South African Journal of Economic and Management Sciences
ISSN: (Online) 2222-3436, (Print) 1015-8812

A high unemployment and labour market segmentation: A three-segment macroeconomic model

Authors:
Philippe Burger[1]
Frederick Fourie[1]

Affiliations:
[1]Department of Economics, University of the Free State, South Africa

Corresponding author:
Philippe Burger,
burgerp@ufs.ac.za

Dates:
Received: 17 Sept. 2017
Accepted: 11 Oct. 2018
Published: 25 Feb. 2019

How to cite this article:
Burger, P. & Fourie, F., 2019, 'A high unemployment and labour market segmentation: A three-segment macroeconomic model', *South African Journal of Economic and Management Sciences* 22(1), a2103. https://doi.org/10.4102/sajems.v22i1.2103

Background: South Africa suffers from an unusually high unemployment rate – officially averaging 25% since 1999Q3. In addition, depending on whether one uses the official or broad definitions of unemployment, since 2008 there are on average between 2 and 3.3 times as many unemployed people as there are people in the informal sector. Hence the question: why do the unemployed not enter the informal sector to create a livelihood?

Aim: To fill this gap we propose a macroeconomic framework that incorporates both formal (primary) and informal (secondary) sectors, as well as involuntary unemployment resulting from entry barriers to the labour market. We believe such a model provides a more suitable basis for macroeconomic policy analysis.

Setting: Standard macroeconomic theories at best provide a partial explanation for the South African unemployment problem, focusing mostly on the formal sector.

Methods: The article uses a theoretical analysis.

Results: The article presents a macroeconomic framework that incorporates both formal (primary) and informal (secondary) sectors, as well as involuntary unemployment resulting from entry barriers to the labour market.

Conclusion: If the assumptions on which the model draws hold in the South African reality, then a solution to the unemployment problem involve policies addressing product and labour market structures and behaviour in the primary sector, as well as policies addressing the numerous barriers to entry, such as borrowing constraints, that potential entrants into the secondary sector face.

Introduction

Few countries have as serious an unemployment problem as South Africa. In the period 2008Q1–2018Q2 the official unemployment rate averaged 25.1%, while the broad unemployment rate (which includes discouraged work-seekers) averaged 34.8% (StatsSA 2017). A well-known peculiarity of South Africa is that, compared to peer-group countries, the informal sector is small relative to total employment (Kingdon & Knight 2004). Almost 17% of employed workers are in the informal sector (see Table 1). In addition, and depending on which unemployment definition is used, since 2008 there have been between 2 and 3.3 times as many unemployed people as informal sector workers (Figure 1).

This raises the following question: if workers do not find employment in the formal sector, why do they become unemployed rather than enter the informal sector? Kingdon and Knight (2004) suggest that there are significant barriers to entry into the informal sector, possibly in the form of capital and skills shortages. South Africa is not the only developing country where barriers to entry into the informal sector appear to exist. Grimm, Krüger and Lay (2011a) and Grimm, Van der Hoeven and Lay (2011b) find significant barriers to entry into the informal sector of many West African countries, as well as Madagascar.

A characteristic of almost all the macroeconomic work on unemployment in SA is that it deals with the formal sector only (Fourie 2011). Meanwhile, evidence from unemployment research in the fields of labour economics and development indicate substantial segmentation in the South African economy: between the formal and the informal economies, within the informal sector, and between the unemployed and the informal and formal economies. Moreover, several labour market barriers exist that prevent people from improving their employment and earnings situation.

The objective of this article is to start bridging the divide between the macroeconomic discourse and the labour and development discourses on unemployment by developing a model that includes labour market segmentation and entry barriers into a theoretical macroeconomic model. A major result of this model is that, given these incorporated features, it explains the existence of persistent high involuntary unemployment in equilibrium.

The labour market component of mainstream theoretical macroeconomic models

Modern macroeconomic theory largely focuses on the formal sector, ascribing unemployment mostly to product and labour market imperfections, as well as hysteresis (see Cahuc & Zylberberg 2004; Carlin & Soskice 2006: chapter 15 for textbook expositions). *What these models do not consider or explain is why those who lose employment then become unemployed and not self-employed*. In the South African case we can expand this question and ask: why do the unemployed not enter the informal sector?

TABLE 1: Composition of the employed (% of total employment).

Year	Shares			
	Formal sector (Non-agricultural)	Informal sector (Non-agricultural)	Agriculture	Private households
2000	58.8	19.7	11.0	10.5
2010	69.5	16.7	4.9	8.9
2017	70.3	16.6	5.0	8.1

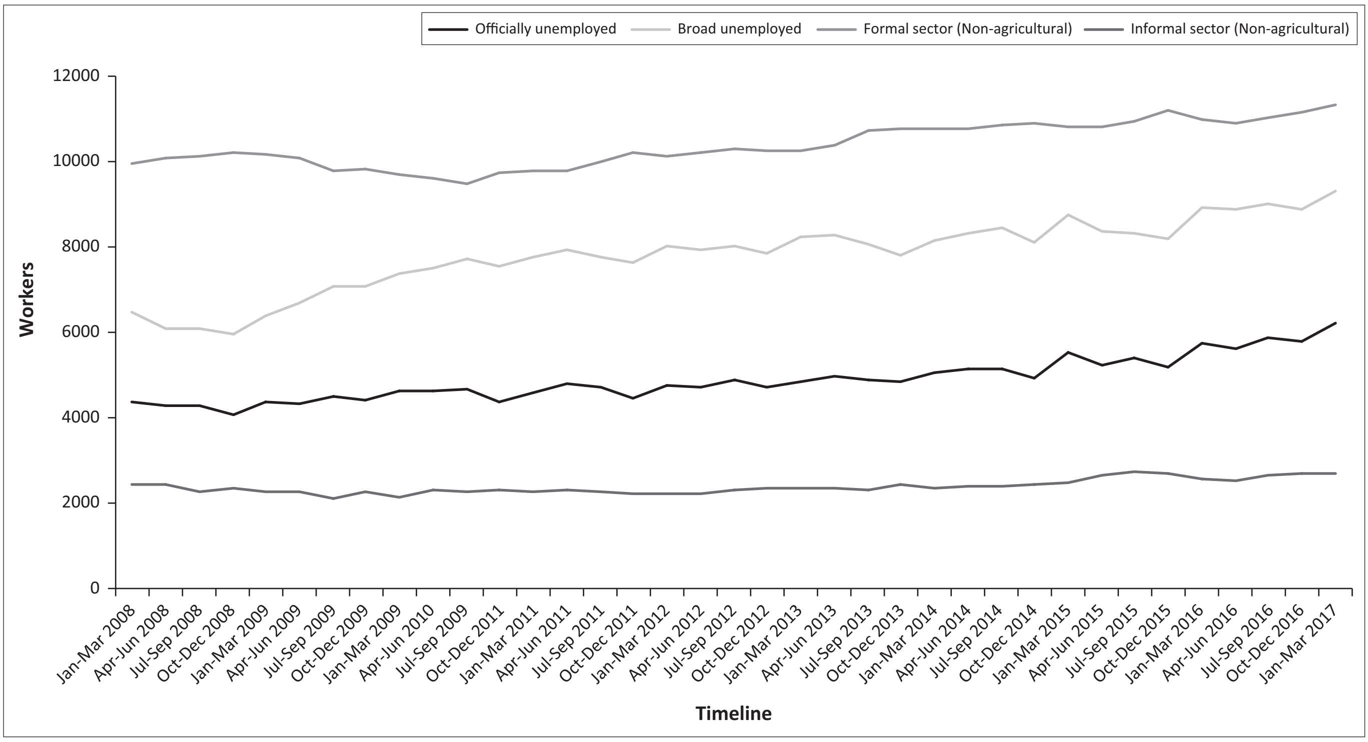
Note: Data for 2000 from the Labour Force Survey. Data for 2010 and 2017 from the Quarterly Labour Force Survey. All data refer to September of the relevant year.

