How flexible are wages in response to local unemployment in South Africa?

GPRG-WPS-015

Geeta Kingdon and John Knight

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was part of the programme of the ESRC Global Poverty Research Group.
HOW FLEXIBLE ARE WAGES IN RESPONSE TO LOCAL UNEMPLOYMENT IN SOUTH AFRICA?

Geeta Gandhi Kingdon and John Knight

Centre for the Study of African Economies
Department of Economics
University of Oxford

April 2005

Abstract:

It is commonly claimed that the South African labor market is unusually inflexible owing to the strength of the unions and the system of centralized collective bargaining. One aspect of labor market inflexibility concerns the responsiveness of wages to local unemployment. Examining this spatial relationship, we find that the elasticity of wages to local unemployment rates in South Africa is -0.1, similar to that found in other countries, including the US and the UK. This is striking because South Africa has a national unemployment rate of over 30%. We find that the wage curve elasticity persists over a much wider range of unemployment rates than in OECD countries, implying that unemployment in South Africa can have a large impact on wages.

Keywords: wages, unemployment, wage curve, South Africa
JEL Classification: J30, J60

Correspondence: Geeta Kingdon, Department of Economics, University of Oxford, Oxford, OX1 3UQ. Tel: 44 1865 271065; Fax: 44 1865 271094.
email: geeta.kingdon@economics.ox.ac.uk

Acknowledgements: We should like to thank Francis Teal for insightful discussions and Simon Appleton, Vandana Chandra, Mary Gregory, John Hoddinott, Dudley Horner, Peter Moll, Andrew Oswald, Francis Wilson and Martin Wittenberg for commenting on earlier drafts. Participants at seminars at the Centre for the Study of African Economies, University of Oxford; SALDRU, University of Cape Town; and Southern Africa Department, World Bank made useful comments. The work was funded under the Employment and Labour Markets Programme financed by the Department for International Development of the UK Government.

Note: The SALDRU data used in this paper can be made available by the authors.
1. Introduction

South Africa has very high unemployment. Depending on the definition used, the latest national unemployment rate is 31% or 42% (StatsSA, 2003). This may well be the result of a rapidly growing labor force, on the one hand, and declining formal employment, on the other: the labor market may lack the flexibility to cope with this divergence (Lewis, 2002; Moll, 1993; Boccara and Moll, 1997; Fallon and Lucas, 1998). Indeed, there are reasons to expect that the South African labor market is unusually inflexible. Based on surveys of company officials, the World Economic Forum’s Global Competitiveness Report (1999: Tables 7.02-7.05, 7.09) ranks South Africa at the bottom of its fifty-nine nation comparison on whether labor regulations on wages, hours or dismissals favor flexibility. The trade union movement is apparently powerful and allied with the government. Moreover, there is an institutional framework – the Bargaining Councils and the Wage Boards – to set minimum wages in parts of the formal sector. However, recent research, examining the union - non-union and bargaining council – non-bargaining council wage differentials, has found this source of inflexibility to be minor (Schultz and Mwabu, 1998; Butcher and Rouse, 2001): these studies estimate the union premium to be similar to that found in other countries, including the US and the UK. Another potential test of labor market inflexibility concerns the responsiveness of wages to local labor market conditions. Our purpose is to measure this by examining the spatial relationship between wages and unemployment. The estimated elasticity of the wage with respect to the local unemployment rate can then be compared with the elasticities found in other countries.

There is accumulating evidence of a negative relationship between local unemployment and wages in various economies. This relationship has been called ‘the wage curve’ and claimed as an empirical ‘law’ of economics by Blanchflower and Oswald (1994). Using mesoeconomic
Blanchflower and Oswald present an impressive array of evidence in favor of the wage curve from the US, UK, and some other OECD countries. The evidence shows that wages are negatively related to contemporaneous unemployment, and that the unemployment elasticity of the wage is approximately -0.1, so that a 10% increase in the unemployment rate leads to a 1% decrease in wages. Carruth and Oswald (1987) and Blanchflower and Oswald (1990) provide theoretical models of the labor market consistent with the wage curve. In a searching review article, Card (1995) does not contest the validity of the wage curve, taking issue instead only with the interpretation of the results and the robustness of the elasticity.

In developing countries, by contrast, the relationship between wages and unemployment has generally been analyzed in the context of probabilistic models of migration (Harris and Todaro, 1970, and the vast literature, mainly theoretical, that followed). Unemployment is viewed as the equilibrating variable in a labor market characterized by segmentation and wage inflexibility: high urban wages attract high urban unemployment. The presumption is that unemployment is a function of the wage and not vice versa. There has been little testing of either causal relationship in developing countries using microeconomic datasets. There are two notable exceptions. Hoddinott (1996), contrary to the prediction of probabilistic migration models, found evidence of a wage curve in urban Cote d’Ivoire, with an elasticity similar to those in OECD economies, and Van der Meulen Rodgers and Nataraj (1999) adduced evidence of a long run wage curve in Taiwan.

The object of this paper is to test for the wage curve relationship in a developing country with very high unemployment, using data from South Africa where the average unemployment rate exceeds 30%. This is several times the typical OECD unemployment rate and twice that in urban Cote d’Ivoire. If a wage curve can be shown to exist in South Africa, this will throw further doubt on the importance in practice of probabilistic models in developing countries.

---

1 Applying microeconomic data and methods to macroeconomic questions.
(Banerjee, 1983). Given that a wage curve is found, are wages as flexible at the high South African unemployment rate as in other countries?

There are two further reasons why the finding of a wage curve in South Africa is of interest. First, when unemployment is very high, its definition itself becomes an issue. In these conditions, the proportion of people actively searching for work is likely to depend on the number wanting work. If many unemployed people stop actively searching for work because they become discouraged, then it may be misleading to measure the unemployment rate by considering as unemployed only those who actively looked for work, i.e. by the narrow definition. The wage curve can assist choice among alternative definitions of unemployment by showing which measure of the unemployment rate is more important in explaining wages. If, as an ILO report on the South African labor market suggests (ILO, 1996, p. 104), those wanting work but not actively seeking (included in the ‘broad’ definition of unemployment) are outside the labor force, we would expect the wage curve to be steeper for narrow than for broad unemployment.

