ARE SKILLED AND UNSKILLED LABOUR COMPLEMENTS OR SUBSTITUTES?

Alberto Behar Nuffield College

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, Hilary 2007

Abstract

Are skilled and unskilled labour complements or substitutes?

Alberto Behar, Nuffield College; thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, Hilary 2007

Using theoretical and empirical approaches, this thesis asks whether skilled and unskilled labour complement or substitute one another in production. We primarily investigate whether an increase in the proportion of workers with skills would raise or lower demand for those who remain unskilled. A secondary issue is the role of factor prices in labour demand.

To study the role of factor prices, we estimate labour demand elasticities and Allen elasticities of substitution between capital and up to five occupations in South Africa. We supplement firm-level data with household survey information and confirm theoretically that the elasticities can be estimated from a cost function under non-constant returns to scale. We show that separable disaggregated inputs can be used to find aggregate elasticities: more skilled and less skilled aggregates are p-complements, so a fall in skilled wages would lead to a rise in demand for less skilled labour. Disaggregated estimates suggest unskilled workers are p-complements with semi-skilled workers but p-substitutes with skilled/artisanal labour.

We investigate the effects of a rise in skill supply on the relatively unskilled by estimating Hicks elasticities of complementarity and factor price. Aggregated estimates suggest more skilled and less skilled labour are q-complements, so an exogenous rise in the supply of skilled labour would raise demand for less skilled labour. Disaggregated estimates suggest skilled/artisanal and unskilled labour are q-complements while semi-skilled and unskilled labour are q-substitutes. The results allow for imperfectly elastic product demand and rigid wages.

Using an endogenous growth model, we show technological progress is skill-biased in the South if it is in the North, resulting in rising wage inequality in developing countries. Assuming skilled and unskilled labour are perfect substitutes, we model expanded educational access as it adds relatively educated cohorts to the labour market. A rising skill composition causes accelerated skill-biased technological change and wage inequality. Relaxing the assumption of perfect substitutability, a one-off rise in skill supply only raises wage inequality if the elasticity of substitution is high, higher than existing empirical estimates.

Contents

List of Figures

List of Tables

Acknowlegements

First and foremost, sincere thanks to my supervisor Margaret Stevens. She has been on call to offer assistance whenever required and her turnaround times for work sent to her should be the envy of many. She has given me the space needed to run with my ideas but also willing to put careful effort into aspects about which I was not sure. Her willingness to adjust to my schedule despite a very tight one of her own, especially in Michaelmas 2006, are most appreciated.

I wish to thank the Department of Economics, including the lecturers and class teachers on the M Phil. Without the grounding offered by that degree, this thesis would not be what it is. I also thank the administrators at the department. I thank my college, Nuffield, for providing a fantastic work environment. I thank its IT support team, the librarians and everyone else in the Nuffield community. Resources were also provided by the University of Cape Town and its Datafirst Resource Unit, including access to data and computing facilities for working on the data. Stating the obvious, I could not have done this work without this institutional support.

I am most grateful for the funding that gave me the opportunity to engage in this research. Financial support came from the Clarendon Fund and Nuffield College, the Department's Doctoral Studentship Scheme (including an anonymous benefactor for the final year scholarship) and the University of Cape Town.

For sustained and invaluable support over the years, I thank T Khan and R Ritz. Furthermore, thank you to C Batista, S Belenzon, H Bhorat, F Bilbiie, T Bold, S Bond, E Budish, C Caceres, S Cowan, Department of Labour, F Ederer, M Ederer, L Edwards, R Evans, E Gerdel, J Giese, M Gregory, J Large, M Leibbrandt, R Meeks, J Muellbauer, N Nattrass, B Nielsen, T Pekkarinen, V Prowse, J Sandefeur, M Sanders, J Seekings, F Teal, M Tse and L Tuthill. I also thank participants at the Gorman Student and the Post-Doc / D Phil workshops in Oxford, The University of Cape Town Economics Departmental Seminars and at the 2005 European Economics Association World Congress.

This is my own work. Any errors and/or omissions remain solely my responsibility. Where the work builds on the work of others, appropriate references have been made in the text. Early drafts of chapters 2-5 were submitted in partial fulfilment of the requirements for an M Phil (Economics) at the University of Oxford. With the exception of chapter 3, which has had minor alterations, these chapters have been substantially revised in a material manner. The revisions pertain not only to presentation, but to additional theoretical modelling and empirical results.

WORD COUNT: The main body of text contains approximately 150 pages and 450 words per page resulting in a word count of 67 500 words. The full thesis, including the front cover, abstract, table of contents, acknowledgments, appendices and bibliography contains approximately 210 pages.

1 Introduction

1.1 Background and motivation

While applicable to a broad range of developing countries, the issues discussed in this thesis are particularly important for South Africa. Following the end of apartheid and the advent of democracy in 1994, a key priority of its government is to reduce poverty and inequality. Granting formerly barred segments of the population access to education is seen as a major way of achieving that aim. Mainly through improved job market prospects, improving the quantity and quality of education received by an individual is regarded as one of the best ways to bring that person out of poverty (Bhorat, Leibbrandt, Maziya, van der Berg & Woolard, 2001; May, 2006). However, this raises the question of what will happen to those who remain unskilled. Will they be indirect beneficiaries or losers from others' receiving education or training?

Within this domestic historical context, forces of economic globalisation have exposed South Africa (and other developing countries) to international competitive pressures. The Global Competitiveness Report views an increase in the skills base, through expansion of both primary and higher education, as a key ingredient for competitiveness in all countries (Lopez-Claros, Altinger, Blanke, Drzeniek & Mia, 2006). The importance of education in developing countries is treated as conventional wisdom, so conventional in fact, that it is offered on some sugar sachets.[1] Some developing countries have at least the intention to expand educational access in general while others are concentrating on particular segments, for example education for women (UNESCO, 2000).

In addition to direct poverty alleviation, the motivation behind the South African government's Joint Initiative on Priority Skills Acquisition (JIPSA) is the view that a shortage of skills is constraining GDP growth and contributing to unemployment. South Africa has an uncharacteristically large unemployment problem. Narrowly defined, this rate is 25% (Statistics South Africa, 2005). Lest there be any doubts as to current thinking on the importance of skills for the broader economy, consider the words of South African Deputy President, Phumzile Mlambo-Ngcuka (2006), at the launch of JIPSA, which is part of the new Accelerated and Shared Growth Initiative for South Africa (AsgiSA):[2]

> *"Yet both unemployment and poverty are still at unacceptably high levels, which mean our growth is not fairly shared. The most fatal constraint to shared growth is skills, and it should be noted that skills are not just one of the constraints facing AsgiSA but*

[1] In Argentina for example, the quote is "La educacion es la base del desarollo", which means "Education is the foundation of development".

[2] AsgiSA is the new flagship of government economic policy, taking over from the 1996 GEAR policy.

a potentially fatal constraint. That fact should be admitted with emphasis. We have to overcome the shortage of suitable skilled labour if our dreams for this economy are to be realised; the task is huge."

As is clear from the quote, and as the name of the program suggests, a cornerstone of AsgiSA is that improvements in living standards are to be shared by all segments of society, in particular the poor. Implicit in the argument for the role of skills in AsgiSA is the claim that expanded and improved educational access, which would equip a portion of the population with skills, will generate enough growth in a way that benefits those who do not get access to those skills.

To what extent skill supply has changed is not clear. Van der Berg (2001) shows that, after 1994, major shifts of fiscal resources from traditionally white to traditionally black schools took place but that this was not accompanied by major changes in teacher:pupil ratios nor outcomes. Chisolm (2004) provides a summary of the intended measures taken. According to the South African Qualifications Authority (SAQA, 2004), the stock of people with tertiary education doubled to approximately one million people between 1992 and 2001. This is countered by indications that about one million people, most of them believed to be skilled, have emigrated since 1994 (Crotty, 2006). While past and present success in expanding educational access may be difficult to establish, the intention of the thesis is to understand what impact a *future* increase in the stock of skills would have, *if* it takes place.

The change of government has brought in new labour legislation (Republic of South Africa, 1995), which arguably has increased the cost of labour, especially for occupations towards the lower end of the skill spectrum. Thus, while potentially reducing wage inequality within the formal sector, this may lead to increased unemployment. We also hear repeated calls for reductions in the cost of capital on the premise that this would stimulate growth and in this way reduce unemployment. This call has recently been renewed by a high profile United Nations sponsored study (Pollin, Epstein, Heintz & Ndikumana, 2006). From a microeconomic perspective, it is possible that a rise in the cost of labour, either in absolute terms or relative to capital, will have adverse effects for labour. Therefore, it is important to understand the role of absolute and relative factor prices.

1.2 Key questions

In asking "Are skilled and unskilled labour complements or substitutes?", the thesis seeks to address two major questions. The dominant issue is what will happen to the unskilled, in terms of demand for their labour, if the skill proportion of the economy rises. A secondary issue is the potential role of factor prices in the labour market. To address these questions, the thesis will provide theoretical and empirical analyses of the interactions between skilled and unskilled labour. The main theory model aims to highlight phenomena affecting developing countries as a whole and some of the modelling required for the empirical analysis will be broadly applicable. However, most of the data used for empirical work will be from South Africa.

1.2.1 The role of skill supply

A dominant issue in this thesis is whether a rise in skill supply will have positive or negative labour market consequences for the unskilled. The precise interpretation of "skills" will vary across parts of the thesis. The theoretical contributions will regularly but not exclusively refer to education levels while the empirical sections have skill levels defined in terms of occupation type. We briefly introduce three arenas in which skilled and unskilled labour could interact.

i) One way, not discussed at length in this thesis, for higher national educational levels to have positive effects on the unskilled is through technology adoption. Nelson & Phelps (1966) advance the idea that new technologies can be better understood and implemented by skilled people. The endogenous growth models of Lucas (1988) and others depict a similar role for human capital. Such technologies improve total factor productivity, which increases the rewards to all factors of production including unskilled workers. The shared productivity effect is a potential source of social returns to education in excess of private returns. A finding that social returns to schooling exceed the private returns might suggest the poor would benefit from human capital acquisition, even if it's not them acquiring it. If, on the other hand, education is merely used as a signal of productivity, not a generator of it, then there are no productivity effects from schooling (Bils & Klenow, 2000) and the only people to benefit are those enjoying higher earnings, not the broader economy. An introductory analysis of the evidence in Bils & Klenow (ibid.), Pritchett (1999), Temple (2001) and Topel (1999) suggests the evidence for positive spillovers to the unskilled is weak (Behar, 2006).

ii) A second way for the unskilled to benefit is if, by removing some of the obstacles to output expansion, additional skilled workers increase GDP and generate demand for all inputs including unskilled workers. This is implied by the quote in section 1.1. However, this positive influence on the demand for unskilled labour must be larger than another potentially negative effect: firms may hire the additional skilled workers to replace unskilled workers, not necessarily to embark on new projects. If this effect is large, then the forces causing unskilled labour to be replaced could outweigh those forces adding to unskilled labour demand. Overall, a rise in skill supply would then raise wage inequality while the effect on GDP would be limited.

We can deduce the sign and magnitude of the overall impact of a rise in skill supply using the Hicks (1970) elasticity of complementarity (HEC) and the associated elasticity of factor price. If the sign is positive, two factors are q-complements. If it is negative, two factors are q-substitutes. One contribution of this thesis is its estimation of empirical measures of these elasticities. We do this for multiple inputs with South African data, using the signs of the estimates to determine whether they are q-complements or q-substitutes.

iii) While the HEC holds technology constant, the directed technological change literature warns how a rise in the skill supply can lead to endogenous *changes* in technology that favour skilled workers. Skill-biased technological change (SBTC) can lead to a rise in wage inequality. The central feature of this literature is that technologies are purposefully developed to suit a particular market. There are separate markets for technologies that complement skilled workers and for technologies that complement unskilled workers. A rise in the supply of skilled workers can make skilled technologies more attractive to adopt and unskilled technologies less attractive. The result is SBTC and potentially increased wage inequality. This suggests the poor might be harmed by a rise in skill supply.

Another contribution of this thesis is its adaptation of such models to a developing country setting. It also models expanded educational access on a cohort-by-cohort basis using a Markov model. Furthermore, it proposes an alternative potential cause of SBTC, arguing it takes place in developing countries simply because it does in technologically advanced nations.

This thesis assumes in general that the rise in skill supply is exogenous, particularly with respect to wages. The justification for this is that a vast proportion of the rise in skills acquisition will come from previously barred segments of society. According to our calculations based on data from SAQA (2004), 99% of the growth in the stock of tertiary graduates came from black people. In assuming exogenous skill supply, the scenario set is one of skill acquisition being constrained by the supply of training rather than demand, which at current wages already exists. This scenario is consistent with the view held by the Department of Labour (1997) in South Africa. In accordance with this view, increased government willingness and ability to supply training would lead to a rise in skill acquisition and the stock of skills.

1.2.2 Factor prices

A secondary question in this thesis is the potential role absolute or relative factor prices may play in determining labour market outcomes, especially for the unskilled. We therefore perform empirical estimations of Robinson (1933) / Allen (1938) / Uzawa (1962) elasticities of substitution and elasticities of labour demand. Furthermore, we can test for capital skill complementarity (Griliches, 1969). Capital skill complementarity (CSC) holds if capital is more easily substitutable for unskilled labour than for skilled labour, such that a fall in the price of capital will have a proportionally worse effect on demand for unskilled labour. CSC is an alternative to rising skill supply as a potential explanation for rising wage inequality. We are also able to investigate whether, in terms of the sign of the elasticity of substitution, skilled and unskilled labour are p-complements or p-substitutes.

1.3 Layout

Chapter 2 begins by describing and contrasting the Robinson (1933) / Allen (1938) / Uzawa (1962) and Hicks (1970) elasticity concepts, tracing out their theoretical links, relevant applications and basic estimation issues. It also describes the questions and assumptions under which one concept or the other is more appropriate. The chapter includes an extension to general technologies of the Uzawa (1962) result, which exploits duality concepts to describe the Allen elasticity in terms of cost function parameters under constant returns to scale only. The chapter continues with an in-depth discussion of the literature on separability, both in general and in the context of the production technology estimated, namely the translog function. The separability discussion aids in determining whether or not certain variables can be omitted for estimation purposes, whether others can be aggregated legitimately and whether we can adopt a value added specification when estimating the production technology. Another theoretical contribution confirms how aggregated price inputs can be used to calculate a price elasticity of demand for aggregate factors, which is analogous to the result for elasticities of substitution in Berndt & Christensen (1973a).

Chapter 3 describes the core manufacturing dataset used for the estimation of these elasticities. The data section also motivates and describes the process by which we combine household wage data with firm-level data, covering both theoretical and practical issues encountered. This includes a brief theoretical analysis of the effects of failing to account for firm-size effects in wage construction.

Chapters 4 and 5 build on the framework established in chapter 2 and use the data from chapter 3 to make two empirical contributions. Chapter 4 estimates cost functions and a system of cost share equations to present highly disaggregated measures of the Allen/Uzawa elasticity of substitution and factor demand elasticities based on Marshall's Rules. These allow us to investigate the potential role of relative factor prices in unemployment. Our focus is on the links between skilled and unskilled labour, so these estimates suggest whether skilled and unskilled labour are so-called p-complements or p-substitutes. If they are p-complements, a fall in the price of skilled labour would lead to a rise in the quantity of unskilled labour demanded. If they are p-substitutes, a fall in the price of skilled labour would to a fall in the quantity of unskilled labour demanded. We also consider the cost of labour relative to capital.

Our analysis is special because it allows for so-called scale or output effects. Holding output constant, a fall in the price of one input means it may be substituted for another input. However, there may be an expansion of output such that, overall, there is a rise in demand for all inputs. We also draw on the restrictions implied by separability to use disaggregated data to estimate aggregated elasticities of substitution. We use a result from chapter 2 to be the first study (to our knowledge) to utilize theory explicitly to construct aggregated elasticities of factor demand.

Chapter 5 estimates production functions to present disaggregated measures of the Hicks elasticity of complementarity and elasticities of factor price. These model the effect of a rise in the supply of one factor on the demand for other factors, holding technology constant. It allows us to test the belief of government officials that a shortage of certain skills is constraining output and that a relaxation of such constraints would lead to a rise in demand for the unskilled. We examine the role of capital but emphasise the role of increased skill supply, being particularly interested in whether skilled and unskilled labour are so-called q-complements or q-substitutes. If they are q-substitutes, a rise in the supply of skilled labour would lead to a fall in unskilled wages. If they are q-complements, a rise in the supply of skilled labour would lead to a rise in unskilled wages.

The chapter presents both aggregated and disaggregated estimates. The default model is one where all factor prices are flexible, such that the effects on unskilled labour are wage effects and not quantity effects. We also allow for the possibility that the unskilled wage is exogenously fixed, so that the effects on unskilled labour are quantity effects. Existing empirical applications of this type assume perfectly elastic product demand. To our knowledge, this is the first to address the issue of imperfectly elastic demand.

Chapter 6 uses a discussion of the literature on SBTC and empirical evidence to suggest SBTC can lead to a rise in wage inequality. Also, it suggests SBTC is driven by profit-driven technology developers/adopters so that (a) the proportion of skilled labour affects the market for skill-biased technologies and (b) developing countries acquire skill-biased technologies because SBTC in technological leaders makes it relatively cheaper to do so. Existing formal modelling of (a) shows a rise in skill supply leads to SBTC and potentially to a rise in wage inequality. Chapter 6 includes a brief formal review of two models in this class. We are not aware of any existing models of (b).

Chapter 7 uses an endogenous growth model to capture the effects of both (a) and (b) on technology choice and evaluates their relative potential for explaining SBTC. It does so on the assumption of perfect substitutability between skilled and unskilled labour to allow us to focus on the key mechanisms. Chapter 8 relaxes the perfect substitutability assumption. Using available estimates of key parameters, we can examine whether or not it is likely a rise in wage inequality would result after a rise in skill supply.

Chapter 9, the final chapter, brings together the results from chapters 4 and 5, searching for consistencies and contradictions. Together with the results from chapter 8, it provides a synthesis on the interactions between global labour demand shifts, domestic skill supply responses and the role of factor prices in analysing labour market outcomes. It offers a concluding answer to the questions posed by the thesis and suggests some possibilities for further research.

2 Elasticities of substitution, complementary, factor demand and factor price for disaggregated and aggregated inputs

"What now emerges is that [Joan Robinson] ought to have the sole right to the Elasticity of Substitution. Mine should have been defined by its reciprocal, which should have been given another name – Elasticity of Complementarity? It should then have been proved that in the two-factor case (alone) one was the reciprocal of the other. There would have been perfect duality between the two concepts. But it is much too late for that." – Sir John Hicks (1970:296)

In chapter 4, we will estimate a multi-factor equivalent of Robinson's (1933) elasticity of substitution due to Allen (1938) and Uzawa (1962) while in chapter 5, we will estimate the Hicks (1970) elasticity of complementarity. This chapter explains how the concepts differ and describes some of the ways they are related via cost/production duality. It suggests why the Allen elasticty of substitution (AES) is best calculated using cost function estimates and why the Hicks elasticity of complementarity (HEC) is best derived from production technology estimates.

There are many candidate elasticities of substitution and complementarity, each one having its own champion (see Stern, 2004; Bertoletti, 2005). With this as a background, the chapter also discusses under which scenarios the HEC or AES might be more appropriate. To estimate the elasticity of substitution using a cost function, it is necessary to generalise to non-constant returns to scale cases a result due to Uzawa (1962). This is done in this chapter. The basic issues of what technology model should be used to estimate the elasticities are introduced.

In this context, the issue of separability, which can justify the aggregation of inputs and/or the use of a value added specification is discussed. This allows us to build on Berndt & Christensen's (1973a) results for aggregated elasticities of substitution to establish equivalent results for elasticities of factor demand, thus confirming Marshall's Rules also hold for aggregated inputs. The key conclusions of this chapter are:

1. Should we wish to model the effects of an exogenous rise in a factor price on endogenous factor quantities, it is appropriate to estimate the AES using a translog cost function. We do this in chapter 4.

2. Should we wish to model the effects of an exogenous rise in factor quantity on endogenous factor prices, it is appropriate to estimate the HEC using a translog production function. We do this in chapter 5.

3. It is legitimate to use a cost function for the AES even if constant returns to scale do not hold (our generalization of Uzawa's result). We apply this in chapter 4.

4. It is legitimate to use value added specifications in cost and production function estimation with our data. We apply this in chapters 4 and 5.

5. We can apply Marshall's Rules to aggregated inputs with disaggregated data (our adaptation of Berndt & Christensen's result). We apply this in chapter 4.

2.1 Two elasticity concepts: Hicks vs Robinson

As Hicks (1970) noted in the quote starting this chapter, there are two very distinct concepts of substitutability between factors, which happen to be closely related in the two-factor context. The aim of this section is to explain the two concepts. Like Hicks (1970), we begin with a very general statement on the demand for a factor before making two alternative sets of assumptions, which allow us to describe the elasticity of factor demand on the one hand and the elasticity of factor price on the other. We will present a definition of substitutability applicable to each of these sets of assumptions.

We refer to the rules developed by Marshall (1920:383) using a linearly homogeneous production function with two factors, say skilled and unskilled labour, $q = f(x_i, x_j)$. With product market price p, wage equals marginal revenue product: $w_i = pf_i$ and $w_j = pf_j$. Hicks (1963:244) writes the own-elasticity of demand for x_i, $\frac{d\log x_i}{d\log w_i}$, as

$$|\lambda_{ii}| = \frac{\sigma_{ij}\left(|\eta| + e_j\right) + s_i e_j\left(|\eta| - \sigma_{ij}\right)}{|\eta| + e_j - s_i\left(|\eta| - \sigma_{ij}\right)} \tag{1}$$

The demand for a factor in an industry is more elastic (high $|\lambda|$) if:

1. It can be easily substituted by the other factor (high σ)

2. Its share of output is higher (high s)[3]

3. The supply of the other factor is more elastic (high e)

4. Product demand is more elastic (high $|\eta| \equiv -\frac{d\log q}{d\log p}$)

Point 1 requires a measure of σ, which is a feature of the production technology. One such

[3] This rule requires some qualification, specifically that $|\eta| > \sigma$

definition, as given in Robinson (1933), is

$$\sigma_{ij}^{R} = -\frac{d \log \frac{x_i}{x_j}}{d \log \frac{f_i}{f_j}}, \tag{2}$$

where she assumes output and the price of the other factor are constant. σ_{ij}^{R} is high if the factor is easily substituted by the other factor. When the supply of the other factor is perfectly elastic, $e \rightarrow \infty$ and we write (1) as:

$$|\lambda_{ii}| \equiv \left| \frac{d \log x_i}{d \log w_i} \right| = \sigma_{ij}^{R}(1 - s_i) + s_i|\eta| \tag{3}$$

This is the own-elasticity of factor demand. Here, σ_{ij}^{R} captures the change in relative demand for the two factors due to the change in relative factor prices[4] at constant output. In a competitive industry, a unit fall in the price of one factor will lower average and marginal cost by that factor's share (see for example Estrin & Laidler, 1995). For a fall in the price of one factor, profit-maximising industry output will rise and so will demand for both factors, by their share. However, as industry output rises, the product price falls, which lowers the marginal revenue product of each factor and mitigates the increase in demand for both factors. The second term, $s_j|\eta|$, captures the output effect. Thus, because all firms in a competitive industry with competitive factor markets experience the same change in factor price, we can measure the overall effect using (3).

A measure of the compensated elasticity of labour demand, which does not allow for output effects, is given by:

$$|\bar{\lambda}_{ii}| \equiv \left| \frac{d \log x_i}{d \log w_i} \right|_q = \sigma_{ij}^{R}(1 - s_i) \tag{4}$$

An alternative to Robinson's measure of σ is given by Hicks (1932,1963), who introduces,

$$\frac{1}{H_{ij}} = \frac{f_i f_j}{f f_{ij}}, \tag{5}$$

where f_{ij} gives the rate of change of the marginal product of one factor for a change in the quantity of the other factor. As presented in Hicks (1963:245), $\sigma = \frac{1}{H_{ij}}$ is increasing in the "facility of substitution". Its inverse H_{ij} measures the percentage change in relative marginal product after a change in the relative quantity of the factors used in production. Hicks' measure assumes the quantity of the other factor and the output price are constant. Thus we can set $e = 0$ to get a

[4] w_j is fixed, so a change in w_i necessarily implies a change in relative factor prices.

measure of the inverse demand for the factor from (1):

$$|\hat{\epsilon}_{ii}| \equiv \left|\frac{1}{\lambda_{ii}}\right| \equiv \left|\frac{d\log w_i}{d\log x_i}\right| = (1-s_i)H_{ij} + \frac{s_i}{|\eta|} \tag{6}$$

We refer to $\hat{\epsilon}$ as the elasticity of factor price (as does Hamermesh, 1993). It describes the change in factor price necessary for firms to absorb the extra supply of that factor in production. Here, H_{ij} captures the percentage change in relative factor prices that must take place after a change in relative factor quantities,[5] assuming output price is constant. After a rise in factor supply, output expands. However, expanded output lowers price, which lowers the marginal revenue product of the factors and therefore means the price of factor x_i must fall by more. The second term therefore acts to make $|\hat{\epsilon}_{ii}|$ higher. If we assume perfectly elastic product demand, then we only have the first term:

$$|\epsilon_{ii}| \equiv \left|\frac{d\log w_i}{d\log x_i}\right|_p = H_{ij}(1-s_i) \tag{7}$$

Inspection of (7) and (4) reveals what appears to be a duality between H_{ij} and σ^R_{ij}. Indeed, Hicks (1963:373) shows that, in the two factor case, $\sigma^R_{ij} = \frac{1}{H_{ij}}$. This is the "perfect duality" referred to in the introductory quote. It is reported in texts like Cahuc & Zylberberg (2004:814) but the simple relationship ends when there are more than two factors.

We will shortly consider a production technology with multiple inputs, where the links between σ^R_{ij} and H_{ij} become more complex. We note that, in the two-input case, these are two distinct *definitions* that happen to be equal. An alternative exposition, for example Allen (1938:343), introduces "the" elasticity of substitution as the indicator of how the marginal rate of substitution changes as we move along a given isoquant. Allen defines it exactly as in (2) and shows this measure equals $\frac{f_i f_j}{f f_{ij}}$. More recent treatments (eg Layard & Walters, 1987:266) repeat Allen's exposition and continue by saying "the" elasticity measure is not $\frac{f_i f_j}{f f_{ij}}$ for multiple inputs.

