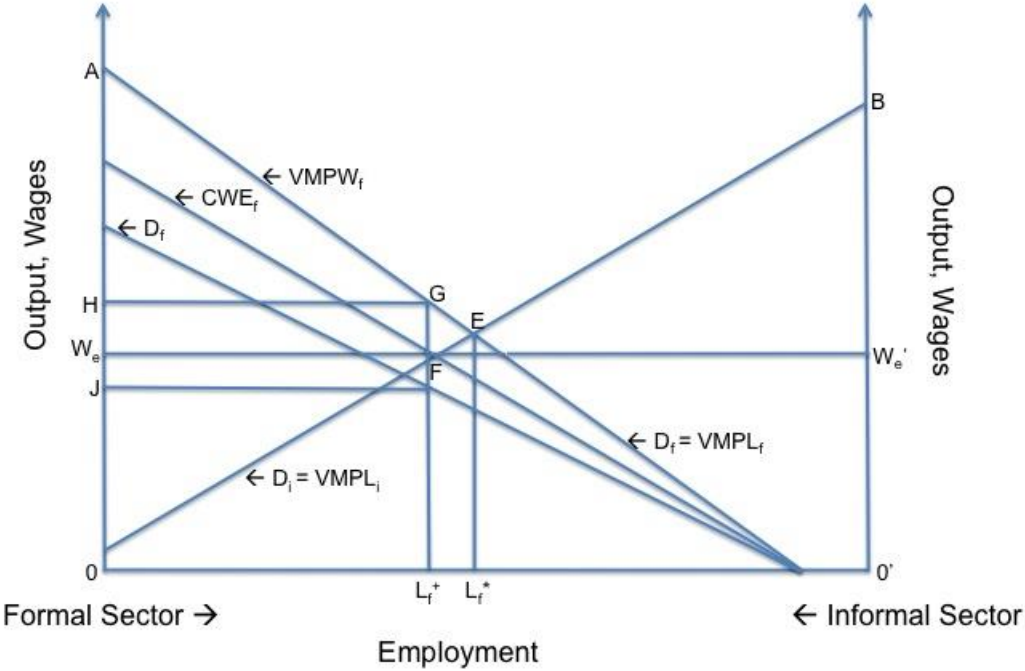
¹⁷ In that model the supply of labour from the traditional to the modern sector (which may at least in part correspond to the informal-formal sector delineation) is perfectly elastic. However, unlike the case under discussion here, the labour supply price does not reflect its value of marginal product, which instead is assumed to be zero. Accordingly, the labour surplus model's implication that the VMPL in the informal sector is below its labour supply price is not built into the discussion here. Below, the possible existence of monopolistic competition and other possible reasons for these two values to diverge is considered.

¹⁸ 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.

¹⁹ See the discussions in, for example, Perry et al (2007), Levy (2008).

the new equilibrium (point F) implies a formal sector employment level of $OL_{f'}$, less than the OL_{f^*} of the benchmark case. 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.

Figure 2d



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; it appears improbable that the elasticities of demand in the two sectors would be such as to produce the opposite result. 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 VMPL curves, which also determine the extent of labour misallocation.²⁰ The likelihood that the SS inefficiency

²⁰ A very low elasticity of demand in either sector guarantees a low misallocation loss. When one sector's labour demand elasticity is low (at the limit zero) this loss rises with the elasticity of the other sector's labour demand. When the elasticity of demand by the informal sector is infinite, formal sector employment will fall to zero if the tax is high relative to formal sector labour demand elasticity. As the elasticity of demand by the informal sector falls, the misallocation loss rises as long as there is no formal sector employment. When the informal sector demand elasticity reaches a certain level, formal sector employment becomes positive and then rises with further decreases of informal

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 SS inefficiency loss is that gap applied to (multiplied by) the number of 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 VMPL curves are linear). Thus if the labour misallocation were to involve 5% of the labour force (probably toward the high end of the plausible range) and formal sector employment were 20% of all employment (a relatively low figure) the ratio of SS inefficiency loss to labour misallocation loss would be 8:1.²¹

How the deadweight loss due to labour misallocation is divided between labour and other factors also depends on the labour demand elasticities and the gap between the CWE and VMPL_f curves. 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 inversely related to the elasticity of each of the labour demand curves. Once again, one factor (here capital) can gain even though society as a whole has lost.

Total loss in this benchmark case is a monotonically increasing function to the degree of inefficiency (as measured by 1-P) and of the level of SS, as measured by the ratio of the SS payment to total labour costs. The CWE_f curve lies below the VMPL_f curve by SS (1-P) percent. The lower is P the greater the loss, SS held constant, and the greater SS the greater the loss, P held constant. Thus, when the SS ratio is high it is especially important that the system be efficient or, looked at from the other side, where inefficiency is destined to be high, it is important that the SS ratio be low.

Policy Implication

When the SS bundle is not optimally constructed, its weaknesses may transform what could have been net benefits into net losses, and in so doing may give rise to a resource misallocation loss on top of the initial SS inefficiency loss. The piggybacking of the second loss on the first gives added reason to take reform of an ineffective SS system very seriously.

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 a lesser amount several different factors might underlie that fact. The simplest case, discussed above, is that in which although all of the resources collected to finance the bundle are used up to provide it to the worker (group

sector demand elasticity (tending to cut the misallocation loss). The height of the loss triangle falls as formal sector employment rises, since with a fixed ratio between CWE and VMPL_f that gap decreases with the level of formal sector employment. Both determinants of the misallocation loss thus create a tendency towards decline in the misallocation loss as informal labour demand elasticity falls. When the formal sector VMPL curve is perfectly elastic, the above pattern is played out in reverse, in the sense that loss rises as informal sector demand elasticity falls from infinity and until formal sector activity begins. Then that loss falls monotonically as formal sector activity rises. When the two demand curves take on more complicated forms, these simple quadratic relationships do not hold.

²¹ The fact that the SS loss per person is one of the determinants of the labour misallocation loss makes it less likely *per se* that the labour misallocation loss could be very large in comparison with the inefficient SS loss, but working against this is the fact that a large misallocation loss cuts formal sector employment and thereby lowers the value of one of the determinants of the SS inefficiency loss (formal sector employment).

of workers) and the worker completely “uses up” the services provided, 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) are then redirected to someone else.²² In this case, the worker’s evaluation of the services, while accurate for himself/herself, does not correspond to their total societal value. Once again, the worker’s perception determines the allocation of labour, but societal loss is now less than in the previous case. At the limit, there might be no loss through bad service design (bad 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 resulting labour misallocation. Once again, GDP would not be the appropriate criterion by which to measure overall economic performance.

A distinction is thus needed according to whether components of the SS bundle of benefits that do not get to the workers in question, (either through their own preference or for other reasons) then go to waste (deadweight loss) or somehow become available to someone else. The other person(s) may be another worker paying for SS or someone else. 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 others (benefits thus being greater than expected by that person), 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 only worker preferences does not provide the whole story, since it is necessary to distinguish cases where the services are not of value to the worker for whom they are designed and those where they are not of value to anyone, so that the resources used for them are wasted from an overall societal point of view.

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 formal sector 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 beside the worker, as just discussed. We refer to the curve that measures this sum as the marginal social benefit from the addition of one more worker in the formal sector (MSB_f) curve.

To recapitulate, as long as no SS-related benefits accrue to others than the workers, total societal welfare will rise or fall according to whether their evaluation of benefits accruing to themselves (assumed to be accurate) is above or below the resources extracted from them. In either case, conventionally measured GDP will fall since worker allocation no longer maximizes GDP but, as noted above, this loss will be outweighed by direct welfare gains from SS in the case where workers value the package more than the wages foregone. Several additional benchmark cases are now worthy of attention

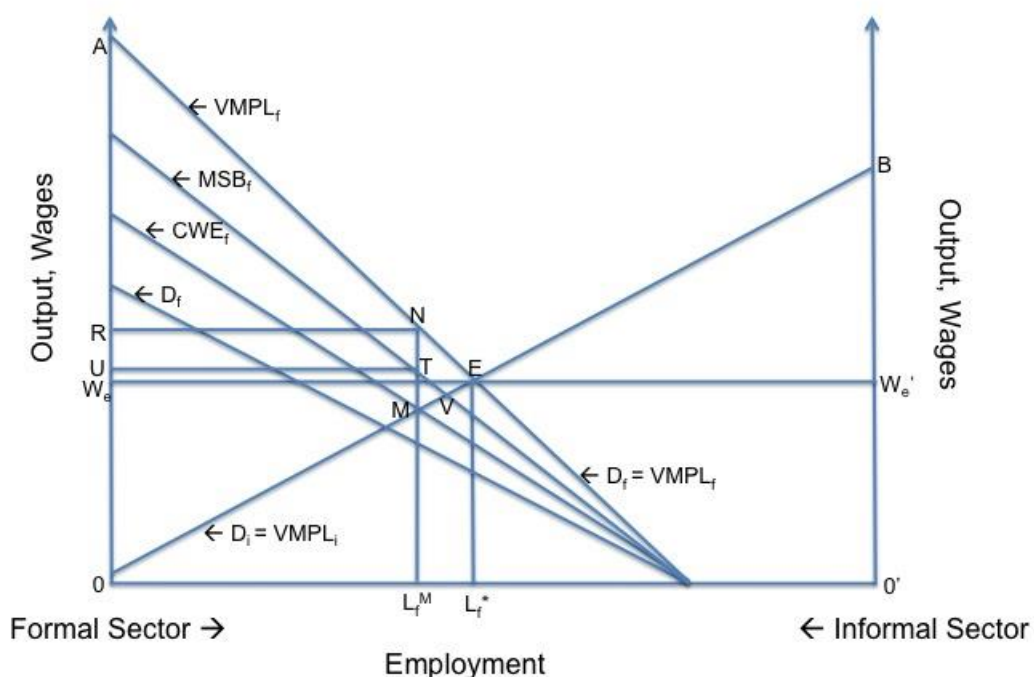
²² In the extreme, they may be stolen by operatives of the SS system or by others, an unfortunate form of redirection, but still creating value for someone.

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) some category of SS affiliated workers but perhaps not the lower income ones, who are often thought 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, the differential between the two values does not constitute a social loss (presumably because services of that value accrue to someone else). In this case the socially optimal allocation of labour would be where $VMPL_i$ equals $VMPL_f$, e.g. the no-intervention equilibrium illustrated in Figure 1. But since the SS package is not worth the foregone wages to the workers, there is a shift of labour towards the informal sector, leading to a deadweight loss caused by the fact that the private gains of the decision-making workers are not fully aligned with societal gains.

Figure 2e illustrates a perhaps more likely case, in which the shortfall between the costs of SS to the worker and the benefits she receives is only partly counterbalanced by net gains to other members of the society. The CWE_f curve lies well below the $VMPL_f$ curve, reflecting her much lower valuation of the SS package than of the wages foregone to acquire it; the gap between CWE_f and the marginal social benefits curve (MSB_f) measures the benefits to other members of society from her involvement in the SS system. The gap between that curve and $VMPL_f$ then measures the marginal deadweight loss to society from the overall inefficiency of the SS system. Since, as always, it is the level of benefits received directly by formal sector workers that determines the allocation of workers between the two sectors, the equilibrium in this case occurs at point M, with L_M 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 of labour, taking the SS system as given, is defined by point V, where the MSB_f curve cuts the $VMPL_i$ curve, and with that allocation societal loss would then have been lower by area MTV than in fact it is.

Figure 2e



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 is 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); whether overall a gain or a loss occurs depends 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.²³ 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 that just discussed, since in terms of the impact on labour allocation between the sectors 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)

²³ Access to severance payments has often provided former paid workers with the funds to start up of a small firm (Cortes et al, 1987). In some cases the worker would have been able to save up on his/her own, but presumably not in all.

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 actual benefits to the worker. This latter variable (not shown) 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*

Some SS systems receive public sector subsidies in the financing of the bundles of benefits they provide. 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 may either increase or decrease the extent of misallocation of labour between the two sectors and may also be expected to encourage inefficiency (or corruption) within the SS system. Absent the latter effects, a subsidy in a situation where in its absence the worker has $P < 1$ because she (correctly) values the SS at less than the foregone wages (i.e. $CWE > \text{current wages} + \text{SS benefits}$) can raise societal welfare by offsetting the distorting effects of the low value of SS benefits on workers' decision (not to work in the formal sector), but if its is too large or the cost of raising the taxes to pay for it is too high or its height has a direct negative effect on the efficiency of the SS system then the opposite will be true. When workers underestimate the SS benefits, either to them or to others, the subsidy is more likely to offset this 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 SS contribution with associated worker benefits on formal sector workers; valuation of the bundle of SS benefits varies across formal sector workers; workers are identical in all other respects;²⁴ no government subsidies to the system

Outcomes are altered in a more complicated way when the valuation of the bundle of SS benefits varies across people (who are still assumed to be identical in terms of productivity).²⁵ 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 the distribution of this variable across workers.

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. When P varies, 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) strictly according to the worker's valuation of the SS package.

²⁴ Another important form of worker heterogeneity is in terms of productivity. In the absence of a MW or some other imperfection, a SS tax imposed on one sector but not the other, in the presence of workers who are heterogeneous in productivity, 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 however, a MW is imposed, the presence of a SS system does matter, as analysed in the next section.

²⁵ Allowing for productivity differences across people complicates the discussion, it now being convenient to think in terms of labour defined in productivity units (see below).

Consider first the plausible case where workers are self-selected into formal employment.²⁶ Note that, 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 $P \geq 1$, labour allocation between the two sectors will be unaltered from the no-SS case, but societal welfare will have increased since all workers in the formal sector (except for 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 their P into the formal sector than if any other selection mechanism is used.

To further illustrate the factors at work here (see Figure 2f) 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 $P=1$ 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.

The equilibrium allocation of labour between the two sectors of this economy involves the condition $W_{Df} = W_{Sf}$, where W_{Df} is the wage at which a given quantity of labour L_f is demanded in the formal sector and W_{Sf} is the wage at which a given quantity L_f is supplied to that sector. Under the assumptions used here that equilibrium condition is

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

where $W_{Sf}(L_j) = VMPL_i(L_j) - (1 - f_1^{L_j}/L_i)$ (SSR) $VMPL_i(L_j)$,

SSR is the ratio of the social security contribution to the basic formal sector wage, (1.0 in the case illustrated here), and

$f_1^{L_j}/L_i$ is the share of workers in the formal sector when worker j joins that sector.

The demand curve D_f of Figure 2f has the same interpretation as in earlier cases, but the supply curve (S_f) of workers to the formal sector is the arc CB, coinciding with the $VMPL_i$ curve at point B but lying below it at all other points. Its position can be deduced as follows. The supply price for any given worker (the base wage he will require to offer his services to the formal sector) is what he could earn in the informal sector (the height of the $VMPL_i$ curve) minus the net gain to him from involvement in the SS system. In the present case, that net gain gradually falls from the full SS contribution (here assumed to be equal to the base wage) for the first person hired away from the informal sector to zero for the last one hired away from that sector (since for that worker the SS contribution is assumed to bring no benefits). The first worker will offer his services to the formal sector at half of what he earns in the informal sector since the SS received (being equal in value to the pay deduction) is worth an amount equal to the base wage. For the last worker, the pay deduction is a pure tax, so he will require a wage equal to his potential informal sector earnings to induce him to shift into the formal sector. (Under the present assumptions, the arc CB would be one quarter below point E directly under that point, since for the worker found there $P=0.5$.)²⁷ The equilibrium, defined by the intersection of D_f and the arc CB occurs at point R, with formal sector employment of L_r and wage of W_r . Under the specific assumptions made here point R will involve less

²⁶ This assumption is plausible because as long as workers can freely choose their sector of employment, self-selection will occur.

²⁷ Put another way, workers moving to the formal sector get a formal sector base wage plus SS benefits. 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 $VMPL_i$ curve (on the rightward axis) at point B.