Secondly, the wage curve can have implications for poverty: a negative relationship may mean that high unemployment in a locality not only has a direct effect on poverty among its households but also an indirect effect via the lower wages of their employed members.

The paper is organized as follows. Section 2 presents a brief theoretical analysis of the relationships between wages and unemployment. Section 3 provides the South African context, and Section 4 discusses the data, model, and tests. The empirical results appear in Section 5, and Section 6 concludes.

2. Unemployment and wages: theory

---

2 This finding is now corroborated by evidence from a large number of OECD countries.
Two main forms of explanation for a negatively sloped wage curve $W = f(U)$ have been suggested in the literature. One is an incentive wage hypothesis, i.e. profit-maximizing firms choose to set the wage above the competitive level in order to influence the behavior of their employees. This is normally presented along the lines of the Shapiro-Stiglitz (1984) efficiency wage model, to which imperfect monitoring and the need to prevent shirking are central. Workers have little incentive to put in effort if the cost of shirking is low. Firms respond by raising the wage above the market level so as to increase the cost of job loss. If local unemployment is high, the difficulties of finding work are great and workers are afraid of dismissal. This acts as a disciplining device, so that it is not necessary for firms to pay such high efficiency wages. The profit-maximizing efficiency wage is therefore inversely related to the unemployment rate. This wage curve has been termed the ‘no-shirking’ curve since it shows the wage just high enough to deter workers from shirking. Another incentive wage mechanism concerns firms’ training costs, which depend on the rate of labor turnover (Stiglitz, 1974). Firms raise wages above the competitive level to discourage voluntary quits. However, high local unemployment increases workers’ expected costs of quitting, so decreasing the profit-maximizing wage and giving rise to a negatively-sloped wage curve.

The second rationalization of the wage curve draws on the union bargaining model. High local unemployment frightens workers and weakens their bargaining power in negotiations over rent-sharing. If trade unions worry about their unemployed as well as their employed members, then negotiating unions may place greater weight on employment and less on wage objectives where local unemployment is high and the chances of finding employment consequently low. The wage curve accordingly has a negative unemployment elasticity.

Labor economists have used the notion of compensating differentials to hypothesize that wages and unemployment are positively correlated across space. In fact, this hypothesis was developed in an African context and formalized in a model by Harris and Todaro (1970): high
wages in a locality attract more workers to it until unemployment rises sufficiently to equalise the ‘expected wage’ across localities. The relationship, representing \( U = f(W) \), is depicted by the curve \( HT \) in Figure 1. However, that is not what researchers have found in other countries: the negative wage curve, interpreted as \( W = f(U) \), is depicted by \( WC \) in the figure. The two relationships are not mutually incompatible. The \( HT \) relationship may indeed prevail in the long run but labor may not be sufficiently mobile for \( HT \) to be observed in the short run. The long run effect can be set aside by the inclusion of locality dummy variables in the wage equation, so permitting the short run wage curve \( WC \) to be isolated (Blanchflower and Oswald, p. 93; Hoddinott, p. 1613-14).

What would be the shape of the wage curve if wages were entirely market-determined? This question is relevant to our hypothesis tests, some of which compare non-market with market outcomes. If labor is perfectly mobile, a single market wage is established in equilibrium (apart from compensating differences for local amenities). However, the (voluntary) unemployment rate is also invariant (apart from differences arising from local tastes). The outcome is likely to be a wage blob rather than a curve. If labor is perfectly immobile, as is assumed in Figure 2, local labor demand curves differ in relation to local labor supply, causing market-clearing wages to differ across labor markets (being \( W_1, W_2, \) and \( W_3 \) in the figure). In response, voluntary (shirk) unemployment may vary inversely with the wage: in market equilibrium a conventional wage curve can result. Imperfect mobility of labor tends to reduce, but does not eliminate, the wage differences between \( W_1, W_2, \) and \( W_3 \), and voluntary unemployment can again generate a negatively sloped wage curve.

Despite the presumption in the literature that the negative wage curve relationship be interpreted as \( W = f(U) \), i.e. the direction of causation is assumed to run from unemployment to wages, it is possible that a negatively sloped wage curve represents the effect of the wage on involuntary unemployment. In Figure 3 the supply curve \( S \) can be downward-sloping (shirk
unemployment) or upward-sloping (search unemployment), but for clarity it is assumed to be vertical, i.e. there is a fixed amount of voluntary unemployment. The three different local demand curves would produce market-clearing wages $W_1$, $W_2$, and $W_3$. However, assume that incentive wage determination, or institutional wage determination, or bargaining (conducted at a non-local level or based on comparisons) raises the wage above the market equilibrium in localities where the latter is low. The lower the market-clearing wage, the greater is the gap. This can then generate a downward-sloping ($cd>ab$) wage curve $WC$ based on involuntary unemployment, as shown in the figure. We call this the intervention-premium interpretation of the negative relationship $U = f(W)$.

3. The South African context

In the apartheid years, the organizational power of African workers and their freedom of movement were heavily curtailed. Permanent urban settlement of rural people was prevented and even temporary labor migration was regulated to keep supply in line with demand. In particular, the residents of the so-called ‘homelands’ - accounting for over half of the African population - faced limited employment opportunities and had negligible bargaining power. Knight (1982) argued that there was much disguised unemployment in the homelands which reflected the above-market wage in the formal sector (governed by incentive wage and institutional wage determination) and the consequent rationing of formal sector jobs. By 1993 - the year of our survey - the influx control regulations had been repealed, all workers were free to organize, and trade unions were generally recognized.