We have used Marshall's encompassing expression to find the special cases of direct and inverse factor demand. We could alternatively begin with the firm's technology to derive the system of factor demands and see how the factor demands change in response to factor prices. Allen (1938:373) does this but relies on what we have called the Hicks definition to reach (3). We will present the multiple input extension by Allen, where $\sigma_{ij} \neq \frac{f_i f_j}{f f_{ij}}$, in the next subsection. It is not a generalisation of "the" elasticity, because there is more than one, but of that of Robinson (1933). We present a full derivation of the inverse factor demand for the macroeconomy in chapter 5, doing

[5] Because the supply of the other factor is assumed fixed, an absolute change in the supply of one factor implies a change in relative quantities.

so for own and cross-elasticities.

Hicks (1970) labelled H_{ij} the elasticity of complementarity. Referring to H_{ij} rather than its inverse lends itself more readily to its role in calculating inverse factor demand (the elasticity of factor price). This terminology has remained unambiguous since 1970 and has been adopted by Hamermesh (1993), Layard & Walters (1987) and others. We will also use this term, reserving, as Hicks suggests in the quotation, the term elasticity of substitution for Robinson's term. Thus we will use notation σ_{ij} to refer to "the elasticity of substitution" and H_{ij} to refer to "the (Hicks) elasticity of complementarity" or HEC.

2.2 Multiple factors: The Allen/Uzawa generalisation of Robinson

In a multi-factor setting, (5) still applies for any two factors and the Hicks Elasticity of Complementarity (HEC) is given by $\frac{ff_{ij}}{f_i f_j}$. We postpone a full derivation until chapter 5 but note f_{ij} and hence H_{ij} can be positive or negative. Similarly, σ_{ij} can be positive or negative. In asking whether two factors are complements or substitutes, in a sense to be made precise, we are essentially concerned with the signs of these expressions. Finding empirical measures such that we can deduce the signs for multiple inputs is the objective of chapters 4 and 5.

Allen (1938) generalises the elasticity of substitution between factors x_i and x_j to the n factor case:

$$\sigma_{ij} = \frac{\mathbf{f}_{ij} \sum_k^n f_k x_k}{|\mathbf{f}| x_i x_j} \tag{8}$$

$|\mathbf{f}|$ is the determinant of the bordered Hessian of equilibrium conditions for a firm's cost-minimizing factor demands, holding output constant. $\mathbf{f}_{ij}$ is the cofactor of f_{ij} in $\mathbf{f}$. By Euler's theorem, the summation term equals q under constant returns to scale. This partial elasticity holds output and the prices of other factor prices constant and is thus a generalisation of Robinson's elasticity. Uzawa (1962) uses the duality between production function $q = f(x_1, x_2, ..., x_n)$ and cost function $C = C(w_1, w_2, ..., w_n, q)$ to express the elasticity of substitution as

$$\sigma_{ij} = \frac{CC_{ij}}{C_i C_j}, \tag{9}$$

where C is total cost, C_i, C_j are first derivatives with respect to the prices of factors i, j - w_i, w_j - and C_{ij} is the cross partial derivative. Uzawa's proof uses a unit cost function, which only uniquely represents the underlying production function under linear homogeneity (Varian, 1992) and his result thus appears strictly applicable to constant returns to scale only. However, many studies use this result in more general settings. For example, of the twelve listed in Chung (1994), only five have a linearly homogeneous production technology. We will find in chapter 4 that the cost data do not appear to be consistent with constant returns to scale. While the validity of (9) under more general technological settings may be "folk knowledge", it is instructive to confirm and document this.[6]

Theorem 1 *For any production function $q = f(x_1, x_2, ..., x_n)$, we can use cost function $C = C(w_1, w_2, ..., w_n, q)$ to express the elasticity of substitution as:*

$$\sigma_{ij} = \frac{CC_{ij}}{C_i C_j} \tag{10}$$

where C_i, C_j are first derivatives with respect to the costs of factors i, j - w_i, w_j- and C_{ij} is the cross partial derivative.

Proof. The conditional factor demands are derived from the cost minimization problem:

$$\min \sum_i w_i x_i \text{ subject to } q = f(x_1, x_2, ..., x_n) \tag{11}$$

The firm's first order conditions are, where μ is the Lagrange Multiplier,

$$w_i = \mu f_i (i = 1, ..., n), \tag{12}$$

$$q = f(x_1, x_2, ..., x_n) \tag{13}$$

and the cost function is

$$C(w_1, w_2, ..., w_n, q) = \sum_i w_i x_i (w_1, w_2, ..., w_n, q) \tag{14}$$

Following Allen (1938) but without assuming constant returns to scale, differentiate the first order conditions with respect to w_1 and divide each equation by μ:

$$\begin{array}{cccccc}
0 & +f_1 \frac{\partial x_1}{\partial w_1} & +f_2 \frac{\partial x_2}{\partial w_1} & +. & +f_n \frac{\partial x_n}{\partial w_1} & = 0 \\
\frac{1}{\mu} f_1 \frac{\partial \mu}{\partial w_1} & +f_{11} \frac{\partial x_1}{\partial w_1} & +f_{12} \frac{\partial x_1}{\partial w_1} & +. & +f_{1n} \frac{\partial x_1}{\partial w_1} & = \frac{1}{\mu} \\
\frac{1}{\mu} f_2 \frac{\partial \mu}{\partial w_1} & +f_{21} \frac{\partial x_1}{\partial w_1} & +f_{22} \frac{\partial x_1}{\partial w_1} & +. & +f_{2n} \frac{\partial x_1}{\partial w_1} & = 0 \\
. & . & . & . & . & . \\
\frac{1}{\mu} f_n \frac{\partial \mu}{\partial w_1} & +f_{n1} \frac{\partial x_1}{\partial w_1} & +f_{n2} \frac{\partial x_1}{\partial w_1} & +. & +f_{nn} \frac{\partial x_1}{\partial w_1} & = 0
\end{array} \tag{15}$$

[6] My thanks are due to Margaret Stevens for the essential role she played in this proof.

By Cramer's Rule,

$$\frac{\partial x_2}{\partial w_1} = \begin{vmatrix} 0 & f_1 & 0 & . & f_n \\ q_1 & f_{11} & \frac{1}{\mu} & . & f_{1n} \\ q_2 & f_{12} & 0 & . & f_{2n} \\ . & . & . & . & . \\ q_n & f_{1n} & 0 & . & f_{nn} \end{vmatrix} \div \begin{vmatrix} 0 & f_1 & f_2 & . & f_n \\ f_1 & f_{11} & f_{12} & . & f_{1n} \\ . & . & . & . & . \\ . & . & . & . & . \\ f_n & f_{n1} & f_{n2} & . & f_{nn} \end{vmatrix}, \tag{16}$$

so

$$\frac{\partial x_2}{\partial w_1} = \frac{1}{\mu}\frac{\mathbf{f}_{12}}{|\mathbf{f}|} \tag{17}$$

As in (8), $|\mathbf{f}|$ is the determinant of the bordered Hessian of equilibrium conditions for a firm's cost-minimizing factor demands, holding output constant, and $\mathbf{f}_{12}$ is the cofactor of f_{12} in $\mathbf{f}$. By (8):

$$\sigma_{21} = \frac{\sum_n f_i x_i}{x_1 x_2}\mu\frac{\partial x_2}{\partial w_1} \tag{18}$$

By the first order conditions,

$$\mu\sum_n f_i x_i = C \tag{19}$$

and, by Shepard's Lemma,

$$x_j = C_j \tag{20}$$

such that $\frac{\partial x_2}{\partial w_1} = C_{21}$. Thus $\sigma_{21} = \frac{CC_{21}}{C_1 C_2}$. ∎

We have generalised the Uzawa (1962) result, which allows us to use cost function parameters to find σ (the AES). We briefly discuss why it can be useful for estimation in section 2.6.

Remark 1 *By Slutsky symmetry (and more directly Euler's Theorem),* $\frac{\partial x_2}{\partial w_1} = \frac{\partial x_1}{\partial w_2}$ *and* $C_{21} = C_{12}$*, so* $\sigma_{21} = \sigma_{12}$.

Corollary 1 *Using equations (18) - (20),* $\sigma_{12} = \frac{C}{x_1 x_2}\frac{\partial x_1}{\partial w_2} = \frac{C}{w_2 x_2}\frac{\partial \log x_1}{\partial \log w_2}$. *But* $s_2 = \frac{w_2 x_2}{C}$. *Thus* $\sigma_{12} = \frac{\bar{\lambda}_{12}}{s_2}$ *or, for any pair* i, j

$$\sigma_{ij} = \frac{\bar{\lambda}_{ij}}{s_j} \tag{21}$$

Remark 2 *The treatment in Layard & Walters (1987:269) defines the multiple-input elasticity of substitution as in (21). Some authors (eg Heathfield & Wibe, 1987:61) assert the relationship between* $\bar{\lambda}$ *and* σ *holds only under conditions of constant returns, but we have confirmed it holds* *for more general technologies.*

σ_{ij} and $\bar{\lambda}_{ij}$ hold output constant. Considering output effects in the context of non-constant returns introduces issues of tractability. In the particular case of increasing returns, we face a theoretical and conceptual difficulty. It is difficult to pin down the optimal output level, because output cannot be optimised with increasing returns to scale. We do not face this issue for a constant level of output and hence are not concerned with this when considering conditional demand estimates. As described in chapter 3, we will be using firm-level data. Our results in chapter 4 will be consistent with an increasing returns to scale technology. How this can be interpreted in an industry context and to what extent we can consider the industry's scale effects is taken up in that chapter. Before proceeding we note that the vast majority of studies (Hamermesh, 1993) consider only conditional factor demands, relying on constant output elasticities. We however make an attempt to allow for non-constant output for elasticities of factor demand and non-constant prices for elasticities of factor price.

2.3 The relationship between H and σ

The exact nature of the duality between H and σ depends on technological assumptions and is widely debated, as shown in Stern (2004) and Bertoletti (2005). For two factors and constant returns to scale, $H = \frac{1}{\sigma}$, as we saw in section (2.1). Hicks (1970) discusses the equivalence for three factors while Sato & Koizumi (1973) do so for n factors using a linearly homogeneous production technology. Syrquin & Hollander (1982) extend this to homothetic non-constant returns. Defining

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & . & 1 \\ 1 & H_{11} & . & H_{1n} \\ . & . & . & . \\ 1 & H_{n1} & . & H_{nn} \end{bmatrix}, \tag{22}$$

$$\boldsymbol{\sigma} = \begin{bmatrix} 0 & 1 & . & 1 \\ 1 & \sigma_{11} & . & \sigma_{1n} \\ . & . & . & . \\ 1 & \sigma_{n1} & . & \sigma_{nn} \end{bmatrix}, \tag{23}$$

$C_y \equiv \frac{\partial C}{\partial y}$, and $\phi \equiv \left(\frac{\partial \ln C}{\partial \ln y}\right)^{-1}$ as the scale elasticity, they obtain:

$$H_{ij} = \frac{\phi}{s_i s_j} \frac{\sigma_{ij}}{|\boldsymbol{\sigma}|} - \frac{\partial \ln C_y}{\partial \ln y} \tag{24}$$

$$\sigma_{ij} = \frac{\phi}{s_i s_j} \frac{H_{ij}}{|\mathbf{H}|} \tag{25}$$

For constant returns to scale, $\phi = 1$ and the second term in (24) vanishes. These expressions nest the constant returns, two input case. In chapter 9, we will use (25) and our chapter 5 estimates of H_{ij} to find implied values of σ_{ij}. We will compare the signs of the implied σ_{ij} with those of the estimated values of σ_{ij} from chapter 4. To our knowledge, this has not been done elsewhere.

Bertoletti (2005) notes the AES relates to the system of factor demands and the HEC relates to the system of inverse factor demands. We have seen how the HEC is related to $\hat{\epsilon}_{ij}$. In terms of the taxonomy introduced by Hicks (1956, 1970), we can say factors i and j are *q-complements* if $H_{ij} > 0$ ($\hat{\epsilon}_{ij} > 0$ if we allow prices to change). They are *q-substitutes* if $H_{ij} < 0$. In contrast, we say factors i and j are *p-complements* if $\sigma_{ij} < 0$. They are *p-substitutes* if $\sigma_{ij} > 0$ ($\lambda_{ij} > 0$ if we allow output to change).

As noted in Hamermesh (1993) and Bertoletti (2005) and found, for example, by Field (1988), it is quite possible for a pair of factors to be found to be p-complements and for that same pair to be found to be q-substitutes (or vice versa). However, Layard & Walters (1987) argue it is likely two factors which are q-substitutes will tend to be p-substitutes and the same applies for complements. Furthermore, while in a two factor setting the pair are necessarily p-substitutes ($\sigma_{ij} > 0$), the multiple factor setting requires only that factor i be a p-substitute with at least one other factor. Similarly, with two factors, the pair are necessarily q-complements. For more factors, the requirement is that each factor be a q-complement with at least one other factor (Sato & Koizumi, 1973).

2.4 Is it wages or quantities that are exogenous?

One reason the AES may be more prominent than the HEC is the nature of the exogeneity assumptions. The view that factor prices are exogenous to a firm and factor quantities are endogenously chosen is preferred to the other way round, certainly at a firm-level. However, neither assumption is entirely correct (Hamermesh, 1993; Fallon & Verry, 1988). At the micro-level, there is evidence for firm-effects on wages, especially firm-size effects (Troske, 1991). Whether the cause is that (i) workers are more productive because of their abilities or a higher capital:labour ratio, (ii) compensating differentials for a less-pleasant environment or (iii) any other effect, wages are at least partly endogenous to the firms (Oi & Idson, 1999). The cost of capital a firm encounters tends to fall as the firm gets bigger, certainly up to a point, because small and/or young firms incur risk premia, and the nature of their activities affects the cost of capital. Firms that are highly leveraged also have risk premia. As one moves to an industry or macro-level, the exogeneity of factor prices becomes less accurate (Hamermesh, 1993). A big rise in demand for a factor at even relatively low levels of aggregation will start affecting factor prices, especially when specific skills are scarce.

Exogenous quantities and endogenous prices are consistent with the context of perfectly inelastic aggregate labour supply. Some may find this hard to reconcile with high unemployment, but the view that these people are structurally unemployed implies they are effectively not part of the workforce. Furthermore, the whole premise on which this thesis is based is a low supply of skills. This is not because of a lack of demand for skills provision (Department of Labour, 1997), which is possibly endogenous to wages, but because of a lack of supply of skills provision. A fixed skill supply, certainly with respect to the wage, is a credible point of departure.

Chapter 3 will present evidence of collective bargaining in South Africa. The wages are at least partially set by those firms in the collective bargaining process, making wages in some measure endogenous for those firms, but there is also extension by the Minister of Labour to other firms, making wages exogenous for the others (Nattrass, 2000). Such wage agreements hamper the responsiveness of wages to quantities, but this is more of an argument about the smooth operation of market forces than about endogeneity or exogeneity. Finally, the issue of collective bargaining is equivocal in this argument, as bargaining can take place over quantity as well as price (Booth, 1995).

To the extent that wages and quantities are neither completely endogenous nor exogenous, it becomes important to see which elasticity concepts are most suitable. It depends on the question one wishes to ask. If we are most interested in the effects of a rise in skill supply on wages, it is clearly the case that the relevant exogenous change is in the supply of skilled labour, through more successful education and training. In this instance, it is to be expected that the relief of the shortages will lead to changes in wages in the economy, even if distorted by labour market institutions. Regardless of whether a rise in the supply of one factor leads to a smooth fall in the wage of another or to a trade union being forced to accept a smaller wage increase, the estimated elasticities provide a useful insight into the gains and losses other factors are likely to see. Thus inverse demand measures based on the HEC would be most appropriate. Furthermore, an adjustment can be made to cope with wage rigidity, as discussed in chapter 5.

If, on the other hand, we are interested in the effects of measures that exogenously affect absolute or relative factor prices, like interest rates or minimum wage legislation, the AES is applicable.

2.5 Other elasticity concepts

We have argued that H is appropriate for measuring the effect of an exogenous rise in factor quantities on endogenous factor prices while σ is more appropriate for measuring the effects of exogenous prices on endogenous quantities. As we discussed, the former can be thought to relate to the system of factor demands and the latter to the system of inverse factor demands (Bertoletti, 2005). There are in fact many more elasticity concepts that fall into either of these groups and that can be further classified according to whether output is being held constant or allowed to vary; that is, whether they are measures of net or gross substitutability.

To be sure, by holding output constant, the AES is a net measure. However, there are gross counterparts of the AES, which have been given a variety of names. Similarly, there are net counterparts of the HEC, which hold output constant. Kim (2000) attaches to one counterpart the name "Antonelli elasticity of complementarity" and advances some of its potential advantages over the HEC. We note that our measure is more appropriate here because, as we will argue in chapter 5, changes in output are part of the effect we wish to capture.

An alternative to the AES is the Morishima elasticity of substitution, forcefully advanced as superior to the AES by Blackorby & Russell (1989). It measures the effect of a change in the ratio of factor quantities after an exogenous change in the price of one of the factors. Like the AES, output is constant, but the AES has ratios in both the numerator and denominator of the derivative. We focus on the AES because it is still most commonly accepted (Bertoletti, 2005), using Marshall's Rules to deal with output effects. This is preferred to any of the so-called gross measures of elasticity because the elasticity of labour demand, both compensated and uncompensated, is a more familiar concept. It is also the most commonly accepted measure used for investigations of capital skill complementarity.[7] [8]

Sato & Koizumi (1973:44) hoped they had ended the confusion started 70 years ago with their "duality" result based on (24) and (25). With reference to the quote at the start of this chapter, they concluded the Hicks (1970) paper was "far from being too late". We record that Stern (2004)

[7] We introduced capital skill complementarity in chapter 1.

[8] This raises the issue of whether capital skill complementarity using one measure implies complementarity using another. This is interesting for future research. If not, robustness of such findings to alernative measures would also be worthy of empirical investigation.

notes ten (!) different elasticity concepts (in terms of both the production and cost function) and that debate on appropriate use rages on (Bertoletti, 2005). Many of the papers cited in this subsection have tried to synthesize the concepts, construct a useful taxonomy and establish when certain elasticities are most appropriate. Despite these noble attempts, we seem to be no closer to a consensus. Furthermore, it is moot whether any differences are material in an empirical context. Data errors and estimation issues are likely to swamp the theoretical nuances discussed.

2.6 Estimating the elasticities: choice of technology

Depending on the functional form chosen to represent the production technology, one can investigate labour demand using direct estimates of ad hoc labour demand equations, specifying the quantity of labour demanded in terms of wages and other variables. This is done by Fajnzylber & Maloney (2001) for Colombia, Mexico and Chile and by Roberts & Skoufias (1996), also for Colombia. One can capture the elasticity of substitution from relative labour demand estimates based on the constant elasticity of substitution cost/production technology (eg Edwards, 2003 and Edwards & Behar, 2006).

One can also try to estimate the technological parameters needed for (5), (8) or (9) directly. To estimate σ, Binswanger (1974a) lists why we might in general prefer using cost functions for (9) to production functions for (8). To satisfy a necessary condition for optimizing behaviour, cost functions must exhibit homogeneity of degree one in prices, which can be imposed to improve estimates without recourse to technological assumptions. Second, cost functions are more consistent with the view that wages are exogenous. This matches the exogeneity assumptions underlying the AES and elasticity of demand.

Production functions, estimated on the basis of exogenous input quantities, are consistent with the exogeneity assumptions underlying the HEC and elasticity of factor price. The fact that input quantities may be endogenous to wages, the production technology and market shocks can lead to biased estimates, but this is an estimation issue discussed in chapter 5. The main reason for using a cost function for the AES and elasticity of factor demand is that they can be far more tractably arrived at than by using production functions. Similarly, production functions are more tractable for estimating the HEC and elasticity of factor price.

A Cobb Douglas function is used in a macroeconomic model of skilled and unskilled labour demand and supply by Du Toit & Koekemoer (2003). Although they claim it was "estimated and validated as representative of the South African production structure" (pg 7), the homogeneity and separability assumptions it carries are too restrictive to go untested in a new study. More importantly, the implication that the elasticity of substitution is unity (Chung, 1994) would completely circumvent one of the central aims of this thesis!

Constant Elasticity of Substitution technologies yield relative factor demand functions that can be estimated and interpreted easily. They allow the elasticity of substitution to differ from one, but the elasticity of substitution is the same between all input pairs (Chung, 1994), which is still a major restriction. For example, Edwards (2003) estimates an equation for the demand for skilled relative to unskilled labour as a function of relative wages, trade and technology variables for the Gauteng Province in South Africa. Edwards & Behar (2006) study the whole of South Africa, allow for a non-homogeneous technology and include measures of tariffs.

If, as in these studies, the elasticities of substitution are not the key parameters of interest but wages need to be included as controls, this specification can be appropriate, especially if one is willing to consider only two factor inputs at a time. Adding factors requires more complex non-linear techniques or step-wise regression to estimate the production function directly (Fallon & Verry, 1988; Hamermesh, 1993). Fallon & Lucas (1998) include capital in their CES function to estimate, with non-linear 3 stage least squares and calibration techniques, demand for black and white labour as proxies for unskilled and skilled labour. The restriction is also particularly problematic as all factors will necessarily be found to be p-substitutes. Again, this would defeat the purpose of this thesis.

More flexible functional forms do not impose a priori technological assumptions like equality of elasticities between all factor pairs. Two functions falling in this class are the Generalised Leontief function due to Diewert (1971) and the Transcendental Logarithmic (translog) function developed by Christensen, Jorgenson & Lau (1973).

This study uses translog functions. While the gains from using a flexible functional form are clear, the reasons for adopting the translog instead of Generalised Leontief are a moderate gain in tractability, especially if constant returns to scale are not assumed, at no apparent cost, and the fact that the majority of international studies appear to use translogs for this application. However, a study wishing to focus on the effects of non-wage or non-input variables – for example the effects of trade liberalisation – might benefit from the easy way in which such variables can be included in a Generalised Leontief specification (Chung, 1994). We note that, in a rare comparison of both technologies, Humphrey & Wolkowitz (1976) get somewhat different elasticity estimates.

The translog function can be viewed as a production function in its own right or as a second order approximation to an underlying technology (Denny & Fuss, 1977). It has been used to estimate H by Mak (2000) using Canadian data. Postponing a more complete discussion of its properties to chapter 5, we write the translog production function as:

$$\log q = \log \alpha_0 + \sum_i \alpha_i \log x_i + \frac{1}{2} \sum_i \sum_j \beta_{ij} \log x_i \log x_j \tag{26}$$

q is a measure of output and x_i, x_j are factor inputs. Differentiating (26) with respect to $\log x_i$ yields

$$\frac{d \log q}{d \log x_i} = \alpha_i + \sum_j \beta_{ij} \log x_j, \tag{27}$$

but

$$\frac{d \log q}{d \log x_i} = \frac{\partial q}{\partial x_i} \frac{x_i}{q} = s_i, \tag{28}$$

which is the share of factor i in output. Therefore:

$$s_i = \alpha_i + \sum_j \beta_{ij} \log x_j \tag{29}$$

The key parameters are β_{ij}. Postponing a derivation until chapter 5, the elasticity of complementarity is calculated as:

$$H_{ij} = \frac{\beta_{ij}}{s_i s_j} + 1 \tag{30}$$

This nests the Cobb Douglas technology when all $\beta_{ij} = 0$, in which case $H = 1$. Teal (2000) estimates elasticities of substitution in Ghana using a translog cost function like:

$$\begin{aligned} \log C \quad = \quad & \log a_0 + \sum_i a_i \log w_i + \frac{1}{2} \sum_i \sum_j B_{ij} \log w_i \log w_j + a_q \log q \\ & + B_q \log^2 q + \sum_i B_{iq} \log q \log w_i \end{aligned} \tag{31}$$

The payment to factor i is w_i. The Allen elasticity of substitution can be written as (Binswanger, 1974a):

$$\sigma_{ij} = \frac{B_{ij}}{s_i s_j} + 1 \tag{32}$$

The factor share of output can also be the factor share of total costs, as shown in chapter 4. The symmetry between (30) and (32) and the similar roles played by β_{ij} and B_{ij} is apparent. We devote more attention to the derivations of (30) and (32) in chapters 4 and 5. We also present expressions for $\epsilon, \hat{\epsilon}, \bar{\lambda}$ and λ for the specific case of the translog technology in those chapters.

2.7 Separability

Separability of the production technology implies we can aggregate certain inputs or exclude inputs to estimate elasticities of substitution or complementarity. It also is needed to justify the value added specification (Chung, 1994). We introduce general definitions of separability before linking them to the legitimacy of the value added specification. For this subsection, the term "factors" refers to labour and capital while "inputs" includes factors and raw materials. Thereafter, we examine separability in translog functions. We discuss the implementation of such tests in the literature before confirming the validity of the value added specification empirically for our data.

2.7.1 Definitions for a general technology

Drawing on Berndt & Christensen (1973a), we have a linearly homogeneous production function $q = f(x_1, ..., x_n)$, where each x_i is a disaggregated input quantity. The n inputs are partitioned into R mutually exclusive and exhaustive subsets $[X^1, ..., X^R]$, which we call partition P. The production function is weakly separable with respect to partition P if the marginal rate of technical substitution between any pair of inputs x_i, x_j from any subset X^S is independent of the quantity of any input outside X^S. That is, $\frac{d}{dx_k}\left(\frac{f_i}{f_j}\right) = 0 \; \forall i, j \epsilon X^S, k \notin X^S$, where as before f_i, f_j are the marginal products of inputs x_i, x_j. Differentiation gives:

$$f_j f_{ik} - f_i f_{jk} = 0 \; \forall i, j \epsilon X^S, k \notin X^S \tag{33}$$

Weak separability is necessary and sufficient for $f(\cdot)$ to be legitimately written as $q = F^W [X_1, ..., X_R]$, where X_S is a function of the elements of X^S only. Strong separability would require condition (33) to hold for all i, j in any subsets X^S, X^T and all k not in any of those subsets. This would mean we could write output as $q = F^S [X_1 +, ..., +X_R]$. An alternative term for strong separability is additive separability (Mas-Colell, Whinston & Green, 1995). For some production technologies, the terms in (33) may depend on the quantities of each input. If they do not, the production function is globally separable (Berndt & Christensen, 1973b).

A result due to Lau (1969) shows weak separability in the production function with respect to partition P implies weak separability in the dual cost function (and vice versa). Formally, the cost function $C = g(w_1, ..., w_n, q)$ dual to $f(\cdot)$ can after partition P consist of R subsets. The cost function can be written as $C = G^W [W_1, ..., W_R, q]$, where w_i are disaggregated input prices and W_S is a function of the prices of the inputs in X^S, which comprise set W^S.