Source: Statistics South Africa (StatsSA), 2009, *Labour force survey – Historical revision: September Series 2000 to 2007*, Statistical release P0210, Pretoria; and Statistics South Africa (StatsSA), 2017, *QLFS data*, Excel datasheet viewed 01 June 2017, from http://www.statssa.gov.za/?page_id=1854&PPN=P0211&SCH=7012

Discussing the informal sector draws segmented labour markets into the discussion. Agénor and Montiel (1999) present a theoretical model incorporating a formal and informal sector. Basically it represents a model with internationally traded and non-traded goods, with the former constituting the formal sector and the latter the informal sector. This model is of limited value in South Africa, as its informal sector, being largely retail-based, is a sector of traded goods.

Another branch of the literature represents the attempts by Layard, Nickell and Jackman (1991:41–44; also Blanchard 2005), as well as the earlier, but theoretically more detailed model of Bulow and Summers (1986). These models include a primary and a secondary sector. The primary sector typically has New Keynesian features (e.g. it is an efficiency wage or union bargaining sector). Excess primary-sector labour supply flows to the secondary sector. While the primary sector in this model is New Keynesian in nature, the secondary sector is surprisingly very New Classical. The secondary-sector labour market is assumed to be market clearing 'in the sense that wages are not high enough to attract a queue of job-seekers, nor do vacancies last long since skill requirements are low' (Layard et al. 1991:42).

Such a market-clearing secondary sector means that those who are not employed in either the primary or the secondary sectors are *both voluntarily and involuntarily unemployed:* they are 'involuntarily unemployed with respect to primary sector' at the going wage there, but simultaneously they are 'voluntarily unemployed with respect to the secondary



Source: Statistics South Africa (StatsSA), 2009, *Labour force survey – Historical revision: September Series 2000 to 2007*, Statistical release P0210, Pretoria; and Statistics South Africa (StatsSA), 2017, *QLFS data*, Excel datasheet viewed 01 June 2017, from http://www.statssa.gov.za/?page_id=1854&PPN=P0211&SCH=7012

Note: Data for 2000–2007 from the Labour Force Survey. Data for 2008–2017 from the Quarterly Labour Force Survey. All data refer to September of the relevant year.

FIGURE 1: The number of employed and unemployed workers ('000).

sector' (i.e. not willing to work at the going wage in the secondary sector). Thus, in the final instance they are voluntarily unemployed. Therefore, the Layard et al. and Bulow and Summers models still leave the question: If actual unemployment is high, why do those who become unemployed in the primary sector, stay unemployed and not all become (self-) employed in the secondary sector? Kingdon and Knight (2004), Grimm et al. (2011a) and Grimm et al. (2011b) suggest that those who wish to enter the secondary sector face significant barriers to entry, possibly in the form of capital and skills shortages. In the mathematical macroeconomic model presented below we incorporate some of these barriers.

A mathematical three-segment barrier model

This section develops a mathematical three-segment model for an economy such as that of South Africa (for more background and explanation, see Burger & Fourie 2015). The section draws on the dual labour market model of Bulow and Summers (1986), which itself is an augmentation of an efficiency wage model – a prominent approach in the New Keynesian class of models. Following Summers (1988) as well as Knell (2014), Pereau and Sanz (2006), Bulkley and Myles (1996) and the suggestion by Bulow and Summers (1986), the article introduces union bargaining into the model to allow for the presence of strong labour unions in the South African economy. Similarly, the presence of high economic concentration and imperfectly competitive product market conditions is an integral part of our augmented model. As mentioned, the Bulow and Summers (1986) dual labour market model explains the *allocation* of workers between the primary and secondary sectors – but not the existence of involuntary unemployment. Drawing on Kingdon and Knight (2004) and Grimm et al. (2011a), the model also incorporates barriers to entry into the informal sector to explain why the unemployed do not enter the informal sector, and remain unemployed.

Step 1: The two-sector model with no involuntary aggregate unemployment

We derive a formal-sector job-offer relationship and an effort supply function.[1] Different from the analysis in Bulow and Summers (1986), this analysis is done in terms of the number of *positions* filled by firms rather than the number of workers demanded, which allows the introduction of factors that will influence the number of positions being filled by firms in the two sectors. Nevertheless, the model is presented in terms of both the number of positions and the positions filled (persons employed).

In addition to these two relationships, the analysis below also presents wage-setting and price-setting relationships. These four relationships are then used to derive equilibrium conditions for the primary and secondary sectors.

The effort supply function

At any given moment firms in the primary sector fill a number of positions (jobs). The total number of jobs available in the primary sector is F_p. Those workers who do not obtain employment in the primary sector are accommodated in the secondary sector (which is assumed to be without entry barriers). In the secondary sector there is equilibrium: the total number of jobs filled is F_s. Thus, although there might be involuntary unemployment in the primary sector, there will not be involuntary unemployment at the aggregate level. The total number of filled positions in the economy (which in this case amounts to the entire labour force) is:

$$F = F_p + F_s \qquad \text{[Eqn 1]}$$

The allocation between the two sectors can be described in terms of the proportion of total positions filled by firms in the primary sector being $p = F_p/F$, while the proportion filled by firms in the secondary sector is $(1 - p) = F_s/F$.

A worker who quits or is laid off in the primary sector, is assumed to move to the secondary sector. The quit rates in the primary and secondary sectors are q_p and q_s; d_2 represents the probability of the worker being laid off when caught shirking (or e.g. low productivity[2]), while d_1 represents the probability of being laid off for shirking while not actually shirking (a false positive). Furthermore, w_p and w_s represent the wage rates in the primary and secondary sectors. Therefore, $(1 - q_p - d_1)w_p$ represents the expected wage of those workers employed in the primary sector (i.e. who have not been laid off and have not quit the primary sector), while $(q_p + d_1)w_s$ represents the expected wage of primary sector workers who are laid off in or quit from the primary sector and move to the secondary sector. (Shirkers are assumed to produce nothing, hence their $PV = 0$ and they are not included.) Likewise, $(1 - q_s)w_s$ represents the expected wage of those workers in the secondary sector who remain in the secondary sector, while $q_s w_p$ represents the expected wage of those workers who quit the secondary sector for the primary sector. Thus, the sum of the present value of expected primary and secondary sector income in the economy is:[3]

$$PV = \left[\frac{(1-q_p-d_1)w_p + (q_p+d_1)w_s}{r}\right]p + \left[\frac{(1-q_s)w_s + q_s w_p}{r}\right](1-p) \qquad \text{[Eqn 2]}$$

1. Concerning the microfoundations of the model, the model assumes a simple utility function, resembling the specification by Bulow and Summers (1986), with infinitely lived agents, where utility, *Ut*, is a function, *f*, of consumption and shirking (or 'non-effort'): $U_t = f(x_p, x_s + \alpha l)/r$, where *x* represents consumption of goods produced in the primary and secondary sectors (subscripts *p* and *s* denote the primary and secondary sectors). In addition, *l* is zero when the worker exerts effort and one if the worker does not exert effort. Non-effort is thus considered to be a consumption good, and it is substitutable for secondary sector goods. Furthermore, *α* is the instantaneous gain in utility from shirking/non-effort, while *r* represents the discount rate. Following Bulow and Summers (1986) we assume risk neutrality (so that $f(\lambda x_p, \lambda x_s) = \lambda f(x_p, x_s)$) and preferences are homothetic and normalised (so that $f(0,0)=0$).

2. For simplicity, quitting and being laid off are modelled to depend on shirking (insufficient work effort or productivity); other factors that determine quitting or being laid off can be modelled analogously. The simplification is not central to the main result of involuntary unemployment present in the full model, but merely facilitates it – involuntary unemployment will depend on the presence of barriers to entry into the secondary sector. Nevertheless, because it is commonly used in international literature, the shirking model is used here.

3. For reasons of simplicity Equation 6 assumes infinitely lived workers and as such uses the simple formula for the calculation of the value of a consol to calculate the present value.