Trade unions play a role in wage determination in South Africa. Their impact was recently explored by Schultz and Mwabu (1998) using the 1993 data set and by Butcher and Rouse (2001) using 1995 October Household Survey data. In both these studies, African unionized workers received a wage about 18-20% higher than that received by African non-
union workers with the same observed characteristics. Although firm-level bargaining occurs as well, minimum wage floors are set in many industries via the agency of Bargaining Councils (BCs) or Wage Boards (WBs). BCs are formed if enough employers in a particular sector and area get together and negotiate. The minimum wage agreements reached through collective bargaining within BCs are usually extended to employers in the industry and area who are not parties to the negotiations. Where BCs operate, workers enjoy minimum wages and the union-nonunion distinction is less important: Butcher and Rouse (2001) estimate that coverage by an BC agreement raises African wages by about 10%.

WB determinations cover some of the industries not organized in BCs or, if a BC exists, some parts of the country not covered by the BC. In 1993 26% of the formal sector employees were covered by BCs or WB determinations, another 31% (in mining and the public services) were also subject to minimum wages, and 54% of the formal sector employees were unionized. The commercial agriculture and domestic service sectors were effectively outside these institutions at the time of the 1993 survey. Although BCs and WBs may limit wage flexibility, it should be noted that most of these institutions are sub-national and impose local rather than national wage minima. For instance, in 1996, with the highest minimum wage for laborers in the clothing industry equal to 100, minimum wages in that industry ranged from 97-100 (BCs in the main metropolitan areas) to 69-76 (BCs in particular non-metropolitan areas) to 30-36 (WBs in particular undeveloped, labor-surplus areas).

4. Data, model, and tests

---

3 Bargaining Councils were previously known as ‘Industrial Councils’. The 1957 Wage Act allows for the establishment of Wage Boards. The Basic Conditions of Employment Act of 1997 provides for the establishment of an Employment Conditions Commission. The Employment Conditions Commission may replace the Wage Boards over time.
The data for this study come from the South African Living Standards Survey, collected jointly by the World Bank and the South African Labor and Development Research Unit (SALDRU) at the University of Cape Town in the second half of 1993. This integrated household survey, patterned on the World Bank’s LSMS surveys, produced cross-section data on 8848 households in 360 clusters. The sample was stratified by province, and used a two-stage self-weighting design in which clusters were selected with probability proportional to size, and an equal number of households were selected from each cluster. The data contain detailed information on sample adults’ labor force participation, employment status, earnings (wages as well as housing and transport subsidy, bonuses, and in-kind payments), hours worked in the past week, job-search activity, occupation, industry, and employer-type. Moreover, data are also available on food prices by cluster and on a number of individual and household demographic characteristics.

There is no *a priori* guidance about what geographical boundaries are appropriate for dividing up the country into ‘local’ labor markets for the purposes of wage curve analysis. In a country such as the US, a meaningful and convenient definition of the ‘local’ labor market may be the state: variations in state labor laws probably lead to ‘different’ state labor markets, data are routinely collected stratified by state, and there are more than 50 states. In South Africa, at the time of the survey, new labor market boundaries were probably emerging as a result of the abolition of apartheid laws such as the Group Areas Act and the Influx Control Act which previously limited the mobility of many Africans. Thus, it is not clear what geographical boundaries constituted meaningful ‘local’ labor markets in 1993. Our decision was governed by the data. There are only 14 provinces by the old definition and only 9 by the new definition. Given that our dataset is a cross-section, the number of province-level unemployment rate observations available is too small for precise inferences to be drawn. Both district and cluster

---

4 The data sources are CSS (1994), pages 2.92, 2.146, 5.3, 5.4; and ILO (1996), p.145.
boundaries are available but the use of the cluster is preferable since the number of clusters in the dataset (360) is about double the number of districts (187). There is a danger that the geographical unit chosen will contain heterogeneous labor market conditions. However, the choice of the cluster as the unit (representing an average population of 21,000, corresponding to a town or a city) is likely to ensure that each cluster forms, or is part of, a single local labor market.

For the link between local unemployment and local wages to be meaningful, it is important that a worker living in a particular local labor market also works in that local labor market. The fact that the local labor market as defined for the wage is potentially broader than that as defined for unemployment introduces a potential source of error. Unfortunately, there is no information in the dataset on distance to work to enable us to identify whether an individual works in the cluster of residence.

The cluster-level ‘broad’ unemployment rate has been calculated as follows: in each cluster, let the ‘broad’ labor force participants comprise all employed persons and all unemployed persons, i.e. those persons who did not work but were either looking for work in the past week or who did not actively look for work in the past week because they believed there were “no jobs or work available”. In accordance with these definitions, the ‘broad’ unemployment rate (URATEB) is the ratio of the number of unemployed persons to labor force

---

5 While cluster characteristics and cluster prices are available in the dataset, there is no information on district characteristics and district prices except by aggregation of cluster characteristics. Furthermore, the district unit aggregates rural and urban areas - several districts have both rural and urban clusters in them - which may represent very different types of labor market. Clusters, being smaller geographical units, are likely to represent more homogeneous labor markets.

6 It need not matter that the cluster may be smaller than, and only part of, the local labor market: it can nevertheless be a representative sample drawn from that labor market. There are on average 50 (43) broad (narrow) labor force participants per cluster. We experimented to see whether the results were sensitive to the size of cluster sample. Restricting the sample to workers living in clusters with greater than or equal to 20 (or 40) sampled workers made little difference to the estimated wage curve elasticities. As a further robustness test, we used a cluster unemployment rate calculated without the individual’s contribution to it. Again, this made little difference to the estimated coefficients and the wage curve elasticity.

7 But if we arbitrarily regard only those who work within 30 (and 60) minutes’ travelling distance from home as working within the cluster, the estimated wage curve elasticity is larger than (roughly equal to) that for the whole sample.
participants\(^8\). Broad unemployment includes the narrowly unemployed (those who searched for work in the past week) and discouraged workers (those wanting work but not actively searching in the past week). The narrow unemployment rate (URATEN) excludes non-searching workers from both the numerator and the denominator. It is thus the ratio of narrowly unemployed persons to persons who either worked or looked for work in the past week. While the narrow definition is the official one, Kingdon and Knight (2000) have argued that the non-searchers’ failure of search actively reflects their low perceived benefits of search in relation to the costs, rather than any lack of desire for employment. Accordingly, we conduct our wage curve analysis using both measures.