We draw on the exposition in Sato (1975) to consider the conditions under which the general production function expressing gross output in terms of a vector of factors $(\mathbf{x})$ and raw materials (x_r) - $q = f(\mathbf{x}, x_r)$ - can be written as $v = v(\mathbf{x})$, where $v = q - x_r$ is value added. Under constant returns to scale and assuming the price of raw materials is constant, Sato (1975:86) shows weak separability is necessary and sufficient for this to hold. In this example, weak separability implies $q = f^p(g(\mathbf{x}), x_r)$ and raw materials are weakly separable iff

$$f_j f_{ir} - f_i f_{jr} = 0 \;\forall i, j \neq r \tag{34}$$

for a general technology.

2.7.2 Separability and translog production functions

Drawing on Berndt & Christensen (1973b), we apply (34) to (26) so that a translog production function is globally weakly separable with respect to input x_r if and only if:

$$s_i \beta_{jr} - s_j \beta_{ir} = 0 \tag{35}$$

Each factor share is positive so a sufficient condition for global separability is:

$$\beta_{ir} = 0 \;\forall i : i \neq r \tag{36}$$

If (36) does not hold for some β_{ir} but does for others, then separability cannot hold. However, if all $\beta_{ir} \neq 0$ we can draw on Berndt & Christensen (1973b) to establish that sperability of factors i, j with respect to r holds if and only if:[9]

$$\frac{\alpha_i}{\alpha_j} = \frac{\beta_{ii}}{\beta_{ij}} = \frac{\beta_{ij}}{\beta_{jj}} = \frac{\beta_{ir}}{\beta_{jr}} \tag{37}$$

This must hold for each factor pair x_i, x_j whose elasticity we wish to obtain. The duality result due to Lau (1969) implies there are obvious analogs for the translog cost function.

One can test for separability econometrically and use the result of Sato (1975) to justify the value added specification. Alternatively, one can test for separability of an input using the Allen elasticity of substitution. x_r is weakly separable from the other inputs if and only if (Berndt &

[9] See Berndt & Christensen (1973b:86). The procedure entails substituting using (29) and Slutsky symmetry $(\beta_{ij} = \beta_{ji})$ to find condition $\alpha_i \beta_{jr} - \alpha_j \beta_{ir} + \sum_m^n (\beta_{im}\beta_{jr} - \beta_{jm}\beta_{ir}) \log x_m = 0$, where n is the number of factors. Imposing the additional restrictions that the terms in brackets equal to zero ensures global separability. Dividing by non-zero β coefficients establishes the result.

Christensen 1973ab):

$$\sigma_{ir} = \sigma_{jr} \; \forall i, j \neq r \tag{38}$$

Kim (1997) produces an analogous result using the Hicks elasticity of complementarity:

$$H_{ir} = H_{jr} \; \forall i, j \neq r \tag{39}$$

Denny & Fuss (1977) as well as Blackorby, Primont & Russell (1977) suggest tests of this class may be too stringent because the tests are implicitly also for whether the translog function is exact or a second order Taylor approximation to an unknown technology. While we can use the elasticity measures as an indicator of separability, testing the restrictions requires more complicated non-linear inference (cf.(30) and (32)) while standard tests can be used on each of the coefficients on the production function.

2.7.3 Implementing the tests

Of the 12 studies listed by Chung (1994), only three test for value added separability while Hamermesh (1993) does not discuss value added separability in his comprehensive survey. A lack of gross output and/or intermediate input data prevent such tests in many cases. In translog production studies, Grant & Hamermesh (1981) express output, taken from annual manufacturing censuses and surveys, in terms of capital and four forms of labour, while Mak (2000) does not have raw materials inputs in her system of share equations. Neither tests the validity of omitting raw materials. In cost studies, Bergström & Panas (1992) and Teal (2000) have value added in their cost share equations but don't mention any tests for separability. Griffin & Gregory (1976) state they assume value added separability. Denny & May (1977) generally reject separability in both their production and cost function estimates. They ascribe differences between their elasticity findings and others to invalid value added estimates in other studies. However, Humphrey & Wolkowitz (1976) find no significant difference between cost function estimates that impose separability and estimates that don't.

For the purposes of this thesis, full output regressions with a constructed measure of raw materials as an additional regressand yield sensible results. They can therefore be used to test separability restrictions. Appendix 1 has the regression output and tests (Appendix 2 describes the variables). When performing multiple tests, it is arguably appropriate to make an adjustment to the statistics used in each test to prevent the false rejection of the null hypothesis. We perform Wald tests which do and don't adjust the probability values required for each restriction. One option for adjustment is the Bonferroni method but we use the Sidak adjustment, which allows for correlation between the combinations being tested (see Statacorp, 2003). The choice of method does not appear to matter; whether adjusted or not, Wald tests of strong separability in the form of (36) are rejected. Tests of weak separability in the form of (37) for each combination of factors do not reject separability, even if we do not adjust for the probability of a type I error. Most p-values are well above 0.90. The tests therefore justify the use of the value added specification in the production function. While, as will be explained, we can construct measures of raw material input quantities, we can not do so for prices, so we use the Lau (1969) result to justify value added separability for our cost function estimates.

2.8 Separability and aggregate elasticities

In this section, we present existing results on the relationships between separability and aggregate measures of σ and H. The results will be employed in chapters 4 and 5 but also allow us to develop analogous results for aggregate elasticities of factor demand, which will be used in chapter 4.

2.8.1 Aggregate measures of σ and H for translog production functions

For us to aggregate factors x_i and x_j, they must be weakly separable from all others in the production function such that, analogous to (35),

$$s_i\beta_{jk} - s_j\beta_{ik} = 0 \ \forall k : k \neq i, k \neq j \tag{40}$$

Berndt & Christensen (1973ab, 1974) show this is equivalent to:

$$\sigma_{ik} = \sigma_{jk} \ \forall k : k \neq i, k \neq j \tag{41}$$

They also show this is equivalent to the legitimate construction of an aggregate index of factors x_i and x_j (quantities or prices). In other words, the elasticity of substitution between some aggregate of x_i and x_j, which we call X_I, and a third input x_k is $\sigma_{Ik} = \sigma_{ik} = \sigma_{jk}$.

Furthermore, Kim (1997) shows the separability of factors x_i and x_j from all other factors of production is equivalent to:

$$H_{ik} = H_{jk} \ \forall k : k \neq i, k \neq j \tag{42}$$

Our dataset has multiple input quantities and prices. We will draw on the results in this section to perform studies of substitution and complementarity between more aggregated inputs. In a simple cross section, it is easy to aggregate by adding the input quantities (chapter 5). For cost functions, we will impose the conditions (40) for estimation (chapter 4).

2.8.2 Aggregation and the elasticity of factor demand for general technologies

While Berndt & Christensen (1973a) make statements about separability that allow us to produce aggregated elasticities of substitution (equation 41)), the same has to our knowledge not been said about the cross and own-elasticities of factor demand. This section will develop such equivalent statements. The results of the theorem presented will have practical importance when we try to use our disaggregated data to calculate aggregated measures of λ and $\bar{\lambda}$.

Assume each of the disaggregated input quantities change by the same proportion and that each of the disaggregated input prices also change by the same proportion. Informally, we can say that the elasticity of an aggregate of one set of input quantities with respect to an aggregate of another set of input prices is the *sum* of the elasticities of *one* of the input quantities with respect to each of the input prices. Equivalently, the elasticity of the aggregate of one set of input quantities with respect to an aggregate of another set of input prices is the elasticity of substitution between the aggregates multiplied by the cost share of the aggregate input whose price has changed. We proceed to state this formally.

To formalise separability, assume we have a linearly homogeneous production function $q = F(x_1, ..., x_n)$, where x_i is a disaggregated input quantity. The n inputs are divided into R mutually exclusive and exhaustive aggregates such that $q = F^P(X_1, ..., X_R)$, where X_I is an aggregate of the input quantities in subset X^I. Call this partition P. All $x_i \in X^I$ are weakly separable from $x_i \notin X^I$. By the result due to Lau (1969), the same partition P implies the dual cost function $C = G(w_1, ..., w_n, q)$, where w_i is a disaggregated input price and q is output, can be written as $C = G^P(W_1, ..., W_R, q)$, where W_I is an aggregate of one or more input prices. All $w_i \in W^I$ are weakly separable from $w_i \notin W^I$.

Write $d\log w_j = \hat{w}\ \forall j \in W^J$ and $d\log x_i = \hat{x}\ \forall i \in X^I$. This means all disaggregated input quantities or prices within an aggregate change by the same proportion. Define the constant output cross-elasticities as follows:

- $\bar{\lambda}_{ij} \equiv \frac{d\log x_i}{d\log w_j}$ is the proportional change in demand for a disaggregated input after a proportional change in the price of another disaggregated input.
- $\bar{\lambda}_{iJ} \equiv \frac{d\log x_i}{d\log W_J}$ is the proportional change in demand for a disaggregated input after an equal proportional change in price of each of the disaggregated inputs within an aggregate.

- $\bar{\lambda}_{IJ} \equiv \frac{d\log X_I}{d\log W_J}$ is the proportional change in demand for each of the disaggregated inputs within an aggregate after an equal proportional change in price of each of all the disaggregated inputs within an aggregate. In other words, it is the aggregate elasticity of factor demand.

Define $S_J = \sum_j s_j \; \forall j \in W^J$ such that the factor share of an aggregate is the sum of the disaggregated shares. Write σ_{ij} as the disaggregated elasticity of substitution, σ_{IJ} as the aggregate elasticity of substitution and σ_{iJ} as the elasticity of substitution between a disaggregated input and an aggregated input.

Lemma 1 *Weak separability with respect to the partition P implies* $\bar{\lambda}_{iJ} = \sum_{j\in W^J} \bar{\lambda}_{ij}$.

Proof. By equation (21), $d\log x_i = \sum_j s_j\sigma_{ij} d\log w_j$ when output is constant. In particular, if only the prices in the aggregate W_J change, $d\log x_i = \sum_{j\in W^J} s_j\sigma_{ij} d\log w_j$. However, $d\log w_j = \hat{w}$ $\forall j \in W^J \rightarrow \sum_{j\in W^J} d\log w_j = \frac{\sum_{j\in W^J} dw_j}{\sum_{j\in W^J} w_j} = \hat{w} = d\log W_J$. Therefore $\bar{\lambda}_{iJ} = \sum_{j\in W^J} s_j\sigma_{ij} = \sum_{j\in W^J} \bar{\lambda}_{ij}$. ∎

Lemma 2 *Weak separability with respect to the partition P implies* $\bar{\lambda}_{iJ} = S_J\sigma_{iJ}$.

Proof. As shown in Berndt & Christensen (1973a), $\sigma_{ij} = \sigma_{iJ} \; \forall j \in W^J$. By Lemma 1, $\bar{\lambda}_{iJ} = \sum_{j\in W^J} s_j\sigma_{iJ}$. Therefore $\bar{\lambda}_{iJ} = S_J\sigma_{iJ}$ ∎

Lemma 3 *Weak separability with respect to the partition P implies* $\bar{\lambda}_{iJ} = S_J\sigma_{IJ}$.

Proof. Using $\sigma_{ij} = \sigma_{iJ} \forall \; j \in W^J$, this follows trivially from Lemma 2. ∎

Lemma 4 *Weak separability with respect to the partition P implies* $\bar{\lambda}_{IJ} = S_J\sigma_{IJ}$.

Proof. If only the prices in W_J change, by equation (21), $\sum_{i\in X^I} d\log x_i = \sum_{i\in X^I}\sum_{j\in W^J} s_j\sigma_{ij} d\log w_j$. By Lemma 2, $\sum_{i\in X^I} d\log x_i = \sum_{i\in X^I} S_J\sigma_{iJ} d\log W_J$. But $d\log x_i = \hat{x} \; \forall i \in X^I \rightarrow \sum_{i\in X^I} d\log x_i = \frac{\sum_{i\in X^I} dx_i}{\sum_{i\in X^I} x_i} = \hat{x} = d\log X_I$. Therefore $\bar{\lambda}_{IJ} = S_J\sigma_{iJ}$. It follows from Berndt & Christensen (1973a) that $\sigma_{iJ} = \sigma_{IJ} \; \forall i \in X^I$ and therefore $\bar{\lambda}_{IJ} = S_J\sigma_{IJ}$. ∎

The results are summarised in the following theorem:

Theorem 2 *Weak separability with respect to the partition P implies* $\bar{\lambda}_{IJ} = \bar{\lambda}_{iJ} = S_J\sigma_{IJ} = S_J\sigma_{iJ} = \sum_{j\in W^J} \bar{\lambda}_{ij}$.

Proof. This follows from lemmata 1-4. ∎

We can also make a statement about aggregated own-price elasticities.

Proposition 1 *Weak separability with respect to the partition P implies $\bar{\lambda}_{II} = -\sum_{J:J\neq I} \bar{\lambda}_{IJ}$.*

Proof. Sato & Koizumi (1973) show $\sum_j \bar{\lambda}_{ij} = 0$. Dividing variables w_j into those that are together with w_i in aggregate W_I and those that are not, we have $\sum_{j\in W^I} \bar{\lambda}_{ij} = -\sum_{j\notin W^I} \bar{\lambda}_{ij}$. By theorem 2, $\sum_{j\in W^I} \bar{\lambda}_{ij} = -\lambda_{II}$ and $\sum_{j\notin W^I} \bar{\lambda}_{ij} = -\lambda_{IJ}$. Therefore$\bar{\lambda}_{II} = -\sum_{J:J\neq I} \bar{\lambda}_{IJ}$. ■

These results, which are easily extended to the uncompensated demand case, show Marshall's Rules hold for aggregated inputs. This confirms aggregate data can be used to calculate cross price elasticities. For our purposes, the practical value is that we can use our disaggregated data to construct an aggregate cross- and own-price elasticity of factor demand by either adding up the factor shares and multiplying by the relevant elasticity of substitution ($\bar{\lambda}_{IJ} = S_J\sigma_{IJ} = S_J\sigma_{ij}$) or by adding up the relevant disaggregated factor demand elasticities; that is, by using the fact that $\bar{\lambda}_{IJ} = \sum_{j\in W^J} \bar{\lambda}_{ij}$. We can use proposition 1 to find the aggregated own-elasticities.

2.9 Summary

We will use data to estimate the parameters of the translog cost function and calculate the Allen elasticity of substitution (σ) as well as the related conditional and unconditional cross-elasticities of factor demand ($\bar{\lambda}$,λ) in chapter 4. This models the effects of exogenous changes in factor prices on endogenous factor quantities. We have confirmed it is legitimate to use a cost function under conditions of non-constant returns to scale. We have also shown that Marshall's Rules hold for aggregated inputs and will use this to produce aggregate elasticities of substitution and factor demand. We will similarly estimate the translog production technology and derive the Hicks elasticities of complementarity (H) and of factor price ($\epsilon, \hat{\epsilon}$) in chapter 5. This models the effects of exogenous changes in factor quantities on endogenous factor prices. We will adopt the value added specification. Each of these chapters will discuss at length the estimation issues that arise in general and in the context of the data used.

3 Data

3.1 Introduction

The chapter begins with a description of the manufacturing firm dataset that forms the basis of the empirical studies in chapters 4 and 5. The more important constructions based on the dataset, notably of raw materials and value added, are disclosed here. There are two potential drawbacks to the firm-level dataset, the first being that it is a cross section. The chapter will establish there are enough candidates to act as controls for firm-specific effects and hence mitigate omitted variable bias.

The second potential drawback is that there are no factor prices, so these must be constructed. After justifying the procedure, this chapter dedicates substantial space to explaining how wage data are estimated for each firm using household survey data. In summary, this is done by predicting wages according to characteristics that are common to both the firm and household surveys. The chapter further explains theoretically the likely effects on the translog estimates of not accounting for firm-size effects on wages. It suggests how the imported wages can be adjusted for firm size. The way the cost of capital is constructed is also presented and the chapter continues by showing how these data are used to calculate total costs, factor cost shares and value added. The constructed wage data and core survey have been used by Edwards & Behar (2006) to examine the links between trade, technology and relative wages in South Africa. We present the list of variables and details of construction where necessary in Appendix 2.

3.2 Data from the firm-level manufacturing survey

The dataset used is from the National Enterprise Manufacturing Survey (NE survey) covering the period of 1998. After adjusting for non-response and outliers, there are about 300 firms with the appropriate variables. Unlike the Greater Johannesburg Metropolitan Council Survey (GJMC survey), the NE survey is national in coverage. For a thorough analysis of the data, see Bhorat & Lundall (2002). Basic descriptive statistics are in Appendix 3. The most notable feature of the statistics is the skewness of the data, with the mean often exceeding the 75th percentile.

The dataset is a single cross section, so variables are required to control for firm-specific effects and avoid omitted variable bias. Fortunately, the NE dataset has a rich set of variables for the purpose. There are nine industries and nine provinces. There is information on whether the firm is a member of a bargaining council or otherwise subject to a bargaining council agreement. Consistent with the view that trade unions are more likely to survive in some industries than others (Booth,

1995) and that they may have a non-price effect on factor quantity, this is useful information. There are also ordinal variables for how much difficulty firms have recruiting workers within each occupation, which may capture something about the nature of the firm's activity and may also have a non-price effect on factor choice. Other variables include the percentage of sales that a firm exports, the age of the firm's equipment, the manager's satisfaction with productivity, the percentage of raw materials the firm imports and the percentage of assets invested in computers. A full list is provided in Appendix 2.

The key variables for the production function are the capital stock and employment numbers by occupation group. The five groups are:

- Managerial/Professional
- Sales/Clerical
- Skilled/Artisan (technicians, welders)
- Semi-skilled (machinery operators)
- Unskilled (labourers, security guards)

In the absence of better information, part-time workers are given a weighting of a half in computing a weighted measure of the total workforce. Capital stock is available in currency (Rand) values. The correct procedure is to adjust this for capacity utilization, but, while this percentage is available for large firms, it is not for small firms. Using data on the actual and maximum average shift length and the number of shifts per week, it is possible to construct a shift capacity utilization variable as a reasonable proxy. The capital input is therefore the capital stock adjusted for shift capacity utilization.

There is information in the data on what percentage of total costs is comprised of raw materials costs, but there is no data on total costs or on raw materials costs. To derive a measure of raw materials costs, it is necessary to assume that turnover equals total costs. Then raw materials as a percentage (p) are multiplied by turnover (q) to get a measure of raw materials costs. Value added is constructed as sales minus the constructed raw materials so that

$$v = (1 - p)q \tag{43}$$

3.3 Wage construction

The National Enterprise dataset does not have wage data. Edwards (2003) tries to control for wages and other industry specific factors by including industry dummies in his labour demand equations. This is inappropriate for a study like this, of which wages are an essential aspect. Therefore, appropriate wage information is transplanted from household survey data, where characteristics common to both the NE survey and household data are used to predict wages by occupation for each firm. This section justifies the procedure and describes it in more detail. Thereafter, it demonstrates the need to adjust wages for firm-size effects and demonstrates a way to do this. The production function estimates do not necessarily require wage data, but the cost function estimates do. Furthermore, the wage data have been used to investigate the links between tariff liberalization and other trade measures, technology and wage inequality using labour demand estimates and the mandated wage equation approach (see Edwards & Behar, 2006).

3.3.1 Why firm-level wages can realistically be represented by supra-firm data

Average wages by industry and occupation are a good approximation to those faced by firms in South Africa. Nattrass (2000) reports that the main wage setting institutions are industrial level bargaining councils (BC), noting that 65% of manufacturing workers are covered by a BC. Furthermore, the Minister of Labour is obliged to extend BC agreements to non-members. Nattrass concludes that extension to non-members is at the core of wage setting in an industry. Also, Moll (1996) shows theoretically how extensions of bargaining council agreements make some firms become more capital intensive and other firms, which tend to be small and labour intensive, leave the industry. This leads to convergence in technologies and wages in the industry.

The NE survey provides data on whether the firm is subject to collective bargaining and/or a BC agreement. On average, over 70% of firms are subject to a BC agreement. Small firms (those with fewer than 50 employees) are almost 100% covered while large firms vary from 32% to 61% by industry in coverage. There is therefore support for convergence of wages in industries and justification for wages being calculated at a supra-firm level. However, the NE survey does not reveal which occupations within a firm are subject to BC wages. One could argue that more skilled wages are less likely to be influenced by collective bargaining, but the household data show that the proportion of trade union membership does not vary materially by occupation. Even if not influenced by bargaining, more skilled people tend to be mobile, which standardises wages across firms through ordinary market processes.

Chapter 2 introduced arguments for why wages may vary because of firm-specific characteristics, which means they are not fully exogenous and therefore not perfectly suited to cost equations. However, constructing the wages using these exogenous factors succeeds in removing the endogenous component from wages, so the process of predicting the wages using other criteria could crudely be described as a two-stage procedure with instrumented wages. Like all such procedures, one still has to worry about the quality of the data.

Predicting wage data has precedence. Teal (2000) generates predicted values from earnings functions using a matching panel: at the same time as firm-level data was collected, employees within the firms were asked questions on earnings and related variables. Classifying workers as skilled or unskilled, Teal generates firm-level wages using the human capital characteristics observed in those workers sampled for each firm, controlling for other factors. Unlike his work, this study unfortunately neither uses matching data nor is able to draw on individuals' characteristics, because neither feature is available in the NE survey.

3.3.2 Data and procedure

We reduced the 1997 October Household Survey[10] sample to include only those 3 500 people working for somebody else in formal manufacturing industries. Definitional correspondence to the NE survey in terms of industry, province and occupation is good, but, as will be explained, the correspondence regarding union membership / collective bargaining is not. Details of the dataset and survey methodology are available in Statistics South Africa (1998).

In the survey, people were interviewed in geographical clusters and stratified by magisterial district. The sample surveyed is not fully representative of the population. We take survey design effects like these into account. People in the 10 households interviewed per geographical cluster don't have independent characteristics, being more likely to have similar features. Therefore, the survey sample variance of the wage would be lower than would be the case in a random sample. Failing to account for clustering often results in standard error estimates that are half what they should be. In contrast, stratification guarantees that this similarity will not happen across strata. It mitigates the chances of there being a non-representative sample and therefore standard error estimates should (correctly) be lower than in the absence of stratification Deaton (1997).

This study accounts for probability weights and clustering but only partially adjusts for stratification. The reason for this is that many magisterial districts (strata) have only one cluster –

[10] The 1998 survey was much smaller due to funding problems. This and an allowance for adjustment lags make the 1997 survey the preferred edition. Inflationary increases are easily dealt with.

many have only one observation – and at least two are needed for variance estimates. A standard procedure for dealing with this is to collapse or merge strata (Statacorp, 2003), but the number of cases to collapse is high in this study. Therefore, compromise stratification by province, which sometimes has close to 100 magisterial districts, is carried out. An aggregate estimate of monthly salary has a Deff statistic, which is the ratio of the estimated variance accounting for survey design to the unadjusted variance, of 2.44. This indicates the importance of dealing with the survey design effects. For each occupation, the characteristics available in both data sources are:

- economic activity (broken down into nine industries)
- province group (the nine provinces were ex post broken down into two groups with similar wages)
- individual trade union membership (household data); collective bargaining and bargaining council membership (firm data)

Construction entails calculating the survey-adjusted means for groupings of people for each occupation. A total of four different wage series are constructed for each of the five occupations. The next few pages describe in some detail each of four wage series, namely wage_{ind}, wage_{all}, wage_{some} and wage_{size}. We will ultimately choose wage_{size} for our regressions, because they take account of firm-size effects on wages.

1) Wage_{ind} only classifies wages by industry. This measure is used mainly as a check against other series and for consistency in preliminary analysis. It provides too coarse a measure of wages to be used in regressions. Because such data are not available elsewhere, a table of average wages by occupation and industry, using only the household data, is provided in Appendix 4. The manufacturing industry data are consistent with the overall manufacturing wage calculated by Statistics South Africa (2000). Using the same survey, they calculate a mean unskilled hourly wage of R7.86 for example, which corresponds roughly to R 1 400 (£100) per month.

2) It is appropriate to classify wages further by location and trade-union membership to generate Wage_{all}. There are nine industries and nine provinces, meaning that, together with a trade-union membership dummy, there are potentially 162 different wages. However, while some means are calculated using a comfortable number of observations, others are based on few data points, sometimes only one. This means the standard errors on the wage estimates can be high (or non-existent). To mitigate this, the nine provinces are divided into two groups, as variation within each of the two groups is low.

Table 1: Selection of categories into which wages for skilled/artisanal workers have been placed

Mean Monthly Salary: Skilled/Artisan (Rands)		
	Estimate	Std Error
Food & Beverages	1562	161
Wood, Pulp & Paper - Prov0	1116	229
Wood, Pulp & Paper - Prov1	1993	169
Chemicals, Rubber & Plastic - Prov0, not unionised	786	152
Chemicals, Rubber & Plastic - Prov0, unionised	2316	264
Chemicals, Rubber & Plastic - Prov1	2067	284
Source: own calculations based on October Household Survey data		

3) However, a third measure (Wage_{some}) is also generated. It achieves more precise estimates by combining some locations and industries and/or not distinguishing by trade-union membership in cases where wages do not differ substantially. Before discussing the process, it is helpful to look at one example of classifications. Table 1 presents six of the fifteen composite groups the skilled/artisan wages are divided into and the associated estimates.

The first row contains wages for all skilled/artisans in the Food & Beverages industry, regardless of location or union membership. The Wood Pulp & Paper industry is subdivided by province group but not union membership (rows 2 and 3). Wages in the Chemicals, Rubber and Plastic industries are subdivided by province group. One group of provinces is further divided into unionised and non-unionised workers (rows 4 and 5) while the other group is not (row 6). In some cases, industries are combined, with the possibility of disaggregation by other criteria. This selection could of course be seen as a prediction from a regression. However, all the explanatory variables are dummies or interactions thereof, so calculating means for each group is an equivalent and more convenient procedure, especially when taking account of multiple interactions.

Classifying the wages for Wage_{some} involves a number of trade-offs. While averages across two or more groups are different, the standard errors may be large, resulting in imprecise estimates. This is often because of a small sample size for that group. One way to proceed is to separate all groups with statistically significant differences in means. However, this is imperfect. An extreme but not infrequent occurrence is that of one observation per group, which generates no standard error and is also outside the confidence interval of another group. Similarly, inference based on very few observations is not reliable. On the other hand, some estimates, even if based on few observations, are so radically different that the groups should be classified separately. The aim is to produce a set of estimates per group with better precision characteristics but sufficient variation to represent the firm-level data. To do this, various combinations are carefully inspected. Factors considered are differences in log wages, the number of observations, and comparisons of the standard errors and confidence intervals of the separate and combined groups. Of course, all the criteria are related.

Comparing the confidence intervals of two groups is naturally akin to performing a two-sample t-test. However, visual inspection is quicker for all the combinations and allows for analysis in conjunction with the other criteria. The choice of confidence interval is a matter of taste in this application, so 85% bands are used. To augment this procedure more formally, standard t-tests, regressions and non-parametric procedures[11] are performed on certain groups.