In equilibrium, labour flows into and out of the primary sector need to be equal. Thus $p(q_p + d_1) = q_s(1 - p)$, so that $q_p + d_1 = q_s(1 - p)/p$. This equality also means that search for work in the primary sector occurs not from a position of unemployment, but from the secondary sector (in the two-sector model there is no aggregate unemployment).

Following Bulow and Summers (1986), we define an *effort supply function*. The effort supply function is stated in terms of α, defined as the instantaneous gain in utility from not exerting effort, as follows:

$$\alpha \leq (d_2 - d_1)\,(PV_p - PV_s) \qquad \text{[Eqn 3]}$$

where $(d_2 - d_1)(PV_p - PV_s)$ represents the gain from non-shirking/effort; PV_p is the present value of primary sector work and PV_s the present value of secondary sector work (recall that non-effort is only possible in the primary sector, the sector that pays a wage premium over the secondary sector wage). This conditional expression shows the premium that firms pay (the right-hand side of Equation 3) to overcome the gain that workers derive from not exerting effort (the left-hand side of Equation 3), thereby ensuring that they exert effort.

As mentioned above, the model in this article combines an efficiency wage model (with its non-shirking component) with a labour union model. As a result α includes also the premium that companies have to pay to ensure the effort of unionised labour (i.e. to ensure that unionised workers limit their strike action or do not strike at all). This will render $\alpha = \alpha_1\alpha_2$, where α_1 is the instantaneous gain in utility from not exerting effort (i.e. from shirking), and α_2 (which is ≥ 1) constituting the premium that unionised workers can extract.[4, 5]

The South African labour market is also characterised by significant spatial distortions resulting from apartheid, where places of residence of black people very often were far removed from places of work (in the primary sector). These distances significantly raise travel costs, which need to be added to the premium that workers require before working in the primary sector. Therefore:

$$\alpha = \alpha_1\alpha_2 + \alpha_3 D \qquad \text{[Eqn 4]}$$

where D represents the distance between place of residence and place of work in the primary sector, and α_3 represents the cost per unit of distance:

- Unions having more power implies a higher value of α_2 and therefore a higher value of α; consequently, the difference between the present values of primary and secondary sector wages will be higher.
- Similarly, the larger α_3 and D, the larger will be the value of α. The inclusion of the term $\alpha_3 D$ means that both distance and the cost per unit of distance impact the reservation wage of workers – and negatively affect job search. If people live far from places of work in the primary sector and have to travel to places of work, they may not be able to afford job search. Note that this particular search and/or entry barrier can be seen as principally due to a financial market failure. Job-seekers find it hard to borrow money to finance their traveling and search costs (intending to repay the loan upon finding a job). Lenders might be unwilling to extend such loans due to both a low probability of finding a job and a low expected wage.

Rearranging Equation 3:

$$\alpha / (d_2 - d_1) \leq (PV_p - PV_s) \qquad \text{[Eqn 5]}$$

Using Equation 2, the present values of being employed in the primary and secondary sectors are:

$$PV_p = \frac{\left[\left(1 - q_p - d_1\right) p w_p + q_s\left(1 - p\right) w_p\right]}{r} \qquad \text{[Eqn 6.1]}$$

$$PV_s = \frac{\left[\left(q_p + d_1\right) p w_s - \left(1 - q_s\right)\left(1 - p\right) w_s\right]}{r} \qquad \text{[Eqn 6.2]}$$

Substituting Equations 6.1 and 6.2 into Equation 5 and normalising on w_p yields:

$$w_p \geq \frac{\alpha r}{\left(d_2 - d_1\right)\left(\left(1 - q_p - d_1\right) p + q_s\left(1 - p\right)\right)} + \frac{\left(q_p + d_1\right) p + \left(1 - q_s\right)\left(1 - p\right)}{\left(1 - q_p - d_1\right) p + q_s\left(1 - p\right)} w_s \qquad \text{[Eqn 7]}$$

Recalling that $q_p + d_1 = q_s(1 - p)/p$ and substituting into Equation 7 yields:

$$w_p \geq \frac{\alpha r}{\left(d_2 - d_1\right) p} + \frac{1}{p} w_s \qquad \text{[Eqn 8]}$$

Equation 8 represents the effort supply function (Equations 3 and 5 above) in a different form that shows the relationship between the *wage* and the *proportion of positions filled* in the primary sector: as p increases, w_p decreases. It also expresses the primary sector wage as the secondary sector wage plus a mark-up. (It still is an effort supply function: the mark-up or premium is what needs to be paid to primary sector workers to ensure effort.) Thus, the relative proportion of positions allocated to primary sector jobs (p) has an impact on the size of the mark-up on the secondary sector wage rate. This is shown graphically in Figure 2.

Note that, as p increases the slope of the relationship becomes flatter, while the intercept decreases (i.e. as p increases, w_p shifts and rotates from w_{p1} to w_{p2}).

4.The premium rate is $(\alpha_2 - 1)$.

5.The South African labour market is also characterised by a clear skills-related stratification of the unemployed, with an oversupply of unskilled workers and a shortage of skilled workers: the unemployment rate among individuals holding post-school degree qualifications is approximately 5%, and among those who have not completed school just below 50% (CDE 2013; Van der Berg & Van Broekhuizen 2012). This article does not include these highly skilled workers into the model simply because when they quit or are laid off they typically do not move to the informal sector, but find employment relatively easily elsewhere in the formal sector.

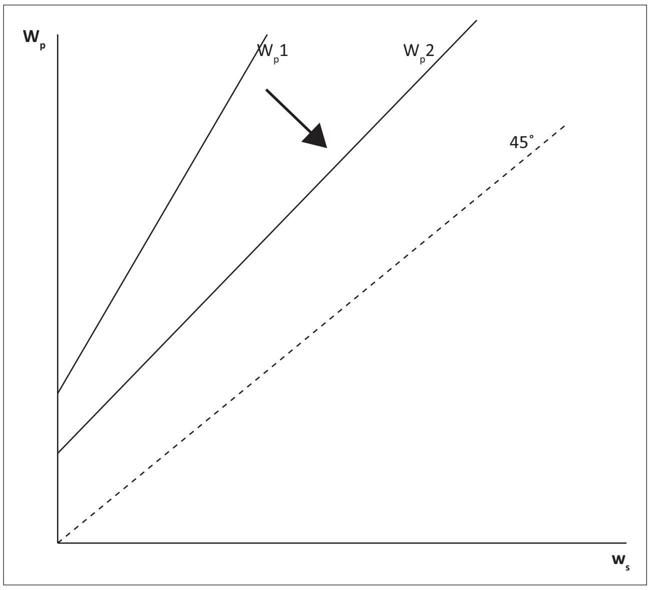


FIGURE 2: The relationship between primary and secondary sector wages.

The price-setting relationship

To derive the price-setting relationship we use the standard textbook equation stating the relationship between wages, the marginal product of labour (and hence the level of employment *E*) and profit mark-up of a monopolistically competitive firm. In Equation 9 this is applied to the primary sector wage:

$$w_p = \frac{\varepsilon - 1}{\varepsilon}\left(MPL\right) = \frac{\varepsilon - 1}{\varepsilon} b\left(E_p\right) \text{ with } b' < 0 \text{ and } wp > 0' \quad \text{[Eqn 9]}$$

with *MPL* being the marginal product of labour and ε the elasticity of product demand in a monopolistically competitive market, thus $(\varepsilon - 1)/\varepsilon < 1$, where $\varepsilon > 1$ to ensure that firms make a profit. *MPL* is defined as a negative function, *b*, of primary sector employment, E_p. Thus, holding ε constant, the primary sector wage becomes a negative function, *g*, of primary sector employment:

$$w_p = \gamma b\left(E_p\right) = g\left(E_p\right) \text{ with } g' < 0,\ \gamma = \frac{\varepsilon - 1}{\varepsilon} \text{and } w_p > 0 \quad \text{[Eqn 10]}$$

where the size of γ relates to the size of the mark-up of a monopolistically competitive form; the higher γ and therefore the closer it moves to 1 (i.e. the closer ε moves to infinity and therefore approaches the perfectly competitive model), the lower the mark-up can be and the less the firm can benefit from its monopolistically competitive position.