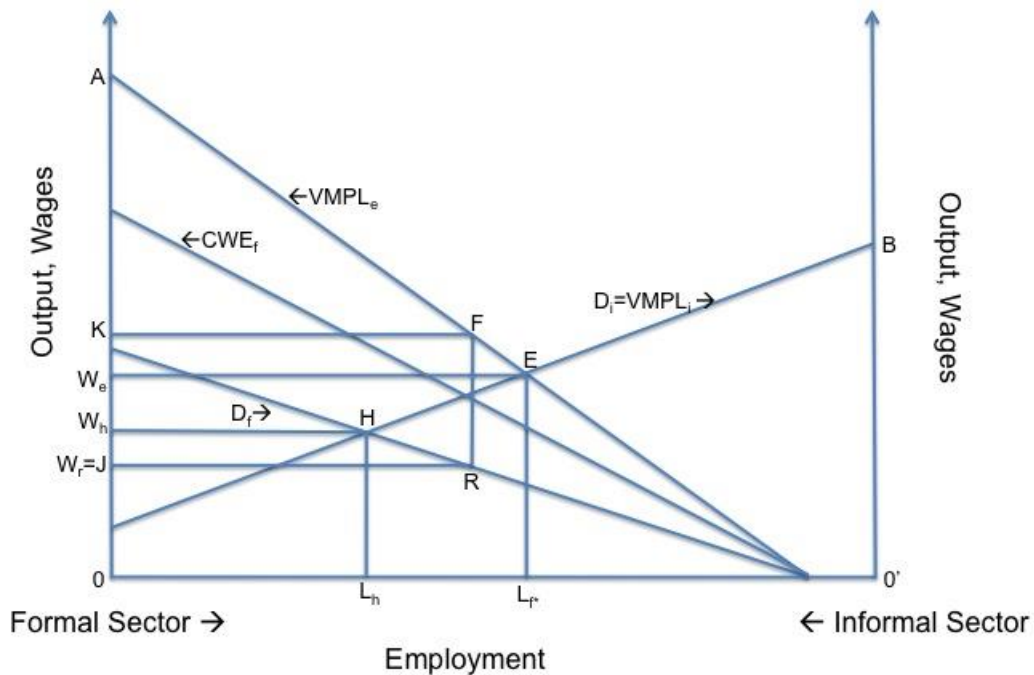
employment in the formal sector than L_{f^*} , the optimal level in the absence of the SS system. The deadweight output loss is EFG and the SS inefficiency loss is something under one quarter of the area FRJK; area FRJK is the total SS deduction imposed on the formal sector workers; the average share lost would be one quarter if formal sector employment were L_f but since it is somewhat less than that the average loss is less than a quarter of the cited area.²⁸

For no misallocation loss to occur in this specific case of worker self-selection it is necessary that half of workers have $P \geq 1$. There is then no SS inefficiency loss either, since only workers with $P \geq 1$ are found in the formal sector. As long as $P > 1$ for some of these workers the SS system will produce a net welfare gain. If more than half of workers have $P > 1$ the labour supply curve to the formal sector, arc CB, will intersect the demand curve D_f at an employment level greater than L_{f^*} so there will be a misallocation loss from too many workers in the formal sector, but it will be partially, wholly or more than wholly offset by the gains from the existence of the SS system (and certainly more than offset if the misallocation loss is small, for example if formal sector employment is only a little above L_{f^*}). As the distribution of P becomes less favourable (shifts down), the SS inefficiency loss naturally rises, and so does the misallocation loss.

If selection to enter the formal sector is at random (with worker values of P , as above, distributed equally over the range 1 to 0), 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 at random in the formal sector opt to work there, as the first workers to join that sector will do since the wage is high. The CWE_f curve (here representing the CWE of the randomly chosen worker, i.e. a worker with $P = 0.5$) will be as shown in Figure 2g, halfway between D_f and $VMPL_f$. However, for formal sector employment levels of above L_h , for which the formal sector offers a base wage less than W_h , the earnings attainable for that worker in the informal sector, those workers with $P = 0$ will opt out of formal sector employment. As more workers enter the formal sector and the base wage goes down farther, a higher and higher share will opt out, i.e. although selected randomly to be offered jobs there they will elect not to accept them. When enough workers are in the formal sector to push the wage down to W_R , the level corresponding to the self-selection case illustrated in Figure 2f, then only those with P above the level required to keep them voluntarily in the formal sector will be there. This is the same group as are found in the formal sector under self-selection. In this specific case, therefore, random selection leads to the same allocation of workers as does self-selection.

²⁸ More precisely, loss would be $0.25 (L_R/L_{f^*})$ FRJK.

Figure 2g



In a situation where worker preference for the SS system varies it makes special sense for their participation in it to be optional. In fact, more and more countries are allowing or encouraging the purchase of private pensions, etc. This possibility also should in principle be included in the analysis but we do not do so here. Results would undoubtedly be further complicated since neither the performance of such systems nor peoples aptitude for choosing optimally are very clear at this time.

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 a degree of self-selection in who winds up in the formal sector. Also relevant is the possibility that not all otherwise formal firms will adopt a SS system.²⁹ As discussed below in Case 6, a more realistic portrayal of an economy in which the issues under discussion play themselves out distinguishes three sectors: clearly formal, clearly informal, and an in-between sector made up largely of intermediate technology small and medium enterprises (SMEs). Application of such social instruments as SS tend to be partial within this last sector (whereas they are virtually complete in the fully formal modern technology sector and nearly absent in the informal sector). The determination of when they are applied and when they are not can then be 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

²⁹ If the presence of SS is the only thing that defines formality, 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.

regulations can be largely 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.³⁰

In broad terms, flexibility in the application of SS regulations can act like worker self-selection into the formal sector, in that it can diminish 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 overlap considerably, 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 is 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 is 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 2f As for Case 2a but a SS contribution with associated worker benefits is imposed on formal sector workers; valuation of the bundle of SS benefits varies across formal sector workers; a sector specific tax is imposed on formal firms.

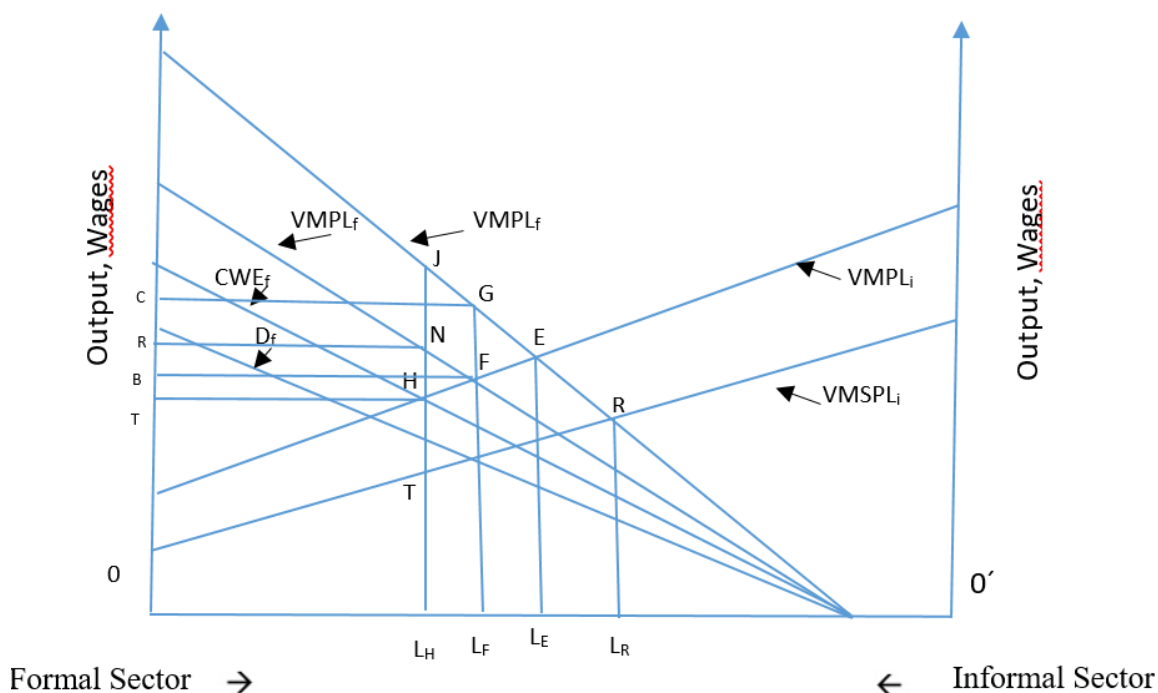
One of the major worries of observers concerned with the informal sector is that, because informal firms typically escape some or all taxes, this constitutes a distortion and encourages excessive informality. By itself, such a tax will indeed lead to sectoral resource misallocation and deadweight loss. Figure 2h portrays a case in which, in the absence of such a tax and of a SS system, equilibrium would occur at point E, where $VMPL_f$ and $VMPL_i$ intersect. If a tax is imposed on formal sector firms—assumed for simplicity here to be a fixed percent of labour costs,³¹ this would lower the labour demand curve to $VMPL_t$, (value of the marginal product of labour minus the tax payable) shift the equilibrium to point F, lower formal sector employment from L_E to L_F and create deadweight loss of GEF. If to this situation is added an inefficient SS system such that CWE_f lies between $VMPL_t$ and D_f , the equilibrium moves to point H, formal sector employment falls to L_H and deadweight loss rises to JEH (sectoral misallocation loss) plus SS inefficiency loss RNHT. The additional loss attributable to the inefficient SS system is greater than it would have been in the absence of the tax on the formal sector.

³⁰ 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.

³¹ In fact, the tax would not be expected to take this form. It could be a sales tax, a value added tax, a profit tax or more likely some combination. It would almost inevitably imply that the VMPL accruing to the firm for a given worker would be reduced by the tax. Thus in the case of a value added tax, the VMPL accruing to the firm would be reduced by the ratio of the tax to valued added, so $VMPL_{ft}$ would lie below $VMPL_f$ by that percent.

general sense, less efficient than the formal one, is reflected in the value of the marginal social product of labour curve ($VMSPL_i$) lying below that of the (unchanged) value of marginal private product of labour curve ($VMPL_i$). Such a situation could result, for example, if monopolistic competition were widespread in the informal sector, such that each unit of sales by one firm lowered the earnings of other competing firms, a case discussed in more detail below. Under these assumptions equilibrium would occur at point H, as in the previous case, whereas the socially optimal outcome would be at point R with formal sector employment of L_R . Total misallocation deadweight loss from the three imperfections is JRT , greater by $HERT$ than in the previous case. Whereas the optimal formal sector employment would be greater than that occurring in the absence of the formal sector tax and the inefficient SS (L_E), the actual employment is less than that.

Figure 2i



The alleged general inefficiency of the informal sector is, however, an empirical question. As discussed in Berry (2011), the belief that the informal sector is less economically efficient than the formal one is often based on a weak data and/or a weak understanding of microeconomics (e.g. a focus on labour productivity rather than on total factor productivity (TFP) as the measure of economic efficiency).³² Two additional points are relevant to this discussion. First, in any consideration of relative efficiency of the two sectors and deadweight loss resulting from too little production in one of them, what matters are not the average characteristics of the two sectors but rather those of the marginal workers (and other factors of production), marginal in the sense that they are the ones whose sector changes as between the options

³² Even those studies that recognize the need to make such comparisons in terms of TFP are not usually able to take account of the fact that, when that measure suggests greater productivity by larger, more formal firms, part of the difference is likely to be due to a higher degree of market power in the former—in other words monopoly power can make a firm look efficient because it can charge a higher price than a firm in a competitive setting would be able to.

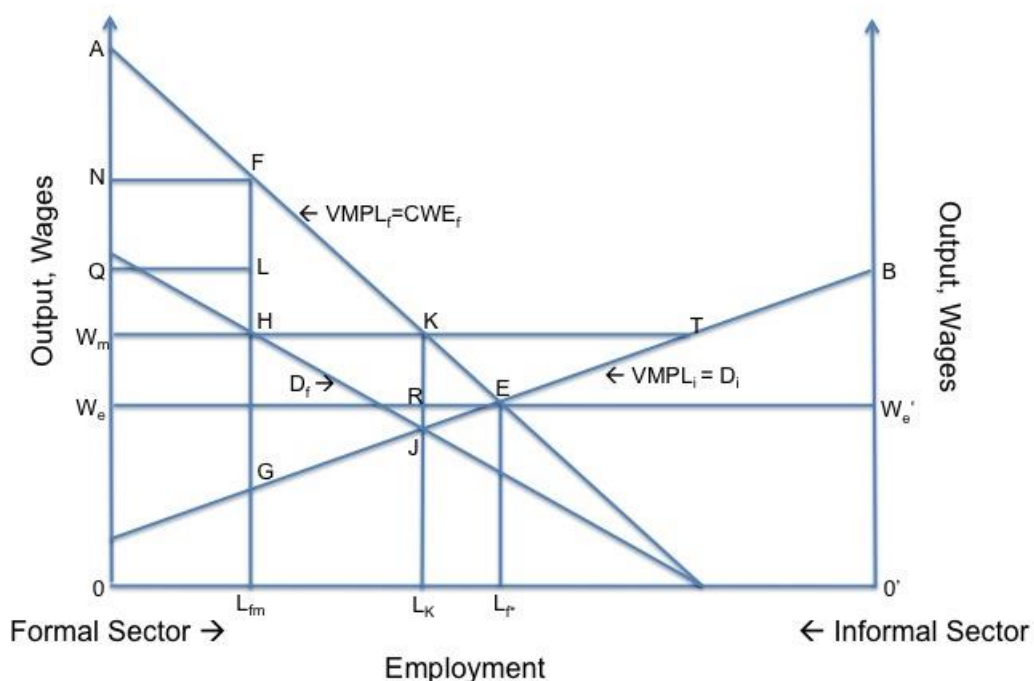
being compared, e.g they would be in the formal sector in the absence of a tax but are in fact in the informal one. Comparisons of average total factor productivity between formal and informal sectors thus tell us little. Second, given that the static deadweight loss resulting from such “distortions” as taxes applied differentially across economic activities tends to be small, their more important impacts may be on growth-related variables like investment rather than on short-term resource allocation (to which deadweight loss estimates pertain). Thus, it is possible that the main negative impact of such taxes is on investment; in that case the final impact could be much larger than implied by the static deadweight loss estimates.

Case 3a A Minimum Wage (MW) above the Equilibrium Wage is Applied to the Formal Sector in the Presence of SS; other assumptions as for Case 2b, including homogenous workers (in terms of preference for SS)

Another frequent real world constraint on the free functioning of the labour market is the existence of a minimum wage (MW), applied *de facto* to the formal sector (e.g. W_m of Figure 3) so as to push the wage above equilibrium; this creates a gap vis a vis the informal sector wage (W_e). In the absence of a SS system this moves the equilibrium to point K and shrinks formal employment (to L_k) with deadweight loss of KEJ—the standard microeconomic prediction for the effect of a MW; there is a fall in informal sector wage (by RJ) and a rise in formal sector one (by RK). The deadweight loss is a monotonically increasing function of the MW until it has risen to the level at which there is no formal sector employment left.³³

³³ 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 from an empirical perspective (e.g. Heckman and Pages, 2004).

Figure 3



For present purposes, the interesting aspect of the MW is the way losses from its presence (and any possible positive or negative redistribution effects) are related to the presence of SS. Given the partial parallel in the effects of an inefficient SS system and those of a MW (both curtail formal sector employment), two points of contrast should be noted. A MW by itself creates deadweight loss exclusively through the misallocation of labour between the two sectors and the associated output decline that results from it. An inefficient SS system, on the other hand, leads to a welfare decline made up of an output decline related to labour misallocation *and* an inefficiency loss within the operation of the SS system.³⁴ However, as long as workers self-select between sectors, a SS system does not lead to queuing. When a MW is imposed, with or without a SS system, queuing occurs since for any given worker there is an earnings gap between the two sectors. As discussed further below, the presence of queuing complicates the analysis of outcomes and creates the possibility of greater loss.

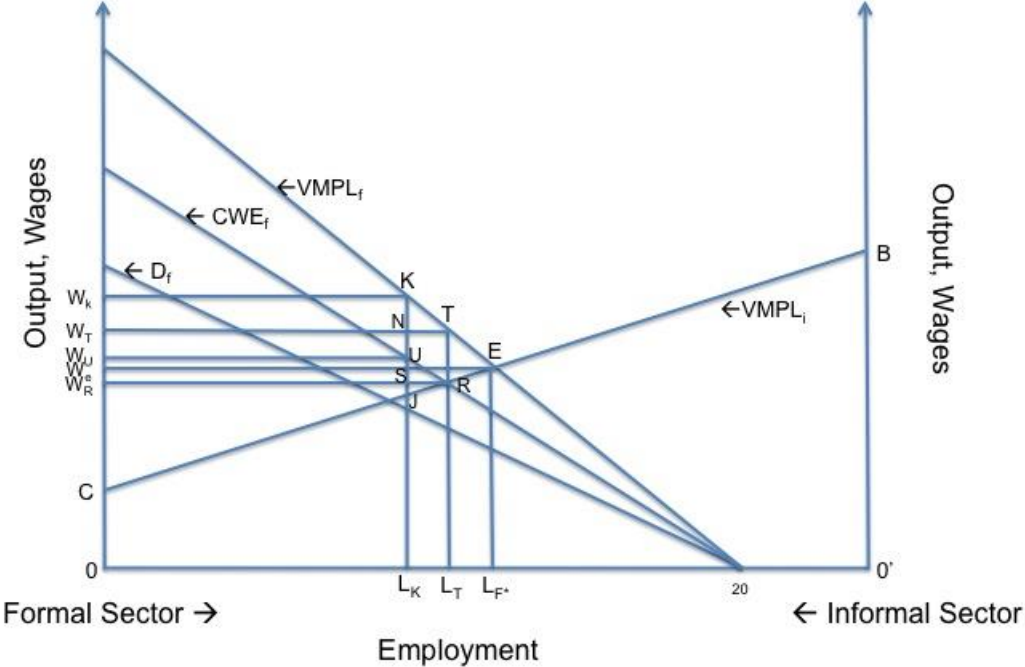
The impact of the presence of an SS system on the welfare effects of a MW depends both on how well that system works and on whether and how its presence affects the choice of the MW to be applied. If there is a tendency to raise the total labour costs to a higher level when a SS system is present (e.g. through a perception that total benefits from work are less than they really are, or because this is one tool selected with the intent of raising total labour benefits—wages plus other components of those benefits) then loss

³⁴ Note that the deadweight loss from a MW that pushes labour costs to any given level is much less than that of an inefficient SS that pushes it to the same level, since the latter incorporates the same misallocation loss plus a substantial SS inefficiency loss.

will be greater; if the opposite pattern holds it will be less. In the presence of SS, the floor placed under wages could in principle refer to total labour costs (including the SS costs), the base wage, the current wage equivalent for the workers (although regulators do not observe this value) or something else.