Going through the above procedure on a case-by case basis therefore produces a set of wages, for each occupation, which partially disaggregates each industry by location and/or trade union membership in a way that optimizes the trade-off between achieving representative wage estimates and having precise estimates. Appendix 5 shows the number of categories for wage_{some} ranges from 7 to 15, with the average number of observations per category ranging from 16.9 to 43.7. The average number of observations for wage_{all} ranges from 3.9 to 14.6. It is important to stress that most groups of wages do not differ substantially from those in wage_{all}, the exceptions being those in wage_{all} based on very few observations and in which one can have very little confidence anyway.

4) The fourth measure of wages is wage_{size}. It adjusts wage_{some} for firm size, as discussed next.

Finally, data from the Trade and Industrial Policy Secretariat shows wages rose by approximately 15% between the time of the household survey and the time of the NE survey. All Wage measures are therefore raised by this percentage.

3.3.3 Adjusting wages for firm size

The start of the chapter suggested there were no firm-specific variables by which the household wages could be classified. However, there is a way to account for firm size. This section (and chapter 4) will show why failing to account for firm size in wages explains some initially poor results. A way to adjust wages for firm size using existing estimates is proposed.[12]

Chapter 2 suggested firm size may affect wages. The following paragraphs explain what impact ignoring this effect may have on translog estimates, concluding that the estimations are more likely to (falsely) reject homotheticity and linear price homogeneity and to overstate returns to scale. Abstracting from individuals' characteristics, wages for occupation i can be seen as a simple function of firm size measured according to value added (q) and a vector of variables available from

[11] Tests of median equality are performed, but they do not factor in survey design. The results do not indicate material differences in classification. Another useful way to compare specific groups is to use Anova and Scheffe's method of comparing the means of each group to those of all the others (Van den Honert, 1997). This method is used but there is also no readily available way to adjust for survey design.

[12] Any exogeneity issues are relatively minor in comparison.

the household survey ($\mathbf{x}$):

$$\ln w_i = \boldsymbol{\beta}_i \ln \mathbf{x} + \gamma_i \ln q, \gamma_i > 0 \quad (44)$$

$$= \ln \hat{w}_i + \gamma_i \ln q \quad (45)$$

In other words, (44) is the wage for an individual with characteristics ($\mathbf{x}, q$) in occupation i while $\hat{w}_i$ is the (perfectly) predicted wage for that individual based on data from the household survey. We substitute (45) into our translog function given in (31) to see how the coefficients involving value added would be affected. Algebra shows a translog cost function such as (31) that does not account for firm effects on wages is the same as:

$$\log C = \sum_i a_i \ln \hat{w}_i + \log q + \frac{1}{2} \sum_i \sum_j B_{ij} \log \hat{w}_i \log \hat{w}_j + \Phi \log^2 q + \Omega \log \hat{w}_i \ln q \quad (46)$$

where $\Gamma = \sum_i a_i \gamma_i + a_q; \Phi = \frac{1}{2} \sum_i \sum_j B_{ij} \gamma_i \gamma_j + \sum_i B_{iq} \gamma_i + B_{qq}; \Omega = \sum_i \sum_j B_{ij} \gamma_j + \sum_i B_{iq}$. The coefficients containing value added may be vastly different to what they are supposed to be. To gauge the likely nature of the bias in a simple setting, we assume the coefficients are correctly estimated. Furthermore, on the assumption that linear price homogeneity and constant returns to scale $\left(\sum_i a_i = 1; \sum_i B_{ij} = \sum_j B_{ij} = 0; B_{iq} = B_{qq} = 0; a_q = 1\right)$ - see Berndt & Khaled (1979) and chapter 4 - hold in the true cost function:

$$\Gamma = \sum_i a_i \gamma_i + 1 \quad (47)$$

$$\Phi = \frac{1}{2} \sum_i \sum_j B_{ij} \gamma_i \gamma_j \quad (48)$$

$$\Omega = \sum_i \sum_j B_{ij} \gamma_j \quad (49)$$

We can't be sure $\sum_i a_i \gamma_i > 0$ as it is not necessarily the case that $a_i > 0$ for all i (Varian, 1992). Therefore, we cannot be sure $\Gamma > 1$, which we would be the coefficient on $\ln q$ under constant returns. However, linearly homogeneous prices imply that, if all the values of γ_i for each occupation are close enough to the average across occupations, the result will tend to be an upward bias on the value added coefficient. If the firm-size effect is equal for all occupations, the bias is γ.

It is not possible to tell what direction the bias will be for Φ. However, if there is an equal firm-size effect for all occupations, price homogeneity implies this will be zero and in fact not biased. If the firm-size effect is not equal for each occupation, there is the possibility of Φ being

Table 2: Estimates of relationship between firm size and wages

Managers	Professional/ technological	Clerks	Sales	Craft	Operators	Labourers	Total
0.089	0.076	0.09	0.066	0.096	0.094	0.031	0.065
Source: Bhorat & Lundall (2002); all estimates except Labourers were significant							

found significant when it actually is not. This would falsely reject a homogeneous technology. A similar analysis concludes the coefficient on Ω may be found significant and therefore falsely reject homotheticity or that linear price homogeneity is rejected by distorted coefficient values.

To understand the likely effects on returns to scale, assume for simplicity a common firm-size effect across all occupations. The assumption of a homogeneous technology is relaxed but homotheticity and price homogeneity are maintained. Returns to scale are given by

$$\left[\frac{\partial C}{\partial q}\right]^{-1} = [\gamma + a_q + B_{qq} \log q]^{-1} \tag{50}$$

One can gauge that omitting the firm size variable will underestimate the denominator by γ on average, so returns to scale will be overestimated. This is intuitive: if wages rise for bigger firms, the returns to scale are less than otherwise. Therefore, including a measure of γ will reduce the estimated returns to scale.

Given the possible problems with ignoring firm-size effects, ways of capturing them must be found. There is no information on the size of the firms which individuals in the household survey work for. One way to proceed might be to assume constant returns to scale or some other degree of homogeneity, estimate the unconstrained model, and see the level of γ necessary to produce the assumed value. Another is to attach values of γ_i to the wage series. Bhorat & Lundall (2002) estimate manufacturing firm-size wage effects using data for the Gauteng Province in South Africa, which are shown in table 2.

Their estimates are rudimentary, using only average firm wages and annual firm sales, but they claim similarity to the US study of Doms, Dunne & Troske (1997). This study does not adjust wages for other factors because no number is available, so the estimates in the table are taken, being adjusted using averages to match the occupations in the NE data. Assuming the unadjusted wages represent those for an average-sized firm, wage_{some} is inflated/deflated accordingly. This fourth wage is denoted by wage_{size}.

3.4 Costs of capital

The Jorgenson (1963) "cost of capital" is:

$$c = p\left(\frac{1-uv}{1-u}\delta + \frac{1-uw}{1-u}r\right) \tag{51}$$

c is the cost of a unit of capital, p is the price of machines, δ is the rate of depreciation, r is a nominal interest rate. u is the corporate tax rate while v and w are the proportions of depreciation and interest expenditure chargeable against income for tax purposes. When capital is aggregated in a static setting, p is normalized to one and (51) is a percentage cost per unit of capital. Setting $v = w = 1$ yields:

$$c = \delta + r \tag{52}$$

Clague (1969) uses this version in his early study of developing countries. In South Africa, v is equal to one and w is at least one because of accelerated depreciation allowances. Because w does not equal one and because firms pay tax even on normal profits, the dropped tax term needs to be reintroduced. We use the slightly ad hoc expression from an industry-level study of capital in South Africa (Fedderke et al., 2001):

$$c = (r - \pi) + \delta + \tau \tag{53}$$

π is the inflation rate measured by the consumer price index and τ is the economy-wide corporate tax rate. Fedderke et al. calculate industry-level values for δ ranging from 11% to 16%.[13] For the real interest rate $(r - \pi)$, they use yields on 10-year government bonds and consumer price inflation, but we use the average prime lending rate and consumer price inflation for 1999.

Furthermore, we adjust the interest rate to account for risk. Adjustments range from –2% for large (>50 employees) and old (> 5 years) firms to + 5% for new small firms.[14] This amendment can be accused of confusing Jorgenson's notion of the cost of capital with the weighted average cost of capital due to Modigliani & Miller (Lau, 2000). Bergström & Panas (1992) use a weighted average cost of capital measure in their study, while Teal (2000) constructs predicted profit rates as a percentage of the capital stock with regressions containing firm- and industry-specific variables. In equilibrium, these two costs should be equal (Lau, 2000). To the extent that this equilibrium does not hold realistically, and because of the practical reality that smaller and newer firms are

[13] I thank Prof Fedderke for providing the data.

[14] 5% is the standard rule of thumb premium for new small ventures in South Africa.

Table 3: Comparison of two measures of value added

Statistic	$V1 = (1-p)q$	$V2 = q - rm$
mean	8	8.71
1st quartile	1.2	0.77
Median	2.8	3.21
3rd quartile	8.75	9.63

likely to have higher borrowing costs, accounting for these risk premia is necessary.

Fedderke et al. (2001) use the nominal corporate tax rate for τ, which was 35% for the fiscal year starting early in 1998 (RSA, 1998). They state it would be ideal to have the effective rates of taxation by industry. Negash (1999) calculates effective tax rates to be about 15% below nominal rates for the 1990s, so a 20% average effective rate is applied to all firms. The resulting range of costs of capital in Appendix 6 vastly improves on standard studies that simply assume a uniform interest rate as the cost of capital, like Maki & Meredith (1987) and Fallon & Layard (1975).

3.5 Total costs, cost shares and value added

Wages are also used in the determination of cost shares and total costs. The vast majority of studies, including but not restricted to Binswanger (1974a), Berndt (1973b), Teal (2000) and Bergström & Panas (1992), derive total cost and/or factor cost shares using factor price and quantity data. This means that variables on the right hand side of the share equation are used to construct the dependent variable, but there is no readily available alternative.

For this thesis, labour costs are obtained by multiplying labour quantities by the constructed wage for each occupation. Capital costs are the cost of capital percentage multiplied by the capacity-adjusted capital stock. Total factor cost (C_f) is the sum of factor costs. Raw material input data is available only as a percentage of total costs (p). Total input cost (C_i), including raw materials, can be calculated as:

$$C_i = \frac{C_f}{1-p} \tag{54}$$

Raw materials costs (rm) are easily calculated using C_i and C_f and subtracted from output to get a second measure of value added, in addition to that given in (43). (43) is the measure of value added we will use, but table 3 considers both measures of value added for comparison.

The measures of central tendency are close but there is moderate dispersion at the 1st and 3rd quartiles. Correlations between the first measure ($V1$) and the wage-based measures ($V2$) vary from 0.89 to 0.91, depending on the wage definition.[15] The similarities are considerable, in spite of

[15] While $wage_{some}$ is used in this table, the results are consistent for all wage definitions.

the difference in calculation, so there are grounds for confidence in the estimates.

In cost estimations, $V2$ would introduce very serious correlation with the dependent variable, which was constructed using the exact same factor prices and quantities. $V2$ would also be highly correlated with the other inputs. Therefore, while useful for comparison with $V1$, $V2$ is not used in regressions. $V1$ is used as the measure of output in the cost function (and in the production function).

As will be explained in chapter 4, we will estimate systems of factor share equations in addition to the cost function. To calculate each factor share, we multiply the factor's wage by its quantity and divide it by the sum of the factors' costs; that is: $s_i = \frac{w_i x_i}{\sum_i w_i x_i}$. Appendix 7 contains some descriptive statistics on the factor shares. Combining the Managerial/Professional and Sales/Clerical categories yields a mean share of 16%, the three less skilled categories combined are 33% and capital is 51%. For Ghana, Teal (2000) calculates shares for skilled labour, unskilled labour and capital of 11%, 29% and 60%. The same table (Appendix 7) shows roughly 5% of labour shares are zero, simply because of zero reported labourers in that category. These firms are at corner solutions that would not be in the feasible production set (Berndt & Christensen, 1973b). The phenomenon is especially pertinent in the translog context, because many calculations require division by one or more factor shares (cf. equation (32)). Berndt (1991) asserts that, instead of actual shares, one should use predicted factor shares from the regression in such calculations (not necessarily because of this problem). This may avoid division by zero, but one condition for profit maximisation is strictly positive factor shares (Berndt & Christensen, 1973a). The post-estimation predicted shares should be tested for this feature, but clearly, under these circumstances, such a test is bound to fail as even relatively precise estimates will yield negative shares for some occupations in some firms.

3.6 Summary

The variables needed for the production estimates are reasonably easily derived. In contrast, costs of capital must be constructed. Also, labour prices for each occupation are predicted by matching firm characteristics with those of a household survey for use in cost function estimates. The three different wage variables derived this way differ in the level of disaggregation. The chapter shows analytically that ignoring firm-size effects on wages can lead to false rejection of homotheticity, false rejection of price homogeneity and/or overestimates of returns to scale. It proposes a method to adjust wages for firm size. Factor shares of cost and total costs are derived using the wage and cost of capital data. Analysis of two constructed value added measures provides preliminary grounds for optimism regarding their reliability.

The next two empirical chapters will use the data described in this chapter and draw on the theoretical material from chapter 2: chapter 4 estimates elasticities of substitution and of factor demand while chapter 5 estimates elasticities of complementarity and of factor price.

4 Allen elasticities of substitution and elasticities of factor demand

4.1 Introduction

According to the narrow ILO definition, South Africa's unemployment rate is about 25% while the expanded definition measures almost 40% of the potential working population as unemployed (Statistics South Africa, 2005). Unemployment has historically been "*literally off the charts*" compared to other developing countries (Nattrass, 2004:90) despite steady economic growth in the last decade. It is a stated aim of the government to try to halve unemployment by 2014 (Hirch, 2004, cited in Pollin, et. al, 2006). While a variety of measures are being attempted to address the unemployment problem, our focus in this chapter is on the role of factor prices.

Specifically, the objective of this chapter is to examine what effects a rise in the cost of some types of labour may have had or will have on employment. We consider this both in absolute terms, relative to the costs of other labour types, and relative to the cost of capital.

On the one hand, South Africa has strong bargaining institutions which, together with the introduction of new labour legislation in 1995,[16] may be raising the costs of labour and contributing to unemployment (Fedderke et. al., 2001). The effects may be particularly pronounced towards the lower end of the skill spectrum. This goes against the World Bank's prescription of adopting wage restraint as a mechanism for employment creation (Fallon & Lucas, 1998).

Together with these rising labour costs, there are frequent calls for measures to reduce the cost of capital. The Congress of South African Trade Unions (COSATU), a junior member of the governing alliance led by the African National Congress (ANC), repeatedly lobbies for cheaper capital. These arguments have traditionally been made on the basis of standard macroeconomic transmission from monetary conditions to fixed investment and output expansion (COSATU, 1998) but have recently been extended to the potential role lower interests rates could have in weakening the currency and stimulating export demand (COSATU, 2005). Similar arguments have been advanced in a highly publicised book funded by the United Nations Development Programme (Pollin et al., 2006). The arguments are also generally based on increased fixed investment and the associated multiplier effects. However, within some circles of the ruling African National Congress (ANC), there is evidence of concern over potentially adverse employment effects of lower capital costs. An ANC discussion document (2005:23) states:

[16] See Republic of South Africa (1995)

"Cheaper capital without reforms to reduce the relative cost of labour is likely to result in higher investment that displaces labour."

This clearly presents the microeconomic substitution argument. Consequently, a microeconomic contribution can be made by estimating the elasticity of substitution between factors. Under certain assumptions, the sign of the parameter can predict whether a fall in the cost of capital relative to that of labour will lead to a relative rise or fall in the demand for labour. Estimates of labour demand elasticities would also suggest to what extent higher wages may be reducing labour demand.

The purpose of this chapter is to estimate the Allen elasticity of substitution (AES) between various labour inputs as well as cross- and own-price elasticities of factor demand in South Africa. Such elasticities are measured between capital and labour inputs disaggregated by skill. This chapter divides the workforce into up to five occupations – managerial/professional, skilled/artisan, semi-skilled and unskilled to produce disaggregated estimates. The occupations are subsequently aggregated into more skilled and less skilled labour. Such measures allow one to address the concerns expressed by the ANC. Furthermore, by differentiating by skill type, we are able to examine the Capital Skill Complementarity (CSC) hypothesis due to Griliches (1969) and hence see whether cheaper capital has relatively stronger negative effects on those with fewer skills.

Importantly, we can also present estimates of uncompensated factor demand. While studies generally provide elasticities that hold output constant (or don't reveal what they claim to be estimating), we can (subject to further assumptions) account for endogenous changes in the optimal level of output. This is important because, although a fall in the price of one factor will lead to a fall in demand for the other if we hold output constant, it may be that the fall in price leads to an expansion in output and demand for all factors such that the overall effect is a rise in demand for the other factor.

Despite the predictions made by microeconomic theory, some are sceptical that relative factor prices have played a role in South Africa's capital intensity and poor employment creation (Kaplinsky, 1995). Pollin et al. (2006) argue measures to reduce wage costs, including wage subsidies, would have a minimal effect. Nonetheless, it would be difficult to argue factor prices have played no role. Traditional elasticity measures are an important contribution to an informed debate.

Despite the importance of such measures for policy, few estimates of disaggregated labour demand elasticities exist in South Africa and for the developing world. According to Fajnzylber & Malony (2001), only two of the nearly 200 studies surveyed by Hamermesh (1993) use establishment data for developing countries. These appear to be limited to Latin America, but Teal (2000) subsequently studies Ghana for two skill types.

This chapter employs firm-level manufacturing data that is highly disaggregated by occupation. It uses a translog cost function and associated factor shares rather than a simpler technology or direct labour demand estimates. Therefore, this chapter presents a thus far undocumented combination for a developing country. Unlike other work, we are able to indicate how these elasticities vary across the firms in the sample. Furthermore, we employ Marshall's Rules to allow for scale (output) effects. We allow for a variety of potential scale effects depending on the industry studied. In addition, we use the implications of separability (see chapter 2) to produce aggregate elasticity measures. With respect to aggregate elasticities of factor demand, this is to our knowledge the first application of an aggregated version of Marshall's Rules.

Having a drawback that is common in developing countries, the firm-level data does not contain wages. Therefore, using the procedure described in chapter 3, household data are used to predict wages for each firm according to characteristics that are common to both the firm and household surveys. Building on the analysis of firm-size influences in chapter 3, wages are adjusted for firm-size effects. In addition to this practical methodological innovation, we can draw on the theoretical results presented in chapter 2, which show we can use cost functions to estimate the Allen elasticities of substitution and constant-output elasticities even if returns to scale are not constant.

Section 4.2 discusses the theoretical background, in particular how we can use a translog cost function to produce the elasticity measures. Section 4.3 summarizes the data used and explains the estimation procedure and stochastic framework. It also discusses particular difficulties with inference on the coefficients and on the elasticities, which are non-linear functions of the parameters.

The preliminary results in section 4.4 are used to analyse and interpret features of the technology estimated, notably homotheticity and returns to scale. It explains the choice of wage variable ultimately employed, finding that wages which allow for firm size have the results predicted in chapter 3 and that the results are superior.

Section 4.5 presents the final estimations used for disaggregated elasticities. It also presents and discusses the range of elasticities calculated. The disaggregated AES estimates suggest capital and all forms of labour are p-substitutes. This means a rise in the price of labour relative to capital will lead to a relative fall in employment. Unskilled workers and skilled/artisanal workers are p-substitutes. This holds even if we allow for constant output under a variety of settings, except for a subsample consisting of exporters. Unskilled workers and semi-skilled workers are p-complements. This highlights the benefits of disaggregation. Uncompensated own-price elasticities for manufacturing range from −0.59 for skilled/artisanal occupations to −0.88 for semi-skilled employees.

The aggregated measures in section 4.6 confirm that a fall in the cost of capital would lead to a fall in employment of both types of labour, even after accounting for output effects (except for exporters). We find no support for the CSC hypothesis and find that more and less skilled labour are p-complements. As is generally found in studies of labour demand (Hamermesh, 1993), demand for more skilled labour is less elastic (–0.88) than demand for less skilled labour (–1.03).

Section 4.7 summarises the results and discusses their policy implications. Before proceeding, it is appropriate to summarise the assumptions employed in this model, which will be contrasted with those in chapter 5. We are assuming factor prices are exogenous and that factor quantities are endogenously chosen by cost minimizing or profit maximizing firms. Our dataset is for manufacturing. Our analysis can be viewed as one of the manufacturing sector and, in some cases, of industries within that sector. The conclusions apply to the broader economy only if manufacturing is representative of all sectors. For constant output elasticities, we do not make any technological assumptions. However, to allow for changes in optimal output by firms within an industry or manufacturing as a whole, we will have to assume locally constant returns to scale for each firm.

4.2 The translog cost function and associated elasticities

4.2.1 Underlying cost function

The translog cost function introduced in chapter 2 equation (31) is:

$$\begin{aligned}\log C &= \log a_0 + \sum_i a_i \log w_i + \frac{1}{2}\sum_i\sum_j B_{ij}\log w_i \log w_j + a_q \log q \\ &\quad + B_q \log^2 q + \sum_i B_{iq}\log q \log w_i\end{aligned} \tag{55}$$

C is cost and w_i, w_j are the prices of factors i, j. q is value added.[17] As we saw in chapter 2 for production functions, the cost share equation for factor i is derived by differentiating the cost function with respect to $\log w_i$. Following Chung (1994),

$$\frac{d\log C}{d\log w_i} = a_i + \sum_j B_{ij}\ln w_j + B_{iq}\ln q, \tag{56}$$

[17] Lau's (1969) duality result suggests empirical support for separability of raw materials in the production function implies support for cost function separability of raw materials. It would be ideal to redo the empirical test on the cost data. We unfortunately do not have information on raw materials prices so we must either rely on that result or, like Griffin & Gregory (1976), assume separability of raw materials from the other inputs.

but

$$\frac{d\log C}{d\log w_i} = \frac{w_i}{C}\frac{dC}{dw_i} = \frac{w_i x_i}{C} \tag{57}$$

(by Shephard's Lemma). Furthermore:

$$\frac{w_x x_i}{C} = s_i \tag{58}$$

Therefore:

$$s_i = a_i + \sum_j B_{ij} \ln w_j + B_{iq} \ln q \tag{59}$$

Consistent with cost minimizing behaviour (Berndt & Khaled, 1979), Slutsky symmetry requires:

$$B_{ij} = B_{ji} \tag{60}$$

Inspection of the cost function reveals linear price homogeneity ($C(\delta\mathbf{w}) = \delta C(\mathbf{w})$) requires:

$$\sum_i B_{ij} = 0 \tag{61a}$$

$$\sum_j B_{ij} = 0 \tag{61b}$$

$$\sum_i a_i = 1 \tag{61c}$$

$$\sum_i B_{iq} = 0 \tag{61d}$$

In addition, technological assumptions can be tested for and if applicable imposed on the cost function and share equations. The returns to scale are given by differentiating the cost function with respect to $\log q$:

$$\frac{d\log q}{d\log C} = \left(\frac{d\log C}{d\log q}\right)^{-1} = \left(a_q + B_q \ln q + \frac{1}{2}\sum_i B_{iq} \ln w_i\right)^{-1} \tag{62}$$

For returns to scale that are independent of factor prices (homotheticity), $B_{iq} = 0\ \forall i$. This can also be observed in (59), where the factor share is no longer a function of output. If homothetic, the cost function is homogeneous of degree r if $B_q = 0$, with $r = \frac{1}{a_q}$. $a_q = 1$ corresponds to constant returns to scale.

4.2.2 The elasticity of substitution and of factor demand

The constant output elasticity of factor demand ($\bar{\lambda}_{ij}$) is the responsiveness of the quantity of factor i to a change in the price of factor j, holding output and all other factor prices constant. We briefly introduced these concepts in chapter 2 but now consider the specific translog case: using the fact that $\bar{\lambda}_{ij} = \frac{w_j}{x_i}\frac{\partial x_i}{\partial w_j}$ and $s_i = \frac{w_i x_i}{C}$,

$$\bar{\lambda}_{ij} = \frac{w_j}{x_i}\frac{\partial}{\partial w_j}\left(C\frac{s_i}{w_i}\right) \tag{63}$$

$$= \frac{w_j}{x_i}\left(\frac{CB_{ij}}{w_i w_j} + \frac{x_j s_i}{w_i}\right) \tag{64}$$

Therefore,

$$\bar{\lambda}_{ij} = \frac{B_{ij}}{s_i} + s_j \tag{65}$$

The Allen elasticity of substitution measures the proportional change in the ratio of factor quantities in response to a change in their relative factor prices. $\sigma_{ij} > 0$ is defined such that the factors are p-substitutes. We know from chapter 2 equation (21) that $\sigma_{ij} = \frac{\bar{\lambda}_{ij}}{s_j}$ and hence

$$\sigma_{ij} = \frac{B_{ij}}{s_i s_j} + 1 \tag{66}$$

Analagously (Binswanger, 1974a), the own-elasticity of factor demand is

$$\bar{\lambda}_{ii} = \frac{B_{ii}}{s_i} + s_i - 1 \tag{67}$$

and

$$\sigma_{ii} = \frac{B_{ii}}{s_i s_i} + 1 - s_i \tag{68}$$

Humphrey & Wolkowitz (1976) suggest σ_{ii} be interpreted as the change in a factor's demand responsiveness after a change in its own price.

To measure the uncompensated elasticity of factor demand (λ_{ij}), we must consider output effects. The optimum choice of output is a function of the industry product price and the industry input prices. In a competitive industry with a homothetic technology, a unit fall in the price of one factor will lower average and marginal cost by that factor's share (Estrin & Laidler, 1995). For a fall in the price of factor x_j, profit-maximising industry output will rise and so will demand for all factors, by their share. However, as industry output rises, the product price falls, which lowers the value marginal product of each factor and mitigates the increase in demand for all factors. For all constant returns to scale technologies, $s_j|\eta|$ captures the scale effect in the equation below:

$$\lambda_{ij} = \frac{B_{ij}}{s_i} + s_j(1 - |\eta|) \tag{69}$$

$|\eta|$ is the elasticity of product market demand. Similarly:[18]

$$\lambda_{ii} = \frac{B_{ii}}{s_i} + s_i(1 - |\eta|) - 1 \tag{70}$$

We will generally use estimated values of B_{ij} and predicted factor shares to calculate the elasticities. For the aggregated estimates, we will use actual shares, as explained in section 4.6.

4.3 Empirical methodology

4.3.1 Data

In chapter 3, we spoke at length about the data, which has about 300 firm-level observations. We have firm-level manufacturing data on capital and up to five labour inputs.