Equation 10 represents the standard primary sector *price-setting relationship* linking employment and wages: given that $g' < 0$, w_p decreases as E_p increases (but the wage cannot turn negative).

The job-offer relationship

The number of positions (F_p) and hence also the proportion of jobs or positions offered by firms in the primary sector, *p*, is a positive function of the marginal product of labour, which itself is a negative function of the level of employment (see the discussion of Equations 9 and 10 above). Suppose, for reasons of simplicity, that this relationship is linear with parameter *h*:[6]

$$p = \frac{h}{\gamma} w_p = \frac{h}{\gamma} g\left(E_p\right) \text{ or } w_p = \frac{p\gamma}{h} \quad \text{[Eqn 11a]}$$

Thus, at higher levels of E_p the real wage is lower (because the marginal product of labour is lower), and hence so is the proportion of positions filled by firms in the primary sector, *p*. Of course, if, for a given level of employment, the marginal product of labour increases – for instance, due to an upgrade in skill levels – the number of positions offered in the primary sector will increase. Thus, the positive sign of *h* means that if workers are more productive, more workers can be employed at a given wage.

Given the role of the marginal product of labour in Equation 11a and its link to the proportion of positions offered, Equation 11a is also a *job-offer relationship* – it links the proportion of jobs or positions being offered to wages. (Below it will interact with Equation 8, the effort supply function, to establish the equilibrium wage and number of positions filled.)

Note that in terms of Equations 10 and 11a there is a positive relationship between *p* and w_p, but a negative relationship between E_p and *p* (given that $g' < 0$): as E_p increases, w_p decreases, causing *p* to also decrease.

The wage-setting relationship

Substituting Equation 11a into Equation 8 yields Equation 12:

$$w_p \geq \frac{\alpha r}{\left(d_2 - d_1\right)\frac{h}{\gamma} g\left(E_p\right)} + \frac{1}{\frac{h}{\gamma} g\left(E_p\right)} w_s \text{ with } g < 0 \quad \text{[Eqn 12]}$$

Equation 12 is a primary sector *wage*-setting equation with its characteristic positive relationship between the level of employment and wages. As E_p increases (and given that $g' < 0$), w_p increases, simply because, as employment in the primary sector increases (and hence as employers offer more jobs), workers can get work easier elsewhere in the primary sector (the probability of getting a job in the primary sector is larger if a larger proportion of total jobs are filled in the primary sector) – hence firms need to offer a higher wage to ensure that they stay, exert effort and do not strike.

Equation 12 interacts with Equation 10, the price-setting relationship between wages and employment, to determine the equilibrium values of wages and employment in the primary sector.

Workers in the secondary sector are just paid their marginal product, which, for simplicity, is assumed to remain constant: with little capital and similar skills and each person more or less operating on their own, they are assumed to have the same marginal productivity.

6. Note that *h* is divided by γ so as to ensure that in Equation 11 *p* is purely a function of the marginal product of labour and not γ: $w_p = g(E_p) = \gamma b(E_p)$, so dividing $g(E_p)$ by γ leaves $b(E_p) = MPL$.

Model summary

The model can be summarised as follows.

First, in p-w_p space there are two relationships (the [ ] indicates the sign of the p-w_p relationship):

A job offer relationship:

$$p = \frac{h}{\gamma} w_p \quad \text{or } w_p = \frac{p\gamma}{h} \; [+] \qquad \text{[Eqn 11a]}$$

An effort supply function:

$$w_p \geq \frac{\alpha r}{\left(d_2 - d_1\right) p} + \frac{1}{p} w_s \; [-] \qquad \text{[Eqn 8]}$$

Secondly, in E_p-w_p space there are two relationships (with $g' < 0$) (the [ ] indicates the sign of the E_p-w_p relationship):

A price-setting relationship:

$$w_p = g(E_p) \; [-] \qquad \text{[Eqn 10]}$$

A wage-setting relationship:

$$w_p \geq \frac{\alpha r}{\left(d_2 - d_1\right)\frac{h}{\gamma} g\left(E_p\right)} + \frac{1}{\frac{h}{\gamma} g\left(E_p\right)} w_s \quad \text{with } g < 0 \; [+] \qquad \text{[Eqn 12]}$$

Equations 8 and 11a – *or* Equations 10 and 12 – can be used to calculate the equilibrium values of w_p, and to calculate the equilibrium values of p. The expressions for w_p and p are:

$$w_p = \left(\frac{\alpha r \gamma}{\left(d_2 - d_1\right) h} + \frac{w_s \gamma}{h} \right)^{0.5} \qquad \text{[Eqn 13]}$$

$$p = \left(\frac{\alpha r h}{\left(d_2 - d_1\right) \gamma} + \frac{w_s h}{\gamma} \right)^{0.5} \qquad \text{[Eqn 14]}$$

To calculate the equilibrium value of E_p note that in equilibrium $E_p = F_p$ and that $p = F_p/F$. So, using Equation 14 and given the value of F, Equation 15 then produces the equilibrium value of E_p:

$$E_p = \left(\frac{\alpha r h}{\left(d_2 - d_1\right) \gamma} + \frac{w_s h}{\gamma} \right)^{0.5} F \qquad \text{[Eqn 15]}$$

Together with the effort supply function 8, the job offer relationship 11a then determines the equilibrium number of positions in the primary sector. Since the proportion of filled positions in the secondary sector is $(1 - p)$, the secondary sector absorbs all those who are not employed in the primary sector and who are willing to work for wage w_s. (This assumption will be relaxed in the next section). Thus, in this model – as in the model of Bulow and Summers – there is no involuntary unemployment.

Step 2: The two-sector, three-segment model with involuntary aggregate unemployment

In this section the model is expanded to contain a third sector or segment that comprises the unemployed. The preference hierarchy follows the model above: workers in the secondary sector prefer the primary to the secondary sector; the unemployed would prefer secondary sector employment to unemployment and primary sector employment to secondary sector employment.

The effort supply function

As in the previous section, we first consider the effort supply function. The effort supply function introduces a role for entry barriers that imply that not all of those who are unable to find a job in the primary sector will be able to find one in the secondary sector.

The model makes a few simplifying assumptions. First, those quitting and being laid off in the primary sector (at rate $q_p + d_1$), move to the secondary sector, while those quitting the secondary sector (at rate q_s) move to unemployment (i.e. nobody moves from the secondary to the primary sector). Those of the unemployed who quit their unemployed status (at rate q_u) move either into the primary or the secondary sector. The unemployed, of course, receive no wage.

As before, the proportion of filled positions (jobs) supplied in the primary sector is p_p, while that of the secondary sector is p_s. A critical difference is that, unlike the two-sector model with no unemployment (where everyone who is willing to work in the secondary sector for a wage equal to their marginal product, w_s, finds employment), in this model the number of filled positions in the secondary sector, p_s, is equal to or less than $(1 - p_p)$; p_s being smaller than $(1 - p_p)$ would result from barriers to entry into the secondary sector. The barriers and obstacles may include physical, financial, human and social capital requirements.