The simplest case is that in which labour cost is the variable to which a floor is applied. Under this assumption, two limiting cases should be noted. First, when a MW of whatever magnitude is imposed in the presence of an (efficient) SS system, the outcome is unaffected by that SS system. The new minimum level of $VMPL_f$ is the same as it would be with the MW alone.³⁵ Second, levels of MW that do not push total labour costs to the firms above those already resulting from the SS system, have no impact. For any given level of an inefficient SS, with its effect of reducing formal sector employment and raising labour costs, there is a MW (referring still to total labour costs) that leads to the same formal sector employment and deadweight loss, i.e. the MW adds no loss to what has already been caused by the inefficient SS system. Suppose, as illustrated in Figure 3a, that $P=0.5$ for all workers; as a result, curve CWE_f lies half way between curves D_f and $VMPL_f$, and the level of formal sector employment falls from L_{f^*} (the equilibrium level in the absence of the SS system) to L_K . Deadweight loss includes that resulting from labour misallocation between the two sectors (TER) and that due to the inefficiency of the SS system ($W_T TRW_R$). If a MW is applied at or below W_T , (the level to which the SS system has pushed total labour costs to the firms), it has no additional impact on the outcome.

Figure 3a



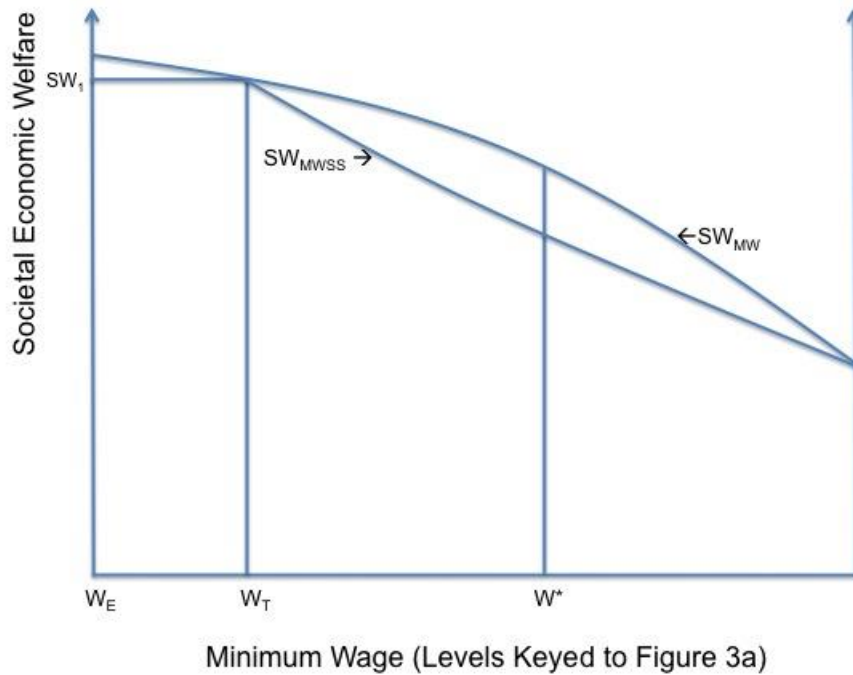
³⁵ As noted above, the presence of a SS system entailing no inefficiency loss (such that $VMPL_f = CEW_f$, as in Figure 3) would not by itself lead to any deadweight loss since the equilibrium would be defined by the intersection of curves CEW_f and $VMPL_i$.

The more interesting (and likely) case is that in which the MW (expressed in terms of total labour costs to the firm) is set above W_T , say at W_K . Its presence then increases the deadweight loss due to sectoral misallocation to KEJ (from TER) and also changes loss due to the inefficiency of the SS system, from $W_T TR W_R$ to $W_K K W U$. Whether this further shrinkage of the formal sector due to the MW raises or lowers SS inefficiency loss depends on the price elasticity of the curves in question.³⁶ The SS inefficiency loss rises on a per worker basis as the wage rises but the number of workers falls, so the total loss falls as long as the portion of the curves involved have a wage (price) elasticity greater than unity. This elasticity falls from high levels to unity at the mid-point of the linear $VMPL_f$ curve and thence to lower levels. Thus, in cases where the initial no-MW no-SS equilibrium occurs within the price elastic part of the labour demand curve, the total SS inefficiency loss falls as formal employment falls. The total deadweight loss (from misallocation of labour *and* SS inefficiency) does, however, rise with the MW, as may be illustrated in Figure 3a. As MW rises from W_T to W_K , the increased sectoral misallocation loss (KTRJ) is partially but not wholly offset by the fact that area NTRS is no longer part of the inefficient SS loss, but since SS inefficiency loss per worker is now greater for the L_K workers still in the sector, there is an unequivocal increase in total loss. This result is general regardless of the level of SS inefficiency.

Knowing that a MW which pushes total labour costs above the level to which an inefficient SS system had raised them adds to total deadweight loss leaves open the interesting question of whether the marginal deadweight cost of a rise in the MW is greater or smaller in the presence of such a SS system than in its absence. The deadweight loss from the two interventions combined is always the sum of the total misallocation deadweight loss (e.g. KEJ of Figure 3a when the MW is set at W_K) plus the SS-inefficiency loss of $W_K K U W_U$ under that same condition). As the minimum wage rises total misallocation deadweight loss rises at an increasing rate and eventually incorporates the whole area BEC, at which point there is no SS-inefficiency loss and the total loss from a MW reaches its maximum; at the other extreme, when the MW is equal to or less than W_T , buy itself it causes no deadweight loss. Thus, as MW rises from W_T to B its marginal welfare cost is less in the presence of SS than in its absence. In terms of Figure 3b, as the MW (imposed in the absence of SS) rises, the societal welfare level with imposition only of the MW (given by curve SW_{mw}) falls at an increasing rate, i.e. the marginal deadweight cost rises with the level of MW. When MW is imposed in the presence of a SS system which has already raised labour cost to the level W_T , increases in the MW up to the level W_T have no impact on societal welfare, which is at level SW_1 . Further increases in MW do lower societal welfare so curve SW_{mwss} (societal level in the presence of a SS system and a minimum wage) is also monotonically decreasing in MW. The gap between these two curves is the additional loss due to SS inefficiency. As noted above, this loss rises with labour costs as long as the wage elasticity of labour demand is below unity and falls when it is above unity. Thus, if W_T corresponds to a point on the $VMPL_f$ curve where elasticity is above unity, that gap will widen until the elasticity of unity is reached (corresponding to a labour cost level of W^* in Figure 3b), then fall. If W_T corresponds to a point where elasticity is less than unity, the gap will narrow monotonically, i.e. the marginal cost of the MW will be less in the presence of an SS system than in its absence throughout the relevant range. In other words, there is a sort of substitution of loss effects at work; because the SS system had already created considerable loss, an increase in the MW does less damage than it would have in a no-SS world. Note, however, that the total deadweight loss is always raised by the MW for levels where it bites.

³⁶ The relevant curves here are $VMPL_f$ and CWE_f . When one curve is below another by a fixed percent for quantity levels, the elasticity of the two curves is the same for each quantity.

Figure 3b



As noted, MWs are in fact not usually defined in terms of total labour cost but rather the base wage. If, at the extreme, the same MW is applied to the base wage in the SS case as in the no-SS case, then the presence of that SS system can greatly magnify the deadweight loss vis a vis that of a MW by itself. Should the MW be applied to the base wage rather than to total labour cost, the deadweight misallocation loss is much greater. For example, in the situation portrayed in Figure 3 that misallocation loss is FEG as compared to KEJ in the absence of the SS system. Total deadweight loss includes also any loss from SS inefficiency; for example, if $P=0.5$ for all workers, the additional loss is NFLQ, half of $NFHW_m$. 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 misallocated³⁷ and partly from the addition of the SS inefficiency loss. In short, the details of the SS system (both its costs and its efficiency) matter much more in the presence of a distorting minimum wage that is defined less with reference to total labour cost than with reference to the base 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 $P \geq 1$ for most or all workers will naturally cut losses vis a vis an inefficient one like that portrayed in Figure 3 (by reducing or eliminating the loss area NFLQ or even turning it into a gain). 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 minus the difference between SS

³⁷ For the first worker misallocated the loss is asymptotically close to zero, but for the last worker (in Figure 3) it is GF.

Although standard microeconomic theory has nothing to say about how the presence of an SS system will affect the setting of the MW, it seems likely that it will often be applied in such a way that total labour costs wind up considerably higher in the presence of an inefficient SS than in its absence. However, where the base wage is the policy variable, as in fact it usually is, it is unlikely to be set as high as it would be in the absence of the SS system, partly because regulators usually take some account of the implications of the level of labour costs.

.....

Case 3b A Minimum Wage (MW) above the Equilibrium Wage is Applied to the Formal Sector in the Presence of SS; valuation of the bundle of SS benefits varies across formal sector workers

As noted above, a MW requires some form of selection into the formal sector, since it leads to queuing for entry there. When workers are identical in all respects there are no efficiency or other identifiable effects from who gets chosen. When workers do differ, such effects arise, whether they differ in preferences vis a vis SS or in terms of productivity. Much then depends on the selection mechanism which determines who gets to work in the formal sector.

Suppose a MW is applied to the base wage in the formal sector (e.g. W_m of Figure 3) thereby producing an equilibrium at point H and shrinking formal employment to L_{fm} . In the case shown there is a large excess supply of labour. Even for workers with $P=0$ (for whom the total benefits from employment would simply be the wage received (W_m) this is superior to informal employment and the excess supply of labour is the quantity HT. If all workers have $P=1$ the excess supply would be everyone currently employed in the informal sector since point F lies above point B. Either way, there is an excess supply of workers available to the formal 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 formal sector employment (i.e. by the 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 for the L_{fm} workers with the highest P, its average value is less than one (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 loss due to the low average P would 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 a MW that raises total labour cost to OP (and the base wage to W_m) is area FEG. If, at the extreme, the MW is applied to the base wage and is kept the same in the presence of the SS system, then much of the increase in total labour cost and the associated loss can be attributed to the presence of the SS system, instead of a misallocation deadweight loss of just KEJ that loss is now FEG. In addition, there is the SS inefficiency loss PFLQ when $P=0.5$ for all workers. This latter area could in theory vary from zero if $P=1$ for all workers to the whole area PFHW_m if $P=0$ for all. If P varies across workers it would be lower (higher) if their selection into the formal sector is according to how high (low) that P level is.

Case 3c. A Minimum Wage (MW) above the Equilibrium Wage is Applied to the Formal Sector in the Absence of SS; workers are heterogeneous in productivity

The impact of a MW may also be affected by the degree of variability of productivity across workers in the same category,³⁸ assumed in the above examples to be zero, but which in fact may be significant. Under perfect markets and no interventions (including a SS system), firm-worker bargaining will lead to the same

³⁸ The meaning of heterogeneity of worker productivity is clarified in the text below.

payment per unit of work to each worker, with the total wage (per unit of time) thus varying across workers according to their productivity. Earnings are allocated on a piece rate basis. When 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 workers between the formal and informal sectors is unrelated to their absolute productivity.

To this point, the unit of labour to which curves $VMPL_f$ and $VMPL_i$ refer could be thought of as either number of workers or amount of work, the ratio of one to the other being assumed fixed. When that ratio is variable, labour market outcomes are less straightforward. The VMPL curves can be expressed either in terms of work units or number of workers, the former being the more directly relevant to firms as they search for profitability but the latter the more relevant for the analysis of institutions like SS and policy interventions like MW. In what follows we distinguish a number of possible cases of interest, retaining the worker as the principal unit of analysis.

Case 3.1. Absolute productivity in the formal sector (AP_f) varies across workers in a capital-augmenting/shrinking way; absolute productivity in the informal sector (AP_i) is constant across workers.

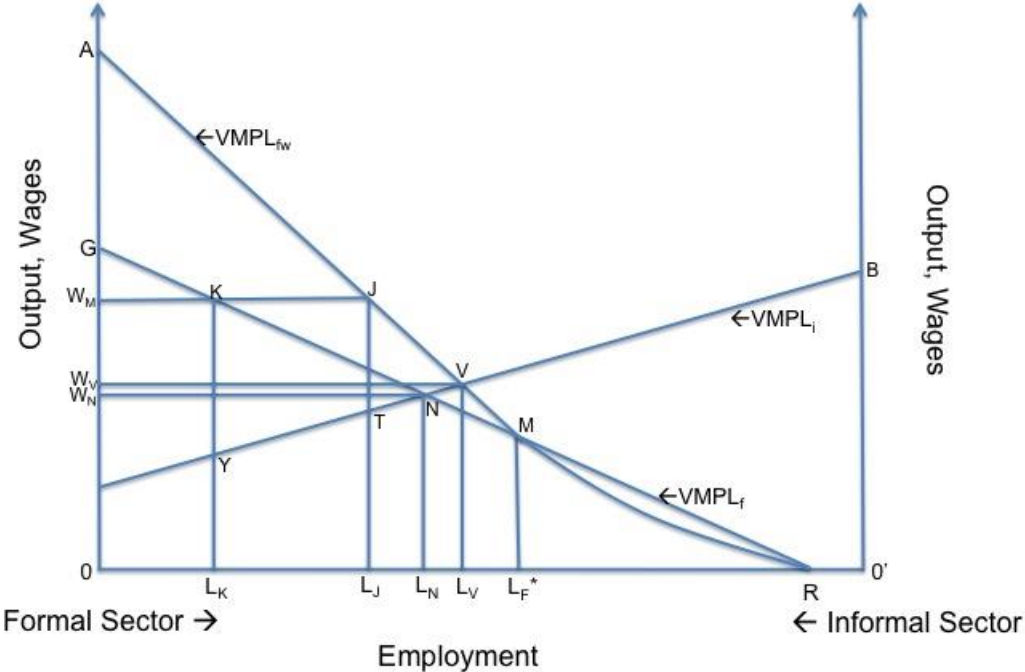
In Figure 4a, curves $VMPL_f$ and $VMPL_i$ refer to the traditional case where productivity in both sectors is constant across workers and hence the curves can refer to either number of workers or number of work units. In the simplest case involving variable worker productivity, absolute productivity in the informal sector (AP_i) is constant across all workers while AP_f falls linearly among workers ranked by that variable; the variability of worker productivity involves different degrees of capital stretching/shrinking rather than of labour stretching/shrinking;³⁹ and $VMPL_f$ and $VMPL_{fw}$ have the same average value across the universe of workers (in other words, it is assumed that an average productivity worker does one unit of work per day). With this set of assumptions the $VMPL_{fw}$ ⁴⁰ curve is steeper than is $VMPL_f$ and the two intersect at

³⁹ A distinction must be made between worker heterogeneity that is capital augmenting/shrinking vs. worker heterogeneity that is essentially labour augmenting/shrinking. In the former case the heterogeneity is formally equivalent to differing levels of capital stock, such that, level of employment in number of workers (or hours) held constant, VMPL is increased by a worker's having above average productivity and decreased by his/her having below average productivity just as if the first worker operated with more capital than average and the latter one with less capital than average. In the latter case one worker's being more productive than another boils down to his/her doing more of the same work as the other person, i.e. he/she simply undertakes more (less) than one standard work unit and this explains his/her above (below) average productivity. In that case the first worker not only has higher VMPL than would a standard worker but, since he/she does the work of more than one average-productivity worker, number of workers and number of work units (definable in clearly conceptualized units of effort, say) no longer coincide on the horizontal axis so the VMPL curve drawn in terms of number of workers falls faster than it otherwise would, i.e. each high productivity worker takes the place of more than one standard worker and thus constitutes a bigger shift to the right in terms of labour input measurement than in the base case. If, on the other hand, the sense in which these workers are high productivity is that they can produce more output with the same amount of complementary factors as can the average worker (e.g. because their labour input is higher quality rather than higher quantity), then the just cited effect is absent and, at least as a first approximation, the curve $VMPL_{fw}$ simply lies above curve $VMPL_f$ by the productivity gap between the two workers. For simplicity, we adopt this latter assumption in Figure 4a. Ideally input heterogeneity would be introduced formally into the production functions involved.