- Managerial/Professional
- Sales/Clerical
- Skilled/Artisan (technicians, welders)
- Semi-skilled (machinery operators)
- Unskilled (labourers, security guards)

We use household data to predict wages for each firm according to a variety of characteristics. This produced a variety of potential wage series, including one which adjusts for firm-size effects. The wages and input quantities are used to calculate total cost and each factor's share of total costs. To obtain estimates of $|\eta|$, we rely on the results in Selvanathan & Selvanathan (2003), who have estimates of consumer demand elasticities for a variety of industries.

4.3.2 Estimation

Our data is a single cross section, so we do not have recourse to panel data methods for dealing with firm-specific fixed effects. We mitigate this problem with a rich set of controls. The controls are industry, location and large-firm dummies, measures for export intensity, raw materials intensity, import intensity, equipment age, the ease firms report in recruiting each type of labour, the productivity dissatisfaction reported by the firm, training expenditure incurred, perceived market

[18]This is the translog example of the general form given in equation (3) in chapter 2. For expositional purposes, the treatment there is in terms of the absolute value of the own elasticity such that it is positive, while here it is negative. Furthermore, the own elasticity was expressed in terms of a cross elasticity of substitution with respect to the (only) other factor. Here, in a multiple factor setting, we use the own elasticity of substitution.

conditions, measures of collective bargaining, firm age, the capital intensity and computer investment expenditure. These are assumed not to affect the elasticities of substitution directly but to control for firm-specific technological features, which if omitted would bias the coefficients and hence the calculated elasticities.

The stochastic specification of the cost function and cost share equations assumes that the firms make random errors in their input selection. In our preliminary investigations, detailed in the next subsection, we estimate (55) only, using OLS. We assume a normally distributed error term with mean 0 and variance ω . The fuller investigations will estimate the cost function together with the system of share equations given by (59). Here, we assume the errors are multivariate normal with mean vector $\mathbf{0}$ and variance matrix $\mathbf{\Omega}$.

The literature appears to have little difficulty with the assumption of normality. However, the values of factor shares are bounded between 0 and 1. In our case, the mean factor share is less than 10% for many factors. This is due in part to the high level of disaggregation we employ. There is truncation of the actual distribution of shares, which is shown starkly by the histograms in Appendix 8. Only capital, which has a mean share of over 50%, resembles a normally distributed share. Jarque-Bera (skewness - kurtosis) tests reject normality at 1% for all shares. Thus, the estimated residuals are unlikely to be normally distributed. It is also highly unlikely that the problem is unique to this dataset.

Existing work does point out that it is likely that errors in input selection will be correlated across the share equations and the cost function such that $\mathbf{\Omega}$ is not diagonal. The diagonal elements of $\mathbf{\Omega}$ may be different, representing different volatilities for each input and share equation. Single equation estimation is therefore inefficient (Berndt, 1991; Guarda, 2000). The cost share equations will therefore be estimated together as a system with the cost function using the Zellner seemingly unrelated regressions (SUR) model, which exploits correlations between the errors in each of the share equations to improve efficiency. Scope for such gains is limited by the fact that the explanatory variables in each factor share equation are identical or at least highly correlated. However, cross equation restrictions do allow for efficiency improvements (Greene, 2003). Many restrictions exist because the cost shares are derivatives of the cost function, so some coefficients are the same. Slutsky symmetry conditions also imply cross equation restrictions.

By construction, the sums of the a_i coefficients across the factor share equations equal unity for each observation. Therefore, the residual cross product and disturbance covariance matrices are singular and prevent estimation (Berndt, 1991). A common response is to impose price homogeneity on the cost function and hence across the share equations. Using (61c), let $a_k = 1 - \sum_l a_l$ where k

refers to capital and l to the labour inputs. This allows for the capital equation to be dropped and the remaining share equations for each of the labour inputs to be estimated as:

$$s_i = a_i + \sum_j B_{ij} \ln \frac{w_j}{w_k} + B_{iq} \ln q + \omega_i \tag{71}$$

We have dropped the capital equation but the choice is arbitrary if the Zellner iterated efficient (IZEF) procedure is used (Berndt, 1991). IZEF is the dominant method in the literature and is the one employed by this study: instead of one or two-step feasible generalised least squares estimates, the procedure iterates over the disturbance covariance matrix and parameter estimates until they converge (Statacorp, 2003).

4.3.3 Inference

For indications of the significance of coefficients and for tests of price homogeneity, technological restrictions, and separability, it is standard to use Wald tests (eg Guarda, 2000; Griffin & Gregory, 1976), which assumes normality. We also use such tests if necessary but with the caveat that they may not be valid if, as we have argued is likely, the residuals are not multivariate normally distributed.

Besides this inference issue, significant regression coefficients (B_{ij}) neither imply nor are necessary for significant elasticities (Anderson & Thursby, 1986). "Significant" can refer to rejecting a null hypothesis of the elasticity being zero, in which case we can be confident the factors are complements or substitutes or can refer to the Cobb-Douglas elasticity of unity. The difficulty lies in the fact that the elasticity estimates are highly non-linear combinations of the coefficients and data (Greene, 2003). Anderson & Thursby (1986) present conditions under which Allen elasticities of substitution asymptotically follow the normal or ratio-of-normals distribution. They find the normal distribution or ratio-of-normals is appropriate only if the means of the actual factor shares are used. They are not appropriate under many alternative calculations found in the literature, including using the logs of the means of the factor price and output variables or the means of the corresponding logs to calculate shares. We do not have the option to use this result as in general we do not use actual factor shares.

Reviews of empirical work on elasticities of substitution make no mention of significance (Chung, 1994; Hamermesh, 1993). Some studies do not report confidence intervals for the estimators at all (Bergström & Panas, 1992; Chung, 1987; Teal, 2000; Guarda, 2000). Others (Binswanger, 1974b) regard the factor shares as fixed and treat the coefficient as the only variable with a confidence interval, incorrectly inferring the elasticity significance from a t-statistic.

One can speculate why some of the recent studies do not present more appropriate confidence measures. Given the number of parameters in the regressions, the number of parameters used in calculating the elasticities, and that the translog is an approximation to complex underlying technologies with few restrictions, achieving precise elasticity estimates is a heavy burden to place on sub-optimal sample sizes. In an analysis of famous pioneering translog studies, Anderson & Thursby (1986) find that confidence intervals for many of the elasticity measures span both the negative and positive orthants, bringing those studies' forceful conclusions into doubt.

Given the highly non-normal distributions of the factor shares, we do not use non-linear procedures like the Delta method in this chapter.[19] For the disaggregated cost estimates in this chapter, we present an alternative informal method of inference. While, like existing studies, we do not apply confidence intervals to the estimates of the coefficients, we do indicate how the elasticities vary across the sample of firms. Other research, by reporting one summary statistic, discards possibly the most valuable benefit of the translog model over CES models: variation across the sample. Besides, it is in some cases more informative to find the elasticity is positive for 95% of the firms in the sample than to know what the average is. Furthermore, elasticity estimates might be precise at the centre of the data, but could be badly behaved at other percentiles. Therefore, the elasticity value for each firm is calculated using the coefficient estimates and each firm's input quantity. We will therefore have a distribution of elasticities. If an elasticity is positive down to the 5th percentile, we can say 95% of firms have positive elasticities. If an elasticity is negative up to the 95th percentile, we can indicate 95% of firms have a negative elasticity.[20]

4.4 Preliminary investigations

The main purpose of this section is to discuss some of the implications of using different wage series, arguing that the wage series which adjusts for firm size is the best. We will motivate why we will drop the Sales/Clerical group from the disaggregated estimates, thus using only four of the five available occupation types. We also discuss and interpret our finding of non-constant returns to scale.

[19] We will use it in chapter 5.

[20] This is not an option for the model presented in chapter 5, as that model assumes constant returns to scale for a representative firm in the economy. All "representative firms" have the same factor share. It would therefore be inappropriate to use such a method in that context.

4.4.1 Results without adjusting wages for firm size

Estimates of (55) using wage_{all} and wage_{some} are completely unsatisfactory. With the exception of the value added variables, there are often no significant coefficients and the estimates yield many high positive own-price elasticities. One example based on wage_{some}, which happens to have a few significant coefficients, is shown in Appendices 9 and 10. Initial estimations, besides the symmetry conditions, do not impose any restrictions on the technology or those implied by profit maximising behaviour, such as price homogeneity. They are usually rejected or "accepted" at uncomfortably low p-values.

The results are unsatisfactory, but it is not clear whether this is poor wage data or an econometric problem. This section will investigate one potential econometric problem, namely multicollinearity. Although this is usually of minor concern to most researchers, the analysis will show that the problem is severe in this study. Nonetheless, the rest of this section will also show that the poor results are not attributable to multicollinearity.

In the preliminary regressions, the coefficients are not robust to the choice of control variable or to slight changes in the subsample, which is symptomatic of the twin problems of multicollinearity and micronumerosity (Gujarati, 1995). Generally, the variance on the coefficient of variable k is higher if (Greene, 2003):

1. the overall regression fit is low
2. the variance of that variable is low
3. the R^2 in the regression of variable k on the other explanatory variables is high (R_k^2)

For the cost equation in Appendix 9, the root mean square error is 0.65 and the R^2 is 0.76. Although not ideal, this is not a serious problem. Depending on the wage definition used, there are a maximum of only 36 different wage levels for each occupation, so there may be micronumerosity. This might favour using wage_{all} over wage_{some}, but makes little difference in practice. The variance inflation factor $VIF = \frac{1}{1-R_k^2}$ measures the third and most common source: multicollinearity. Multicollinearity is to be expected, given the way the wage variables are constructed by a common set of characteristics. Furthermore, translog estimates, particularly with a high number of factors, are especially vulnerable to multicollinearity. VIF values are not easily derived after systems estimation, but a single equation value is informative, and as shown can be used as a diagnosis tool. The illustrative statistics in Appendix 11 are staggering, even after accounting for the fact many of the terms are squared terms. They show the top ten statistics and the average. The maximum VIF is 211473. The average of 18702 includes the control variables, which generally have values of less than 2.

One way to mitigate multicollinearity is to drop higher order terms in the factors, which is tantamount to a Cobb Douglas specification, possibly with higher order value added terms (as found in Greene (2003)). While not useful for calculating elasticities, estimating this simpler model provides a useful diagnosis tool: estimates succeed in mitigating multicollinearity, as the average VIF falls to 1.84. However, these estimates yield negative shares for some factors and measures of returns to scale are still implausibly high, having a median value of (also) 1.84 for a quadratic specification and 2.75 for a (significant) cubic specification. It is clear that multicollinearity, although serious, is not the cause of the poor results, so it remains to be seen whether adjustments for firm size yield improvements.

4.4.2 Preliminary results adjusting wages for firm size

Single equation estimates using firm-size adjusted wages accept price homogeneity and homotheticity while those that used other wage measures reject both restrictions. Using Wage_{size} for systems estimation alters the estimates of the technology in the way predicted in chapter 3 and produces a meaningful improvement in the results. As shown in Appendices 12 and 13, median own-price elasticity estimates are now all negative except for the sales/clerical value of 0.41. We compare returns to scale calculated here with those calculated based on estimates that don't use firm-size adjusted wages. Appendix 14 shows how the previous average of 2 falls to an average of 1.6. This latter value is still quite high but the figure in Appendix 15 shows the values do approach unity as firms get big in terms of value added.

4.4.3 Interpretation of increasing returns

Even with the improved wage data, which does reduce the measured returns, we still have predicted returns to scale well above unity. These were also found by Guarda (2000) (most studies impose constant returns). If we are to interpret the firm-level data as reflective of the broader manufacturing industry, what could explain this? Part of the phenomenon could be technological; South Africa is a small market and it is a stylised fact that many industries are not exploiting potential economies of scale. Another explanation for high returns to scale is the falling cost of capital for bigger firms, driven by the risk-adjustments imposed. This uncovers a tension between cost and production estimates. Returns to scale are supposed to be a technological phenomenon in a world of exogenous factor prices, yet bigger firms do enjoy lower costs of capital so they do have lower relative costs.

This introduces theoretical difficulties because firms producing below their optimal scale of output should merge. Put differently, having firms below their minimum efficient scale is inconsistent with long run perfectly competitive equilibrium. In a panel data setting, we might explain part of this phenomenon by survivorship bias: firms tend to start out small and either exit or grow and become more efficient. Consistent with this, we have a cross sectional correlation of 0.38 between the logarithm of firm age and the logarithm of turnover.

Romer (1986) and others develop theoretical models where, at a firm-level, returns to scale are constant but, because of knowledge spillovers to other firms from investment, industry returns are increasing. An alternative explanation is that firms are near the optimal sizes for their industries, and that they experience at least locally constant returns to scale. We will need to rely on this interpretation when calculating uncompensated labour demand elasticities, assuming any induced output changes are small. If this assumption is invalid and returns to scale genuinely are not constant, then the adjustments for output will not be valid.

While many authors directly impose constant returns to scale in their estimates (Chung, 1994), mainly to improve efficiency, we do not do so here. We found that imposing such restrictions, together with all the other restrictions, tends to straitjacket the data and produces nonsensical results. In general, we impose restrictions only if they are not rejected by the data.

4.5 Disaggregated elasticities

This section presents the results of the chapter for the disaggregated inputs. We perform a brief diagnostic on the final regressions before turning to the parameters of interest, the elasticities. We present Allen elasticities of substitution before proceeding to constant output labour demand. Thereafter, we mention the material implications of adjusting for non-constant output. Our main findings are that capital and all occupations are p-substitutes: a rise in the cost of any labour type relative to capital will lead to a fall in relative employment levels. We also find most pairs of occupations are p-complements, with the exception of artisanal and unskilled labour.

4.5.1 Final cost function and share estimates

Having decided to employ wage_{size}, we proceed to estimate the system (71) together with the stochastic version of cost function (55). We impose all the restrictions implied by the theoretical equality of coefficients in the share equations and cost function, with the exception of a_i, which is the constant in share equation s_i. This is because the equations may still suffer from measurement error and other specification issues. Many of the biases of these imperfections are deposited on the constant (Wooldridge, 2002), so restricting these catchments for error would spill the biases throughout the system. The constants are therefore left free and the relevant elasticities are calculated from the cost equation.

In addition, the sales/clerical factor is dropped. This naturally assumes separability of the form (35). Of all the factors, this is the one one should be most comfortable dropping. After all, it is hard to believe that the number of sales people or clerks a company employs will have any impact on the relationship between other factors, especially the production workers on the factory floor. The own-price elasticity is persistently positive, as was shown in Appendix 13 for example. Part of the reason for the bad specification is that this is quite a diverse group in terms of skill-level, so wages are more likely to be inaccurate in this occupation. Also, the role of this diverse group varies more than usual across firms, so the control variables are less able to refine it. Furthermore, in systems estimation, errors in one equation transmit themselves to other parts of the system, despite this being mitigated as explained in the previous paragraph. Therefore, the damage to other results from including the Sales/Clerical occupation is most likely greater than any damage from excluding it. The final disaggregated cost equation is in table 4.

Presenting the share equations would reveal very little additional information. However, basic diagnostics for the whole system are presented in Appendix 16. Notably, all the share equations are highly significant. These statistics are however unreliable because of non-normally distributed residuals. This is no surprise given the earlier discussion and the statistics are shown in Appendix 17.

Contrary to the single equation results, more precise systems estimates reject the assumption of homotheticity, so the constraints are no longer imposed. There are two possible explanations for this. One is that the wage adjustment is not accurate enough and poor data are still causing false rejections of homotheticity. In particular, the firm-size effects on wages imposed in construction may not be large enough. Söderbom & Teal (2004) produce firm-size effect estimates for African firms of up to 0.15, which are more than twice the average we took from Bhorat & Lundall (2002). Another is that factor shares are genuinely a function of output. As discussed earlier, bigger firms have cheaper capital and therefore may employ more of it or it could be a genuine technological feature. We fail to reject the hypothesis that all $B_{ij} = 0$. Overall, however, the regression fit is good, with a pseudo-R^2 of 0.85 for the cost equation.

The variables controlling for firm specific effects are entered in levels and not interacted with

Table 4: Cost Function Parameter Estimates

Dependent variable: Cost Variable	coefficient	p	Variable	coefficient	p
Constant	4.11	0.04	ind2	0.20	0.41
Capital	0.25	0.73	ind3	0.50	0.05
Man/Prof	0.27	0.44	ind4	-0.25	0.19
Skil/Art	0.08	0.72	ind5	-0.05	0.79
Semi	0.17	0.73	ind6	0.42	0.15
Un	0.25	0.37	ind7	0.11	0.63
value added	0.29	0.00	ind8	0.07	0.81
0.5*Capital^2	-0.33	0.07	ind9	-0.31	0.05
Capital*Man/Prof	0.06	0.28	loc2	0.21	0.43
Capital*Skil/Art	0.08	0.20	loc3	-0.33	0.08
Capital*Semi	0.15	0.17	loc4	-0.29	0.12
Capital*Un	0.05	0.48	loc5	-0.75	0.00
0.5*Man/Prof^2	0.03	0.41	loc6	0.74	0.08
Man/Prof*Skil/Art	-0.03	0.08	loc7	0.70	0.04
Man/Prof*Semi	-0.03	0.43	loc8	-0.23	0.58
Man/Prof*Un	-0.03	0.34	loc9	-0.29	0.11
0.5*Skil/Art^2	0.02	0.37	exports / output %	0.24	0.25
Skil/Art*Semi	-0.07	0.05	raw materials / cost %	0.01	0.00
Skil/Art*Un	0.01	0.84	recruitment ease Man/Prof	0.1	0.05
0.5*Semi^2	0.01	0.94	recruitment ease Sale/Cler	-0.05	0.23
Semi*Un	-0.05	0.26	recruitment ease Skil/Art	-0.07	0.11
0.5*Un^2	0.03	0.58	recruitment ease Semi	0.01	0.82
0.5*(value added)^2	0.13	0.00	recruitment ease Un	0.02	0.81
(value added)*Cap	0.01	0.78	training expenditure	0.00	0.01
(value added)*Man/Prof	-0.02	0.00	market conditions index	-0.01	0.17
(value added)*Skil/Art	0.00	0.96	firm size > 50 employees	0.37	0.00
(value added)*Semi	0.01	0.43	computer investment / output %	-3.33	0.00
(value added)*Un	0.00	0.76	ownermanaged	-0.61	0.00
Observations	307		productivity dissatisfaction	0.052	0.02
"R^2"	0.85		collective bargaining	0.00	0.96
Homotheticity		0.02	firm age	0.04	0.09
Joint significance of B_{ij}		0.31	cap/lab ratio indicator	1.40	0.00

Table 5: Disaggregated Elasticities of Substitution

$\sigma_{ij} = \sigma_{ji}$		j				
		Capital	Man/Prof	Skil/Art	Semi	Un
	Capital	-1.62*	2.19*	2.91*	2.73*	1.74*
	Man/Prof	2.19*	-5.96	-5.77	-1.46	-2.04
i	Skil/Art	2.91*	-5.77	-7.53	-7.28*	1.79*
	Semi	2.73*	-1.46	-7.28*	-5.78*	-2.44*
	Un	1.74*	-2.04	1.79*	-2.44*	-5.94*
* indicates sign consistent for at least 95% of firms in sample						

any of the input prices.[21] As they are not of direct interest, we keep the discussion brief. Of the included variables that are significant, the indicator of raw materials as a percentage of costs is highly positively significant, suggesting firms using a large component of their inputs might be less efficient. Firms with older equipment and which themselves are older also tend to have high costs, which might be expected. Firms incurring higher training expenditure tend to have higher costs, but this can be a natural correlation for big firms. The firm size dummy is also significant. We find the dummy for owner managed firms, which we take to be those with only one manager, is significantly negative. In the production estimates, this variable is significantly positive. In both cases, this suggests such firms are more efficient. Firms that invest more in computers tend to be those with lower costs while the indicator of the capital labour ratio is positive, suggesting higher technology firms are not necessarily more efficient. This study is not intended to address issues of efficiency in firms so we take these correlations as mild indicators only.

4.5.2 Allen elasticities of substitution

The Allen Elasticities of substitution are calculated for each firm using (66) and (68) and the median values are presented in table 5.

We indicate with an asterisk values that have that sign for at least 95% of firms in the sample.[22] A positive coefficient denotes a pair of factors are p-substitutes. A fall in the cost of one factor relative to the cost of the other will lead to a rise in its relative quantity. A negative coefficient denotes a pair of p-complements.

The table suggests capital is a p-substitute with all occupations, as shown in the first row/column. A relative fall in the cost of capital will lead to a fall in employment relative to utilization of capital.

[21] Some variables were included in the equations for each factor; for example, the ease of recruiting managers/professionals was included in the share equation for that occupation. Strictly speaking, this corresponds to an additional interaction term between Man/Prof and this control in the cost equation.

[22] This was discussed in section 4.3.3.

Table 6: Disaggregated Compensated Elasticities of Factor Demand

λ_{ij}		j				
		Capital	Man/Prof	Skil/Art	Semi	Un
	Capital	-0.96*	0.18*	0.18*	0.40*	0.19*
	Man/Prof	1.28*	-0.56	-0.32*	-0.20	-0.20
i	Skil/Art	1.77*	-0.42*	-0.56	-0.99*	0.19*
	Semi	1.60*	-0.12*	-0.43*	-0.80*	-0.26*
	Un	1.03*	-0.16*	0.12*	-0.34*	-0.65*
* indicates sign consistent for at least 95% of firms in sample						

This holds across at least 95% of the firms in the sample. There is no indication of the Griliches (1969) capital-skill complementarity hypothesis in these statistics. It does not appear to be the case that capital and less skilled occupations are more substitutable than the more skilled occupations. The least skilled occupation has the lowest value ($\sigma = 1.74$) of all the labour types.

The suggestion that all forms of labour seem roughly equally substitutable for capital suggests separability, which has two methodological implications. First, studies of labour/capital substitution do not incur a great cost by aggregating various forms of heterogeneous labour. Second, should data constraints prevent the use of costs of capital in studies of intra-labour elasticities, omitting capital would not affect the estimates badly.

Most occupations share a common substitute – capital – but are themselves p-complements. This is the focus of the thesis. The finding that, in general, various skill types tend to complement each other, is a key result. The result is important because it uses multiple inputs to distinguish from two-factor studies like those of Edwards (2003), which by construction will find skilled and unskilled labour to be substitutes. While the previous paragraph suggested some simplifications to the model need not be overly damaging, only using two factors including a labour aggregate can be very misleading.

The values imply that wage restraint by one occupation, by allowing relative wages to fall relative to the cost of capital, would increase employment of that occupation and the other occupations. Not all skill types are p-complements, however, which we explore in depth with the compensated labour demand elasticities.

4.5.3 Compensated elasticities of factor demand

Table 6 presents the median conditional elasticities of labour demand, which are calculated using equations (65) and (67).

All own-price elasticities are negative, although not always across 95% of firms in the sample. The fact that these are for constant output elasticities verifies that this is consistent with cost minimizing behaviour and that the negative effect is not artificially created by a scale effect. The fact that the own-elasticity is not negative for 100% of firms suggests some of the predicted factor shares are outside of a firm's feasible production set (see Berndt & Christensen, 1973b). Holding output constant, a 1% rise in the unskilled wage will lead to a 0.65% fall in unskilled employment. The results suggest any recent rises in wages, particularly at the lower end of the skill spectrum, have contributed to poor employment performance.

The results also imply that, in general, a rise in the wage of one labour type has had negative employment consequences for the other labour types. For example, a 1% rise in semi-skilled wages would lead to a 0.34% fall in unskilled employment. The exception to this finding is skilled/artisanal workers, where a 1% rise in skilled/artisan wages would lead to a 1.2% rise in unskilled employment, holding output constant. This demonstrates the utility of disaggregation.

4.5.4 Uncompensated elasticities of factor demand

The results so far do not allow for output effects. On the assumption of locally constant returns, we can take these into account. Inspection of equations (69) and (70) shows the scale or output effect is $-s_j|\eta|$. With the exception of capital, the factor shares are all quite small (see Appendix 7 for the average actual shares and their distribution across firms). This suggests that the scale effect will tend to be quite small for disaggregated occupations.

To allow for scale effects, we need measures of $|\eta|$. Selvanathan & Selvanathan (2003) produce estimates for a variety of industries in manufacturing and other sectors. Two manufacturing industries that overlap with the industry definitions in our dataset our clothing and furniture. The elasticity value for clothing is $|\eta| = 0.423$, which is relatively inelastic. Furniture demand is the most elastic at $|\eta| = 0.947$. When we allow for output effects in these industries, not one of the signs is changed from the uncompensated elasticities. This includes capital, which has a large factor share. Because the numbers are not materially different to the constant output elasticities, we leave the results in Appendix 18.

To allow for extreme elasticities, we consider the case of exporters, specifically all those manufacturing firms who export more than 20% of their output. Behar & Edwards (2004) suggest the elasticity of demand for manufactured exports could be as high as $|\eta| = 6$. In this case, it is not surprising that the output effect dominates the substitution effect, such that every element in the table in Appendix 18 is negative.[23]

[23] The larger the predicted output effect, the less "local" the assumption of locally constant returns becomes, which makes the assumption less appropriate.

Table 7: Disaggregated Uncompensated Elasticities of Factor Demand

λ_{ij}		j				
		Capital	Man/Prof	Skil/Art	Semi	Un
	Capital	-1.25*	0.14*	0.15*	0.32*	0.14*
	Man/Prof	0.99*	-0.60	-0.36*	-0.28*	-0.25*
i	Skil/Art	1.48*	-0.46*	-0.59	-1.07*	0.13*
	Semi	1.3*	-0.16*	-0.45*	-0.88*	-0.31*
	Un	0.74*	-0.19*	0.08*	-0.42*	-0.70*
* indicates sign consistent for at least 95% of firms in sample						

Finally, we take a naive average of the industry elasticities in Selvanathan & Selvanathan (2003), although they are not all straightforwardly classified as being in the manufacturing sector, to obtain an average elasticity of $|\eta| = 0.5$ for manufacturing. The corresponding uncompensated elasticities are presented in table 7.

Unsurprisingly, the own-elasticities are more negative than in table 6, though not by much. All the signs from the uncompensated elasticities are preserved. This is particularly important in the case of labour and capital, where the sign is positive. This means the substitution effect is not outweighed by the output effect. In other words, despite leading to a rise in manufacturing output, a fall in the price of capital would lead to a fall in demand for all labour types. Furthermore, we still find that both managers/professionals and semi-skilled workers are p-complements with unskilled workers: a fall in the price of either skill type would lead to a rise in unskilled employment. This is no surprise as a negative compensated elasticity implies a negative uncompensated elasticity. The Skilled/Artisanal and Unskilled groups are p-substitutes: a fall in the price of skilled/artisanal workers would lead to a fall in unskilled employment, despite positive output effects.