Grimm et al. (2011a) present a small model in which the barrier to entry results from the borrowing constraint of the potential secondary sector entrant interacting with the minimum scale of capital, K^*, needed to generate a higher return. Note that the capital, K, typically includes physical capital, but the concept can also be expanded to include human capital (i.e. the basic education and training needed to be employed by or operate a small enterprise). Thus, below the minimum scale the return to capital is very low. The question a potential entrant into the secondary sector faces is whether or not the minimum scale of capital is lower than her borrowing constraint. The borrowing constraint originates from asymmetric information: lenders do not know whether borrowers will in fact acquire the capital with their borrowed funds and thus be in a position to generate a return in excess of what the borrower needs to pay the lender for the borrowed funds. Thus, if the borrowing constraint is lower than the minimum scale, then the return to capital is small, and the entrant will have to use her total return to cover the cost of capital, r_K; there will be no profit left after paying the cost of capital. Hence, investment will not take place and the entrant will not enter the secondary sector. If, however, the minimum scale is lower than the borrowing constraint, investment will take place and returns to capital will exceed capital cost (this high return will of course fall to zero as the scale of capital is expanded and the marginal product falls with the expansion in scale). In their model (Grimm et al. 2011a:S30) the secondary market entrant would maximise her profit, π, subject to a borrowing constraint, with output produced by a simple production function where $y = f(K)$, yielding output y produced with capital K when $K > K^*$, and capital producing just enough output to cover its cost when $K \leq K^*$:

$$\text{Max. } \pi = y - rK \qquad \text{[Eqn 16]}$$

Subject to:
$y = f(K)$ if $K > K^*$
$y = r_K K$ if $K \leq K^*$
and $K \leq B^*$ [Eqn 17]

The capital stock is chosen so that $f'(K) = r$ if $B^* > K^*$. If $B^* \leq K^*$, that is, the borrowing constraint is binding, then the entrant is indifferent between different levels of capital, since capital has a zero profit when $0 < K < K^*$ – hence, one can expect no investment to occur. Thus, one could argue that those potential entrants whose borrowing constraint is lower than the minimum scale capital, $B^* \leq K^*$, will not enter the secondary sector, and will move to unemployment. The proportion of potential entrants for whom $B^* > K^*$, will be defined as θ.

Note that in the two-sector model of the previous section all those workers who were unable to find jobs in the primary sector were able to find a job in the secondary sector if they were willing to work for a wage equal to the marginal product of their labour. However, in the three-segment model of this section, barriers to entry into the secondary sector means that only a fraction, θ, of those who are unable to find jobs in the primary sector are able to enter the secondary sector. Therefore:

$$p_s = \theta\,(1 - p_p) \qquad \text{[Eqn 18]}$$

That fraction, θ, is itself a function of the barriers of entry – the higher the barriers to entry, the lower the fraction. In terms of Equations 16 and 17, the lower B^* is and the higher K^* is, the higher is the barrier to entry into the secondary sector and therefore the lower θ will be.

This implies that $(1 - p_p - p_s)$ is the proportion of positions that the primary and secondary sectors would have supplied, *had there not been barriers to entry in the secondary sector*. It also means that, in this model, p_p and p_s are expressed as ratios of $F_p + F_s, + U$ (which now comprises the labour force), with U being the involuntarily unemployed.

With the above, and similar to Equation 2, the sum of the present value of expected primary, secondary and tertiary sector income in the economy is (where the zeros represent the zero wage earned by the unemployed):

$$PV = \left[\frac{(1-q_p-d_1)w_p + (q_p+d_1)w_s}{r}\right]p_p + \left[\frac{(1-q_s)w_s + q_s(0)}{r}\right]p_s + \left[\frac{(1-q_u)(0) + q_u p_p w_p + q_u p_s w_s}{r}\right](1-p_p-p_s) \qquad \text{[Eqn 19]}$$

In equilibrium, outflows from the primary sector need to equal inflows into the primary sector from the third segment (unemployed). Thus, $(q_p + d_1)p_p = q_u p_p(1 - p_p - p_s)$, which also means that $qu = (q_p + d_1)/(1 - p_p - p_s)$.

In addition, the outflow from the secondary sector needs to equal inflow into the secondary sector from both the primary sector and the unemployed segment. Thus, $q_s p_s = (q_p + d_1)p_p + q_u p_s(1 - p_p - p_s)$, which (after reorganising) implies that $(q_p + d_1)p_p = q_s p_s - q_u p_s(1 - p_p - p_s)$ (which also equals $q_u p_p(1 - p_p - p_s)$ – see previous paragraph).

Assuming that the unemployed receive no income, it means that in this case too $\alpha/(d_2 - d_1) = (PV_p - PV_s)$ (compare Equation 5). The present values of primary and secondary work are:

$$PV_p = \frac{(1-q_p-d_1)p_p w_p + q_u p_p(1-p_p-p_s)w_p}{r} \qquad \text{[Eqn 20]}$$

$$PV_s = \frac{\left((q_p+d_1)p_p w_s + (1-q_s)w_s + q_u p_s(1-p_p-p_s)w_s\right)}{r} \qquad \text{[Eqn 21]}$$

Therefore:

$$\frac{\alpha r}{d_2-d_1} \leq (1-q_p-d_1)p_p w_p + q_u p_p(1-p_p-p_s)w_p - \left((q_p+d_1)p_p w_s + (1-q_s)w_s + q_u p_s(1-p_p-p_s)w_s\right)$$

which after normalising on w_p yields:

$$w_p \geq \frac{\alpha r}{(d_2-d_1)\left((1-q_p-d_1)p_p + q_u p_p(1-p_p-p_s)\right)} + \frac{(q_p+d_1)p_p + (1-q_s) + q_u p_s(1-p_p-p_s)}{(1-q_p-d_1)p_p + q_u p_p(1-p_p-p_s)}w_s \qquad \text{[Eqn 22]}$$

Using the equilibrium condition that $(q_p + d_1)p_p = q_u p_p(1 - p_p - p_s)$ (which also means $q_u = (q_p + d_1)/(1 - p_p - p_s)$), Equation 22 simplifies to:

$$w_p \geq \frac{\alpha r}{(d_2-d_1)p_p} + \frac{(q_p+d_1)(p_p+p_s) + (1-q_s)}{p_p}w_s \qquad \text{[Eqn 23]}$$

Now recall that $p_s = \theta(1 - p_p)$ and substitute it into Equation 23 to yield:

$$w_p \geq \frac{\alpha r}{(d_2-d_1)p_p} + (q_p+d_1)(1-\theta)w_s + \frac{\left((q_p+d_1)\theta + (1-q_s)\right)w_s}{p_p} \qquad \text{[Eqn 24]}$$

Equation 24 represents the effort supply function in the three-segment model. As was the case with the two-sector model with no involuntary unemployment, an increase in p_p would cause w_p to decrease and the slope of the effort supply function becomes flatter the larger p_p becomes. Note that, unlike in Equation 8, the quit rates do not disappear from Equation 24. The reason for this is that the existence of barriers to entry into the secondary sector cause θ in Equation 24 to be smaller than 1 (i.e. $\theta < 1$).[7]

The job-offer relationship and the price-setting and wage-setting relationships

Equations 9 to 11a remain unchanged, with Equation 11b subscripted for the primary sector:

$$w_p = \frac{\varepsilon - 1}{\varepsilon}\left(MPL\right) = \frac{\varepsilon - 1}{\varepsilon} b\left(E_p\right) \text{ with } b' < 0 \text{ and } wp > 0' \qquad \text{[Eqn 9]}$$

$$w_p = \gamma b\left(E_p\right) = g\left(E_p\right) \text{ with } g' < 0,\ \gamma = \frac{\varepsilon - 1}{\varepsilon} \text{ and } w_p > 0 \qquad \text{[Eqn 10]}$$

$$p_p = \frac{h}{\gamma} w_p = \frac{h}{\gamma} g\left(E_p\right) \quad \text{or } w_p = \frac{p_p \gamma}{h} \qquad \text{[Eqn 11b]}$$

Therefore, there is a positive relationship between p and w_p, but a negative relationship (given that $g < 0$) between E_p and p_p (as E_p increases, w_p decreases, causing p_p to also decrease). Equation 10 represents, again, the price-setting relationship, while Equation 11b represents the job offer relationship.

Substituting Equation 11b into Equation 24 yields the detailed wage-setting equation:

$$w_p \geq \frac{\alpha r}{\left(d_2 - d_1\right)\frac{h}{\gamma} g\left(E_p\right)} + \left(q_p + d_1\right)\left(1 - \theta\right) w_s + \frac{\left(\left(q_p + d_1\right)\theta + \left(1 - q_s\right)\right) w_s}{\frac{h}{\gamma} g\left(E_p\right)} \quad \text{with } g' < 0 \qquad \text{[Eqn 25]}$$

As E_p increases (and given that $g' < 0$), w_p increases.

Model summary

The model can be summarised as follows.