⁴⁰ Note that, with labour augmenting/shrinking heterogeneity of worker productivity, the $VMPL_{fw}$ curve, with workers arrayed by productivity, is not a standard labour demand curve, since the latter is defined in terms of units of work. The height of $VMPL_{fw}$ does, however, provide the information on the maximum offer firms will make to acquire the services of the marginal worker; it does not imply the same offer for other workers (as in the case of a standard

point M, where L_f equals half of the total labour force $00'$.⁴¹ Instead of an equilibrium at point N, as in the case of constant worker productivity, that equilibrium now occurs at point V, the intersection of curves $VMPL_{fw}$ and $VMPL_i$ and with a modestly higher wage rate (W_v) than in the traditional case (W_N) and a greater level of formal sector employment (L_v rather than L_N).⁴² Total societal benefits from the worker heterogeneity are measured by area AVNG, the amount by which total income or GDP exceeds that in the constant worker productivity case. This gain reflects the fact that heterogeneity *per se* is beneficial when an economy can allocate groups of workers to the sectors where they are most productive.

Figure 4a



The concept of a single, simple equilibrium wage, while it remains clear in the case of labour-augmenting demand curve).

⁴¹ For simplicity we have thus far assumed that the $VMPL_f$ and $VMPL_i$ curves are linear. Since by assumption the ratio of the height of the latter to that of the former gradually falls but does not reach zero, this implies that curve $VMPL_{fw}$ is or becomes non-linear before it and $VMPL_f$ intersect the horizontal axis. Since the precise monotonic function relating the two curves does not affect the arguments, we take account of the eventual non-linearity of $VMPL_{fw}$ in the region below M, where it intersects curve $VMPL_f$.

⁴² These last two results are not general, but depend on the fact that, in the case shown, since less than half of all workers are employed in the formal sector, the average productivity of the last one to be hired is above average. This, however, is a plausible assumption since so many countries do have a higher level of informal than of formal employment.

productivity differences across workers (and is simply the equilibrium payment for a unit of work) becomes more complicated when those differences are capital-augmenting. In this case the $VMPL_{fw}$ curve slopes down both for the standard reason--the increasing amount of labour applied to a fixed amount of other factors, and because workers are arrayed from most productive to least productive. For any given total number of workers employed, including the equilibrium level OL_v , the intramarginal workers are worth more to the firms than is the marginal worker. As long as there is a good level of competition on the demand side, and relative productivity between any two workers is the same for all formal sector firms, the market may be expected to pay each worker according to whatever productivity he/she has.⁴³ In effect, as in any labour market, there will be submarkets according to quality (reflecting training, innate skills, whatever matters to productivity). At the limit, the market will allocate efficiently, as it may be expected to in the case where workers differ only by quantity of work of which they are capable. The area between this $VMPL_{fw}$ curve and the homogeneous labour curve $VMPL_f$ will be the return to the above average human capital of the workers involved. At the other extreme, if rules or conventions dictate that all workers receive the same daily wage, that wage would be given by the height of this curve.

When the labour market is less competitive and/or worker productivity differences vary by firm, bargaining rules and practices will evolve to establish a wage structure, with only the wage of the marginal worker being firmly established by the market.⁴⁴ As much as all of the potential "surplus" resulting from their above average productivity could be bid away by them, making the labour share of GDP undefined in this situation.⁴⁵ The bargaining process might involve "n" bilateral monopoly situations, with each of the "n" workers having a reservation price determined by productivity in the informal sector and each employer having a maximum bid for a given worker.

Assume now that a minimum wage is imposed at the level W_M (Figure 4a). In the traditional case the deadweight loss would be KNY and in the new case it is the substantially smaller amount JVT . The reduced deadweight loss results from the steeper $VMPL_{fw}$ curve.⁴⁶ Since GDP has been raised by the worker heterogeneity, the deadweight loss is smaller relative to GDP both because it is smaller in absolute terms and because GDP is larger. When AP_f varies only moderately across workers, the scenario will naturally be

⁴³ When worker productivity varies across firms it then matters to the position of the $VMPL_{fw}$ curve and to total output which worker winds up with which firm, so if matching is not efficient this implies a lower curve than otherwise, and other results discussed here must be altered and complicated.

⁴⁴ Standard factors explaining differences in productivity across workers are innate skills, experience, education and training, drive, etc. Some of these phenomena are more likely to be at work in the case of labour augmenting differences and others in the case of capital augmenting differences. Some productivity differences related to differences in education/training are formalized by the definition of one's occupation, and access to that activity may be governed by regulations. In such cases, it is normal to think of separate labour markets. More interesting are the differences which do not provide the basis for a taxonomy of types of workers or for segmented labour markets but nevertheless create differences across workers that are relevant to employers. There is a large literature on the implications of worker heterogeneity and on how effectively markets "match" worker skills to firm needs (e.g. Woodcock, 2010; Shimer and Smith, 2001). The task of efficient allocation of workers across firms is relatively simpler when productivity differences across workers are the same regardless of which firm they work for. When this is not the case, the matching task is somewhat harder.

⁴⁵ Further, as previously noted, when allocation of workers across firms is not perfectly efficient part of the potential output defined by the area under $VMPL_{fw}$ is in fact lost.

⁴⁶ In the figure, the gap between the minimum wage and the prior equilibrium is a little larger in the traditional case but even if this gap were held constant as between the two cases, the loss would be less under the variable worker productivity case.

little different from the base case of worker homogeneity; when it varies widely, the difference will be correspondingly greater. Note also that the MW could be imposed at a level somewhat above the original homogeneous labour equilibrium (i.e. up to level W_v) without affecting efficiency, since the new equilibrium wage is higher than the old one because of the heterogeneity.

More generally, analysis of the impact of a MW (and hence also of a MW-SS combination) is complicated by the fact that, whereas the gap between a MW and the market wage can be defined clearly in a situation of homogeneous labour and a single market wage, this is no longer true when labour is heterogenous in productivity. The MW that, with heterogeneous labour, may be considered comparable to a homogeneous labour case in which the MW is 20% above the pre-MW formal sector market wage, is not obvious; possibly a reasonable choice would be one that is 20% above the previous average economy wide wage, but there is no simple answer.

In what are for present purposes variants of this case, AP_f and AP_i may be positively but not perfectly correlated (thus RP_{fi} varies across workers) or they may be negatively correlated. These situations are qualitatively similar to Case 3.1 since the ranking of workers by AP_f remains the same. The latter case is notable for the fact that with the steeper $VMPL_{iw}$ curve, the deadweight loss from a MW is greater than in Case 3.1.

Where labour heterogeneity takes the form of labour quantity augmenting/shrinking the impact of a MW may be reduced even more significantly. While the comparison between the level of a MW and the equilibrium wage that would have obtained in its absence is straightforward in the homogeneous labour case, this is less so in the heterogeneous labour augmenting/shrinking case since there is no single equilibrium wage per period of time worked, but rather per unit of effective work.⁴⁷ With the minimum wage defined in terms of units of time, its impact now depends on the distribution of workers by productivity, on where the MW is set, and on the size of the sector on which it is imposed.

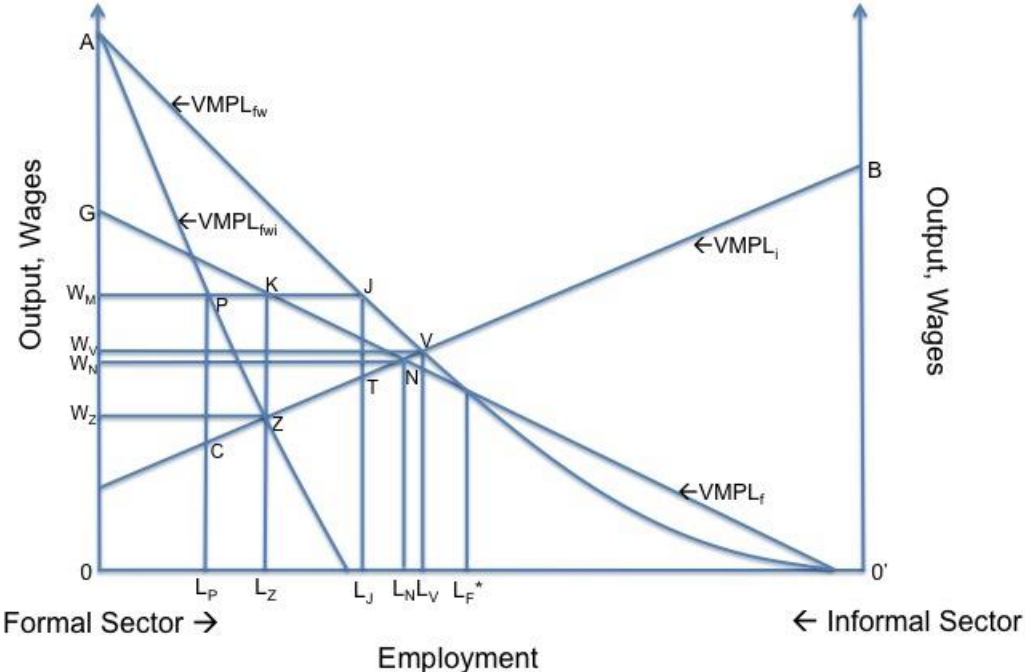
Prior to that imposition there is an equilibrium wage per unit of work (labour input). In this setting a MW (value per day) can be imposed at a level somewhat, perhaps substantially, above the average daily wage corresponding to the per work unit equilibrium wage, since many of the work units in the formal sector may be undertaken by persons with above average productivity per unit of time. Before its imposition there is no reason for those with high AP_f levels to be found in the formal sector, since the criterion for employment there is level of RP_{fi} ; imposition of the MW is thus likely to promote some movement of workers between the sectors.

To take a simple case, suppose that the most productive worker does 50% more work than the average (equal to that of all workers in the homogeneous labour case) and that by the 50th percentile worker labour productivity is at the average. This formal sector VMPL curve defined in work units (call it $VMPL_{fwi}$ to distinguish it from the corresponding curve in the capital augmenting/shrinking case) will, accordingly, have a vertical intercept 50% higher than that of the $VMPL_f$ curve (Figure 4b); but it will, as noted above, slope downward more steeply than the latter curve and than the $VMPL_{fw}$ curve with the same vertical intercept but corresponding to the capital-augmentation case just discussed; its slope will also rise with L_f . By the time $L=L_f^*$ the VMP of the last worker on the $VMPL_f$ curve (and on the $VMPL_{fw}$) has reached the level L_f^*M ; on the $VMPL_{fwi}$ curve that height has been reached when L_f is equal to GML_f^*/AML_f^* as much; this is

⁴⁷ The imposition of a MW defined in units of work would have the same impacts and could be analyzed in the same way as one defined in terms of period of work where productivity does not vary across workers.

the ratio of the above average productivity workers that it takes to do the same amount of physical work as the first $0L_f$ workers in the homogeneous worker case. As a result, the $VMPL_{fwi}$ curve has reached the same vertical distance at $L_f = L_{f*} (GML_{f0}/AML_{f0})$ as curve $VMPL_f$ has reached at L_{f*} . In terms of number of workers, the formal sector needs less than otherwise to achieve a given level of output. Under a free labour market, where wages reflect productivity, earnings in the formal sector will vary by worker. Equilibrium occurs at point Z, with formal sector employment of L_z ; and a wage rate of informal sector workers (and of the marginal formal sector worker) of W_z . This wage is lower on a per unit of work basis than in the benchmark case,⁴⁸ essentially because the presence of variable productivity across workers in the formal sector has the same effect as an increase in the overall supply of labour (work units) and thus pushes the wage rate down.

Figure 4b



When a MW is imposed on a per-day basis, moving the equilibrium to point P and formal sector employment to L_p , the loss of formal sector employment is less than in the benchmark case (with no variability of worker productivity and curve $VMPL_f$) because curve $VMPL_{fwi}$ is steeper than the benchmark $VMPL_f$ curve; the deadweight loss (PZC) is also less. In this case, though a MW *per se* creates less societal loss than in the benchmark case, the variability of worker productivity has by itself had the impact of lowering formal sector employment while raising GDP. It has the similar effects to those of the

⁴⁸ In the informal sector, where productivity does not vary across workers, the wage per worker and that per unit of work are the same, but in the formal sector this is not the case, and only the marginal worker earns the same as the informal sector workers.

introduction of labour-saving but output-enhancing technological change. The final outcome, before or after imposition of the MW, involves a higher GDP than does the benchmark case; in the absence of the MW, for example, GDP is measured by $0AZB0'$ rather than the $0GNB0'$ of the benchmark homogenous labour case.⁴⁹ That final outcome also involves a lower formal sector employment and a lower wage for informal sector workers and for the marginal formal sector worker. Some (probably most) of the other formal sector workers will earn more than before however, since their productivity is higher; the most productive will, as noted, earn amount $0A$.

A MW that sets a floor under the wage per unit of time may or may not diminish the capacity of the formal sector 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 formal firms can hire the profit maximizing number of units of work without raising the payment per unit of work, then their profitability will be unaffected, as will their total output and their total employment measured in work units. More workers will now remain in the informal sector but since they are less productive than average, 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 high enough 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.⁵⁰

The smaller reduction in formal sector employment and the tendency to smaller deadweight loss in cases involving variability of worker productivity results from two distinct phenomena:

- i) variability allows formal sector firms to select high productivity workers (high AP_r) when a MW forces them to pay higher daily wages;
- ii) the elasticity of the VMPL curve is lower than in the case of homogeneous workers. When that elasticity approaches zero, deadweight loss and employment reduction do as well. The MW can be raised much farther without significant damage when the curve is very inelastic. Since

⁴⁹ That total output is greater in the presence of the variability of worker productivity in the absence of MW follows from the fact that more total work units are being expended in production, and they are being allocated efficiently between the two sectors. That the same result holds in the presence of a MW is more complicated to show, though it is clearly the case whenever deadweight loss is less in the worker heterogeneity case. Note though that since the equilibrium wage per unit of work is lower under worker heterogeneity, a given MW constitutes a larger percentage and absolute increase in the per worker wage than in the homogeneous case.

⁵⁰ When workers differ among themselves both in terms of productivity and of preference with respect to SS, both these factors help to determine their socially optimal allocation between sectors. We do not consider in detail how the relationship between the two features of workers would affect the impacts of MW and SS, but it is clear that the potential for reduction of deadweight loss through worker heterogeneity would be greatest when high productivity is correlated with a high P .

variable worker productivity promotes a steep VMPL curve, it increases the likelihood that these losses will be small

At the same time, variable worker productivity lowers the number of people likely to be hired in the formal sector and increases inequality in the economy, both inequality among workers as a group and inequality between lower income workers (those in the informal sector and the less productive formal sector workers) and everyone else (composed of capital owners and the higher productivity workers).

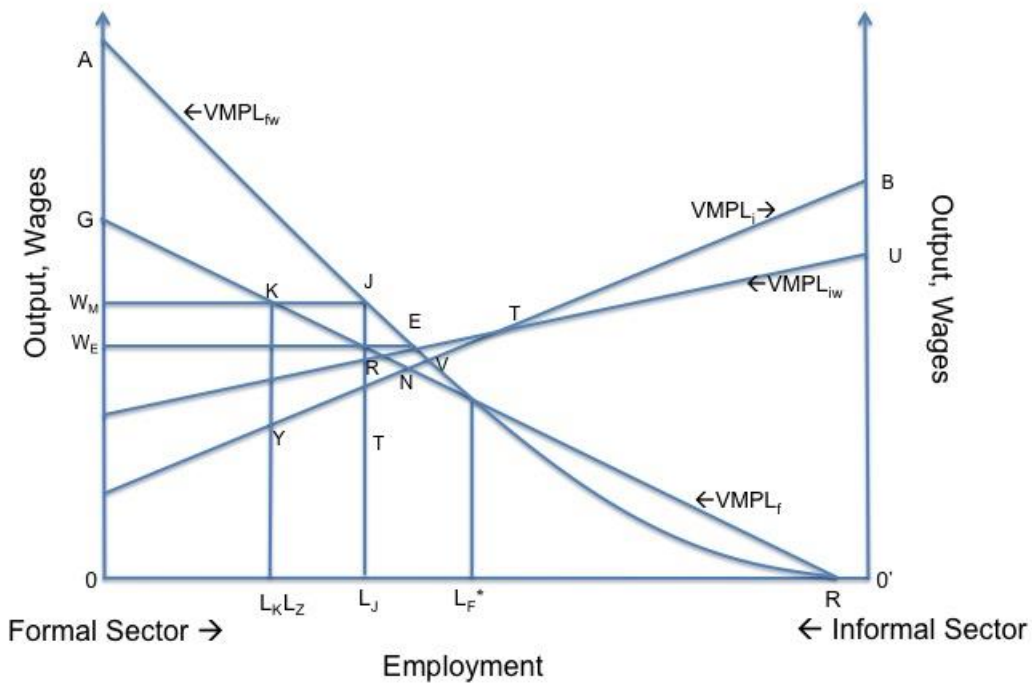
We now return to a consideration of alternative assumptions about the relationship between AP_f and AP_i under the simpler assumption that productivity differences across workers are of the capital augmenting/shrinking nature. In Case 3.1 the ranking of workers by productivity is the same whether AP_f or RP_{fi} is used since these two variables are perfectly correlated. The following cases drop that assumption.