Having revealed some of the heterogeneity presented by disaggregated estimates, we proceed to aggregated estimates in the next section.

4.6 Aggregated elasticities

This section presents measures for capital, more skilled labour and less skilled labour. Aggregating input quantities can be straightforward in a cross section - we just add them - but is not as easy for input prices. We choose instead to estimate with the disaggregated input prices and shares but impose the separability conditions discussed in chapter 2 in the form of equation (35). Next, we present the aggregate Allen elasticities of substitution. We apply the method from section 2.8.2 for finding factor demand elasticities in both constant output and non-constant output settings.

4.6.1 Aggregating wages

The heterogeneity of results between skill types would suggest aggregation would be inappropriate. However, the elasticities of substitution between capital and the labour types seem close, which would imply separability of capital from the labour types (cf. (41)). Furthermore, many simulation based or Computable General Equilibrium (CGE) models could benefit from a single elasticity between skilled and unskilled labour.

A formal testing procedure would stop at our failure to reject the null hypothesis that $B_{ij} = 0$ $\forall i, j$ in table 4, which is a failure to reject so-called strong global separability of all inputs (this is a feature of the Cobb Douglas Function). In other words, the testing procedure would stop before investigating the separability of particular sets of inputs. For our purposes, it legitimizes the aggregation of labour types.

We aggregate the Managerial/Professional and Skilled/Artisanal groups into more skilled labour and the Semi-skilled and Unskilled groups into less skilled labour. We omit the Sales/Clerical occupation. The distinction we use differs from the white collar / blue collar division available in most datasets. While aggregation of input quantities is intuitively straightforward, at least in a cross-sectional sense, aggregation of prices is not. Researchers using a dataset that contains "blue collar wages" often take the aggregation procedure at face value. Just because the Berndt & Christensen (1973ab) and Lau (1969) results imply the existence of a valid price index, it doesn't solve the problem of how best to perform the aggregation.

This problem is made explicit in our dataset, where we have the disaggregated wages. One might conjecture that an average of the prices, weighted in some way by their relative shares would be appropriate. This corresponds to the conditions for separability laid out in Berndt & Christensen (ibid.) and reproduced from chapter 2:

$$s_i B_{jk} - s_j B_{ik} = 0 \ \forall k : k \neq i, k \neq j \tag{72}$$

We operationalise our separability assumptions by running the same *disaggregated* regression as in the previous section, but by imposing these restrictions.[24] For simplicity, we do not calculate elasticities for each firm but use the sample average for the restrictions and for the elasticity calculations. We impose restrictions such that our cost function[25] $C = G(w_1, w_2, w_3, w_4, w_5, q)$ can be written as

[24] An alternative would be to impose non-linear restrictions of the form shown in chapter 2, but this would require an entirely different estimation procedure.

[25] By continuing to drop the Sales/Clerical group and omit raw materials, we continue to assume they are also separable, but this is not shown.

Table 8: Cost function parameter estimates with separability imposed

Dependent variable: Cost		
Variable	Coefficient	p-value
Constant		
Capital	0.09	0.90
Man/Prof	0.26	0.42
Skil/Art	0.15	0.42
Semi	0.30	0.46
Un	0.20	0.41
value added	0.28	0.00
0.5*Capital^2	-0.27	0.10
Capital*Man/Prof	0.06	0.19
Capital*Skil/Art	0.05	0.19
Capital*Semi	0.10	0.17
Capital*Un	0.07	0.17
0.5*Man/Prof^2	0.03	0.24
Man/Prof*Skil/Art	-0.03	0.09
Man/Prof*Semi	-0.03	0.06
Man/Prof*Un	-0.02	0.06
0.5*Skil/Art^2	0.03	0.12
Skil/Art*Semi	-0.03	0.06
Skil/Art*Un	-0.02	0.06
0.5*Semi^2	0.02	0.74
Semi*Un	-0.05	0.26
0.5*Un^2	0.03	0.51
0.5*(value added)^2	0.13	0.00
(value added)*Cap	0.01	0.47
(value added)*Man/Prof	-0.02	0.00
(value added)*Skil/Art	-0.004	0.45
(value added)*Semi	0.004	0.62
(value added)*Un	0.005	0.48
Observations		307
"R^2"		0.85
Homotheticity		0.00
Joint significance of $B_{ij} = 0$		0.09

$C = G^P(W_M, W_L, W_5, q)$, where W_M is an aggregate of w_1 and w_2 (managers/professionals and skilled/artisanal workers), W_L is an aggregate of w_3 and w_4 (semi-skilled and unskilled workers), $W_5 = w_5$ is capital (for notational consistency) and q is value added. W_M is more skilled labour while W_L is less skilled labour. There are six restrictions like (72) in total, three for when i, j are w_1, w_2 with respect to the other three inputs and three for when i, j are w_3, w_4 with respect to the other three inputs. The regression results are presented in table 8 and basic diagnostics for the system can be found in Appendix 19.

The test of homotheticity is rejected at 1%. Unlike the previous estimates, we find the B_{ij} jointly significant at 10%. The beta coefficients corresponding to the interactions between more and less skilled labour are individually significant. Like the regression in table 4, the predicted

Table 9: Aggregated Elasticities of Substitution

σ_{IJ}	
Factor pairing	elasticity
Capital / More	2.40
Capital / Less	2.19
More / Less	-1.71

shares and residuals have non-normally distributed residuals, so Wald tests like these should be treated circumspectly. We omit to present the controls because they add no new information.

4.6.2 Allen elasticities of substitution

The Allen elasticities were confirmed to be exactly equal for separable inputs, for example $\sigma_{15} = \sigma_{25} = \sigma_{M5}$ We present the three aggregated elasticities in table 9.

As one might expect, the elasticity between capital and each of the skill types falls between the elasticities between capital and each of the disaggregated components. Both types of labour are still found to be p-substitutes for capital. A rise in the cost of labour relative to capital would lead to a relative fall in its employment quantity. We also find that capital skill complementarity does not hold. The explanation lies partially in the disaggregated estimates from earlier, where the semi-skilled and skilled/artisanal workers were highly substitutable for capital.

Preserving the general sign pattern from before, we find more skilled and less skilled labour are p-complements. The generally lower absolute values relative to the disaggregated estimates are to be expected as more aggregated inputs are less easy to substitute for one another. The results suggest wage compression artificially created by bargaining institutions at the low end of the skill spectrum, which raises less skilled wages relative to more skilled wages, leads to a relative fall in less skilled employment.

4.6.3 Compensated elasticities of factor demand

We can draw on section 2.8.2 to calculate aggregated labour demand elasticities. For all cross-elasticities with respect to the price of capital, which is an aggregate of only one input, it is clear that, using the notation from the previous section, $s_5\sigma_{15} = s_5\sigma_{25} = s_5\sigma_{M5}$, so $\bar{\lambda}_{15} = \bar{\lambda}_{25} = \bar{\lambda}_{M5}$. This is the special case of theorem 2, where an aggregate consists of only one input here. Similarly, for w_3, w_4, which are in aggregate W_L, $\bar{\lambda}_{L5} = s_5\sigma_{L5}$. Applying the theorem for changes in the price of more skilled labour W_M, we have $\bar{\lambda}_{IM} = S_M\sigma_{IM}$ $(I \neq M)$. Similarly, $\bar{\lambda}_{IL} = S_L\sigma_{IL}$ $(I \neq L)$. We use proposition 1 to get $\bar{\lambda}_{KK} = -\bar{\lambda}_{KL} - \bar{\lambda}_{KM}$; the procedures for the two labour types are analogous. Table 10 reports the compensated elasticities.

Table 10: Agregated Compensated Elasticities of Factor Demand

λ_{IJ}		J		
		Capital	More	Less
	Capital	-0.94	0.35	0.58
I	More	1.26	-0.80	-0.45
	Less	1.14	-0.25	-0.90

Table 11: Aggregated Uncompensated Elasticities of Factor Demand: exporters

λ_{IJ}		j		
		Capital	More	Less
i	Capital	-4.23	-0.49	-0.92
	More	-2.04	-1.70	-1.95
	Less	-2.15	-1.10	-2.40

We note that the estimate for capital is very close to that in the disaggregated estimates. The labour elasticities are high compared to those reported in Hamermesh (1993) but, in concurring with his review, more skilled labour demand is less elastic than less skilled labour demand. These aggregate own-elasticities suggest wage push would have contributed to decreased employment levels. In aggregate, we find that the two labour types are p-complements: a fall in the price of skilled labour would lead to a rise in unskilled employment.

However, capital and both labour types are p-substitutes. While a fall in the price of skilled labour leads to a rise in unskilled employment, it leads to a fall in demand for capital. Furthermore, a fall in the price of capital will lead to a fall in demand for labour. The coefficients exceed unity for both labour types, so the employment effects would be relatively large. While $\bar{\lambda}_{IJ} < 0 \rightarrow \lambda_{IJ} < 0$, it remains to be seen whether the p-substitutability of capital and labour still holds after allowing for output effects.

4.6.4 Uncompensated elasticities of factor demand

We allow for output changes by using the aggregate analogue of λ_{ij}, assuming locally constant returns. In a high $|\eta|$ case, where we compute the elasticities for the subsample of firms who export more than 20% of their output, it is not surprising that all coefficients are now negative. A fall in the price of any factor would lead to a rise in demand for any of the others, as shown in table 11.

Despite the very large product demand elasticity imposed ($|\eta| = 6$), the factor demand elasticities elasticities are not extraordinarily large, with the possible exception of the own-elasticity for capital. Taking an average elasticity of $|\eta| = 0.5$ across all manufacturers, we still preserve the signs from the compensated elasticities. This is shown in table 12. The own-elasticities are immoderately more negative and the complementarity between labour types is enhanced. Most

Table 12: Aggregated Uncompensated Elasticities of Factor Demand: manufacturing

λ_{IJ}		j		
		Capital	More	Less
i	Capital	-1.20	0.28	0.45
	More	0.99	-0.88	-0.58
	Less	0.89	-0.32	-1.03

importantly, despite potential increases in output, a fall in the cost of capital would not lead to a rise in employment.

4.7 Conclusion and policy implications

This chapter has used the theory introduced in chapter 2 and the data constructed in chapter 3 to estimate Allen elasticities of substitution. It also presents results of conditional (constant-output) and unconditional factor demand elasticities. To our knowledge, we are the first to use translog functions to present unconditional labour demand and to apply the implications of separability to aggregate elasticities. While the shortage of establishment-level wage data prevails for developing countries, this study, along with that of Teal (2000), presents an option for supplementing it with data from other sources.

Both the disaggregated and aggregated estimates suggest occupations share a common substitute – capital. This generally holds despite allowing for output effects. Our results therefore support the concern that policies which lower the cost of capital relative to labour will lead to a fall in employment in favour of capital acquisition in manufacturing. This microeconomic contribution based on manufacturing evidence suggests that, despite potential scale effects, calls for a lowering of interest rates by trade unions may be against the interests of their constituents.

Taken literally, the corollary to these implications is that, to raise employment, there should be a rise in interest rates! This microeconomic contribution is not designed to confront standard macroeconomic transmission arguments for output expansion under the assumption of insufficient aggregate demand, so we by no means wish to make such a strong claim. We do not consider any productivity benefits from investment. The direct productivity effects of an increase in the quantity of capital on various types of labour is discussed in chapter 5. Furthermore, the potential effects of technological change are presented later in the thesis. However, the results at the very least suggest some caution is called for by those advancing lower interest rates as a solution to mass unemployment. Furthermore, other policies besides interest rates can affect the cost of capital, as discussed in section 3.4. Measures to increase the cost of capital relative to labour, such as ending generous depreciation allowances on machinery (see Fedderke et. al., 2001) have been implemented.

On the basis of these results, such measures would have positive employment consequences.

We do not find support for capital skill complementarity so we do not predict the employment effects to be worse for those with fewer skills. The own-price labour demand elasticities suggest those with lower skills may suffer disproportionately from any increases in the cost of labour. Aggregate estimates show the own-wage elasticity of demand is higher for less skilled workers than for more skilled workers. Allowing for output effects on the assumption of locally constant returns makes the elasticity as high as unity for manufacturing as a whole.

Exploiting the opportunity to disaggregate labour into four occupations, we reveal that unskilled labour is a p-complement with semi-skilled workers but a p-substitute with skilled/artisanal labour. These results hold for the vast majority of firms in the sample. With the exception of exporters, our results hold for uncompensated cross-elasticities. This means that a rise in the price of unskilled labour would decrease employment for both unskilled and semi-skilled workers as firms would switch to managerial/professional and/or skilled/artisanal workers (and/or capital). The same applies to a rise in the price of semi-skilled labour. One implication is that wage restraint by semi-skilled labour would have positive employment benefits for unskilled workers and vice versa. This introduces the possibility of a coordination problem. To what extent it is addressed by trade unions representing both unskilled and semi-skilled workers is a matter of further empirical and theoretical research.

Finally, and most importantly in terms of the thesis as a whole, our aggregate studies suggest more skilled and less skilled labour are p-complements.

5 Does training benefit those who do not get any? Elasticities of complementarity and factor price

5.1 Introduction

"Government's ambition to grow [the] manufacturing base risks being stillborn unless the country addresses a worsening skills crisis." - Paton (2003:18)

The importance of skills to manufacturing and other industries is a topical issue in developing countries, as evidenced by the above quote from a lead article in the *Financial Mail.* The statement that artisans, for example welders or tool-makers, are "essential to every aspect of manufacturing ... production" and that their shortage will "severely hinder ... ability to deliver on ... capital investment projects" (pg 18), articulates the widely held belief that artisans and other occupations are complements in production; that there are limited opportunities for substitution by other occupations and that the main effect of shortages is to lower output and thereby demand for all factors.

Together with these apparent shortages, there is often joblessness. Much of South Africa's unemployment appears to be structural in that an oversupply of unskilled labour exists alongside estimates of as many as 500 000 vacancies for skilled workers (The Economist, 2004). These unfilled vacancies are evidence of skills shortages constraining output: filling them would allow production and employment to rise for all occupations.

Both these observed features of the economy imply that skilled and unskilled labour are (Hicks) complements and not substitutes in production. If they are complements, a rise in the supply of skilled workers (in the face of excess demand for skilled labour) has benefits for all occupations, including the unskilled. If skilled and unskilled labour are substitutes, then unskilled labour will be worse off if there is a rise in the supply of skilled labour. In particular, if vacancies for skilled workers are being partially filled by less skilled workers, then improved availability of the first-choice factor will result in these suboptimal substitutes losing out.

Increasing the skills of the workforce is regarded as a key requirement for reducing wage inequality (Bhorat et al., 2001). People who acquire such skills or training are likely to earn productivity linked wage increases (Fallon & Lucas, 1998), but this raises the question of what will happen to those who remain unskilled? If skilled and unskilled labour are substitutes, training aimed at a limited subsection of the labour force may actually worsen wage inequality.

The South African case is an interesting one to study because, in addition to the labour market characteristics described, serious attempts are being made to increase the supply of skills: In order to encourage firms to train their workers, the South African Skills Development Act of 1998 introduced a system where firms incur a tax on payroll, which is reduced if they equip workers with skills in cooperation with Sector Education Training Authorities (SETAs). Approximately 46 000 people were enrolled in such programmes at the start of 2004 and the number is rising (Mdladlana, 2004). It is important to gauge whether the types of skills being produced are those most conducive to growth and most beneficial to the unskilled.

Chapter 2 introduced the Hicks (1970) elasticity of complementarity (HEC), which measures the percentage change in the ratio of endogenous factor prices to an exogenous change in their relative quantities. Similarly, the cross-elasticities of factor price measure the percentage change in a factor price in response to an exogenous change in another factor's quantity. If the effect is positive, the factors are said to be q-complements. If the effect is negative, the factors are q-substitutes

An important early application was performed by Grant & Hamermesh (1981) to examine the interactions between youths, white women and other workers in the United States. Field (1988) investigates the HEC between free and slave labour. Appelbaum & Kohli (1996) use the HEC between imports and local inputs in the context of a broader study. Vere (2001) estimates the parameters for skilled and unskilled labour over time in Taiwan. The study closest in spirit to ours is by Mak (2000), who studies whether workers with different education levels are q-complements or q-substitutes in Canada.

This chapter uses a translog production function to measure Hicks elasticities of complementarity and cross-elasticities between capital and five labour occupations using the same detailed firm-level data as the last chapter. The relationships of most interest are within the production occupations as this is where the SETAs hope to contribute the most. We also produce estimates using more aggregated groups, dividing the labour force into more skilled and less skilled groups.

All documented empirical work assumes perfectly elastic product demand, which may cause it to find two factors are complements when they are actually substitutes. Therefore, this chapter allows for imperfectly elastic demand. It also allows for the possibility of one factor having rigid wages. Furthermore, it employs the "Delta" method to calculate p-values for the elasticities, which are non-linear functions of the technological parameters.

The key finding is that a rise in the supply of the skilled/artisan occupation will increase unskilled wages while a rise in the supply of semi-skilled workers will reduce unskilled wages. (Skilled/artisanal and unskilled labour are q-complements while semi-skilled and unskilled labour are q-substitutes.) The results are robust to relatively inelastic demand in the product market.

Accounting for rigid unskilled wages preserves the relationship, which suggests a rise in the supply of skilled/artisanal workers would raise unskilled employment while a rise in the supply of semi-skilled labour would lower unskilled employment. The findings are consistent with the view that a shortage of artisans is holding back production and that relieving the shortage will raise demand for unskilled labour. Aggregated results suggest capital complements more skilled labour but not less skilled labour and that a rise in the supply of more skilled labour would raise demand for less skilled labour, but these results are somewhat more susceptible to product demand elasticity and flexible wage assumptions than the disaggregated results are.

This chapter derives the elasticity of factor price in the context of a model for the macroeconomy and shows how it can be used to allow for rigid unskilled wages in section 5.2. Section 5.3 shows how the elasticities are calculated using translog production functions. Section 5.4 describes the firm-level data, identification strategy, specification issues and estimation. In particular, it considers endogeneity bias in the contexts of fully flexible and rigid wages. Section 5.5 presents the disaggregated results, section 5.6 discusses the aggregated estimates and section 5.7 offers concluding comments.

5.2 A macro model of exogenous input quantity changes

This section derives the elasticity of factor price in the case when all wages can adjust freely before introducing an adjustment for the case when the wage of one factor is rigid.

5.2.1 Model with fully flexible wages

We have an economy that uses n exogenously supplied factor inputs, with factor prices adjusting to ensure full employment. Economy-wide output (Y) is determined by factor input quantities (X_i) according to a linearly homogeneous technology utilized by all h representative firms in the economy.

$$Y = f(X_1, ..., X_n) = hf(x_1, ..., x_n) \tag{73}$$

If not in a small open economy, the price (P) received by firms is determined by output.

$$P = P(Y), \frac{\partial P}{\partial Y} < 0 \tag{74}$$

Profit-maximising firms pay each input a wage (w_i) equal to its marginal revenue product, which is a function of the supply of all the inputs in the economy.

$$w_i = P(Y)f_i(X_1, ..., X_n), \quad f_i = \frac{\partial f}{\partial X_i} > 0 \tag{75}$$

A change in factor supply has two effects on wages. First, it changes overall output and hence prices and, second, it changes the marginal rate of technical substitution given by the production technology, as shown respectively by the first and second terms of (76):

$$\frac{dw_i}{dX_j} = \frac{dP}{dY} f_i f_j + P f_{ij}, \text{where } f_{ij} = \frac{\partial^2 f}{\partial X_i \partial X_j}, \ f_{ii} < 0 \tag{76}$$

Converting to elasticity form:

$$\begin{aligned} \frac{d \log w_i}{d \log X_j} &= \frac{X_j f_i f_j}{w_i} \frac{dP}{dY} + \frac{P f_{ij} X_j}{w_i} && (77) \\ &= \frac{X_j f_i f_j P}{w_i f} \frac{1}{\eta} + \frac{f_{ij} f}{f_i f_j} \frac{f_i f_j P X_j}{f w_i}, && (78) \end{aligned}$$

$\eta < 0$ is the elasticity of demand in the product market. The Hicks elasticity of complementarity between factors i and j is:[26]

$$H_{ij} = \frac{f_{ij} f}{f_i f_j} \tag{79}$$

Factor j's share of output is:

$$s_j = \frac{f_j X_j}{f} \tag{80}$$

Together with equation (75), this can be used to rewrite (78) as:

$$\frac{d \log w_i}{d \log X_j} = \hat{\epsilon}_{ij} = s_j \left(H_{ij} - \frac{1}{|\eta|} \right) \tag{81}$$

$\hat{\epsilon}_{ij}$ is the elasticity of factor price. It is the multiple factor equivalent to the own-elasticity given by equation (6) in chapter 2. A rise in the supply of a factor, for example, works through 3 channels. (i) Because output is determined by the supply of factors, a rise in supply of a factor necessarily leads to a rise in output and demand for all other factors. ii) However, this effect is mitigated because a rise in output leads to a fall in product price and hence a fall in factor demand. iii) Furthermore,

[26] This is the same value at economy and firm levels, because $f(X_1, .., X_n) = hf(x_1, ..., x_n)$, $f_i(X_1, ..., X_n) = f_i(x_1, ..., x_n)$ and $f_{ij}(X_1, ..., X_n) = h^{-1} f_{ij}(x_1, ..., x_n)$ for a linearly homogeneous technology.

the nature of the technological relationship between factors means that, holding output constant, $f_{ij} \lesseqgtr 0$ for $n > 2$ and $i \neq j$.

As $\eta \rightarrow \infty$, there are no price effects, yielding the elasticity of factor price ϵ_{ij} when price is constant. This captures effects i) and iii) only (Sato & Koizumi, 1973) and is most suited to small open economies.

$$\epsilon_{ij} = s_j H_{ij} \tag{82}$$

Equations (81) or (82) can be interpreted as the change in factor returns necessary for the economy to generate factor demand equal to the new factor supply; that is, to accommodate the change in supply of one factor and keep demand for all factors equal to their (unchanged) supply. ϵ_{ij} will tend to produce elasticities that are higher than $\hat{\epsilon}_{ij}$ so, unlike other documented work, this study allows for product price effects. Another potential adjustment to ϵ_{ij} accommodates the possibility that the wage of one factor is rigid, as discussed next.

5.2.2 Allowing for rigid wages

We started with the assumption that all factor prices adjust to ensure full employment. Given that wages may be rigid and that we do in practice see unskilled unemployment in South Africa and other developing countries, one can adjust for this using the methods first used by Johnson (1980). In the simple case, where only one factor's wage is completely rigid, one can calculate the effect of an exogenous change in the quantity of another factor on the quantity of that factor. This means the HEC can be used to infer effects on employment rather than wages for a particular factor. Following Grant & Hamermesh (1981), assume all factors' prices are flexible except for unskilled labour, which has wage w_u. All factor quantities are fixed except unskilled labour, which has quantity X_u. In this model, we do not allow for changes in product price and set $P = 1$ on the assumption of perfectly elastic product demand. The marginal productivity conditions are:

$$w_u = f_u(X_u, X_2, ..., X_n) \tag{83}$$

$$w_i = f_i(X_u, X_2, ..., X_n), \quad i = 2, ..., n \tag{84}$$

Differentiating the equations with respect to X_j and solving the resulting system:

$$\frac{dx_u}{dx_j} = \frac{-f_{uj}}{f_{uu}} \tag{85}$$

$$\frac{dw_i}{dx_j} = \frac{-f_{iu}f_{uj} + f_{ij}f_{uu}}{f_{uu}}, \quad i = 2, ..., n \tag{86}$$

Using equation (80), $X_i = \frac{fs_i}{f_i}$ and, by equation (79), $f_{ij} = \frac{f_i f_j H_{ij}}{f}$. Hence:

$$\frac{d\log x_u}{d\log x_j} \equiv \rho_{uj} = \frac{-H_{uj}s_j}{H_{uu}s_u} \tag{87}$$

$$\frac{d\log w_i}{d\log x_j} \equiv \epsilon_{ij}^{\rho} = \frac{s_j\left(-H_{iu}H_{uj} + H_{ij}H_{uu}\right)}{H_{uu}}, \quad i = 2, ..., n \tag{88}$$

As presented in Grant & Hamermesh (1981), equations (87) and (88) demand a burdensome calculation of coefficients and p-values. However, simply using (82) can aid in computation:

$$\rho_{uj} = \frac{\epsilon_{uj}}{-\epsilon_{uu}} \tag{89}$$

$$\epsilon_{ij}^{\rho} = \epsilon_{ij} - \frac{\epsilon_{iu}\epsilon_{uj}}{\epsilon_{uu}} \tag{90}$$

ϵ_{uj}, ϵ_{ij}, ϵ_{uu} and ϵ_{iu} represent what the elasticities would have been if all wages were flexible. To gather some very informal intuition, note the first term on the right hand side of (90) is the conventional elasticity. The second term, in a two input context, can be written as approximately equivalent to $\frac{d\log w_i}{d\log x_u} * \frac{d\log x_u}{d\log w_u} * \frac{d\log w_u}{d\log x_j}$. This product of derivatives, through chain rule arguments, shows how changes in the quantity of a factor would affect unskilled wages (last term) but instead affect unskilled quantities (middle term), which in turn affect the wage of factor i (first term). The intuition for (89) is more straightforward. Informally writing $\rho_{uj} \approx -\frac{d\log w_u}{d\log x_j}\frac{d\log x_u}{d\log w_u}$, we are multiplying the conventional elasticity of factor price by the (two-input) own-elasticity of factor demand such that the price effect becomes a quantity effect. Cancelling the numerator and denominator and removing the negative sign[27] yields the intended result in terms of quantity changes only.

It is therefore possible to infer the effects of a change in the quantity of a factor on the quantity of unskilled labour (equation (89)) or on the prices of other factors, taking unskilled wage rigidity into account (equation (90)). However, before performing any calculations, it is necessary to estimate the relevant technological coefficients.

5.3 Elasticities and translog production functions

To find the elasticities of interest, we need to estimate the parameters of the underlying technology. The translog production function can be viewed as a second order Taylor approximation to an unknown technology - with the coefficients being the first (α_i) and second (β_{ij}) derivatives of the function in the approximation - or as an exact production function in its own right (Denny &

[27] The negative sign preserves the sign when transforming a price effect into a quantity effect. For example, a move predicted to lead to a fall in the wage should lead to a fall in the quantity. The own elasticity is negative and, without the minus sign, would transform the effect into an increase in quantity.