First, in p-w_p space there are two relationships (the sign within [ ] indicates the sign of the p-w_p relationship):

A job offer relationship:

$$p_p = \frac{h}{\gamma} w_p \quad \text{or } w_p = \frac{p_p \gamma}{h} \ [+] \qquad \text{[Eqn 11b]}$$

An effort supply function:

$$w_p \geq \frac{\alpha r}{(d_2 - d_1) p_p} + \left(q_p + d_1\right)\left(1 - \theta\right) w_s + \frac{\left(\left(q_p + d_1\right)\theta + \left(1 - q_s\right)\right) w_s}{p_p} \quad [-] \qquad \text{[Eqn 24]}$$

which is distinguished by the presence of θ (a function of barriers to entry) and quit rates.

Secondly, in E_p-w_p space there are two relationships (with $g' < 0$) (the [ ] indicates the sign of the E_p-w_p relationship):

A price-setting relationship:

$$w_p = g(E_p)[-] \qquad \text{[Eqn 10]}$$

A wage-setting relationship:

$$w_p \geq \frac{\alpha r}{\left(d_2 - d_1\right)\frac{h}{\gamma} g\left(E_p\right)} + \left(q_p + d_1\right)\left(1 - \theta\right) w_s + \frac{\left(\left(q_p + d_1\right)\theta + \left(1 - q_s\right)\right) w_s}{\frac{h}{\gamma} g\left(E_p\right)} \quad \text{with } g' < 0\ [+] \qquad \text{[Eqn 25]}$$

which is also distinguished by the presence of θ and quit rates.

In a similar fashion as in the previous section, Equations 11b and 24, and 10 and 25 can be used to calculate the equilibrium values for w_p, p_p and E_p:

$$w_p = \left(q_p + d_1\right)\left(1 - \theta\right) w_s + \frac{\left(\left(-\left(q_p + d_1\right)\left(1 - \theta\right) w_s\right)^2 + 4\left(\frac{\alpha r}{d_2 - d_1} + \left(q_p + d_1\right)\theta + \left(1 - q_s\right)\right) w_s \gamma / h\right)^{0.5}}{2} \qquad \text{[Eqn 26]}$$

$$p_p = \frac{h\left(q_p + d_1\right)\left(1 - \theta\right) w_s}{\gamma} + \frac{\left(\left(-h\left(q_p + d_1\right)\left(1 - \theta\right) w_s / \gamma\right)^2 + 4\left(\frac{\alpha r}{d_2 - d_1} + \left(q_p + d_1\right)\theta + \left(1 - q_s\right)\right) w_s h / \gamma\right)^{0.5}}{2} \qquad \text{[Eqn 27]}$$

$$E_p = \left(\frac{h\left(q_p + d_1\right)\left(1 - \theta\right) w_s}{\gamma} + \frac{\left(\left(-h\left(q_p + d_1\right)\left(1 - \theta\right) w_s / \gamma\right)^2 + 4\left(\frac{\alpha r}{d_2 - d_1} + \left(q_p + d_1\right)\theta + \left(1 - q_s\right)\right) w_s h / \gamma\right)^{0.5}}{2}\right) F \qquad \text{[Eqn 28]}$$

Note that, unlike their two-sector equivalents (Equations 13–15), Equations 26–28 contain θ (a function of barriers to

7. That the first term containing q_p would equal zero if $\theta = 1$ is straightforward to see. In the case of the second, recall that $(q_p + d_1)p_p = q_u p_p(1 - p_p - p_s)$, which means $(q_p + d_1) = q_u(1 - p_p - p_s)$, with $(q_p + d_1)$ appearing in the second term on the right-hand side of Equation 24 that contains q_p. If $\theta = 1$ then $p_p + p_s = 1$, so that $q_u(1 - p_p - p_s) = 0$, which also means $(q_p + d_1) = 0$.) In the literature (cf. Campbell & Orszag 1998:121), higher levels of employment and wages are associated with a higher quit rate – higher employment levels imply a higher probability of finding a job again once the worker quits (more about this in section 4, which compares the two models).

entry) and the quit rates. The implications of these are discussed in the next section. Together with the effort supply function, the job offer relationship then determines the equilibrium number of positions in the primary sector. In addition, recalling that $p_s = \theta(1 - p_p)$, one can calculate the employment level in the secondary sector:

$$E_s = \theta\,(1 - p_p)\,F \quad \text{[Eqn 29]}$$

In the three-segment model the unemployed are involuntarily unemployed. Those who end up in the third segment and who cannot re-enter either the primary or the secondary sectors, due to the presence of barriers to entry into both the primary and secondary labour markets, find themselves involuntarily unemployed.

Using Equations 28 and 29, one can calculate the total equilibrium employment level in the economy ($E_p + E_s$), which equals the equilibrium level of positions filled, $F_p + F_s$. Hence:

$$U = F - (F_p + F_s) \quad \text{[Eqn 30]}$$

is the number of involuntary unemployed.

A comparison of the two models

The two-sector, three-segment model shows how the two-sector model can be expanded from a model that merely explains the allocation of labour between the primary and secondary sectors, to a model that caters for the possibility of involuntary unemployment on the aggregate level. The key difference centres on the following. In the two-sector model, workers who quit or lose a job in one of the sectors, circulate back to a job in the other sector. In the three-segment model, workers who quit or lose a job in one of the two employing sectors do not necessarily find a job again and may end up being unemployed. Some workers might also never have worked (and remain unemployed).

The main reason why workers end up unemployed is the existence of barriers to entry such as a lack of physical and human capital as discussed above. (If there are no barriers to entry into the secondary sector, the three-segment model reverts to the two-sector model.) To compare the two models, compare Equations 13–15 and 26–30. Compared to the two-sector model, the presence of the quit rate q_p in the three-segment model's Equations 26–30 implies higher equilibrium values for w_p, p_p and E_p.[8]

In the literature (cf. Campbell & Orszag 1998:121), higher levels of employment and wages are associated with a higher quit rate – higher employment levels imply a higher probability of finding a job again once the worker quits. In two-sector model equilibrium, quit rates (as well as d_1, i.e. the probability of being laid off for shirking while not actually shirking) do not affect w_p, p_p and E_p because in equilibrium the flow into the primary sector equals the flow out of the primary sector – those who quit find jobs in the secondary sector and are replaced, in turn, by workers moving from the secondary to the primary sector.

However, because of entry barriers in the secondary sector in the three-segment model, the flows into and from the primary sector are not necessarily equal. This implies a relationship between quitting and w_p, p_p and E_p. In the three-segment model, barriers to entry mean that $\theta < 1$ (θ being a function of barriers to entry B). If $\theta = 1$, then all the terms containing q_p in Equations 26–28 would disappear by virtue of being equal to zero,[9] which will also mean that q_p would have no effect. Thus, in this model the presence of barriers to entry (which cause $\theta < 1$) also ensure that q_p has an effect on w_p, p_s and E_p. Higher levels of employment in the primary sector imply that should a worker quit, the probability of ultimately finding a job again in the primary sector is higher, which, in turn, may engender a greater willingness on the part of primary sector workers to quit. Hence the positive relationship between quit rates and p_p and E_p.

Unlike the two-sector model where all workers are employed either in the primary or the secondary sector, in the three-segment model $p_p + p_s \leq 1$ with $\theta < 1$. The higher the barriers to entry B, the lower p_p *and* p_s will be, hence (using Equations 26, 27 and 28), the lower w_p and E_p will be.[10] Thus, barriers to entry mean fewer positions will be filled in both the primary and secondary sectors; employment will thus be lower. It also means wages in the primary sector will be lower than in the two-sector model.

Furthermore, note that the higher the quit rate q_s from the secondary sector, the lower are w_p, p_s and E_p. In the three-segment model, quitting from the secondary sector means that the worker moves towards unemployment, while in the two-sector model it means that the worker circulates back to the primary sector. For given quit rates from the primary and tertiary sectors ('tertiary quitting' being quitting from unemployment and thus moving back to either primary or secondary sector employment), a higher quit rate in the secondary sector means a higher probability of ending up without a job, even if one starts out in the primary sector. Thus, a higher quit rate from the secondary sector depresses wages, employment and the number of jobs in the primary sector.