For the purposes of estimating the wage function, we include in our analysis persons aged 16-64 who were resident in the cluster in which their household was situated (i.e. not oscillating migrants), in wage employment, and for whom wages, hours worked, cluster food prices, and cluster characteristics are available. This yields a sample of 6498 individuals.\(^9\)

The wage function estimated is of the form:

$$\ln W_r = \beta X_{ir} + \gamma U_r + D_r + \epsilon_{ir}$$

(1)

where \(W_r\) is the hourly wage rate for person \(i\) observed in the local labor market \(r\), \(U_r\) is the unemployment rate in labor market \(r\), \(D_r\) is a vector of province dummies, \(X_{ir}\) is a set of measured characteristics of individual \(i\) such as gender, education, marriage status, race, region (i.e. urban/rural and homeland), occupation, and potential experience, and \(\epsilon_{ir}\) is an error term.

The logarithm of the unemployment rate is not used on the right hand side of the wage function

---

8 An individual-specific cluster unemployment rate was also calculated without the index-individual’s contribution to the cluster unemployment rate, i.e. taking the index individual out of both the numerator and the denominator of the calculation. Using this rate made virtually no difference to the wage curve results.

9 Of all ‘resident’ persons aged 16-64 who reported working in the past week in wage employment (n=8118), earnings information is missing or incomplete for 941 individuals and hours worked information is missing for 144 persons. Moreover, information on cluster food prices is missing for one cluster leading to the loss of 53 wage employee observations. Finally, cluster characteristics are missing for 36 clusters altogether, which leads to the loss of a further 482 observations.
because in some 7% of clusters the computed unemployment rate was zero.\textsuperscript{10} In order to allow for the potential endogeneity of unemployment in the wage equation, we shall also conduct an instrumental variable estimation.

Figures 4 and 5 present a first look at the cluster level data and help to check whether the shape of the wage curve is sensitive to outliers. It seems that the cubic form apparent in the simple correlation between wages and unemployment in Figure 4 occurs because of a few outliers. Figure 5 constrains both URATEB and wage to lie within two standard deviations of their respective mean values at the cluster level. It shows that when the few outlying observations are removed, the relationship is first downward-sloping and then approximately flat at very high rates of unemployment. However, we have not as yet controlled for other influences on the wage and, in order not to impose a particular form on the unemployment-wage relationship \textit{a priori}, we experiment with specifications containing quadratic, cubic, and quartic terms of $U_r$.

Some specification issues bear discussion. The inclusion of regional fixed effects in the wage equation is generally preferred because it allows the permanent component of the wage to have a correlation with the permanent component of unemployment, and uses only the deviations of unemployment and wage from their average values to estimate the wage curve elasticity. Card emphasizes this as the reason why the inclusion of region dummies was an important issue in Blanchflower and Oswald’s estimation of the US wage curve.\textsuperscript{11} However, given institutional features that prevented free movement of people until recently in South

\textsuperscript{10} Experimentation with imputing alternative small positive values to the zero unemployment rate observations showed that the estimated wage unemployment elasticity was quite sensitive to the choice of value imputed.

\textsuperscript{11} Card (1995) states that for the US the inclusion of the state fixed effects matters because average levels of unemployment across states are weakly \textit{positively} correlated with average wages, whereas ‘transitory’ wages and unemployment rates are strongly \textit{negatively} correlated. As a result, the US wage curve elasticity tends to be a small negative (sometimes even positive) unless location dummies are included. In the UK data, the addition of region dummies does not affect the estimated wage curve elasticities, perhaps reflecting the greater degree of ‘permanence’ in the geographic patterns of UK unemployment, as noted by Pencavel (1994). Region dummies may have had an important effect on the US wage curve elasticity also because there were no regional price deflators available.
Africa, we expect a large degree of ‘permanence’ in the geographical pattern of unemployment in South Africa and do not expect regional dummies strongly to affect the wage curve elasticity. Since our dataset is a cross-section, only a single value for the cluster unemployment rate is available and, as a result, it is not possible to use cluster dummies. However, nine province dummies are available and these are used as crude regional fixed effects.

Since the cost of living can vary substantially among regions, a specification that uses price-deflated wages is preferable. In the SALDRU data, price information by cluster is available only on food prices. A price index is created by weighting cluster food prices by their weight in the Consumer Price Index and assuming that non-food prices are uniform across the country. This index is then used to deflate wages. Deflating wages by an index of food prices only (i.e. assuming that non-food prices vary across clusters as much as food prices) would over-correct for regional price variations since food prices tend to vary more across regions than the prices of other commodities.¹²

Turning to hypothesis testing, the explanations for the wage curve are best tested by estimating the unemployment elasticity of the wage for particular sub-groups. We compare the wage curves of those workers for whom a theory appears particularly apposite and of other workers. The shirking theory might apply more to workers whose performance is more difficult to monitor or for whom the threat of dismissal is more credible. These might, for instance, be workers in discretionary jobs or workers in whom the firm has not invested heavily. By contrast, the labor turnover theory appears most relevant to workers in whom the firm has invested heavily. Thus alternative versions of the incentive wage theory make different predictions. The bargaining theory is more likely to apply to unionized workers who engage in local bargaining and less to workers subject to non-local wage determination or to workers without organization or power. The wage curve of the relevant group should be downward-
sloping whereas that of the residual group need not be. Unfortunately, the prediction for the residual group is ambiguous. If the wage is market determined, the wage curve tends towards the vertical whereas if the wage is governed by non-local bargaining, the wage curve tends towards the horizontal.