Fuss, 1977). In this chapter, the measure of output (q) is value added and it is a function of six disaggregated inputs. We assume representative firms make random errors in their use of the inputs available to them, resulting in error term ω, which is normally distributed with a mean of zero and a constant variance. The error term must be orthogonal to the inputs, an issue which is discussed in section 5.4.2.

$$\log q = \log \alpha_0 + \sum_i \alpha_i \log x_i + \frac{1}{2}\sum_i \sum_j \beta_{ij} \log x_i \log x_j + \omega \tag{91}$$

Slutsky symmetry conditions $\beta_{ij} = \beta_{ji}$ are imposed in the construction of the variables. The main advantage of the translog function for this study is that there are no assumptions imposed on the elasticities. Furthermore, technological features are not assumed but tested for and, if accepted, imposed on the system. Equation (91) is homogeneous of degree k if:

$$\sum_j \beta_{ij} = \sum_i \beta_{ij} = 0 \text{ and } \sum_i \alpha_i = k \tag{92}$$

If $k = 1$, there are constant returns to scale (Chung, 1994). We saw in section 2.6 that differentiating (91) with respect to $\log x_i$ yields the share of factor i in output:

$$s_i = \alpha_i + \sum_j \beta_{ij} \log x_j \tag{93}$$

It is common to estimate the system of equations (93) to improve efficiency characteristics (Berndt, 1991). However, in the data used for this chapter, factor shares are not available,[28] so the parameters estimated in (91) are used to predict s_i and calculate elasticities in a translog context: observe that $w_i = \frac{q}{x_i} s_i$, so

$$\frac{d \log w_i}{d \log x_j} = \frac{x_j}{w_i} \frac{d}{dx_j} \left(\frac{q}{x_i} s_i \right) \tag{94}$$

Differentiating (91), holding the product price component of q constant and recalling $\frac{dq}{dx_j} = w_j$:

$$\frac{d \log w_i}{d \log x_j} = \frac{x_j}{w_i} \left(\frac{q\beta_{ij}}{x_i x_j} + \frac{w_j s_i}{x_i} \right) = \frac{\beta_{ij}}{s_i} + s_i \left(\frac{w_j x_j}{q} \right) \left(\frac{q}{w_i x_i} \right) \tag{95}$$

Therefore:

$$\epsilon_{ij} = \frac{\beta_{ij}}{s_i} + s_j \tag{96}$$

[28] In the cost chapter, we employed constructed wages, which were needed for the cost function, to construct cost shares. We have opted not to use such data in this chapter.

By (82):

$$H_{ij} = \frac{\beta_{ij}}{s_i s_j} + 1 \tag{97}$$

If all $\beta_{ij} = 0$, we have a Cobb Douglas production function in (91), constant factor shares for each factor in system (93) and $H_{ij} = 1$. Furthermore (Binswanger, 1974a):

$$\epsilon_{ii} = \frac{\beta_{ii}}{s_i} + s_i - 1 \tag{98}$$

$$H_{ii} = \frac{\beta_{ii}}{s_i^2} + 1 - \frac{1}{s_i} \tag{99}$$

The elasticities of factor price presented so far assume a constant price in the context of perfectly elastic product demand. All studies discussed in the introduction to this chapter seem to operate on this assumption but, as indicated in chapter 2, this fails to account for imperfectly elastic product demand. We will indicate why this is important and how we propose to deal with this in the results section. To obtain a measure of the production technology, we use firm-level data.

5.4 Empirical issues

The default stochastic specification (91) assumes firms face exogenous variation in the inputs available to them. In addition, they make errors in the use of the inputs available, which we assume to be normally distributed with a mean of 0 and a constant variance ω. Implicit in this specification is the assumption that the errors are orthogonal to the inputs and other controls for technology. These assumptions are a necessary condition for valid estimation by ordinary least squares (OLS). Stochastic specifications with alternative assumptions and estimation are considered from section 5.4.2 onwards.

5.4.1 Firm-level data for a macro model

The data used was discussed in chapter 3. We have six input quantities, value added[29] and a number of controls. Key variables are capital stock and employment numbers by occupation. The five occupations are non-production workers, divided into Managerial/Professional and Sales/Clerical, and production workers, divided into Skilled/Artisan (technicians, welders), Semi-skilled (machinery operators) and Unskilled (labourers, security guards). Capital stock is available in currency (Rand) values and is adjusted for shift capacity utilization. We have a number of variables that could act as controls and/or instruments - as discussed in section 5.4.2.

[29] As discussed in chapter 2, we performed the tests of $\frac{\alpha_i}{\alpha_j} = \frac{\beta_{ii}}{\beta_{ij}} = \frac{\beta_{ji}}{\beta_{jj}} = \frac{\beta_{ir}}{\beta_{jr}}$ required to find raw materials are weakly (non-additively) separable from the other inputs and consequently adopt value added as our output measure.

Using firm-level manufacturing information for a macro model presents a number of challenges, in particular valid identification. One concern is whether one can use firm-level data to make inferences about the macroeconomy. The crucial factor is the assumption of constant returns in the model and the support for that assumption in our production data, which we find in section 5.5.[30] Constant returns allows us to consider many representative firms and one single firm for the economy interchangeably. This means we can use variations across firms in their inputs to estimate the parameters of a macro production function.[31] In other words, if we were to know the parameters in (91) (that is, if we did not need to estimate them) and if the firms in the economy have identical linearly homogeneous production functions, then we can use equations (96)-(99) to make statements about economy wide effects. Of course, we don't know the parameters and that is why we need to estimate them.

One potential drawback of using manufacturing data is that the firms in the survey may not be sufficiently representative of firms in the broader economy. Using evidence from manufacturing for the whole economy would be invalid if the production technology is fundamentally different in manufacturing compared to other sectors. However, manufacturing is the largest sector in the South African economy (Bhorat & Hodge, 1999), contributing about 20% to GDP (Smith, 2003). It also includes a particularly heterogeneous range of economic activity by international standards, ranging from the beneficiation of primary commodities to relatively service intensive industries (Wood, 1995). This means manufacturing is quite representative of broader economic activity. Consistent estimates for manufacturing should therefore be informative regarding the broader economy. Using the data to make inferences about manufacturing alone would require relaxing the assumption of perfectly exogenous labour supply at the cost of far greater complexity. Being fully aware of these issues, Grant & Hamermesh (1981) also employ cross-sectional manufacturing data. Borjas (1986) yields the same results from estimates for the whole economy and for manufacturing alone in the United States.

Finally, we need to estimate our parameters for the manufacturing firms consistently. While exogeneity of the factors is arguably valid for the economy as a whole, it is not justified at the firm level. This raises the potential for inconsistent estimation of the parameters, an issue discussed at length next.

[30] In chapter 4, our cost data was not consistent with constant returns. This disparity is of interest in itself and its interpretation is briefly discussed in that chapter and in chapter 9.

[31] In the cost chapter, we could use the manufacturing data to make statements about the manufacturing sector or even industries within it.

5.4.2 Endogeneity bias

One issue with estimating production functions is the potential for endogeneity bias. This is of general concern and is not limited to the application in this study. A specification where exogenous inputs generate endogenous output ignores the possibility that inputs are chosen by the firm to maximise profit. If the assumptions of exogenous factor inputs, which are consistent with the production function and elasticity concepts, held perfectly, there would be no problem. Deviations from the assumption have the implication that econometric estimates would be biased. Firm-specific technologies or other firm-specific influences are a source of bias, but we believe there are enough suitable variables to act as controls for these effects reasonably well. Without claiming such controls are a perfect resolution to the problem, this is not the issue discussed here. The issue discussed here is that the inputs are chosen based on output decisions, not the other way around.

Because of the importance many researchers attach to this source of endogeneity, and their taste for Two Stage Least Squares (2SLS) and other instrumentation procedures, considerable space is devoted to it. This section motivates why OLS with proxies is the best option available to us. It begins with a detailed discussion of the problem and possible methods for addressing it. The conceptual differences between using proxies and instruments are presented before performing estimations with both approaches. The very clear conclusion is that 2SLS does not necessarily improve the bias characteristics of the coefficients but grossly amplifies the standard errors. Formal tests, despite having low power, suggest no need for 2SLS in the first place. We also discuss the use of instruments in the special case where we model one wage as rigid and one factor input as endogenous.

The endogeneity problem and potential responses The discussion will be restricted to a Cobb Douglas case for simplicity, but applies equally to more general production functions (Levinsohn & Petrin, 2003). Consider a production function for a firm,

$$q_i = \boldsymbol{\alpha}\mathbf{a}_i + \beta_l l_i + \beta_k k_i + u_i, \tag{100}$$

where q_i, l_i, k_i are output, labour and capital for firm i at time t (we suppress the time subscript). The technology is common across firms in the sense that β_l, β_k are common to all firms, but each firm has a vector of specific features ($\mathbf{a}_i$) that affect the intercept of the production function. For the purposes of this discussion, assume we satisfactorily capture these intercepts through a variety of control variables.

The error term u_i comprises various components unobserved by the econometrician. The first is a factor the producer observes in time to affect his choice of inputs. A common example is the weather. Another, which is more readily applicable to manufacturing, is changes in various market conditions ($\mathbf{m}$) (Griliches & Mairesse, 1995). For simplicity, assume none of these can be observed or controlled for by the econometrician. The error term also comprises errors on the producer's part. He can make mistakes in predicting the nature of external shocks (p) and be unable to adjust his inputs once p is observed. Due to incompetence or institutional inertia, he may not be able achieve the optimal input combination (e). e and p are therefore unrelated to the observed input quantities (Zellner, Kmenta & Dreze, 1966). Attaching coefficients of unity to these last two sources, we have:

$$q_i = \alpha \mathbf{a}_i + \boldsymbol{\beta}_l l_i + \beta_k k_i + (\boldsymbol{\delta}\mathbf{m} + e + p) \tag{101}$$

The producer chooses his inputs according to the conditions in $\mathbf{m}$ so there is a correlation between the error term and some regressors. Griliches & Mairesse (1995) observe a number of responses to this problem in the microeconomic context. The first is denial. Grant & Hamermesh (1981), Johnson (1980), Field (1988), Appelbaum & Kohli (1997) and Mak (2000) make no mention of the issue, nor does Hamermesh (1993) in his empirical review. While many of these authors neglect to state it explicitly, it appears they use either the SURE approach (discussed in the cost chapter) or estimate the production function only, using OLS. It is also arguably the case that this source of bias is limited in a cross sectional setting, as opposed to time series.

Second, people can estimate cost functions. Given that production functions are being estimated for reasons already documented, this is not an option. The third is to assume $\mathbf{m}$ can't be predicted by the producer. If so, the inputs are not conditioned on any component of the error term and hence OLS estimates are not biased (Zellner et al., 1966). Griliches & Mairesse note that, despite this being published in *Econometrica* by respected authors, people feel "guilty" (pg 5) about using this assumption.

The fourth option is to try to control for $\mathbf{m}$ through proxies. The proxy approach has regained impetus recently, with the use of lagged investment (or raw materials) inputs as proxies. The reasoning is that investment is a function of $\mathbf{m}$ (and the capital stock). If this input demand function is invertible, the inclusion of investment therefore captures $\mathbf{m}$ (Levinsohn & Petrin, 2003). Identifying the parameters on the other inputs is simple, but identifying the capital and/or raw material input parameters requires non-parametric techniques. This approach holds much promise, but seems more appropriate for multiple period data. The reason is that, according to Levinsohn

& Petrin, investment is monotonically increasing in **m** only if **m** is stochastically increasing in past values. Monotonicity is required for invertibility. The use of investment as a proxy seems aimed at productivity or other shocks to **m** within firms across time. This monotonic relationship between investment and **m** across firms does not hold and is therefore not appropriate in cross-sectional space.

The fifth approach is two-stage-least-squares (2SLS). Before exploring this in more detail, we mention the sixth option, which is to employ panel data methods using firm- or time-specific fixed effects and 2-period-lagged input values as instruments (Levinsohn & Petrin, 2003). Panel methods are not an option for this study so no further attention is paid to this approach.

Conceptual comparison of using 2SLS and controlling for m Returning to the possibility of using 2SLS, the key issue is whether instruments for the observed endogenous variables are appropriate or whether the data available are more suitable as proxies/controls for the unobservables causing the endogeneity bias. The principal reference for the ensuing discussion is Wooldridge (2002). As mentioned, the source of endogeneity bias is the correlation between a factor input and the error term. Instrumenting for l entails regressing l on the set of relevant exogenous variables $\mathbf{w}$, obtaining $\hat{l}$ and using $\hat{l}$ in equation (101). $\mathbf{w}$ must include at least one variable w not in the structural equation that is partially correlated with l net of the other variables. In fact, there must be at least one w for each endogenous variable. In this study, there are far too many inputs and interacted inputs for this requirement to be met, so the aim is to find instruments for those thought "most" endogenous or perhaps to instrument for first order variables only, leaving the higher order ones uninstrumented.

The consistency of the coefficient estimates under IV estimation critically relies on the assumption that the w variables are not correlated with the error ($u\prime$) in the instrumented equation. At the same time, they must be sufficiently correlated with the instrumented variable for desirable efficiency characteristics. In practice, standard errors tend to be large as good instruments are hard to find. Furthermore, even a low correlation between w and $u\prime$, especially if w is a weak instrument, can result in biases that are larger than those under OLS, particularly in small samples.

As an alternative to 2SLS, the source of endogeneity can be viewed as an omitted variable problem in the sense that **m** is not observed. Rather than remove the endogeneity bias by adjusting l, it can be removed from u_i by finding one or more substitutes for each phenomenon in **m**. For each m in $\mathbf{m}$, let one substitute be z such that:

$$m_i = \phi_0 + \phi_1 z_i + r \tag{102}$$

where r is a random error term uncorrelated with z. Substituting into (101):

$$q_i = \alpha \mathbf{a}_i + \boldsymbol{\delta\phi}_0 + \beta_l l_i + \beta_k k_i + \boldsymbol{\delta\phi}_1 \mathbf{z}_i + (\boldsymbol{\delta}\mathbf{r} + e + p) \tag{103}$$

$\mathbf{z}$ must be redundant in (101) in the sense that it would add no explanatory power to q if $\mathbf{m}$ was hypothetically included. It is in practice comfortably assumed that each z is not correlated with $\mathbf{e}$ or $\mathbf{p}$. However, the need for the other factor inputs to be uncorrelated with $\mathbf{r}$ is a requirement that cannot be verified, so one must either rely on theoretical priors or judgement to conclude this correlation is lower than the original one between the factor inputs and $\mathbf{m}$.

Instruments must be uncorrelated with $\mathbf{m}$ while proxies must be highly correlated with $\mathbf{m}$, so candidate variables can either make good proxies or good instruments, but not both. The issue is therefore whether to estimate (103) by directly including $\mathbf{z}$ or, alternatively, to estimate:

$$q_i = \boldsymbol{\alpha}_q \mathbf{a}_i + \boldsymbol{\beta}_l \hat{l}_i + \beta_k \hat{k}_i + \boldsymbol{\mu}_q \mathbf{z}_q + v_q \tag{104a}$$

$$\hat{l}_i = \boldsymbol{\alpha}_l \mathbf{a}_i + \boldsymbol{\mu}_l \mathbf{z} + v_l \tag{104b}$$

$$\hat{k}_i = \boldsymbol{\alpha}_k \mathbf{a}_i + \boldsymbol{\mu}_k \mathbf{z} + v_k \tag{104c}$$

where $\mathbf{z}_q$ is a strict subset of $\mathbf{z}$. Any controls in the equation for output would also be in the first stage regressions for the factor inputs.

The candidate variables for addressing these endogeneity issues are in Appendix 2. One can suggest that a variable for the effects of market conditions on hiring would make a good proxy for market conditions but would not remove the endogenous component from the input if included in the input's instrumenting equation. This makes it favourable as a control. Industry dummies and variables on export quantities, while controlling for an aspect of the technology, would also account for some differences in market conditions. Valid instruments are harder to find; recruitment difficulty or wages are themselves endogenous to market conditions – being higher when demand is high – and would at best only partially purge the factor input of its endogeneity.

Empirical comparison of OLS and 2SLS estimates We turn now to some preliminary results. This exercise is in a Cobb Douglas context.[32]. In Appendix 20, we see a comparison of OLS and 2SLS. The OLS includes variables acting as proxies for omitted variables that might be correlated with the inputs. All the significant variables seem to be those controlling for technology or efficiency. Those designed to control for cross-firm market or other shocks that co-determine output and input demand, notably market conditions, are not significant. They could be bad proxies or there could simply not be much cross firm variation in this sample. The factor inputs are generally significant and have reasonably-sized and correctly-signed coefficients; four are highly significant.

The 2SLS have all six inputs instrumented for using some of the proxies from OLS as well as wages.[33] The coefficient (factor share) for skilled/artisans is negative and the coefficient for managers/professionals seems implausibly large. This is consistent with the variables being used as instruments themselves being correlated with the error term. No variable except one industry dummy and, marginally, the owner managed dummy is significant. The location dummies are not jointly significant and the factor inputs are not significant either.

A summary of the first-stage regressions for each input provides an indication of instrument quality. The R^2 ranges from 0.46 to 0.55. Of the few significant variables, only training expenditure and market conditions are genuine instruments as the others are also in the second stage. The significant market conditions variable could mean it is a good instrument and poor proxy, but it was only significant for some inputs, suggesting market conditions are more relevant for some inputs than others. The wage variable for managerial/professional labour was incorrectly signed and the wage variables were insignificant instruments. Because of the way they were constructed, our factor prices might be particularly weak instruments and thus seriously dent the reliability of the instrumentation procedure .

The result is very wide confidence bands on the inputs, which can be inspected in the regression output in Appendix 20. In fact, they are so wide that those in the OLS regression fall well within the bounds of the IV regression for every variable. The importance of efficiency to this study is reason enough to abandon the IV approach. Furthermore, doubts about the validity of some instruments, supported by the nature of the coefficients, provide no reason to believe the IV estimates are less biased than OLS with proxies.

[32] This is used because it is much easier to compare two sets of parameters and to distinguish plausible coefficients from unlikely ones. The differences in estimates will be so stark that they will overwhelm any distortions that may result, in this context, from not using the fuller translog specification.

[33] It is strictly speaking theoretically inappropriate to use wages because they are modelled as endogenous. In standard applications, wages can have a degree of endogeneity. Ironically, with the exception of the firm size adjusted wages, which we do not use here, the wages constructed are trully exogenous in our study, though they are also particularly weak instruments.

These conclusions are reinforced (and explained) by formal tests, which are reported in Appendix 20. For details of the exact procedures, see Wooldridge (2002) and Statacorp (2003). The Hausman specification test for endogeneity compares the regressions, finding a p-value of 0.90 for the null hypothesis that the variables in equations (103) and (104a) do not differ systematically. In other words, there is no endogeneity bias and there is no need to consider IV in the first place. The results are consistent with the view that, as is more likely for a cross section, the conceptual issues raised are less relevant and/or firm-specific effects are reasonably captured by the controls included. The difficulty with this test is the underlying conjecture that the IV regression coefficients are consistent. The possibility for this not being the case is now tested.

A Hausman type test runs an IV regression, stores the residuals and regresses these residuals on all the instruments. There should be no explanatory power for the residual if the instruments are exogenous. The test statistic is based on the R^2 from this regression and has a χ^2 distribution. The test for the residuals from (104a) does not reject the exogeneity assumption at the 20% level. Whether the instruments are exogenous "enough" given the weakness of the instruments is a matter of judgement, but, because only moderate rejections of exogeneity can result in large biases if instruments are weak (Wooldridge, 2002), 20% is possibly too low for inclusion. If, on the other hand, one is satisfied with the exogeneity assumption, then the first test is valid and there is no need for 2SLS in the first place.

Given the poor efficiency characteristics, the risk of bias, the formal tests, and of course the fact that any instrumentation procedure cannot deal with all the factor inputs, 2SLS is not an effective solution. One must conclude that OLS with correct controls is a better approach.

In simple settings and/or assumptions that the explanatory variables are not correlated, it may be possible to suggest the direction of any biases caused by any remaining endogeneity. For example, lucky or better run firms will tend to grow larger, produce more output, and demand more of the easily adjustable inputs, which suggests an upward bias in the coefficients. To this end, adding the coefficients on the inputs for OLS gives returns to scale of 0.98, which would suggest any upward bias on the coefficients is limited. However, given the use of proxies, the highly collinear nature of the data, and the fact that higher order terms will be used in the translog estimation, it will be very difficult to have a more precise insight into potential biases. Greene (2003:86) notes, "Although expressions can be derived for these biases in a few of these cases, they generally depend on numerous parameters whose signs and magnitudes are unknown and, presumably, unknowable."

Endogeneity bias with rigid wages When we model all our wages as exogenous, 2SLS encounters further particular difficulties in our model. The first is that our wage variables are likely to be weaker than in other datasets, because of the way they had to be constructed. The second difficulty is that we would have far too many variables to instrument for. The third is that we are modelling wages as endogenous and quantities as exogenous. For wages to be valid instruments, they must be exogenous. While a fortunate by-product of our wage construction is constructed exogeneity, this is inconsistent with our theoretical model.

With respect to the third difficulty, it could be argued that, in the rigid wage model, the factor is endogenous by assumption and must be instrumented accordingly. In other words, the rigid wage model says we must instrument for the input whose wage is rigid. It also says that input's price is exogenous. With respect to the second difficulty, instrumenting for only one input reduces the number of instruments required. With respect to the first difficulty, having a relatively low number of instrumented variables will mitigate the efficiency loss. This makes a case for applying 2SLS and instrumenting for the unskilled wages.

Like Johnson (1980) and Grant & Hamermesh (1981), we initially use OLS for the rigid wage case as for the flexible wage case. We also endogenise the factor with a rigid wage. The difficulty in this application is the (still) large number of endogenous variables that need valid instruments: the 6-factor example requires 7 instruments (for itself, its square and its interaction with the other 5 factors). In addition to the instrument difficulties discussed above, it would be theoretically appropriate to use only that factor's wage as a valid instrument. This is because the model implies that wages, except the rigid wage, are endogenous to output and therefore by definition invalid instruments.

Instead, we perform a manual two stage procedure, where unskilled labour is identified using the unskilled wage. The predicted unskilled wage is then interacted with the other factors to yield predicted higher order terms before use in the second stage regression. In other words, instead of instrumenting for every term involving the unskilled wage, which would have to be done in the first stage of 2SLS, we only predict the first order term for unskilled labour and use that predicted value to construct predicted values for all the terms involving unskilled labour.

Such manual procedures should dictate regressing the endogenous variable on all exogenous variables including the identifying instrument(s) (Wooldridge, 2002). This is done but the higher order terms involving unskilled labour are excluded. Omitting the interaction terms involving unskilled labour in the first stage and not performing first stage regressions to identify them means the coefficients in the second stage regression might be inconsistent.

This manual procedure described refers to the disaggregated input case. We also perform it for aggregated inputs, instrumenting for the less skilled labour aggregate rather than unskilled labour. However, it is also feasible, in terms of the number of instruments required, to perform standard 2SLS for the aggregated specification, treating less skilled labour and its interactions with the other two factors as endogenous. The findings will be discussed in the results sections.

5.4.3 Inference

Inference presents particular challenges because the elasticity estimates are highly non-linear combinations of the coefficients and data (Greene, 2003). Significant regression coefficients neither imply nor are necessary for significant elasticities (Anderson & Thursby, 1986). Reviews of empirical work using translog cost or production functions make no mention of significance with respect to elasticities (Chung, 1994; Hamermesh, 1993). With respect to translog estimates of H and ϵ, Vere (2001) is the only person who presents confidence intervals for elasticity estimates.

The cost chapter used variations across firms in the sample to give a measure of the distribution of elasticities. In our model for this chapter, which assumes constant returns and a representative firm, we imply that every firm has the same factor shares. In the context of the macroeconomy, there is only one share of economy-wide output for each factor. It would therefore be inappropriate to use such variations. Instead, we apply the "Delta" method to the elasticity estimates and use it to present p-values. The Delta method calculates Taylor approximations to underlying distributions of functions of parameters (Greene, 2003). Taking for example an elasticity between factors s and u and using (93) and (96),

$$\hat{\epsilon}_{su} = \frac{\hat{\beta}_{su}}{\hat{\alpha}_s + \sum_j \hat{\beta}_{sj} \log x_j} + \hat{\alpha}_u + \sum_j \hat{\beta}_{uj} \log x_j,$$

so we can see the estimated elasticity (denoted with a "hat") is a function of many estimated parameters. The Delta method accounts for confidence intervals on all the parameters in the elasticity calculation. It is important to realise that the p-values can be sensitive to the distributions of the underlying parameters in finite samples, and must be treated as indications only. Estimating equation (91) with six inputs is already asking much of the data, even before using the parameters in further complex elasticity calculations. Despite this, the data do yield some meaningful results, as discussed next.

5.5 Disaggregated Results

We start with a brief discussion of preliminary estimates before presenting the regression used as the basis for the disaggregated elasticities. After analysing the disaggregated elasticity estimates, which assume fully flexible wages, we discuss the case where unskilled wages are assumed fixed.

5.5.1 Disaggregated regression results and diagnostics

As we discussed in chapter 2, we tested for the value added specification using the applicable restrictions in chapter 2 and found raw materials are non-additively separable. Thus, as suggested earlier in this chapter, we estimate (91) with q as value added, not turnover. The value added specifications tended to produce a better fit in any case.

For initial diagnostics, the histogram of the residuals suggests they are approximately normally distributed (see Appendix 21). Although the Jarque-Bera test rejects normality, the residuals are far closer to normal than those observed in the cost chapter. Tests of heteroskedasticity were found to be significant. Inspection of the residuals revealed the translog approximation does not fit nearly as well for very small firms as for the rest of the sample. As shown in Appendix 22, residuals for smaller firms tend to be negative. This may also partially explain why we produce returns to scale of 1.13 which are found to be significantly different from unity at 5%. The crude but effective solution is to drop firms with a value added of less than R1 million (about £ 75 000) at the cost of 30 observations. Cross plots (Appendix 23) also revealed a high positive mean error when the managerial/professional category has one worker, which indicates there are efficiency gains enjoyed by individual- or owner-managed firms. In addition to the variables contained in the original dataset, we constructed a dummy for owner-managed firms. As expected, the dummy was found to be highly significant. Steps like these were successful at improving the fit of the data, lowering heteroskedasticity and producing returns to scale estimates of 1.06, which were insignificantly different from unity.

We suggested that the general direction of endogeneity bias on the input coefficients might be upwards. The unrestricted estimate of 1.06 for returns to scale suggests the bias might be limited. Imposing the restriction of constant returns, besides enhancing efficiency, further limits the potential endogeneity bias. The results are shown in table 13.