8. Why is this so? With $\gamma > h$ in all realistic scenarios, a higher q_p means that the third term on the right-hand side of Equations 26–29 that contains q_p ($4(q_p + d_1)\theta\, w_s\gamma/h$) will always be larger than the second term that also contains q_p, for instance $-(q_p + d_1)(1 - \theta)w_s)2$ in Equation 26, leaving the net effect of these two terms as a positive value. With the first term on the right-hand side also containing q_p, the net effect of the three terms on the right-hand side containing q_p will be positive, meaning higher equilibrium values for w_p, p_p and E_p. (The only exception to this scenario would be the primary sector goods market approximates an almost perfectly competitive market, contrary to the assumptions of this model.)

9. Why the first two terms containing q_p would equal zero if $\theta = 1$, is straightforward to see. In the case of the third, recall that $(q_p + d_1)p_p = q_u p_p(1 - p_p - p_s)$, which means $(q_p + d_1) = q_u(1 - p_p - p_s)$, with $(q_p + d_1)$ appearing in the third term on the right-hand side of Equations 26–28 that contain q_p. If $\theta = 1$ then $p_p + p_s = 1$, so that $q_u(1 - p_p - p_s) = 0$, which also means $(q_p + d_1) = 0$.

10. The logic is as follows: Higher barriers mean a lower θ, and the lower θ, the higher will be the first term on the right-hand side of Equations 26-28 containing θ, but also the lower will be the second and third terms on the right-hand side of Equations 26–28 containing θ. The effect of the second and third terms will exceed that of the first, which means that the net effect of these three terms on w_p, p_p and E_p in a case of a lower θ is negative. With both p_p and θ being lower, p_s will also be lower.

A graphical representation of the models

Figure 3 is a graphical presentation of the models discussed above. It shows employment in the two employing sectors on the horizontal axis and real wages W on the vertical axis. Primary sector employment is measured rightward from the vertical axis (marked W_p), while secondary sector employment is measured leftward from the vertical axis (W_S). N_N represents the working-age population. Distance e shows those who are not economically active.

Suppose, to start off, there is a perfectly competitive labour market with no market power and no efficiency wages. The wage paid in the primary and secondary sectors would be equal (i.e. there is no real distinction between the primary and secondary sectors). L^S_{PC} and L^D_{PC} represent labour supply and demand in a perfectly competitive (subscript C) labour market among firms in the primary sector, while L^S_{SC} and L^D_S represent labour supply and demand in the secondary sector. L^D_S is horizontal, following the simplifying assumption that the marginal product of labour in the secondary sector is constant.[11] Because the markets are perfectly competitive, wages in the primary and secondary sectors would be the same, $W_{PC} = W_{SC}$, with E_{PC} and E_{SC} being the corresponding employment levels in the primary and secondary sectors. The distance marked a represents those workers who would be voluntarily unemployed – they could always find work at the prevailing wage W_{SC} (i.e. if they are willing to reduce their reservation wages).

Now suppose the economy is Neo-Keynesian, with market power and efficiency wages in the primary sector. This produces the two-sector Neo-Keynesian model (subscript K), still with no barriers to entry into the secondary sector. The wage-setting (WP_p) and price-setting (PS_p) relationships in the primary sector will, due to effort behaviour, establish a wage W_{PK} that is higher than W_{PC}. Employment in the primary sector, at E_{PK}, will be lower compared to the perfectly competitive case, at E_{PC}. The difference in the number of workers being employed in the primary sector equals distance b in Figure 3: $b = E_{PC} - E_{PK}$. Workers who are not accommodated in the primary sector are diverted to and employed in the secondary sector. Thus, labour supply in the secondary sector is L^S_{SK} and b' (the horizontal leftward displacement from L^S_{SC} to L^S_{SK}) equals distance b (the quantity of workers relocated from the primary sector). Notice that in this model distance a equals distance $b + c$, since all these unemployed workers can find employment in the secondary sector at wage W_{SC} should they wish to (i.e. if they lower their reservation wage); they are voluntarily unemployed.

Next we introduce barriers to entry into the secondary sector (for simplicity we ignore barriers to entry into the primary sector). Given the nature of 'effort behaviour' in the primary sector, as before a quantity of workers equal to b will not be accommodated in the primary sector (compared to the perfectly competitive case). However, the presence of barriers to entry in the secondary sector means that labour supply in the secondary sector will be at L^S_{SB} – lower than the previous case's L^S_{SK}. A quantity of workers equal to distance d will be involuntarily unemployed. This constitutes the third sector or segment in the model.

Unlike the case of the perfectly competitive market where workers can simply offer their labour at a lower wage, in a market with efficiency wages (with firms paying a wage to ensure effort), firms in the primary sector set both wages and prices. Hence, workers cannot increase employment in the primary sector by offering to work for a lower wage. In addition, even if unemployed workers are willing to work in the secondary sector for a wage equal to the marginal product of labour, barriers to entry prevent them from doing so.

The workers represented by distance c still are *voluntarily* unemployed. Even in the case of a perfectly competitive market, their reservation wage would have been above the market wage – they would have preferred unemployment even in the case of a perfectly competitive market. Note that the quantity of workers $b + d$ are willing to work in either the primary or the secondary sector at a wage of $W_{PC} = W_{SC}$, but are prevented from doing so due to the payment of efficiency wages in the primary sector and the existence of barriers of entry in the secondary sector.

Conclusion and potential policy implications

To create a theoretical model that explains the dual nature of the South African labour market (with its formal and informal sectors) and the simultaneous existence, indeed persistence, of very high unemployment, this paper draws on the dual labour market model of Bulow and Summers (1986) and the suggestion by Kingdon and Knight (2004), as well as work by Grimm et al. (2011a) that show that barriers to entry into the informal sector exist. Following the latter authors, such barriers are defined as the interaction of a borrowing constraint (itself the result of the asymmetric information faced by lenders in financial markets) and the minimum scale of capital needed to earn a high return.

The model shows:

- How a primary sector characterised by efficiency wage and labour union behaviour, as well as a mark-up due to high transport cost, can explain the dual nature of the labour market.

11. Assuming a constant marginal product of labour for the secondary sector is not an altogether unrealistic assumption. Berry (2001:7) argues that large and medium enterprises (in our model operating in the primary sector) usually have an amount of capital that complements a number of workers. As the number of workers increase, it might lead to a decrease in the marginal product of labour. However, by their very nature, firms in the informal sector are very small, and the capital needed is replicable on a small scale (i.e. in the extreme case of one-person firms – own-employment – it is not the case that for instance a second worker is added to a given set of capital in a single small firm, but rather that the second worker can set up his or her own firm and replicate the capital – each worker is therefore the first worker and there is not really a second worker that can decrease the marginal product of labour. A similar point can be made for firms employing say two or three workers since with two or three workers, there is not much scope to decrease the marginal product of labour, particularly if the capital is replicable on a small scale. Berry (2001:7) argues that the flat marginal product of labour and thus the flat labour demand for informal sector workers has been well verified, given the expandability of the informal sector. Of course, as the discussion below will indicate, there might be financial constraints on acquiring that minimal amount of capital, which might limit the size of the effective labour supply.