Case 3.2: Absolute productivity (AP) differs across workers in the same (capital augmenting/shrinking) way in both sectors.

The simplest variant of case 3.1 involves the assumption that (AP) differs across workers in the same (capital augmenting/shrinking) way in both sectors, in other words that RP_{fi} is constant across workers, so their ranking by AP_f is the same as their ranking by AP_i . The effects of the imposition of MW^{51} are similar to those reported for Case 3.1. Figure 4c portrays this situation; it is similar to Figure 4a except that curve $VMPL_{iw}$ must be added to take account of the non-constancy of worker productivity in the informal sector. In this case, the ratio of the height of $VMPL_{iw}$ to that of $VMPL_i$ is the same as ratio of the height of $VMPL_{fw}$ to that of $VMPL_f$. The $VMPL_i$ and $VMPL_{iw}$ curves cross (point T) at the same formal sector employment level (L_{f*}) as do $VMPL_{fw}$ and $VMPL_f$; $VMPL_{iw}$ is flatter than $VMPL_i$ since the workers who are most productive in the informal sector are (by assumption) the first ones to be withdrawn from that sector. The intersection of $VMPL_{fw}$ and $VMPL_{iw}$ (point E) gives the optimal allocation of labour between the two sectors. The wage rate received by the last worker hired into the formal sector is W_E but, as noted above, without specification of the wage determination process for others, what the intramarginal workers in each sector will earn remains indeterminate. In a perfect market with wages per unit of work the same across workers and worker pay varying by worker AP level, the most productive one would earn AG/GO times what an average productivity worker would.

⁵¹ As always, labour productivity depends on the availability of other factors. In the short run the fixed supply of those other factors helps to give the demand curve for labour its downward slope. In the present case, labour productivity also depends on individual worker characteristics.

Figure 4c



The societal gain from worker heterogeneity in this case amounts to $AETNG$ minus TBU , the latter loss occurring since the workers left in the informal sector are now mostly of below average productivity; this net societal gain is the difference between the previous welfare level of $OGNB0'$ and the new level $OAEU0'$

If under these conditions a MW of W_m is imposed on the formal sector, it constrains employment (in terms of workers) to the level L_J , with deadweight loss JER determined in the usual way, but defined by curves $VMPL_{fw}$ and $VMPL_{iw}$. Note however, that the loss is considerably less than it would have been had the same MW been set in the homogenous productivity case. Employment would have been cut to OL_K and loss would have been KYN . The smaller loss when workers are heterogeneous follows from the fact that for all workers of above average productivity the $VMPL_{fw}$ curve lies above the $VMPL_f$ curve so that for this range of workers it now takes a higher MW to prevent any given worker from being employed in the formal sector.

As noted for Case 3.1, where AP_f varies only moderately across workers, the scenario will be little different from the base case. However, for any given degree of variation of AP_f (slope of curve $VMPL_{fw}$) the reduction of deadweight loss will be greater than in Case 3.1 since with curve $VMPL_{iw}$ lies above curve $VMPL_i$. The loss area is reduced, since the workers drawn to the formal sector have a higher productivity in the informal sector than in the base case and the cost of misallocation of the workers between sectors is lessened.

Case 3.3 Where AP_f and AP_i vary in different ways across workers; more complicated sectoral reshuffling
Consider now the more general case where relative worker productivity as between the formal and

informal sectors differs due to imperfectly correlated variability of AP_f and AP_i across workers.⁵² The ranking of workers now differs as between AP_f and RP_{fi} and it is RP_{fi} that determines the “comparative advantage” of a given worker in the formal sector.⁵³ Assuming that these features of worker productivity are known to the workers and to potential employers, and in the absence of any imperfections in the market or any interventions like MW or SS, workers would be selected into the formal sector according to the level of their RP_{fi} . Absolute productivity would play no role in where a worker found a job; in each sector 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.

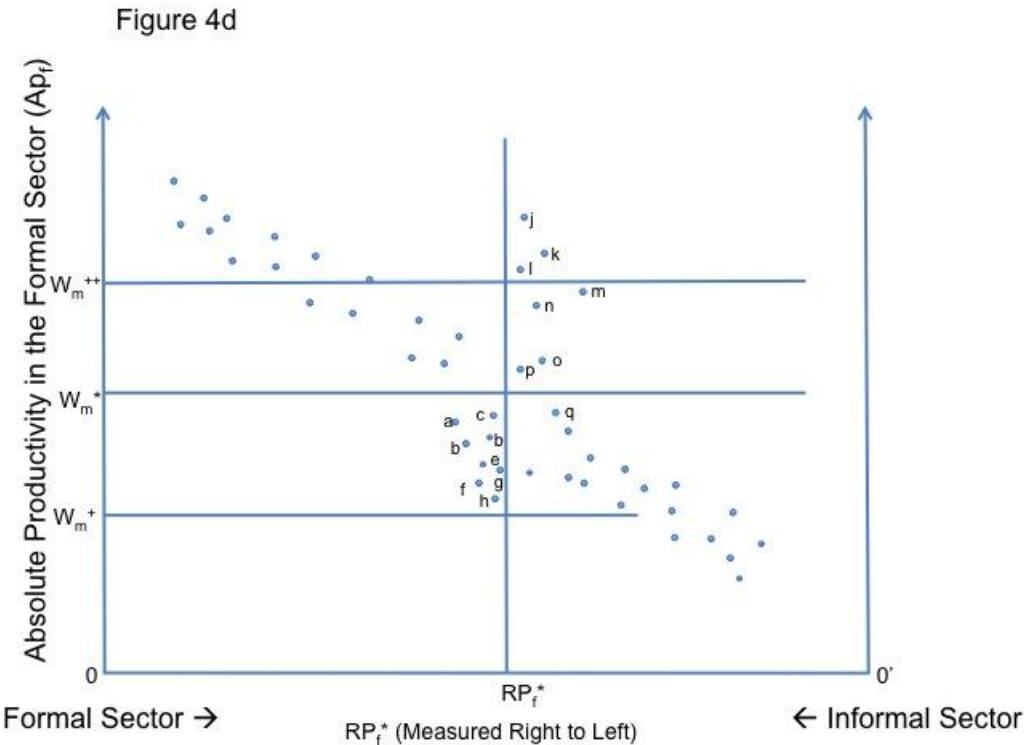
Under these circumstances, the imposition of a MW will normally disturb the prior equilibrium, raise the labour cost per unit of work in the formal sector and curtail employment there.⁵⁴ If there is a large enough pool of people available who, earning the pre-MW payment per unit of work, would also be earning the new MW per day, then it will have no more than a modest impact. Under some circumstances the daily MW can be set well above the previous daily earnings of many people in both the formal and informal sectors without greatly affecting the prior equilibrium, since they can now be shuffled between sectors in such a way that the only ones earning less than MW per day after that imposition are moved (by market forces) to the informal sector. This outcome is especially likely when, even though AP_f varies widely across workers in the vicinity where the MW is set, the RP_{fi} of these workers varies little. Figure 4d shows such a situation. Workers are ranked by AP_f , measured on the vertical axis, while RP_{fi} is measured horizontally from right to left; each dot represents one worker and shows that worker’s AP_f and RP_{fi} . Assume that in the pre-MW equilibrium, ORP_{fi}^* is the number of workers employed in the formal sector, implying that all those to the left of the vertical line above point RP_{fi}^* are so employed. Now assume a MW is imposed at a level W_m , such that eight workers previously hired in that sector (workers “a” through “h”) have AP_f too low to justify payment of that wage. They will move to the informal sector. Workers “j” through “p” already satisfy the AP_f requirement and are now likely to move to the formal sector, and worker “q” and perhaps another one or two might do so. How many shift to the formal sector depends on the extent to which total labour demand falls there because units of work have become somewhat more expensive. Recall that any worker’s $VMPL_f$ determines the maximum wage offer a firm is willing to make to her. That variable is affected both by the worker’s AP_f ranking (which determines the $VMPL$ corresponding to that worker, i.e. the height of curve $VMPL_f$ at the horizontal position corresponding to that worker) and her personal AP_f , which determines how much higher or lower $VMPL_{fw}$ is vis a vis $VMPL_f$ at that point. The removal of workers “a” through “g” from the formal sector has the effect of raising the $VMPL_{fw}$ values for each worker farther down the AP_f ranking, thus creating the likelihood that worker “q” will move to the formal sector and the possibility that others will as well. In any case, it is evident that in this scenario neither employment nor output will be much affected. Replacement of workers “a” through “g” with workers “j” through “p” for example, will moderately lower GDP, since the replacements have lower RP_{fi} that those they replaced so more informal sector output will be given up per unit of formal sector output

⁵² In the special case where productivity in the formal sector is constant across all workers but RP_{fi} varies, workers are selected into the formal sector by the value of RP_{fi} (the higher it is, the lower the opportunity cost from their leaving the informal sector). Any MW above the daily wage previously earned by formal sector workers will force up the cost of labour and lead to deadweight loss.

⁵³ In case 3.2, with RP_{fi} constant across workers, all were equal in terms of the economic logic of their being employed in the formal sector, as would also be the case of homogeneous labour.

⁵⁴ As in the standard case of homogenous labour, the elasticity of labour demand in the formal sector will be an important determinant of deadweight loss.

they produce than was the case originally for workers “a” through “g”. But since the differences in RP_{fi} are not large, and the number of workers being shuffled between the sectors is modest, this output loss will also be small. For the number reshuffled to be large, the RP_{fi} differences would have to be small, *ceteris paribus*, so the likelihood of a big effect on output would still be low. Note that a MW imposed at level W_{m+} will have no impact since all of the workers already employed in the formal sector earn above that level and the set of included workers are those with the highest RP_{fi} . On the other hand, a high MW like W_{m++} will curtail both employment and output more markedly as many workers are pushed out of the formal sector while a few workers could be transferred into it.



To reiterate, the extent to which the daily MW raises the formal sector wage actually paid in unit of work terms and thereby creates a deadweight loss depends on the degree to which those former formal sector workers whose cost to the employer is raised by the MW can now be replaced in the formal sector by others whose lower relative productivity (RP_{fi}) than those they replace is combined with high absolute productivity; in other words it depends on whether there is a relatively high frequency of people with this set of characteristics (as in the case illustrated in Figure 4d). When AP_f and RP_{fi} are strongly (positively) correlated across workers then their selection into the formal sector is very close to what it would be if only absolute productivity were the criterion and there is little opportunity or need to curtail deadweight loss through reshuffling of people between the sectors.⁵⁵

⁵⁵ Prior to the implementation of the MW in this economy, workers are, as noted, selected into the formal sector based on their RP_{fi} , and independent of their absolute productivity.

At the limit deadweight loss, might be totally avoided by a sectoral reshuffling of workers. This would occur when there is a group of workers with identical RP_{fi} , some of whom were initially employed in the formal and some in the informal sector; in other words, the cut-off level of RP_{fi} for inclusion in the formal sector matches that of these workers, but only some were previously in the formal sector. Within the group there is a range of AP_f (and hence of AP_i) and an exchange involving the movement of some of those with higher AP_f to the formal sector and the reverse movement for some of those with low AP_f will leave total output in each sector and overall economic efficiency the same as before. Total units of work in each sector will stay the same.⁵⁶ Alternatively, deadweight loss will be avoided if the MW is set below the daily earnings of everyone currently employed in the formal sector, which means in turn that it must be well below the average earnings in that sector. In such a case, there will be no reshuffling of workers between the sectors as a result of the MW.

The normal outcome in this sort of situation, however, would be some shrinkage of the formal sector's employment and output and some deadweight loss. The new feature, vis. vis. cases of worker homogeneity or cases like 3.1 and 3.2 where the ranking of workers was not different as between AP_f and RP_{fi} , is that worker reshuffling can diminish those negative effects of a MW.

If one assumes a two-good economy (one good produced by the formal sector and another by the informal sector) then a condition for equilibrium is $RP_{fi} = P_f/P_i$, where P_f and P_i are the prices of the formal and informal sector goods respectively.

⁵⁶ To think in terms of a concrete example, assume for simplicity that absolute productivity in the formal sector varies within the range 0.8 to 1.25 with a rectangular distribution,⁵⁶ and that relative productivity R_f/R_i 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 that 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.

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 not 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 $R_f/R_i = 1.0$ 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.⁵⁶ But if so many of the $R_f/R_i = 1.0$ workers were already in the formal sector that the above reshuffling could not be achieved without diminishing employment there, then the equilibrium would be affected and economic inefficiency would result.

Among those situations where worker ranking by AP_f differs significantly from that by RP_{fi} , thereby creating the possibility that worker heterogeneity may make a big difference (either to deadweight loss or to its minimization through reshuffling), two are specially worthy of attention. In one (extreme) situation AP_f and RP_{fi} are negatively and closely correlated (with workers ranked by AP_f , their AP_i values fall faster than the AP_f values). The workers whose selection into the formal sector maximizes profits are thus those with the lowest AP_f values. Imposition of a MW induces large scale (perhaps total) exodus of those workers formerly employed in the formal sector since only those with high AP_f can now be hired there. Some now move from the informal to the formal sector, though at a high opportunity cost in terms of informal sector output foregone. Although precise measurement of the costs requires more detailed specification of the relevant functions, it seems clear that this is a case in which loss could be high, since the only people who can be hired in the formal sector are those least appropriate (in terms of their opportunity cost) to work there, and though there is an exchange of workers between the sectors it is accompanied by a high efficiency cost. The degree of loss is determined by the structure of worker heterogeneity, the production functions of the formal and informal sectors, and the structure of demand for the products of the two sectors. If the VMPL curves are quite elastic in each sector and the products are close substitutes for each other, a high MW will terminate production in the formal sector, but the overall loss from this will be small. Where either the production functions involve low factor substitutability or demand for each group of products is quite inelastic, then output composition may not change much, a great deal of worker reshuffling will occur, and deadweight loss will be high due to the efficiency cost of that reshuffling. In terms of real-world likelihood with respect to the characteristics of the labour force, this extreme case is fortunately, not very probable. A reasonable guess is that, although there is a good deal of variation in worker rankings by AP_f and by AP_i , there is likely to be more variability in the former than the latter, where presumably less complicated skills are required, and, barring the effects of learning by doing, there is likely to be a positive correlation between the two. As long as this condition is met, the dangers associated with rigid production functions or demand functions are reduced.

An intermediate case would be that in which there is no correlation between AP_f and AP_i and hence between AP_f and by RP_{fi}). As discussed above, with no imperfections in the labour market, prior to the imposition of a MW workers will be allocated to the formal sector according to their RP_{fi} . When a MW is imposed, some shuffling of workers will occur with only those with a high enough personal AP_f (together with the position of the standard work unit-based $VMPL_f$ curve) to imply a marginal product of labour above or equal to the MW remaining as candidates for formal sector employment. An efficiency loss accompanies this reshuffling of workers, but it simultaneously reduces the deadweight loss associated with the MW by reducing the output decline in the formal sector that results from that MW.

As in the simpler cases considered above, the MW raises the average wage per unit of work in the formal sector because in effect it brings about a shrinkage of the supply of work units to that sector, and the resulting labour reallocation lowers the wage 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.⁵⁷ 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

⁵⁷ This combination of outcomes is likely to increase income inequality in the country.

arrangements) this useful option will be cut off.

Although it is evident that the presence of labour heterogeneity greatly complicates the analysis of the impact of a MW and hence also of a MW-SS combination, with different scenarios leading to significantly different outcomes, in the majority of real-world situations it may well reduce the deadweight loss from these interventions. The comparison is hard to make for many reasons, beginning with the fact that whereas the gap between a MW and the market wage is a simple enough concept in a situation of homogeneous labour and a single market wage, this is no longer true when labour is heterogeneous; there is for example no obvious counterpart to a 20% gap between the minimum and the equilibrium wages in the homogeneous labour case. In terms of real effects, labour heterogeneity basically allows formal sector employers to reshuffle workers so that the MW does not bind or binds less, thereby reducing deadweight loss due to less-than-optimal output in the formal sector; however, it also creates the possibility that, output constant, too few workers will be employed in the formal sector, and that the wrong ones will be, i.e. under a situation of excess labour supply to the formal sector, the lowest opportunity cost workers will not generally be chosen to work there. In light of these countervailing effects, whether the loss is less with labour heterogeneity than with homogeneity clearly depends on the specifics of the case.