Predictions are also possible about the steepness of the wage curve. A lower risk of dismissal requires a higher wage rate to prevent shirking. This wage premium is dependent on unemployment: the higher is unemployment, the greater the expected cost of becoming unemployed, and thus the smaller the required premium (Blanchflower and Oswald, 1994, pp 76-7). Thus, the more secure workers feel in their jobs (for instance, on account of firm-specific training), the higher is the wage required, and the steeper is the wage curve. The same form of argument applies also to quitting (greater investment in firm-specific capital) and striking (more powerful local collective action). In each case, a higher wage is required to compensate for the additional cost that workers can impose on the employer; in each case this compensation falls as rising unemployment reduces expected income in the alternative state. Finally, what if efficiency wage behavior, institutional wage setting, or non-local bargaining raise the wage above the competitive level in low-wage areas, so generating involuntary unemployment and a negative function \( U = f(W) \)? We would expect the wage curve to be flatter, i.e. the market distortion to be stronger, in the case of workers with characteristics which make them less prone to market wage determination. In South Africa these are the uneducated, the unionized, the production workers, and those employed in the public sector.

5. Empirical results

---

12 For example, Kanbur and Grootaert (1994) find that in Cote d’Ivoire, non-food prices were roughly uniform across the country but food prices varied substantially.
In line with much of the literature on comparator countries, we begin with OLS specifications. Later in this section, the cluster unemployment rate is treated as endogenous. Table 1 presents two different specifications of the OLS wage function, one with and the other without the squared term of $U_r$, defined both broadly and narrowly. Robust t-values are reported rather than raw, to take account of the fact that the regression errors may be correlated across individual workers within each cluster. There is no significant relationship between narrowly measured unemployment and wages but that there is a significant quadratic relationship between broadly measured unemployment and wages.

The usual Mincer-type relationships are found between wages and productive characteristics: for example, wages increase with experience but at a decreasing rate. Marginal returns to education increase with years of education, a result similar to Moll’s (1996b) finding that the returns to primary education are very low in South Africa. Each of the race groups Africans, Coloureds, and Indians earn significantly lower wages than their white counterparts even after controlling for gender, region, productive characteristics, and occupation, and the wage disadvantage is greatest for Africans.

Service occupations are the base category: workers in professional, clerical, and production occupations receive higher wages, but farming workers and laborers receive significantly lower wages than service workers. Male, urban and married workers receive

---

13 See Deaton (1997, p. 77) and Moulton (1990).
14 The quartic $U$ variable was insignificant in all experiments. Though the cubic term of URATEB was significant at the 1% level, this turned out to be due to the influence of outliers. Upon exclusion of the very few waged worker observations (29 out of 6498) that lived in clusters with a URATEB of $\geq 80\%$ (more than three standard deviations above the mean of URATEB at the individual level), the cubic term became insignificant. This confirms that the quadratic shape of the wage curve seen in Figure 5 is robust to the control for other factors that influence the wage, such as personal traits, occupation, employer-type, and region of residence. The results shown in the tables include all observations, including outliers.
15 Experimentation with cubic and quartic terms in experience showed that their inclusion made virtually no difference to the estimated wage curve elasticity.
16 Inclusion of workers’ occupation, sector of work (public), union membership, and location etc. here (as in many standard earnings functions) carries no presumption that these factors are necessarily exogenous. The estimated wage curve elasticity is highly robust to the exclusion of these variables (Table 2).
significantly higher wages than their female, rural, and unmarried counterparts. Controlling for productive characteristics and institutional factors, homeland workers receive significantly higher wages than non-homeland workers, though their raw, unstandardized hourly wages are much lower. This standardized wage premium could be interpreted as a compensating differential for working in low-amenity homeland areas.

As expected *a priori*, public sector and union workers are significantly better paid than private sector and non-union workers. Province dummies are important, signaling that there is significant regional variation in pay levels. Finally, cluster variables are significant influences on wages. Lack of a tarred road in the cluster and greater distance to various facilities (such as post office, bank, restaurant, phone, etc.) reduce wages.

[Table 2 about here]

The elasticity of wage with respect to the local unemployment rate estimated from these and similar models is presented in Table 2. The first row provides basic evidence on the wage curve in South Africa as a whole. Using the quadratic specifications in Table 1 as the basis, the broad unemployment elasticity of the wage evaluated at the mean is nearly \(-0.11\). In words, a doubling of broad unemployment generates an 11% reduction in wages. However, the narrow unemployment elasticity of the wage is close to zero.

The remaining rows of Table 2 provide further estimates of the wage-unemployment elasticity (WUE) under a number of alternative specifications. It shows that departures from the basic specification make little difference to the estimated WUE of about -0.11 using broad unemployment, and of about zero using narrow unemployment, that is, the departures are not statistically significant.\(^{17,18}\) For instance, we see that whether or not the equation is corrected

\(^{17}\) The full set of regressions underlying Table 2 is available from the authors on request. The exclusion restrictions for identifying the lambda term in the selectivity-corrected wage equation in the last row of Table 2 were household non-earned income, number of dependants, and number of other workers within the household. The selectivity term was well identified but selectivity-correction made only a small difference to the size of the wage curve elasticity.
for sample selectivity, and whether or not wages are deflated, make no notable differences to the WUE using broad unemployment.

These findings shed new light on the on-going debate about the appropriate definition of unemployment in South Africa. The ILO Report (ILO, 1996) argued that the broad measure exaggerates unemployment by including in the labor supply persons who did not actively look for work in the past week. The results here show that broadly measured unemployment has a greater impact on wages than narrowly measured unemployment.\(^{19}\) In other words, discouraged workers are taken into account by wage-setters and their numbers in a locality do influence local wage determination. These appear to be persuasive grounds for retaining the broad definition as the more appropriate concept of unemployment in South Africa.