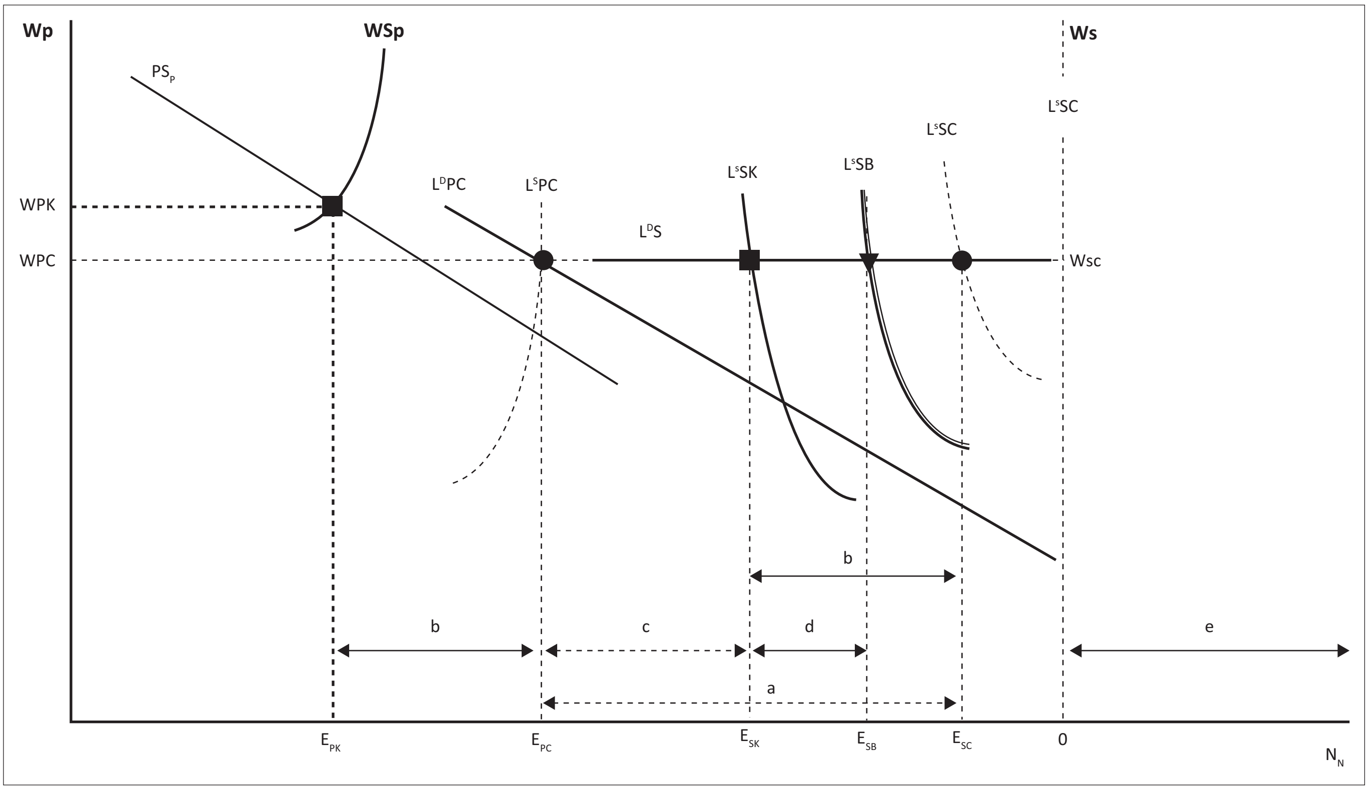


FIGURE 3: Unemployment in the theoretical three-segment model.

- How barriers to entry faced by potential entrants into the secondary sector can prevent workers from entering the secondary sector. This constrains the effective supply of labour to the secondary sector.
- How, as a result, these workers end up being (involuntarily) unemployed in a long-term macroeconomic equilibrium. The secondary sector does not simply absorb all those who cannot find employment in the primary sector.

From a policy point of view, the above suggests that there is no single or 'silver bullet' solution to address the unemployment problem. The solution is not as easy as, for instance, simply decreasing wage levels to render labour cheaper. Indeed, if the assumptions on which the above model draws hold in the South African reality, then a solution to the unemployment problem involve policies addressing product and labour market structures and behaviour in the primary sector, as well as policies addressing the numerous barriers to entry, such as borrowing constraints, that potential entrants into the secondary sector face.

Acknowledgements

The authors wish to thank REDI3x3 for the generous funding of this project, as reflected in Burger & Fourie (2015), which is a more comprehensive source to consult. The usual disclaimers apply.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Authors' contribution

Both authors contributed equally to the article.

References

Agénor, P.R. & Montiel, P., 1999, *Development macroeconomics*, 2nd edn., Princeton University Press, Princeton, NJ.

Berry, A., 2001, *The role of the small and medium enterprise sector in Latin America: Implications for South Africa*, TIPS Working Paper 5, viewed 05 June 2017, from http://www.tips.org.za/files/421.pdf

Blanchard, O. 2005, *European unemployment: The evolution of facts and ideas*, Working Paper 11750, National Bureau of Economic Research, Cambridge, MA.

Bulkley, G. & Myles, G.D., 1996, 'Trade unions, efficiency wages and shirking', *Oxford Economic Papers* 48(1996), 75–88. https://doi.org/10.1093/oxfordjournals.oep.a028562

Bulow, J.I. & Summers, L.H., 1986, 'A theory of dual labor markets with application to industrial policy, discrimination and Keynesian unemployment', *Journal of Labor Economics* 4(3)(Part 1), 376–414.

Burger, P. & Fourie, F.C.V.N., 2015, *Macroeconomic policy and South African unemployment: Developing a three-segment macroeconomic model*, REDI3x3 Working Paper 6, viewed from http://www.redi3x3.org

Cahuc, P. & Zylberberg, A., 2004, *Labor economics*, MIT Press, Cambridge, MA.

Campbell, C. & Orszag, J.M., 1998, 'A model of the wage curve', *Economics Letters* 59(1998), 119–125. https://doi.org/10.1016/S0165-1765(98)00018-4

Carlin, W. & Soskice, D., 2006, *Advanced macroeconomics*, Oxford University Press, Oxford.

CDE (Centre for Development and Enterprise), 2013, 'Graduate unemployment in South Africa: A much exaggerated problem', *CDE Insight*, viewed April 2013, from http://www.cde.org.za/publications/jobs-growth/83-jobs-and-growth/403-graduate-unemployment-in-south-africa-a-much-exaggerated-problem

Fourie, F.C.V.N., 2011, *The South African unemployment debate: Three worlds, three discourses?*, Working Paper 63, SALDRU, University of Cape Town Or: Working Paper 1, REDI3x3, viewed 05 June 2017, from http://www.redi3x3.org

Grimm, M., Krüger, J. & Lay, J., 2011a, 'Barriers to entry and returns to capital in informal activities: Evidence from Sub-Saharan Africa', *Review of Income and Wealth* 57, Special Issue, S27–S53. https://doi.org/10.1111/j.1475-4991.2011.00453.x

Grimm, M., Van der Hoeven, R. & Lay, J., 2011b, *Unlocking potential: Tackling economic, institutional and social constraints of informal entrepreneurship in Sub-Saharan Africa: Main findings and policy conclusions*, International Institute of Social Studies, Erasmus University.

Kingdon, G.G. & Knight, J.B., 2004, 'Unemployment in South Africa: The nature of the beast', *World Development* 32(3), 391–408. https://doi.org/10.1016/j.worlddev.2003.10.005

Knell, M., 2014, 'Efficiency wages, staggered wages, and union wage-setting', *Oxford Economic Paper*, 66(3), 848–870.

Layard, R., Nickell, S. & Jackman, R., 1991 and 2005, *Unemployment: Macroeconomic performance and the labour market*, Oxford University Press, Oxford.

Pereau, J.-C. & Sanz, N., 2006, 'Trade unions, efficiency wages and employment', *Economics Bulletin* 10(4), 1–8.

Statistics South Africa (StatsSA), 2009, *Labour force survey – Historical revision: September Series 2000 to 2007*, Statistical release P0210, Pretoria.

Statistics South Africa (StatsSA), 2017, *QLFS data*, Excel datasheet viewed 01 June 2017, from http://www.statssa.gov.za/?page_id=1854&PPN=P0211&SCH=7012

Summers, L.H., 1988, 'Relative wages, efficiency wages, and Keynesian unemployment', *American Economic Review* 78(2), 383–388.

Van der Berg, S. & Van Broekhuizen, H., 2012, *Graduate unemployment in South Africa: A much exaggerated problem,* Working Paper 22/2012, University of Stellenbosch, viewed from http://resep.sun.ac.za/index.php/research-outputs/stellenbosch-working-papers/wp2012/