To summarize, labour heterogeneity involving the same productivity differences regardless of sector of employment tends to cut formal sector employment by making the $VMPL_f$ curve less wage elastic, and at the same time to reduce deadweight loss from a MW because of that lower elasticity. A MW set the same percent above the marginal formal sector wage as the homogeneous case MW is set above the equilibrium wage will cause less deadweight loss. This will be true, *a fortiori*, if the heterogeneous case MW is set the same percent above average wages in the economy, assuming these to be lower in the informal than in the formal sector. But if it is set the same percent above the average pre-MW wage in the formal sector this may constitute a considerably higher wage than when the reference point is the marginal wage in the formal sector.

The second complication arising when labour is heterogeneous relates to possible shuffling of workers between sectors when a MW is applied when AP_f and AP_i are not perfectly ordinally correlated with each other. Three factors are simultaneously at play: the output and formal sector employment loss that would occur from a given MW imposed on the formal sector with no possibility of worker reshuffling; the amount of reshuffling that happens and the degree of its impact towards moderating deadweight loss and formal sector employment reduction; and the efficiency loss associated with the reshuffling. Generalizations are difficult because these factors may work in opposite directions. For example, the curtailment of formal sector employment by a MW would be greatest when the $VMPL_f$ is elastic and productivity differences among workers are small, but not much deadweight loss might result from its imposition in that case. Heterogeneous labour then permits a large reshuffling such that formal sector output and employment do not fall much, but the reduction in deadweight loss is small because that loss was not large in the first place.

Since the heterogeneity of productivity among workers has distinct effects impacting on economic efficiency in opposite directions, it is important to consider whether the empirical conditions that lead each one to be large are probable ones or not. That consideration is, however, beyond the scope of this paper.

Heterogeneity of Worker Productivity and SS

Given the complexity of cases where a MW is imposed under conditions of worker heterogeneity, it would

not be surprising if the addition of inefficient SS systems to the mix accentuated the complexity. Some generalizations are possible, however. First, in any situation where worker entry is not determined by self-selection, the door is open to loss through misallocation of workers, as discussed extensively above. With AP_f or RP_{fi} the criterion of selection into the formal sector and in the presence of queuing due to the MW, loss will be less when P (preference for the SS system) is closely correlated with that selection criterion, since that will minimize SS-inefficiency loss. Second, to the extent/under those circumstances when worker heterogeneity reduces the loss effects of MW imposition because it makes the relevant $VMPL_f$ curve less elastic it is likely to reduce SS-inefficiency loss as well because less workers are involved. When the heterogeneity reduces loss by inducing a reshuffling of workers between the sectors it will further lower SS-inefficiency loss if the workers shuffled to the formal sector have higher P than those leaving it, but it will raise SS-inefficiency loss in the opposite case.

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 absence of a self-selection system, when worker supply prices can no longer be that selection mechanism, is likely to increase the total loss since to the other sources is now added the deadweight loss associated with allocation of the wrong workers to each sector. The extent to which self-selection can still play a role thus becomes an important aspect of any situation of the sort under discussion here. Another posited explanation for inter-sectoral wage differentials is a system of “efficiency wages.”⁵⁸ (Mazumdar, 1959; Basu and Felkey, 2009). An efficiency wage mechanism does not by itself prevent a trade-off between wages and SS benefits but is more consistent with worker self-selection into the formal sector than are some other scenarios. 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 at above the going wage. Such a view gains plausibility from the fact that better fed workers are more productive or that a wage advantage over others 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. Any combination of these factors and others cited in the literature could be at play. In any case, an important implication of the presence of an efficiency wage mechanism in a given scenario is that a binding MW will typically cause a smaller deadweight loss than otherwise. In its presence, a substantial wage differential between a higher productivity sector and the rest of the economy may be largely voluntary from the firms’ perspective; at the limit the MW may be 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 curtail formal sector employment. But 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 are therefore likely to undertake a conscious search for such 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. This process may bring both benefits--if for example it encourages fruitful training and learning, and costs, if it widens income gaps between different groups of workers.

⁵⁸ 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.

How does SS fit an efficiency wage scenario? Associated losses from SS inefficiency will tend to be less because employment is less than otherwise. SS inefficiency loss will be less if the workers selected into the formal sector have higher P than those left outside it; this could occur as firms undertake their selection with a view to future worker satisfaction and/or if they design, operate, and perhaps adjust their SS system with worker satisfaction in mind.

Open Unemployment

Open unemployment is part of any economy, and quite important in some labour markets. 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 imply 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 formal-informal segmentation is the Harris-Todaro (HT) 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, another source of loss must be added to those considered in the cases analyzed above—resulting from 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 inter-sectoral wage gap and the associated suboptimal level of formal sector employment causing a deadweight loss associated with the gap between the VMPLs in the formal and informal sectors, that loss is now measured by $VMPL_f$, 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.⁵⁹

It has not been easy to settle the 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 could be argued that workers move from higher productivity rural activities to lower productivity urban informal ones in order to be better positioned for formal sector urban jobs. This latter possibility can only be analyzed in a more complete model. A context like that analyzed here, in which there is queuing for desirable jobs, also complicates the concept of efficient (job search) unemployment since the search may well be efficient from the private point of view of those who engage in it, but not from a societal point of view, if the process is close enough to a zero sum game such that the gains of the winners are nearly offset by the losses of the losers (those who continue to work, say in rural or urban informal activities) and through not focussing enough on the job search wind up not getting the attractive jobs. Indirect costs of queuing may involve a public policy bias towards spending on urban rather than on rural populations as argued by Lipton (1977), when urbanization is per se an expensive venture Lewis (1977).

Case 4. Product Market Imperfections in one or Both Sectors

⁵⁹ This outcome would be parallel to that of the “credentialist” school’s interpretation of the way in which the private returns to education may overstate the societal benefits because education is used as a criterion for the selection of candidates for occupations in which it is not necessarily the most logical criterion, and its use in that way induces more people to pursue education in order to be candidates for the jobs in question (Dore, 1976).

Thus, far perfect competition has been assumed with respect to the product markets of both sectors and the capital market. In real life, perfectly or nearly perfectly competitive markets are an exotic anomaly, arguably characterizing under 5% of any economy. It is thus necessary to consider how much the present analysis must be modified when the assumptions about how these markets function are shifted in the direction of reality.

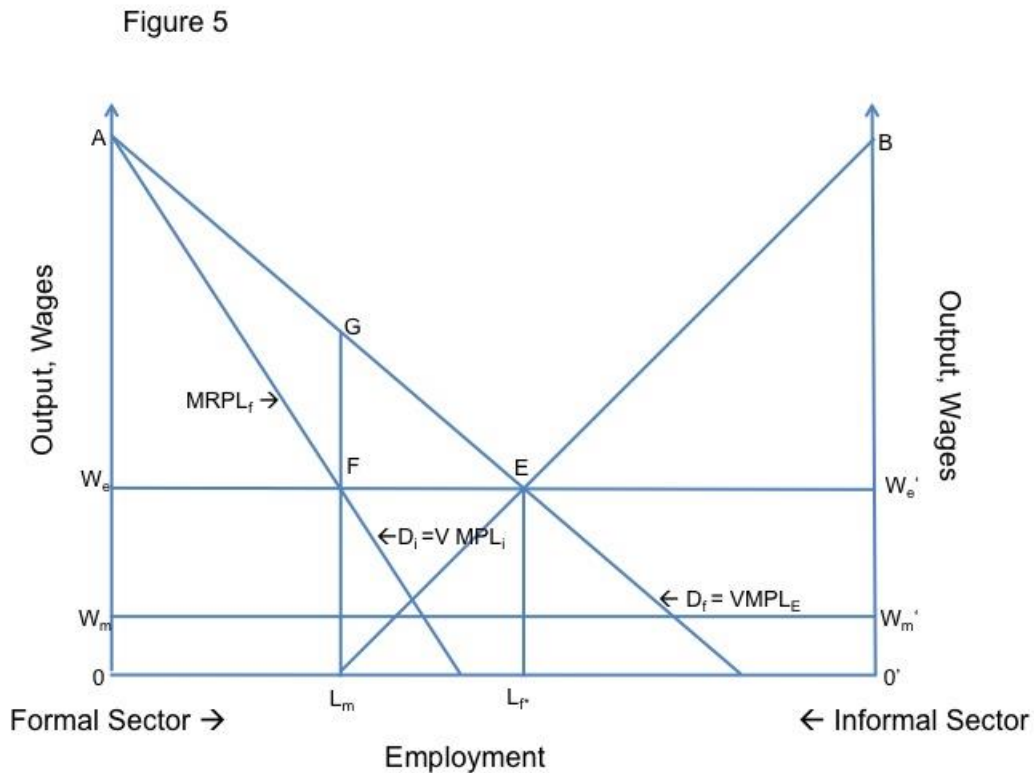
When the assumption of perfect competition in both the formal and the informal sector is dropped, the coincidence of the firm demand curves for labour and their VMPL curves disappears. As a result, no clear results on the impacts of additional imperfections like the MW can be deduced from theory except in special cases.⁶⁰ The outcomes depend, *inter alia*, on the nature of the 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 or main form of non-perfect competition is monopoly. For a single monopoly, the demand for labour (the MRPL curve) is by definition below the VMPL curve, and the ratio between these two variables gives a commonly used index of monopoly power. When all firms in a sector are monopolies, the curve showing the total demand for labour will be correspondingly below the horizontal summation of the VMPL curves. When some firms 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). Deadweight loss is likely to be accentuated if the degree of monopoly is greater in one sector than in the other, which is to be expected as between the formal and informal sectors. Any differences in the degree of market imperfection between the sectors will interact with other imperfections, like an inefficient SS system or a MW, with the serious possibility that imperfections will reinforce each other to create large deadweight loss. An extreme case of differing market power between the two sectors would arise where a set of monopolies make up the formal sector and a set of competitive firms and industries make up the informal sector. In this case, too few workers and all other factors as well will be located in the formal sector, leading to potentially large deadweight loss, whose magnitude depends inversely on the elasticity of demand for the monopoly goods.

An illustrative feel for the sort of outcome that 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, for example) can be obtained by treating the group of monopolists as if they were just one firm, without monopsony power in its labour and other inputs markets. The socially optimal allocation of labour between the two sectors is then at point E of Figure 5 where the two VMPL curves intersect. But since the monopolist's demand curve for labour is $MRPL_f$, equilibrium 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)

⁶⁰ The "theory of the second best" applies here; its general message is that whereas simple microeconomic theory (of the sort found in textbooks) can say useful things about the impacts of a market imperfection when it is the only one from which the economy suffers, such theory is much less useful (if useful at all) in predicting the impact of additional imperfections when one or more already exist (Lipsey and Lancaster, 1956-57).

sector. The dead weight loss is EFG, wages in both sectors are at W_m instead of W_e .^{61 62}



The imposition of a minimum wage on the monopoly sector will further curtail output and employment, leading to a further decline in wages in the informal sector.⁶³ The same 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 is likely in this case, and it therefore suggests a high level of labour misallocation, it is often tempting to lay the blame (or

⁶¹ 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.

⁶² 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.

⁶³ 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.

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 *and/or the exercise of* monopoly power in the formal sector can make that sector smaller,⁶⁴ perhaps markedly smaller, than it should be, and sorting out their relative roles in producing this outcome becomes important. In the presence of market power in product or other (than labour) input markets, one must countenance the possibility that the main 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.⁶⁵

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 at a point of time 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 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 a great deal of 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. Ambiguity with respect to the degree of monopoly power in an economy implies that it is also hard to predict how much monopoly and other forms of market power are likely to affect labour demand.⁶⁶

⁶⁴ Monopoly power has a direct impact on sales, employment and other inputs used, since limiting supply is necessary to keep prices up. It also provides firms with a wider range of choice among technologies since they are less impelled to maximize profits than if they were pure competitors—they may thus be more akin to “satisficers” (Baumol, 1959); with easy access to capital either from own profits or through credit, they are likely to opt for more capital intensive technologies, which will further reduce their demand for labour.

⁶⁵ Monopoly power can be based on bureaucratic regulations, on economies of scale, on superior technology (e.g. one that has not yet been widely diffused 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 of either speculation in assets markets or 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 national monopolies with larger multinational oligopolies which allocate many resources to developing and maintaining market power in each country where they operate. This makes it hard to judge whether overall market power has risen or fallen.

⁶⁶ 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. If one assumed that half of the allegedly labour income accruing to the top 5% of the personal income distribution in the US as of 1998 were in fact disguised capital income, then the total capital (i.e. non-labour) share would be about 10 percentage points higher as a share of total income, since that top 5% have about 30% of total income (Piketty and Saez, 2003, 10) and about two-thirds of it is reported as labour

Appropriate Policy Response

When it is high levels of market power rather than things like labour regulations that explain the smallish size of the formal sector, its expansion through the entry of additional firms may not occur unless that power is directly targeted by policy. Where bureaucratic barriers to entry (like licences with no logic in economics) are the problem, the simple answer is to get rid of them. However, when the market power has some other origin, its contribution to deadweight loss and to low formal sector employment are unlikely to be much affected without a direct attack on the basic problem--the exercise of the monopoly power and the associated low labour absorption. When that power rests on economies of scale, the theoretically appropriate step to reduce deadweight loss is to regulate the firm, for example through price controls, though this may not do much for employment creation. When it rests on unfair market practices, these should be outlawed, with the preferred objective of facilitating entry of small and medium firms that will be employment creating. Increasing formal sector employment may be impossible if its limited 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.

Under these conditions raising informal sector labour productivity may be the best way to raise incomes at the lower end, in particular when the informal sector firms involved are sufficiently independent of the monopoly sector. But firms with market power 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. A general equilibrium analysis is usually necessary to judge what informal sector developments would benefit those employed there.

If the predominant market structure is *monopolistic competition*, the marginal revenue product (MRPL) of one more worker is again below the VMPL of the firm (reflecting its monopoly aspect), but in equilibrium the representative firm has zero abnormal profits. 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 damaging about 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 formal sector monopoly confronts informal sector perfect competition, discussed above.

Monopolistically competitive markets, like much of retail commerce, may be expected to provide a more elastic demand for labour than do other non-pure competition 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 easy or natural product differentiation. It is a structure that 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 it does at least constitute a sort of last resort earnings-

income (ibid, 16).

sharing scheme.

The implications of monopoly or market power in an economy depend not only on the type of market power, as just discussed, but also 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 bulk of 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 monopoly. 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, then the whole system (chain) 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 serious problem if the exercise of that power severely curtails formal employment (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 found there may be essentially competing against each other for a modest share of the economy's aggregate demand, in which case any increases in productivity that they achieve may simply lead to lower prices that benefit the formal sector monopolists.

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 providing certain types of firms with better access to capital and savings opportunities than others have. A general presumption is that the informal and usually small-scale sector is relatively deprived of capital and that large firms have good access.⁶⁷ Meanwhile, labour is disproportionately located in the informal sector, partly due to labour market imperfections discouraging employment in the formal sector. Factor allocation between formal and informal sectors is thus distorted, with too much capital in the formal sector and too much labour in the informal one.⁶⁸ This leaves it theoretically ambiguous whether there is too much output in the formal

⁶⁷ 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 typically 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 somewhat (but probably only a little) less sensitive to the degree of misallocation of labour than it would otherwise be.

⁶⁸ Note that when the movement of both labour and capital between a formal sector and an informal one are taken

or the informal sector. Clearly, the optimal scenario from a static efficiency perspective 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 presence of 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 composition inefficiency within each of the two sectors.⁶⁹ 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 the capital market imperfection is a given (i.e. is not subject to policy) its presence does not directly affect the analysis of the effects of SS, MW, etc. in any qualitative sense. It does contribute to a VMPL_f curve with a high vertical intercept and a steep slope, and thus affects the overall equilibrium reached in the economy. Since the elasticity of labour demand in the formal sector is a determinant of the deadweight loss associated with movement of workers between sectors, this capital market imperfection does increase 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.

Conclusions from the two-sector static 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 from a SS system (additional to any design inefficiency loss from that system depends on there being some actual or potential segmentation 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 play a role, and on the variance in absolute and relative productivity among workers.⁷⁰
- iii) product and capital market imperfections, apart from the deadweight loss to which they give rise directly, are likely to decrease the labour misallocation loss directly attributable to distortions like those under discussion (MW and SS), because both of them typically contribute to an inelastic formal sector demand for labour. Interventions like MW and SS do tend to amplify the loss from these underlying imperfections, but under some circumstances there may be only a marginal payoff to worrying about labour market imperfections when the ones that really count are elsewhere. Case by case analysis is essential to arriving at accurate diagnosis and good policies.