In order to throw light on the mechanisms underlying South Africa’s wage curve, Table 3 presents the broad unemployment elasticities of wage for different groups of workers, disaggregated by education, experience, sector, gender, race, and region.\(^{20}\) In South Africa, the pay of many public sector workers is set centrally.\(^{21}\) We would expect workers whose wages are negotiated centrally to have a flatter wage curve than workers who engage in local level bargaining. In line with this hypothesis, Table 3 shows that the wage curve is substantially flatter for public than for private sector workers. A comparison of union and non-union workers

---

\(^{18}\) Choice of too large a unit of analysis – most unlikely to be true of the cluster – may dilute the effect of local unemployment on wages. Experimentation with the district (as opposed to cluster) unemployment rate showed that the point estimate of the coefficient on the district broad unemployment rate (DURATEB) was lower but insignificantly different from the point estimate of the coefficient on the cluster broad unemployment rate (URATEB). However, the coefficient was imprecise and insignificant at the 5% level. This is not surprising since the standard error of DURATEB was much higher (than that of the cluster URATEB) because of fewer observations. The estimated wage curve elasticity using DURATEB was \(-0.06\) (t-values: -1.70 on DURATEB and 1.67 on the square of DURATEB). The use of the province also produces a negative elasticity (-0.126), but it is very imprecisely determined due to the small number of provinces. The main result is robust to the choice of unit.

\(^{19}\) Although it is possible that discouraged workers are a proxy for demand-side factors that depress the wage, the most plausible demand-side factors – location and remoteness – are already controlled by means of the province, region and cluster-characteristics variables.

\(^{20}\) When estimating wage equations for different worker groups in Table 3, selectivity issues arise since some of the sub-samples are potentially non-random. However, it was not possible to fit selectivity-corrected wage functions for each worker group owing to the unavailability of variables that could plausibly predict choice and also be excluded from the corresponding wage equations.
shows that unionized workers’ wages do not respond negatively to local unemployment rate at all. The bargaining hypothesis implies that the wage curve exists when local bargaining takes place but not if bargaining is national or regional: non-local bargaining insulates workers’ pay from local labor market conditions. In South Africa, much union bargaining is non-local and unionized workers are in a better position to ensure that Bargaining Council non-local wage agreements are implemented by recalcitrant employers who were not party to the Bargaining Council wage negotiations. Thus, the results for unionized workers (proxying bargained wages) are as expected.

[Table 3 about here]

Because education and training are normally complementary, and because educated workers are likely to be in discretionary jobs, education can serve as a proxy for both firm-specific skills and difficulty of monitoring. The same is likely to be true for age, length of employment experience, and more skill-intensive and responsible occupations. Table 3 shows that the wages of older, more experienced, and more educated workers are less sensitive to local unemployment than are those of younger, less experienced, and less educated workers. These former characteristics are liable to encourage efficiency wage setting so as to avoid shirking, whereas the latter are more likely to produce market wage setting. The results are thus in line with one version of the incentive wage rationalization of the wage curve.

In summary, the fact that younger, less experienced, less educated, non-unionised, and private sector workers tend to have steeper wage curves suggests that the free-market tendency (proxied by youth, low skill, and weak organization) is towards a vertical wage curve, and that non-local bargaining (proxied by union membership and public sector) is towards a horizontal wage curve. The elasticities for older, experienced, and educated workers are consistent with

---

21 But not in all cases, e.g., provincial administrations and local governments set their own wages and these take account of local labor market conditions.
the efficiency wage rationalization of the wage curve, and the elasticity for private sector workers with the bargaining explanation. However, these results are no more than suggestive of the forces that might be shaping the wage curve.

The fact that women have a very flat wage curve may be due to our use of the overall cluster unemployment rate, rather than the gender-specific unemployment rate. If women’s and men’s jobs are not substitutes (as may be the case in certain occupations and industries, e.g. mining, domestic work and health care), then the relevant unemployment rate to use is the gender-specific one. There is support for this hypothesis. Using gender-specific unemployment rates, the WUE is marginally lower for males than for females: -0.040 for males (t = -2.8, 3.2, for the linear and squared term respectively) and -0.063 for females (t = -2.7).23

Table 3 also shows that the wages of rural workers are about as flexible with respect to local labor market conditions as those of urban workers. By contrast, whereas a highly significant negative elasticity is obtained for non-homeland areas, homeland wages are not negatively related to unemployment in South Africa. The wages of non-African workers are apparently substantially more responsive to unemployment than are those of African workers but this arises from their enforced spatial concentrations. As will be seen later, non-Africans are concentrated in low unemployment regions, i.e. regions where the wage curve is steeply downward sloping, whereas Africans are concentrated in high unemployment clusters where the wage curve is flatter24.

22 Union membership information is available only for regular wage workers and not for casual wage workers. We have assumed that none of the casual workers is unionized. However, if some are unionized, as is likely, then the true UEW for non-union workers should be greater (i.e. a bigger negative) than the reported figure.
23 For women, the wage-unemployment relationship is linear. Although it may be desirable to use gender-specific unemployment rates, computing these implies that only about half the number of labor force participants are available for each gender and sample sizes become too small for reliable estimates of cluster unemployment rates.
24 We considered separately conducting the analysis for Africans and non-Africans throughout the paper, but decided against this on both statistical and economic grounds. First, although the wage functions for Africans and non-Africans are significantly different (F-value = 4.94), this is true also for each of the other divisions shown in Table 3. Second, by 1993 the different races in a particular area were operating in a common labor market, standardising for worker characteristics. It is not the case, for instance, that conditional African wages depend on the local unemployment rate for only Africans.
Table 1 showed that South Africa apparently has a U-shaped wage curve, with the minimum point occurring at a broad unemployment rate of 36%. Is the nature of the wage-unemployment relationship significantly different as between South Africa and other countries? In their analysis of wage curves in the UK and USA, Blanchflower and Oswald (1994) did find that a substantial part of the wage curve slopes upward, and they said that this “might, in principle, reflect Harris-Todaro forces, although such an interpretation would be highly speculative” (p. 106). They suggested that the upward-sloping part of the wage curve in UK and USA rests on small numbers of observations in the high unemployment portion of the curves and on wage functions in which it was not possible to control fully for regional and industry fixed effects (Blanchflower and Oswald, 1993, p. 245). They also argued that since the upward-sloping part occurs at very high rates of unemployment which are not observed in the actual data, one could reasonably ignore that part of the wage curve. However, one cannot dismiss the upward-sloping part in the South African wage curve because a significant minority (27%) of wage employees live in clusters in which the broad unemployment rate is greater than 36%.

[Tables 4a and 4b about here]

To explore further the nature of the wage-unemployment relationship, we sought to divide the country into low and high unemployment regions. We expected unemployment to be quite different in rural and urban parts of the country. However, Tables 4a and 4b show that the rural-urban division does not capture labor market segmentation well and that the clearest labor market fault-line occurs along ‘homeland’-non homeland lines. Broadly measured unemployment is catastrophically high in the (former) homeland regions, and it is much lower in non-homeland regions. Consequently, we fitted separate wage functions for the two regions (not reported but available from the authors). A Chow test rejected the pooling of the two samples.
The first row in Table 5 shows that ‘homeland’ South Africa - with its very high unemployment rates - exhibits no significant relationship between local unemployment and wages. This flatness of the wage curve suggests that at the very high levels of unemployment found in homeland areas, relatively small changes in unemployment do not affect wages. This may be because the high levels of unemployment reflect long duration of unemployment (rather than high inflow into unemployment) and the obsolescence of human capital and skills of the long-term unemployed mean that they provide less competition for jobs, so that their presence in the labor force exerts little downward pressure on wages.

Non-homeland South Africa, by contrast, displays a significant convex relationship between unemployment and wage, with a WUE of -0.11. The minimum point occurs at a high broad unemployment rate of 36%. Since the vast majority (87%) of the non-homeland workers live in clusters which have a unemployment rate below 36%, the downward-sloping part of the wage curve is the more relevant.

The wage curve elasticities estimated for different groups of workers in homelands and non-homelands are presented from the second row onwards in Table 5. Since in the case of the homeland areas, there is no significant non-linear relationship between unemployment and pay, and since in non-homelands, the downward sloping part of the wage curve is the more relevant, we refer to the elasticities in the linear columns. The elasticities for different groups of workers in homelands are almost invariably insignificant. The elasticities for different groups of workers in non-homelands South Africa show that the results are quite similar to those for South Africa as a whole in Table 3, and bear the same interpretations. There is, however, one interesting difference. In rural non-homelands, with their low unemployment rates, wages are three times as responsive to unemployment as in urban non-homelands. This is likely to be because wages tend to be less determined by bargaining in rural non-homelands - home of South African
commercial agriculture - because employers are smaller, there is little unionization of agricultural workers, and the dominant rural industry (farming) was not covered by an Industrial Council in 1993.

[Table 6 about here]

Table 6 sets out the wage curve elasticities by industry/sector in non-homeland areas, the part of South Africa for which a negative relationship exists.\(^{25}\) In all sectors except the two smallest (electricity and restaurant/entertainment), there is a large negative wage curve elasticity. The elasticity is low - a small negative - in ‘other sectors’, i.e. in educational, legal, medical, and armed forces sectors where public sector workers are concentrated and centralized bargaining is likely to determine wages. The elasticity is high (a large negative) in the agriculture, fishing, and forestry industry where there is little institutional wage determination. The elasticity is also high in transport and construction sectors where the large majority of workers are non-unionized. While some sub-industries within these sectors are covered by Bargaining Council agreements, the relevant Bargaining Councils are very localized and their wage agreements are likely to take local market wages and local unemployment rates into account.

We are interested not only in the shape of the wage curve but also in the size of the wage curve elasticity over a range of unemployment. Table 7 shows that in non-homelands, the unemployment rate can be tripled from 10% to 30% without reducing the wage curve elasticity below about -0.1. In South Africa as a whole, the unemployment rate can be increased two-and-a-half times from 10% to 25% without reducing the size of the wage curve elasticity below about -0.1. In other words, the wage-unemployment elasticity is robust over a large range of unemployment.

\(^{25}\) While one would expect that workers in industries where Bargaining Councils (BCs) exist will have flatter wage curves than other workers, in practice it is not straightforward to test this because BCs are organised for sub-industries rather than for the broad industrial categories available in the dataset. Moreover, most BCs cover small areas, not large regions, so they may apply to some but not all workers in a given industry.
Until now both the dependent variable and all the independent variables except the unemployment rate and cluster characteristics were measured at the individual level. Unemployment was measured at the aggregate, *i.e.* cluster, level. The fact that the unemployment variable $U_r$ in equation (1) has no $i$ subscript has an important implication. It is reasonable to expect that individuals living in the same locality may share some common unobservable characteristics that would lead the regression disturbances to be correlated within the cluster. Following Moulton (1990), it can be shown that even small levels of such correlation can cause conventional standard errors of $U_r$ to be significantly downward-biased, resulting in spurious findings of statistical significance of $U_r$. All the regressions reported thus far show t-values based on standard errors corrected for this possible source of bias. However, as a further test of the robustness of the wage curve elasticity, we estimate a model using cluster means of characteristics instead of individual observations. Making the level of aggregation the same on both sides of the wage function eliminates the downward bias on the standard errors.

The wage functions using cluster-level means of hourly wages and individual traits are presented in Table 8 and the elasticities computed from these regressions are contained in Table 9. Table 9 shows that, when aggregate data are introduced, the estimated wage-unemployment elasticity falls from about -0.11 to about -0.07 for all South Africa, and the coefficient on unemployment ceases to be significant. However, for non-homeland areas, the elasticity is virtually unaffected and the coefficient of unemployment is still significant at 1%.

A further issue of importance is potential simultaneity bias: wages and local unemployment may be jointly determined. A correction for the simultaneity bias in $U_r$ requires plausible instruments. The instruments most often used in the literature are lagged local unemployment rates but, given our cross-section data, these are not available to us. However, if
wages do influence unemployment, the effect is expected to be positive (the higher the institutionally-determined wage is above the market-clearing level, the higher is unemployment). In that case, any negative effect of unemployment on wages is likely to be underestimated in OLS estimation of the wage curve, i.e., the OLS estimate of the negative wage-unemployment elasticity is expected to be closer to zero than to the true negative elasticity. Thus, even if the strategy below for correcting for simultaneity bias is unconvincing, our OLS estimate of -0.11 gives us a lower bound on the South African wage curve elasticity. Since despite bias towards zero, our OLS estimate of the wage curve elasticity is -0.11 and is highly significant, this suggests that wages are at least as responsive to local unemployment rates in South Africa as they are in the OECD countries for which wage curves have been estimated; they may well be more so.