Case 6. Beyond the Two-Sector Static Model

into account, formality itself becomes endogenous in the sense that firms can move in either direction when a specific factor market becomes more competitive.

⁶⁹ This inefficiency results from the underpricing of capital relative to labour in the formal sector and the opposite distortion in the informal sector.

⁷⁰ The impact of this last factor may, however, be more complex than discussed here.

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. While this practice has some heuristic value as a first level of discussion, it is inappropriate and potentially quite misleading as a base for serious policy conclusions. Nearly 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. As soon as the conceptual framework goes beyond a sectoral dichotomy to a larger number of groups, the analysis changes in a qualitative sense; it is no longer automatically the case, for example, that if one sector is too big for the overall efficiency of the economy another one must be too small. It now requires that one know both which sector is too big and at the expense of which other sector. More knowledge is required in order to judge which sectors should grow and whether and how this may best be done.

Distinguishing a small number of categories in terms of which to frame the analysis can be useful; the questions are, how many categories to distinguish and how to draw the most meaningful 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 or two. The optimal dividing lines are also likely to depend on the economic structure of the country. Meanwhile, use of 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 to the high levels of risk and uncertainty they often face, it is clear that a static model may not be adequate.

Consider first the number of sectors to be distinguished. A three-sector approach to some of the issues related to SS and SP is of obvious interest since a common and useful classification of firms and the workers they hire involves three groups—microenterprise (or informal) firms, small and medium enterprise and large enterprise.⁷¹ The first group is mainly composed of family enterprises (in the sense that most or all workers are family members) and size is usually 1-4 or so workers; small and medium enterprises are larger (by many definitions 5-100 or so workers) and although most workers are not family members the organizational structure is still simple, with one person often being responsible for all of the major managerial tasks; large enterprise is structurally more complicated and some of the firms are incorporated. The nature of product competition and other forms of interaction between each pair of groups tends to differ. In the two-sector analysis used thus far, a labour tax that raises wage costs in the formal sector increases the size of the informal sector with the efficiency loss described. But in a three sector analysis⁷² 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.⁷³ Because this will likely reduce total labour absorption in these two sectors taken together, more workers will shift to the informal sector as an

⁷¹ See the large literature on the “missing middle”, e.g. Mazumdar and Sharkar (2008).

⁷² The distinctions among sectors in a model can reflect the fact that they are internally different from each other and/or the fact that policy treats them differently. Usually both these factors are at play, since policy is differently applied to sectors that are structurally different from each other. That is the case here.

⁷³ 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.

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, it is clear that to conceptualize the process as one in which the impact will be to discourage activity in the fully formal 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. Usually the real world 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 be based on this high labour productivity sector since that is not likely to be the shift that takes place.⁷⁴

A large literature, dating back at least 3-4 decades, has argued that the intermediate technology firms (typically small/medium in size) often 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 contestable and the truth must vary considerably from country to country and from setting to setting, 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. In other words, even if TFP is not higher for this group than for larger, more modern firms, its impact on the labour market outcomes is almost certain to be more positive.

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 specific question being asked, the formal sector is not a very heterogeneous one, the analysis can be folded back into that simpler two-sector framework. But it should not start there.

Unsurprisingly, much of the movement of labour into and out of the informal sector appears to involve 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 and the informal sector. The analysis of the impacts of SS and of a MW should also be based on the SME sector whenever, as is usually the case, the policy issue mainly involves the response pattern of that sector.

The need for a three-sector conceptualization of the economy complicates the analysis of the impacts of any given policy intervention, from MW, to SS to SP. 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 complementarity,⁷⁵ and how these interventions affect the firms in their interactions with one or

⁷⁴ This sort of oversimplistic assumption was made in the studies undertaken by the McKinsey Global Institute (e.g. Farrell, 2004).

⁷⁵ An important mechanism which often involves two or even all three of the size-modernity categories is subcontracting. Incentives to induce large firms to undertake lots of sub-contracting can be useful.

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 markets work, this is likely to be most true of this factor market. Similarly, the characteristics of product markets may play a big role. Finally, an important and understudied aspect of the picture involves the degree and character of teh

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

Case 7. Allowing for risk, uncertainty, macroeconomic instability, and firm dynamics

As noted above, many informal and small formal sector firms are anything but “static” over time; risk and uncertainty play an especially large part in their lives and those of their workers. How do these facts affect the likely implications of SS and SP systems on employment and on economic efficiency? How should they 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 continued existence and success; and their need to be risk averse in order to minimize the likelihood of failure/collapse. One implication of these characteristics is that getting locked into various aspects of formality is risky for any firm which faces a significant chance of needing 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; the risk to a firm from becoming more formal is correspondingly 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 uncertainty. One is the rules on worker dismissal. Under certainty (with respect to worker performance, macroeconomic evolution and the firm’s overall future), these are just one more determinant of average labour costs. However, when there is uncertainty on those or other 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 of 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, more complicated and possibly more negative; in any case, it warrants analysis in its own right.

Case 8. Formalization Policies

The above discussion has considered how a variety of market imperfections, government intervention and other factors affect the allocation of labour (mainly) and other factors (secondarily) between a formal and an informal sector, or between more formal and less formal points on the formality spectrum. Much of the literature around policies like SS and MW has its origins in the concerns that such policies will discourage formal sector firms from hiring as much labour as would be socially optimal. It is static in character, and focuses on how these policies affect resource allocation. A different literature focuses on the question of whether too few small informal firms graduate into the “formal” sector by growing and adopting the aspects of formality like paying taxes and satisfying labour legislation (including SS). This literature shares with the first one the position that too few resources are in the formal sector, but focuses on why firms

opt for informality rather than formality and what can be done about this. Naturally, some of the concerns about elements of formality that discourage employment of existing formal firms (like a high MW) also come into the interpretations of why informal firms do not become formal voluntarily. Those who argue the desirability of more formalization than typically occurs and believe this process should be as voluntary as possible are likely to concur with those who feel that SP *per se* will discourage desirable formalization.⁷⁶

Case 9. Integrating the Complexities: An Illustrative Possible Scenario within which to analyze the impacts of SP

At a theoretical level, the number of permutations of assumptions that can be considered as possible scenarios within which the effects of SS and SP would play themselves out in a context including a sizable informal sector is excessively large and leads to quite varying results according to which are chosen. Accordingly, it is important to consider what the theory implies under the more realistic sets of assumptions. For many countries, this set would include:

- i) three sectors distinguished by the positively correlated features of formality, technology and size, i.e. large, formal; SME, partially formal; small, informal;
- ii) a SS system and MW that are applied systematically in the large/modern sector, not applied in the very small/informal sector and partially applied in the intermediate/small-medium sector
- iii) institutional disincentives to formal sector labour hiring, such as the just cited SS and MW and possibly other elements of the labour regulatory package, 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, with access improving the higher a firm is 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 groups of firms across the different categories, coexisting with some direct and indirect competition across the categories;
- vii) varying degrees of competition from imports in the domestic markets of the goods and services produced;
- 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 its 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.⁷⁷ Labour

⁷⁶ For a detailed discussion, see Berry (2011).

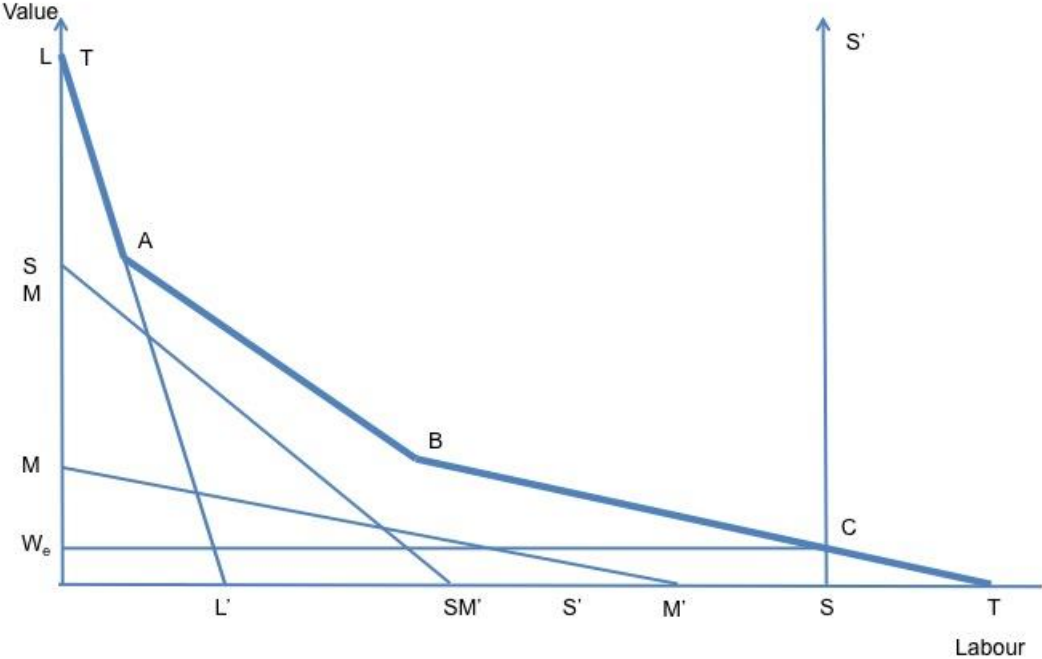
⁷⁷ Numerous econometric studies have estimated the elasticity of labour demand and/or the impact of changes in labour regulations. Unfortunately, few have attempted to isolate different groups of firms, e.g. distinguish between larger more technologically modern formal firms and smaller less modern ones. 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.

regulations create a degree of disincentive to formal sector employment; this effect may be small in the larger, more modern sector but important in the smaller partially 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.

A simple way of portraying the role of the three groups of firms in overall labour demand is illustrated in Figure 6. The labour demand curves of larger, more modern firms start higher than those of smaller, less modern firms but are also steeper (less elastic). At the other end of the spectrum is the microenterprise sector which has a very flat labour demand curve--the curve starts low since not even for a very small labour input is productivity high. The aggregate labour demand curve of large enterprise (LE), small and medium enterprise (SME) and microenterprise (ME) are portrayed as curves LL', SMSM', and MM' respectively. The total demand curve (TT') for labour is the horizontal summation of these three curves, shown as the heavy line. Most of the

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.

Figure 6



potential employers at very high wages on this total demand curve are modern firms, the bulk of the middle of the TT' curve corresponds to the demand of SMEs, and the majority towards the bottom of it are

quite small, low technology microenterprises. Were there enough of the complementary factors (capital, natural resources) to generate a TT' curve far enough to the northeast in the figure to cut the labour supply curve (SS') at a high wage this would of course be desirable; this is the situation in developed countries where the main component of the TT' demand curve corresponds to the relatively capital intensive, high-technology firms which make up the modern sector. In developing countries, the availability of complementary factors is too limited and hence the size of the modern sector too small for this outcome to emerge. Under these conditions, a country which allocates a high share of capital to the very modern sector is likely to have the majority of the population working with very little capital and hence low labour productivity (demand for labour). The equilibrium wage would be quite low; workers in the modern sector would probably be able to bargain away some of the profits in that sector through labour legislation or collective bargaining. Another economy, endowed with the same amount of capital and other non-labour resources (and the same number of workers) but which allocated more of the capital to the SME sector rather than the LE sector, would have a labour demand curve which was lower for small quantities of labour (i.e. at points closer to the vertical axis) but higher for larger amounts of labour; it would therefore normally have a higher equilibrium wage than the first economy, though perhaps less very high wages of the type resulting from favoured workers bargaining away some of the very high profits of favoured sectors. In most cases the lowest part of the demand curve is in effect a demand for own-labour on the part of low income self-employed people. The paid wage rate for hired workers in LE and in SME will be higher than these individuals can generate as own income.

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 middle sized intermediate-technology firms) then total formal sector demand for labour is likely to be small and inelastic. 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 technology sector. With respect to this goal, it may be very important whether certain labour regulations constrain the growth of this particular set of firms. It may be equally or more important that policies can contain the monopolistic practices of the large-scale formal sector if these are the main impediment to the growth of smaller formal 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.

Case 10. Adding Social Protection Policies (SPs)

Concern has been expressed that social policies with widespread or universal coverage (e.g. old-age pensions, welfare systems, etc.), when applied in the presence of already existing imperfections like an badly designed SS system or labour legislation like a MW , may have negative effects by discouraging formalization, lowering GDP and cutting formal and total employment (e.g. Levy, 2008; Chapter 8).⁷⁸ The appearance of SP programs changes the net benefits implicit in having a formal sector job and the

⁷⁸ When the SP in question is the only “imperfection” in the system (e.g. because SS does not introduce one and there is no MW), the likely deadweight loss resulting will usually be small unless the SP payments are implausibly large. If the $VMPL_f$ and $VMPL_i$ 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, the loss will be greater.

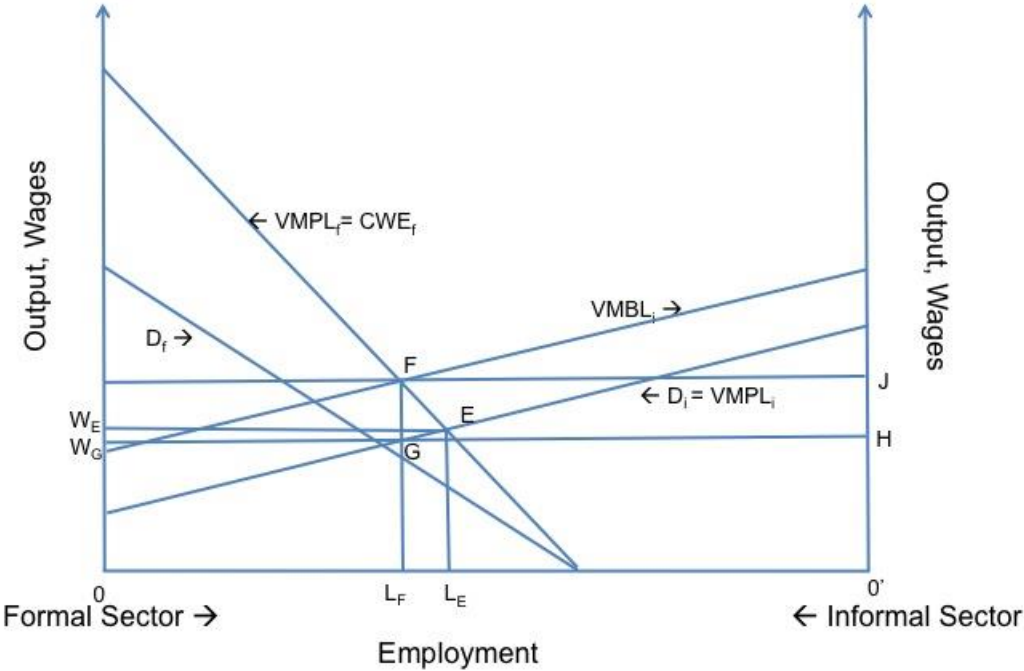
associated SS package and can thus have a variety of impacts, many of which have been identified in the literature. In what follows we review briefly the efficiency and other impacts that may be expected under various of the scenarios discussed above.

Case 10.1. Two sectors; the SS system is efficient (all formal workers have $P=1$); there is no MW; access to SP is contingent on working in the informal sector.

In this simple case where access to SP is contingent on working in the informal sector, 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 7a the curve $MVBL_i$ shows the value of the total benefits from informal sector employment, the sum of earnings (given by $VMPL_i$) and the SP benefits.

With the SP as the only source of misallocation of labour between the two sectors (since the SS system introduces no inefficiency because of poor design), the deadweight loss is likely to be small in relation to the gross SP benefits and hence also stands a good chance of being less than the net SP benefits if the SP program is well designed. In Figure 7a) the pre-SP equilibrium is point E, at the intersection of the $VMPL_f$ (or CWE_f) and $VMPL_i$ curves. The SP program raises the marginal benefits from involvement in the informal sector to curve $MVBL_i$ so the equilibrium moves to point F, formal sector employment falls from L_E to L_F , and there is a misallocation deadweight loss of FEG. The equilibrium wage in the formal sector rises to W_F , as do the equilibrium labour-based benefits in the informal sector, though the actual wage there falls to W_G . Total earnings of formal sector workers plus those engaged in the informal sector rise significantly, at the cost of formal sector capital incomes and the taxpayers who finance the SP.

Figure 7a

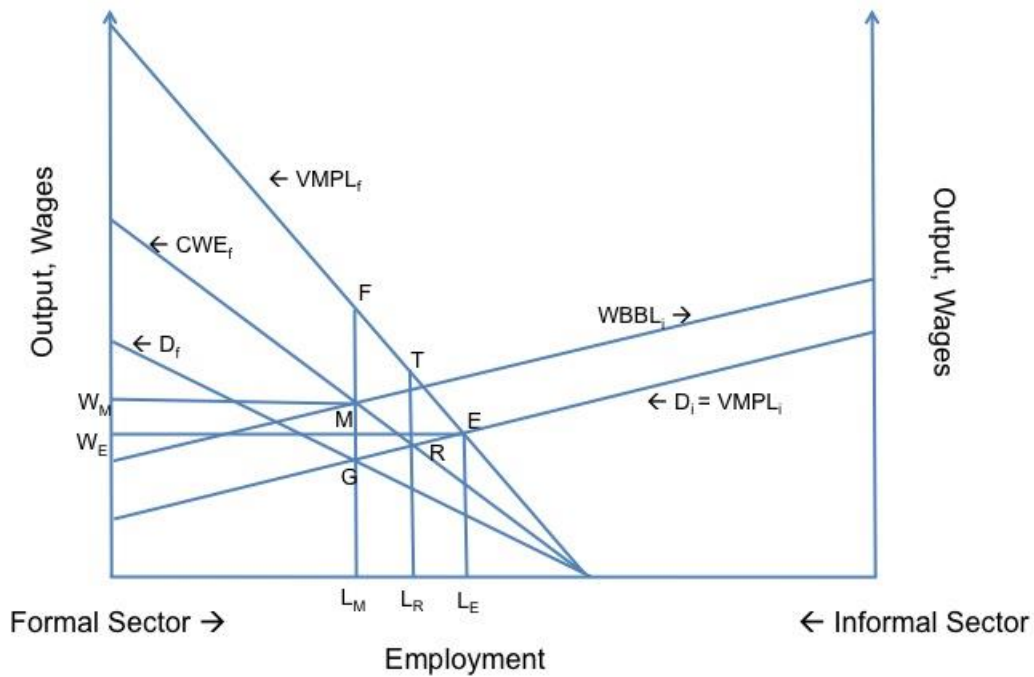


The total outlay on the SP program is FJHG; if the benefits accruing exceed the cost by more than the ratio of FEG/FJHG, then society as a whole has gained. If distance LFLE is say one tenth of LFO', this ratio would be one twentieth; it would thus be quite feasible that the benefits from the SP exceed the costs if it were reasonably well designed. This comparison is, however, conceptually more complicated than that involving SS, where workers pay and get the benefits. SP involves an income transfer so the necessary comparison is between the welfare lost by one group and that gained by a different group. If the marginal utility of income is higher for the receiving group (the underlying logic of SP programs in the first place) the point here is that it may not have to be much higher to lead to a static welfare gain for society. The main worry might be that there would be growth costs if the required taxation and/or lowering of formal sector profits cuts investment too much.

Case 10.2. Two sectors; the SS system is inefficient (all formal workers have $P=0.5$); there is no MW; access to SP is contingent on working in the informal sector.

When, as portrayed in Figure 7b (similar to Figure 2d) but with the SP benefits taken into account through curve $VMBL_i$, equilibrium occurs at point M, where CWE_f intersects $MVBL_i$. In the presence only of the formal sector SS system, equilibrium was at point R (intersection of CWE_f and $VMPL_i$) and formal sector employment was $OL_{R,i}$. With the addition of the SP program more workers (L_M L_R) now shift to the informal sector, leaving just OL_M in the formal sector. Disregarding the impact of SP *per se* on the benefiting workers as a separate matter, there is now a total welfare loss due to overproduction in the informal sector equal to area FEG, up from the previous level of TRE. If the elasticity of the relevant formal sector curves is less than unity in the relevant range, this increase of FTRG in output loss will be partially offset by a decrease in the loss associated with the imperfectly designed SS bundle (as discussed above in the context of Case 3a). The overall impact of the policy on societal welfare depends also on the direct benefits of the SP expenditures. If SP benefits have the same weight in the social utility function as the benefits given up to fund them, then area JGN plus RFK measures the net loss from the policy. Unless the SP program is badly designed, its benefits will exceed its direct (or opportunity) costs; if this difference exceeds the just-cited marginal deadweight loss it causes due to resource misallocation, then a net social benefit accrues.

Figure 7b



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,⁷⁹ then the shift towards the informal sector has both the effect of lowering GDP through misallocation of labour between the sectors and of raising the overall efficiency of social spending. Thus, the GDP loss from such shifts will be either partly, wholly or more than wholly offset by the gain from that greater social spending efficiency. Presumably the best outcome of all, of course, is where the SS system is reformed as it enters into competition with the SP system.

A situation of special interest is that in which the formal sector SS system is subsidized by the state. In that case, in the absence of other distortions, that subsidy will have led to an above optimal sized 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 formal sector SS system causes some loss if P is under 1 and/or if there is a state subsidy to the SS. If only the latter, and $P = 1$, then some SP (equal to the subsidy to SS) will tend to rectify the deadweight loss.

⁷⁹ Either in absolute terms or, if total labour supply falls, in relative terms.

If $P < 1$ and there is no SS subsidy, the resulting shift of workers into the informal sector is socially efficient since, as discussed above, GDP is the wrong indicator of societal welfare; it should be GDP minus waste associated with the SS policy. When SP arrives on this scene, it encourages more workers into the informal sector, further lowers (incorrectly measured) GDP but also lowers the value of output net of SS-inefficiency losses.

Case 10.3. Two sectors; the SS system is inefficient (all formal workers have $P=0.5$); there is no MW; access to SP is not contingent on working in the informal sector.

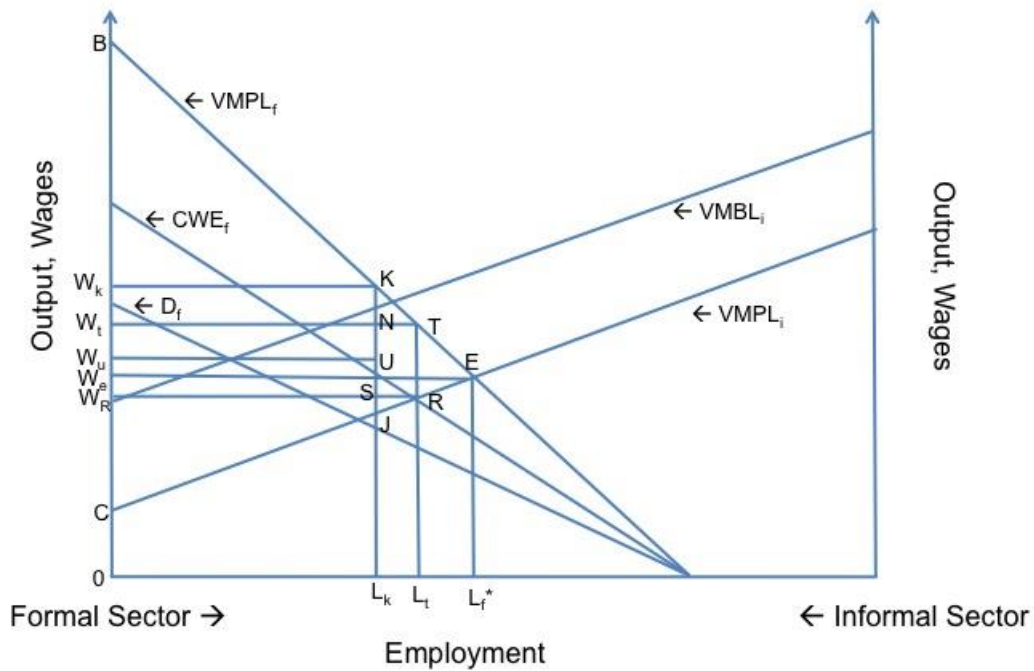
When the SP payment is not contingent on employment in the informal sector but remains either unavailable or uninteresting to those working in the formal sector, the main likely effect in the labour market is that the supply of labour will decrease. The welfare impacts of such a shift cannot easily be shown in the simple figures used above (where a perfectly wage-inelastic total 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 $VMPL_i$ curve will now be closer to the leftward axis, the equilibrium point will also shift to the left and employment will fall in both sectors, with wages rising in both. 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. 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. A likely scenario is that those who drop out of the labour supply as a result of the SP system are people who would otherwise be working due to poverty.

Case 10.4. Two sectors; the SS system is inefficient (all formal workers have $P=0.5$); there is a MW; access to SP is contingent on working in the informal sector.

The above discussion leads us back to an earlier proposition: the sort of quantitative shift that is likely to occur toward the informal sector will probably 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 true. 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-selection of workers into the formal sector infeasible. Empirically speaking this makes it important to know to what extent such self-selection does operate.

Consider a modification of the case portrayed above in Figure 3a, where the SS system is inefficient and a MW is imposed with resulting deadweight and welfare loss and with queuing for entry into the formal sector. If to this scenario is added an SP contingent on working in the informal sector (See Figure 7c), the marginal benefits curve from informal activity rises to $MVBL_i$. In the absence of the MW, this SP would have led to a reallocation of labour toward the informal sector but in the presence of that MW and the queuing that results, it will not, unless it is quite high, have any effect on labour allocation, and hence on GDP. It will simply raise incomes in the informal sector and lower those of whomever pay the taxes to fund the program. It will lower the excess supply of labour available to the formal sector at the MW from HT to HV . This outcome is generalizable to all situations in which the pre-SP level of excess supply or queuing for formal sector jobs exceeds the amount by which the $MVBL_i$ curve lies to the left of the $MVPL_i$ curve.

Figure 7c



Since the impact of an efficiency wage system bears considerable similarity to that of MWs that lead to queuing, it is of interest to consider whether the effects of an SP may be absent in that case as well. The answers will depend in part on which factors led to the existence of the efficiency wage system. If it is primarily the benefits to hiring firms from paying their workers more than they can attain elsewhere, that benefit will probably be at least partially eroded by the higher incomes now attainable through informal sector employment, so the SP may lead to upward pressure on formal sector wages and a resulting decrease in employment.

Overview for the two-sector perfect competition cases

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. See below, however, for a discussion of a three-sector model, a more plausible conceptual framework within which to analyze these issues.

Case 10 5 Approximation to Reality: Three sectors; product and capital market imperfections

No attempt is made here to consider how the existence of SP will affect labour market and other outcomes under each of the more interesting possible scenarios described above. Rather, several general points are made.

First, there is a significant likelihood that SP policies, whether linked to informal sector workers or not, will have very little effect on fully formal sector employment and hence on deadweight loss related to a less-than-optimal level of employment in that sector. This outcome can be due to one or more of: low labour demand elasticity (which may be due to non-competitive market structures), queuing due to a MW, or other factors at work that prevent the SP from having an effect on the allocation of labour to this sector. This means that the main determinant of how SP affects outcome variables will be its impact on the allocation of resources between the informal and the SME partially formal sectors, where SS and MW are only partially applied, and where market power usually takes the form of monopolistic competition.

Second, the lack of information available on the *modus operandi* of the small-medium scale partially formal sector leaves open many possible outcomes. At one end of the spectrum, the SP may make it harder for SMEs to remain competitive and grow, for example if it raises the wage rate these firms must pay, and it may make formalization less attractive if the SP has to be given up when that occurs. At the other end of the spectrum, some types of SP may raise informal firm productivity and earned incomes without having the just cited negative effects. They may facilitate informal firm growth into SMEs, whether more or less formalized. 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. Thus, it is possible that a SP implemented together with, say, a 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 there by former formal sector workers as they shift to the informal sector) this makes the SP policy more beneficial than it would otherwise have been.

Third, in principle it is possible that, with SS and the MW applied to middle-sized intermediate technology firms there would be queuing for jobs there and that the efficiency cost of this would be considerable because of a relatively elastic demand for labour. But, as with the fully formal sector, if there is queuing to enter this partly formal sector, SPs are likely to have little or no impact on the intersectoral allocation of resources.

In short, it is pivotal to have a much more detailed understanding of the dynamic interactions among the three sectors, but especially between the informal firms and the smaller partially formalized ones, in order to predict with any accuracy the likely impacts of SPs of different types.

From a methodological perspective, a key point is that 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 that:

- i) capital market imperfections normally imply 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) product market imperfections normally to imply that too few of all resources are being used in the less competitive components of the formal sector but that attempts to correct this

- misallocation through labour market, SS or SP policies are likely to be very much a second-best approach.
- 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.

Summary.

A main message of the above discussion is that the analysis of the impact of SP policies on labour allocation and overall economic efficiency is a very complicated matter, which cannot be safely analyzed in a two sector (formal-informal) model, nor in one that assumes labour homogeneity or perfect competition in the product and capital markets. None of these assumptions approximates reality and each can lead to erroneous conclusions. The less satisfactory message is that analysis of any given case is likely to be complicated enough to pose serious barriers to the attainment of reliable conclusions, even qualitative let alone quantitative ones. Nonetheless, such analysis is worthwhile; through better understanding of those aspects of economic reality to which the results are particularly sensitive (such as labour heterogeneity) it will be possible to move towards more useful results from a policy perspective.

For labour market functioning to have direct policy implications on the desirability of SP, it has to be the case that labour allocation between the two (or among the three) sectors will respond to the presence of SP—true for example in non queuing cases where the elasticity of formal demand for labour is not close to zero elasticity.

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, it is essential to remember the GDP ceases to be the right aggregate welfare indicator under many of the settings to which SSs and SPs give rise.

References

Basu, Kaushik and Amanda J. Flekey (2008) "A theory of efficiency wage with multiple unemployment equilibria: how a higher minimum wage law can curb unemployment" *Oxford Economic Papers*, Vol. 61, pp. 494-516.

Baumol, William J. (1959) *Business Behavior, Value and Growth*. New York: Macmillan.

Berry, Albert (1969) "A Note of Welfare Comparisons Between Monopoly and Pure Competition", *The Manchester School*, Vol. 37, pp. 39-57.

Berry, Albert (2011) "To Formalize or not to Formalize: Trying to Ask the Right Questions." Unpublished manuscript.

Cortes, Mariluz, Albert Berry and Ashfaq Ishaq. *Success in Small and Medium-Scale Enterprises: The Evidence from Colombia*. Oxford and New York ????: Oxford

University Press, 1987.

Dixit; Avinash K. and Joseph E. Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity." *The American Economic Review*, Vol. 67, No. 3, pp. 297-308.

Dore, Ronald (1976) *The diploma disease: education, qualification, and development*. Berkeley: University of California Press.

Farrell, D. (2004) The hidden dangers of the informal economy. *The McKinsey Quarterly* 3.

Harberger, Arnold C. (1954) "Monopoly and Resource Allocation" *American Economic Review*, No. 44, No. 2, pp.77-87.

Harris, John and Michael Todaro (1970) "Migration, unemployment and development: a Two-sector analysis" *American Economic Review*, Vol. 60, No. 1.

Heckman, James J. and Carmen Pagés (editors) *Law and Employment: Lessons from Latin America and the Caribbean*. Chicago and London: University of Chicago Press.

Lewis, W. Arthur (1954). "Economic Development with Unlimited Supplies of Labor,". *Manchester School of Economic and Social Studies*, Vol. 22, pp. 139-91

Lewis, W. Arthur (1978). *The Evolution of the International Economic Order*. Princeton, New Jersey: Princeton University Press.

Levy, Santiago (2008) *Good Intentions, Bad Outcomes: Social Policy, Informality and Economic Growth in Mexico*. Washington, D.C: The Brookings Institute

Lipsey, Richard G. and Kelvin Lancaster (1956-57) "The General Theory of the Second Best." *The Review of Economic Studies*, Vol. 24, No. 1, pp. 11-32.

Lipton, Michael (1977) *Why Poor People Stay Poor: Urban bias in world development*. Cambridge: Harvard University Press.

Mazumdar, D. 1959. The marginal productivity theory of wages and disguised unemployment. *Review of Economic Studies* 26: 190-197.

Mazumdar, Dipak and Sandip Sharkar (2008) *Globalization, labour markets and inequality in India*. Abington; New York: Routledge.

Perry, Guillermo E., William F. Maloney, Omar S. Arias, Pablo Fajnzylber, Andrew D. Mason and Jaime Saavedra-Chanduvi (2007) *Informality: Exit and Exclusion*. Washington, D.C. The World Bank

Piketty, Thomas and Emmanuel Saez (2003) "Income Inequality in the United States: 1913-1988" *Quarterly Journal of Economics*, Vol. CXV111, Issue 1.

Reynolds, Lloyd G. (1969) "Economic Development with surplus Labor: some Complications" *Oxford Economic Papers*, Vol. 21, No. 1, 89-103.

Shimer, R. and L. Smith (2001) "Matching, search and heterogeneity" *B. E. Journal of Macroeconomics*, 1(1), Article 5.

Woodcock, Simon D. (2010) "Heterogeneity and Learning in Labor Markets" *B. E. Journal of Macroeconomics*, 10 (1), Article 85.