A number of cluster characteristics are used as instruments for the cluster unemployment rate, including whether the community has roads that become impassable at certain times of the year, whether it has any public transport passing by it, and distance to nearest transport. A priori it is expected that condition of roads and availability of transport will affect the cost of job-search and thus the local unemployment rate, but should not affect wages other than through their impact on unemployment. The instrumental variables for $U$, were generally insignificant in the wage equation when local unemployment rate was included, but are well correlated with unemployment. The standard tests proposed by Bound et al. (1995) – the test of overidentifying restrictions, the F-test of joint significance of instruments in the first stage regression, and partial R-squares of the regression of the endogenous variable on the identifying instruments, reported in Table 10 – all indicate that the instruments are of high quality. Table 10 shows that in each case (homeland, non-homeland, and all South Africa) the elasticity

\footnote{The full set of results, including first stage estimates, are available from the authors upon request.}

\footnote{Angrist, Imbens, and Rubin (1996) show that two conditions must be fulfilled for the identification of causal effects using instrumental variables: the correlation of the instrument of $U$ with the error term in the wage}
estimated from the 2SLS model is larger (a bigger negative coefficient) than that from the corresponding OLS model. This is consistent with the existence of an even steeper wage curve than previously reported in Tables 3 and 5. In the non-homeland case, the 2SLS result is highly significant, both in the linear and quadratic forms.

[Table 10 about here]

6. Conclusion

The wage curve has been the subject of intensive discussion, primarily using OECD data where unemployment rates are typically 5-12% (Blanchflower and Oswald, 1994, p. 297). The robustness of the wage curve, both across countries and across ranges of unemployment rates, has become an interesting question. In this paper we look for a wage curve in South Africa, where the average unemployment rate is very high. We test for the wage curve relationship \( W = f(U), f' < 0 \), which contrasts with the conventional assumption about labor market behavior in developing countries, \( U = f(W), f' > 0 \). There is no evidence in favor of the latter. The negative relationship is robust to the inclusion or exclusion of the regional dummy variables that might standardize for a long run positive reverse relationship. Even in the mid-1990s, after the relaxation of controls on mobility, the Harris-Todaro model appeared not to be relevant to South Africa. This finding presents a further challenge to one of the best accepted theoretical models of the way the labor market operates in developing countries with high unemployment.

South Africa has one of the most interesting labor markets in the world. Its powerful labor market institutions and strong labor unions would suggest lesser flexibility of wages to local unemployment than in OECD countries. Yet we find that the South African wage curve elasticity is no weaker than in OECD countries, and that it is robust over a wide range of unemployment. Tripling unemployment from 10% to 30% reduces wages by approximately
30%. The wage curve was subjected to a range of robustness tests - allowing for possible endogeneity, dividing the workers into high/low unemployment rate areas, dividing the sample into different groups of workers, and allowing for the correlation of errors across individuals within regions - but none of these reduces the wage curve elasticity below about -0.10 over the range of unemployment between 10% and 30%. At rates of unemployment above 30%, the wage unemployment elasticity falls to zero. The South African labor market displays considerable wage-flexibility in response to local labor market conditions across this large country.

Our estimates of the elasticity of the wage curve for various sub-groups generally correspond to those in other studies. Like Blanchflower and Oswald (1994, pp. 148-67, 258, 261, 342, 349-51; and Hoddinott (1996, pp. 1618-19), we found that younger, less experienced, less educated, non-unionized, and private sector workers tend to have steeper wage curves. This suggests that the free-market tendency (proxied by youth, low skill, and weak organization) is towards a vertical wage curve, and that non-local bargaining (proxied by union membership and the public sector) is towards a horizontal wage curve. The elasticities for older, experienced, and educated workers are consistent with the efficiency wage rationalization of the wage curve, and the elasticity for private sector workers with the local bargaining explanation. It also appears that unions, often bargaining at the non-local level, limit the impact of local unemployment on wages in South Africa.

The paper illuminates the important debate about the relevant definition of unemployment in South Africa. The South African government has recently adopted the narrow concept as the official definition and has de-emphasized estimates of broad unemployment (StatsSA, 1998). Yet, our analysis indicates that the broad definition of unemployment is the appropriate one for labor market analysis in South Africa since local wage determination takes zero. The test results reported in Table 10 show that both conditions are fulfilled.
discouraged workers into account as genuine labor force participants. Although even the narrow estimate of unemployment is high and justifies policy priority, the alliance between the ruling party and the trade union movement (representing the employed) might well lead to a playing down of the unemployment issue. Our argument in favor of the broad measure of unemployment therefore has policy relevance. Moreover, it is clear that unemployment contributes to poverty both directly and indirectly through its effect on the wage level.

Over the decade 1990-2000 the South African labor force grew by 3.4% per annum whereas formal sector employment fell by 1.2% per annum and formal sector real wages rose by 3.3% per annum (StatsSA, 2002, pp. 7.21, 8.20). This descriptive evidence suggests that, over time, there is a positive relationship between wages and unemployment. By contrast, our evidence of a wage curve suggests that, across space, there is a negative relationship between unemployment and wages. This apparent contradiction would be consistent with an exogenous rise in the wage curve over time, possibly associated with increased bargaining power or greater job security. This could generate a new equilibrium higher up the aggregate labor demand curve, involving a higher wage and more unemployment.

Fallon and Lucas (1998) have produced time series estimates showing the South African wage elasticity of demand for labor to be large and negative (-0.7) so that, over time, a 10% increase in wages is predicted to lead to a 7% fall in employment (p. 11). This high wage elasticity of demand for labor implies that wage flexibility in response to unemployment would help significantly to raise employment and thus to lower unemployment. However, the remarkable pace of divergence (by 4.6% per annum over the 1990s) between the rise in the labor force and the fall in formal sector employment would put great strain on any labor market, however flexible it might be in response to marginal changes.
References