The fit is good overall, with an R^2 of 0.96. Furthermore, the coefficients on all the higher order terms are jointly significant, which rejects the null hypothesis that the production function is a Cobb Douglas technology. The coefficients predict (statistically significant) positive factor shares

Table 13: Production Function Parameter Estimates

Dependent variable: value added					
Variable	Coefficient	p-value	Variable	Coefficient	p-value
Constant	-0.90	0.01	ind2	0.31	0.04
Capital	0.19	0.00	ind3	0.16	0.33
Man/Prof	0.09	0.34	ind4	0.58	0.00
Sale/Cler	0.43	0.00	ind5	0.15	0.33
Skil/Art	-0.02	0.71	ind6	0.24	0.15
Semi	0.11	0.08	ind7	0.26	0.07
Un	0.20	0.00	ind8	0.03	0.85
0.5*Capital^2	0.09	0.01	ind9	0.23	0.12
Capital*Man/Prof	-0.07	0.10	loc2	0.24	0.35
Capital*Sale/Cler	0.04	0.26	loc3	0.09	0.42
Capital*Skil/Art	-0.02	0.56	loc4	0.16	0.18
Capital*Semi	-0.05	0.12	loc5	0.34	0.19
Capital*Un	0.00	0.97	loc6	-0.38	0.32
0.5*Man/Prof^2	-0.04	0.69	loc7	-0.82	0.04
Man/Prof*Sale/Cler	-0.01	0.88	loc8	-0.17	0.62
Man/Prof*Skil/Art	0.02	0.71	loc9	0.12	0.31
Man/Prof*Semi	0.04	0.41	exports / output %	0.27	0.16
Man/Prof*Un	0.06	0.21	raw materials / cost %	-0.02	0.00
0.5*Sale/Cler^2	0.05	0.56	recruitment ease Man/Prof	0.04	0.59
Sale/Cler*Skil/Art	-0.04	0.45	recruitment ease Sale/Cler	0.08	0.18
Sale/Cler*Semi	-0.03	0.57	recruitment ease Skil/Art	-0.02	0.79
Sale/Cler*Un	-0.02	0.60	recruitment ease Semi	0.00	0.99
0.5*Skil/Art^2	-0.09	0.08	recruitment ease Un	0.17	0.09
Skil/Art*Semi	0.07	0.05	training / output %	0.38	0.00
Skil/Art*Un	0.05	0.09	market conditions index	0.02	0.10
0.5*Semi^2	0.05	0.28	firm size > 50 employees	0.47	0.00
Semi*Un	-0.09	0.00	computer investment / output %	2.60	0.01
0.5*Un^2	0.00	0.96	owner managed	0.37	0.02
Number of observations		239	Separability (equations (36) and (37)):		
"R squared"		0.96	Exclude Sales/Clerical		0.53
Joint test on all $\beta_{ij} = 0$		0.02	Aggregate Man/Prof & Skil/Art		0.77
			Aggregate Semi & Unskilled		0.69

Table 14: Hicks Elasticities of Complementarity

$H_{ij} = H_{ji}$		j					
		Capital	Man/Prof	Sale/Cler	Skil/Art	Semi	Un
	Capital	-1.81	-1.13	1.68	-0.23	-3.01	1.07
	Man/Prof	-1.13	-5.14	0.85	2.10	3.89	4.47
i	Sale/Cler	1.68	0.85	-1.22	0.03	0.11	0.14
	Skil/Art	-0.23	2.10	0.03	-20.82	11.87	7.83
	Semi	-3.01	3.89	0.11	11.87	-3.28	-13.87
	Un	1.07	4.47	0.41	7.83	-13.87	-10.93

or the representative firms, as shown in Appendix 24, so the monotonicity conditions ($f_i > 0$) are confirmed.

The control variables are not of direct interest but they contribute explanatory power and suggest many of the firm-specific effects are controlled for. We have included them purely as levels effects and have not interacted them with any of the input variables. The industry and location variables capture some technology specifics as well as market conditions relevant to the product and region. The control for export intensity suggests exporting firms tend to be more efficient while the raw materials intensity captures an aspect of the firm's technology. The recruitment ease variables are only significant at 12%; these are thought to be correlated with general market conditions, while the ordinal variable for perceived market conditions by firms is itself significant. Training expenditures, computer investment intensity and a dummy for whether a firm is owner managed or not are all significant. The regression output also includes tests of separability based on restrictions (36) and (37), which are not rejected. The implications of this will be discussed in the aggregated section. For now, we start with analysis of disaggregated elasticities based on this regression.

5.5.2 Disaggregated elasticities

Fully flexible wages Table 14 presents the HEC estimates calculated using equations (93), (97) and (99).

The results show, for example, that a 1% rise in the ratio of managers or professionals to unskilled workers would raise the ratio of unskilled to managerial/professional wages by 4.47%. Relative rises of this occupation would in fact help all other forms of labour, making them q-complements. Pairs which are q-substitutes are more rare: capital and skilled/artisanal labour are an example. The results suggest skilled/artisanal and unskilled labour are big complements while semi-skilled and unskilled labour are substitutes.

Table 15 presents ϵ_{ij} - the percentage change in the price of factor i after a 1% rise in the

Table 15: Elasticities of Factor Price

ϵ_{ij}									
i	j	coef	p	η	i	j	coef	p	η
Capital	Capital	-0.30	0.19	.	Skil/Art	Capital	-0.04	0.92	.
	Man/Prof	-0.22	0.41	.		Man/Prof	0.41	0.48	-0.48
	Sale/Cler	0.67	0.01	-0.59		Sale/Cler	0.01	0.99	-39.5
	Skil/Art	-0.02	0.92	.		Skil/Art	-1.89	0.01	.
	Semi	-0.22	0.26	.		Semi	0.87	0.12	-0.08
	Un	0.09	0.54	-0.94		Un	0.64	0.11	-0.13
Man/Prof	Capital	-0.19	0.42	.	Semi	Capital	-0.49	0.35	.
	Man/Prof	-0.99	0.06	.		Man/Prof	0.75	0.34	-0.26
	Sale/Cler	0.34	0.38	-1.17		Sale/Cler	0.04	0.95	-9.49
	Skil/Art	0.19	0.49	-0.48		Skil/Art	1.08	0.11	-0.08
	Semi	0.28	0.28	-0.26		Semi	-0.24	0.71	.
	Un	0.37	0.12	-0.22		Un	-1.14	0.14	.
Sale/Cler	Capital	0.28	0.01	-0.59	Un	Capital	0.17	0.56	-0.94
	Man/Prof	0.16	0.36	-1.17		Man/Prof	0.86	0.16	-0.22
	Sale/Cler	-0.48	0.02	.		Sale/Cler	0.16	0.72	-2.46
	Skil/Art	0.00	0.99	-39.5		Skil/Art	0.71	0.09	-0.13
	Semi	0.01	0.95	-9.49		Semi	-1.01	0.05	.
	Un	0.03	0.73	-2.46		Un	-0.90	0.03	.

quantity of factor j, assuming perfectly elastic demand in the product market, as well as p-values calculated using the Delta method.[34] η will be explained shortly. It is encouraging to report that all own-elasticities are negative and, with the exceptions of capital and semi-skilled labour, significantly so. Some coefficient values are virtually zero, suggesting some factors are neither complements nor substitutes. Other coefficients are imprecisely estimated, partially because of the estimation and inference procedure, but also because the technology does not exhibit a strong pattern of complementarity.

There are many parameters so we focus on unskilled wages. We find many low p-values for the effects on unskilled wages. Taken literally, the results suggest that a 10% rise in the supply of skilled/artisanal labour would lead to a 7.1% *rise* in unskilled wages while a similar rise in the supply of semi-skilled labour would lead to a 10.1% *fall* in unskilled wages. These effects are significant and are a key result of this chapter. In fact, skilled/artisanal labour complements all other factors, which strongly supports the claim that the shortage of this labour type is hampering output growth.

The assumption of perfectly elastic demand could be justified given that South Africa can be

[34] The Hicks elasticity involves far more parameters than the elasticity of factor price. This is especially challenging because two predicted factor shares, each of which contains many paramteres, are *multiplied* by each other. The reliability of such p values would be greatly compromised. We found very high p values for the majority of estimates, despite finding low p values for the elasticity of factor price.

considered a small open economy. There are indications that even manufactured exports have near perfect elasticities (Behar & Edwards, 2004). Nonetheless, if product demand elasticities are not perfect, the estimates will tend to be too positive. In particular, some factor pairs thought to be complements may indeed be substitutes. No documented studies attempt to allow for this. One reason may be that, at a macro level, the elasticity is a difficult concept to pin down, let alone estimate.

We propose the following solution: rather than impose a value for η, we calculate values of η that would make $\hat{\epsilon}_{ij}$ equal to zero. By (81), $\eta = -H_{ij}^{-1}$ for strictly positive factor shares. If the elasticity is η or less elastic, then a positive elasticity would become negative. Obviously, for $\epsilon_{ij} < 0$, $\eta > 0$, so the calculation is not presented. In other words, we ask how inelastic demand must be to overturn a result that two inputs are complements. The more inelastic demand must be, the less likely it is that we are incorrectly predicting two factors are q-complements.

The calculations in the table yield many relatively high threshold values of $|\eta|$ that could easily exceed the true value. If similar thresholds were to be found with other data, this suggests that many studies finding two factors are complements may be doing so misleadingly. This may be so even for relatively large positive values of ϵ, as the case of capital and sales/clerical workers demonstrates. In contrast, demand would have to be highly inelastic for the managerial/professional and unskilled labour coefficient to change sign. Furthermore, the calculations strongly suggest skilled/artisanal and unskilled labour are complements even after accounting for imperfectly elastic product market demand.

Rigid wages The bottom right corner of table 16 presents the employment response of unskilled workers to changes in the quantities of other factors (equation (89)) based on the OLS regression in table 13. A 10% rise in the quantity of skilled/artisanal workers would lead to a 7.9% rise in unskilled employment while a rise in semi-skilled workers would reduce unskilled employment. The low p-values are telling given the large number of parameters involved in the calculation.

The rest of the table presents the factor price responses of the other five factors (equation (90)). There are a few sign switches relative to table 15, but these were for small coefficients that were not significant. Many of the changes occur for semi-skilled workers, largely because they are large substitutes for unskilled labour, showing that elasticities of factor price could be misleading if there is even one rigid wage. Note that a rise in supply of semi-skilled workers could even lead to a rise in its own wage. Equation (90) shows this is more likely for factor $i = j$ if the technology is such that, with completely flexible wages, the own-elasticities of factors i and u are low, and the two

Table 16: Elasticities after adjusting for rigid unskilled wages

i	j	coef	p	i	j	coef	p
ϵ^{ρ}_{ij}							
Capital	Capital	-0.28	0.21	Skil/Art	Capital	0.09	0.81
	Man/Prof	-0.14	0.61		Man/Prof	1.03	0.16
	Sale/Cler	0.68	0.01		Sale/Cler	0.13	0.82
	Skil/Art	0.05	0.82		Skil/Art	-1.38	0.05
	Semi	-0.32	0.24		Semi	0.14	0.84
Man/Prof	Capital	-0.11	0.61	Semi	Capital	-0.71	0.32
	Man/Prof	-0.64	0.35		Man/Prof	-0.34	0.79
	Sale/Cler	0.40	0.36		Sale/Cler	-0.16	0.86
	Skil/Art	0.48	0.16		Skil/Art	0.18	0.83
	Semi	-0.13	0.79		Semi	1.05	0.54
				ρ			
Sale/Cler	Capital	0.28	0.01	Un	Capital	0.19	0.54
	Man/Prof	0.20	0.33		Man/Prof	0.96	0.19
	Sale/Cler	-0.48	0.01		Sale/Cler	0.18	0.72
	Skil/Art	0.03	0.83		Skil/Art	0.79	0.08
	Semi	-0.03	0.85		Semi	-1.13	0.12

factors are large complements or large substitutes.

Alternative calculations with endogenised unskilled labour are presented in table 17, based on regression output in Appendix 24. As we explained earlier, we perform a manual two stage procedure. In the first stage, unskilled labour is predicted using other factor types and their higher order interactions (not those involving unskilled labour), unskilled wages and various controls. The second stage regression, which also imposed constant returns, uses predicted unskilled labour and its constructed interactions with the other factors. p-values are not valid given this procedure, but the results based on point estimates are consistent with those based on OLS estimates. Of the 30 calculations, only 5 signs differ. Most notably, the effects of semi-skilled and skilled/artisanal labour on the unskilled is preserved.

The key disaggregated finding is that a rise in the supply of skilled/artisanal workers will lead to a rise in demand for the other occupation types, suggesting that a shortage of this skill type is constraining output. In particular, we find that the unskilled would benefit from the training of skilled/artisanal workers. In contrast, they would not benefit from the training of semi-skilled workers. Such a finding would only have been possible with this highly disaggregated data. We now turn to more aggregated studies.

Table 17: Elasticities after adjusting for rigid unskilled wages (manual two stage procedure)

i	j	coef	i	j	coef
ϵ_{ij}^{ρ}					
Capital	Capital	-0.13	Skil/Art	Capital	0.06
	Man/Prof	-0.21		Man/Prof	1.04
	Sale/Cler	0.24		Sale/Cler	0.22
	Skil/Art	0.03		Skil/Art	-0.88
	Semi	0.07		Semi	-0.44
Man/Prof	Capital	-0.23	Semi	Capital	0.09
	Man/Prof	0.27		Man/Prof	-1.30
	Sale/Cler	0.49		Sale/Cler	-0.35
	Skil/Art	0.54		Skil/Art	-0.28
	Semi	-1.07		Semi	1.83
			ρ		
Sale/Cler	Capital	0.13	Un	Capital	-0.27
	Man/Prof	0.24		Man/Prof	1.80
	Sale/Cler	-0.29		Sale/Cler	0.48
	Skil/Art	0.06		Skil/Art	0.89
	Semi	-0.14		Semi	-1.91

5.6 Aggregate results

With a view to adopting a parsimonious more aggregated production, we tested for weak separability with respect to the Sales/Clerical group (cf. table 13) and consequently drop it. We do this because the skill content of this occupation is not clear and can vary enormously. Furthermore, we test whether the Managerial/Professional and Skilled/Artisan occupations are separable from each of the other occupations. Failure to reject this test means we can validly aggregate these occupations into the more skilled group. Similarly, we can validly aggregate semi-skilled and unskilled labour into the less skilled group (Fuss, 1977). In other words, it is legitimate to write the disaggregated translog function (91) as:

$$\begin{aligned} \log q &= \log \alpha_0 + \left[\sum_i^3 \alpha_i \log X_i + \frac{1}{2} \sum_i^3 \sum_j^3 \beta_{ij} \log X_i \log X_j \right] \\ &\quad + \left[\alpha_k \log X_k + \frac{1}{2} \beta_{kk} \log^2 X_k \right] + u \end{aligned} \tag{105}$$

X_1, X_2 and X_3 refer to the aggregates of capital, managerial/professional and skilled/artisanal labour, and semi-skilled and unskilled labour respectively while X_k is the Sales/Clerical occupation, which we omit.[35] While the cost chapter presented aggregation issues, here each aggregate is the

[35] The Sales/Clerical occupation was strongly separable and can thus be presented in a relatively straightforward manner. When testing for the validity of the value added specification, the separability tests suggested raw materials were weakly separable. When only weakly separable, including raw materials would not be as straightforward and would only serve to obfuscate.

Table 18: Production Function Parameter Estimates (aggregated inputs)

Dependent variable: value added		
Variable	Coefficient	p-value
Constant	-1.25	0.00
Capital	0.46	0.00
More	0.43	0.00
Less	0.10	0.20
0.5*Capital^2	0.12	0.00
Capital*More	-0.00	0.90
Capital*Less	-0.12	0.00
0.5*More^2	-0.01	0.84
More*Less	0.01	0.68
0.5*Less^2	0.10	0.03
"R squared"		0.89
Joint test on all $\beta_{ij} = 0$		0.00

sum of the disaggregated input quantities. This translog example of a separable partition is called a Cobb-Douglas aggregation of two translog functions in Blackorby, Primont & Russell (1977).

Aggregation may be desirable for various reasons. Aggregate estimation is less demanding of the data. The vast majority of models employ a single elasticity between skilled and unskilled labour, so such an estimate would be needed to make predictions using those models. This pertains to applied computable general equilibrium (CGE) models in particular. While many training programmes are aimed at the lower end of the skill spectrum, it is still important to understand what general rises in the broad skills of the population would yield for the relatively less skilled.

Table 18 presents the estimates of the translog function with capital, more skilled labour (managerial/professional & skilled/artisanal) and less skilled labour (semi-skilled and unskilled). Control variables are not presented as they yield no new information. Despite the simpler specification, the fit is still good. Tests reject the null hypothesis that all $\beta_{ij} = 0$ and monotonicity is satisfied. Table 19 presents Hicks elasticities of complementarity,[36] which intuitively show capital and more skilled labour tend to complement each other in production while capital and less skilled labour appear to substitute each other in production. This is analogous to the concept of capital-skill-complementarity employed by krusell et al. (2000), where a rise in the quantity of capital increases the marginal product of skilled labour more than it increases the marginal product of unskilled labour. We also find that, even at more aggregated levels, occupations of a higher skill type tend to complement lower-skilled occupations. This finding masks the distinction between skilled/artisanal and managerial/professional labour on the one hand and semi-skilled labour on the other (table

[36]Out of the three pairs of H_{ij} $(i \neq j)$, at least two pairs must be positive (Sato & Koizumi, 1973). These results meet that minimum as opposed to the maxium of three.

Table 19: Hicks Elasticities of Complementarity (aggregated inputs)

$H_{ij} = H_{ji}$		j		
		Capital	More	Less
	Capital	-1.07	0.96	-0.60
i	More	0.96	-1.25	1.10
	Less	-0.60	1.10	-1.17

Table 20: Elasticities of Factor Price (aggregated inputs)

ϵ_{ij}				
i	j	coef	p	η
Capital	Capital	-0.25	0.10	.
	More	0.44	0.01	-1.04
	Less	-0.19	0.29	.
More	Capital	0.23	0.01	-1.04
	More	-0.57	0.00	.
	Less	0.34	0.00	-0.90
Less	Capital	-0.14	0.27	.
	More	0.50	0.00	-0.90
	Less	-0.36	0.02	.

14), but suggests that a general rise in the economy's education/skill levels will benefit those who remain relatively less skilled. Table 20 presents the elasticities of factor price and the value of η that would make a positive coefficient negative. It is encouraging to report that the own-elasticities for capital in tables 15 and 20 are similar and relatively precisely estimated. The own-elasticities for the aggregated labour types are also significantly negative. The results suggest there would be a (significant) positive effect on less skilled wages after a rise in the quantity of more skilled labour. However, the threshold values for η are lower than in the disaggregated case. This suggests it is not as easy to conclude more and less skilled labour are q-complements.After adjusting for the possibility that the less skilled labour market has rigid wages, we use the OLS results to find in table 21 that many of the coefficients are close to zero and are not significant. However, we still see a rise in the supply of more skilled labour would lead to a rise in less skilled employment.[37]

We also perform estimations that allow for endogenous less skilled labour and its higher order terms. We present the results from a manual two-stage procedure in Appendix 26 and the calculated elasticities in Appendix 27. We present conventional 2SLS results in Appendix 28 and the calculated elasticities in Appendix 29. These are accompanied by further explanation of the method used.

[37] The symmetry in results is no coincidence. According to Sato & Koizumi (1973), $\sum_j \varepsilon_{ij} = 0$. This can be verified in all the calculations presented. The result implies:
$\sum_j \varepsilon^{\rho}_{ij} = 0$. In the 3 factor case, it is easy to show. Using equation (90) ,
$\varepsilon^{\rho}_{ij} = -\varepsilon^{\rho}_{ii}$ implies:
$\varepsilon_{ii} + \varepsilon_{ij} = \frac{\varepsilon_{iu}(\varepsilon_{ui}+\varepsilon_{uj})}{\epsilon_{uu}}$.
Therefore, using Sato & Koizumi's result, it implies:
$\varepsilon_{iu} - \frac{\varepsilon_{iu}\epsilon_{uu}}{\epsilon_{uu}} = 0$.

Table 21: Elasticities after allowing for rigid wages (aggregated inputs)

i	j	coef	p
ϵ_{ij}^{ρ}			
Capital	Capital	-0.18	0.49
	More	0.18	0.49
More	Capital	0.09	0.52
	More	-0.09	0.52
ρ			
Less	Capital	-0.39	0.39
	More	1.39	0.00

In summary, both conventional 2SLS and the manual procedure analogous to the six-input case yielded implausible results, including own-elasticities of factor price as high as 17.83. In the 2SLS case, where they are valid, p-values were all 0.85 or higher. Returning to the OLS-based results, capital and labour could be neither complements nor substitutes in the presence of rigid wages, in contrast to the flexible wages case. However, the results still suggest a rise in the supply of skilled labour would lead to a large and statistically significant rise in the quantity of unskilled employment.

Thus, we conclude that, on balance, a general rise in the skill supply will benefit the unskilled and also that, to a lesser extent, a rise in the quantity of capital would benefit the skilled more than the unskilled.

5.7 Concluding comments

The key finding is that a rise in the supply of skilled workers / artisans will lead to a rise in demand for unskilled labour while a rise in the supply of semi-skilled workers will lead to a fall in demand for unskilled labour. These results are statistically significant, account for imperfectly elastic product market demand and allow for rigid unskilled wages. For other factor combinations, assuming perfectly elastic demand can predict factors are complements when they are in fact substitutes, while not accounting for rigid wages can also lead to the incorrect sign. Estimates of aggregated inputs find more skilled and less skilled labour are q-complements, although this depends on how elastic product demand is.

The estimated values have some clear policy implications. The government has long been concerned about a shortage of artisans. These results produce empirical support for its concern. The findings suggest that there is indeed a shortage of artisans, that addressing this shortage would expand output, and that the unskilled would benefit. There is therefore cause for serious concern because prospective artisans make up a low proportion of those being trained in South Africa

(Paton, 2003). By producing semi-skilled workers, these training programs may raise unskilled unemployment.

Of further concern is the claim by trade union officials that it is only so-called soft skills which are being provided by firms so that they can reclaim their skills levy. The government is also frustrated at the inability of its skills strategy to produce the intended skills. The Trade and Industry Minister commented on his failure to understand why we have a "shortage of welders and boilermakers" before calling for a re-introduction of the traditional apprenticeship system (Ntuli, 2006).

While the mechanism for developing such skills are yet to be found, the high priority attached to it is evident from the government's new growth strategy, AsgiSA (Mlambo-Ngcuka, 2006). AsgiSA also includes a major public capital expenditure plan with the financing to back it up (Manuel, 2006). To the extent that we can view such capital as similar in nature to the capital used by firms, which is a major assumption, our estimates suggest such a programme would relieve bottlenecks and expand output by enough to raise demand for skilled labour. However, it would not lead to a rise in demand for unskilled labour.

Our disaggregated estimates found significant coefficients for the effects of capital on managerial/professional and sales/clerical labour but not for the production occupations. To the extent that it is a means of expanding output, our results suggest the programme will be effective. As a targeted means of increasing demand for those with lower skills, this may not be an effective strategy.

A general rise in the proportion of people with the education associated with higher-skilled occupations, be it through universities or other tertiary institutions, would also in general benefit the unskilled. Chapter 4 found more and less skilled labour are p-complements. We will discuss the relationship between those elasticity estimates, which were derived from a cost function and the estimates from this chapter, which are derived from a production function, in chapter 9. For now we report our key aggregate finding is that skilled and unskilled labour are q-complements.

6 Observed patterns of SBTC and wage inequality in developing countries: the roles of skills supply and international technology

6.1 Introduction and background

The evidence for shifts in labour demand favouring skilled workers in the United States (US) and other developed countries is well documented. Moreover, a consensus has developed that this shift is due to skill-biased technological change (SBTC) (Katz & Autor, 1999). Despite a steady rise in the supply of skilled workers in the US, wage inequality increased in the second half of the 20th century. There is evidence that developing countries, particularly middle income countries, have also experienced technological change that favours skilled workers (Berman & Machin, 2000) and that some have seen rises in wage inequality (Berman, Bound and Machin, 1998).

Bhorat & Hodge (1999) find similar patterns for South Africa using a decomposition analysis similar to that of Berman, Bound & Griliches (1994): within-sector changes to production methods led to a rise in demand for skilled labour and a fall in demand for unskilled labour. The findings are consistent with evidence documented in Edwards (2001), Fedderke, Shin & Vaze (2003) and Edwards & Behar (2006).

Chapters 7 and 8 will model skill-biased technological change and wage inequality in developing countries. The purpose of this chapter is two-fold. First, it reviews some of the literature supporting three premises, which we will capture in our models. The three premises are:

1. SBTC can be a major contributor to the skill premium. This premise comprises the claims that SBTC is a major source of labour demand shifts, that labour demand shifts can explain differences in wage inequality and that labour market institutions, for example collective bargaining, do not necessarily weaken this link.

2. Technological change can be skill-biased because of the incentives facing profit-driven technology suppliers. That is, technological change is directed. One factor influencing the attractiveness of adopting or developing a skill-biased technology relative to a technology that favours unskilled workers is the relative supply of skills in an economy. We argue that this factor can help predict the effects of expanded educational access. In particular, expanded educational access can increase the skill premium and harm those who remain uneducated.

3. The technologies adopted by developing countries ("the South"), although subject to similar incentives, are heavily influenced by those developed in technologically advanced countries ("the North").

In addition to substantiating these claims, the second purpose of this chapter is to trace out the key components of two models in the directed technological change literature. Kiley (1999) builds a model which assumes skilled and unskilled labour are perfect substitutes. This type of model will be used in chapter 7. Acemoglu (1998, 2002ab) relaxes this assumption and chapter 8 builds a model in which skilled and unskilled labour are imperfect substitutes.

This chapter proceeds in the following manner. A brief sketch of the observed wage and employment experiences in the US and Europe serves as a motivation and background for the literature. This is done is section 6.2. After outlining alternative explanations, we focus on the role and nature of SBTC in section 6.3. Two sources of SBTC in an economy are identified. The first is domestic: SBTC is driven by rises in the relative skill supply. The second is driven by international technological developments. In section 6.4, we outline existing models of the effects of skill supply on SBTC and wage inequality, but there appear to be no models for the second effect. Section 6.5 introduces how we will build the second effect into the models and how they can be adapted to a developing economy in chapters 7 and 8. Section 6.6 summarises the key points of the chapter.

6.2 Differing relative employment and wage outcomes: alternative explanations

6.2.1 Observed differences in wage and employment patterns by skill

Katz & Autor (1999) document increased wage inequality in the US and UK in the latter part of the 20th century, despite a rise in the relative supply of skilled labour. It is generally agreed that the second half of the 20th century saw a bigger rise in wage inequality in the US and UK than in Continental Europe (Nickell & Layard, 1999; Acemoglu, 2003a; Sanders, 2005).

While in recent times European total unemployment has risen faster than in the US, the evidence on relative unemployment is mixed. Nickell & Layard report that unskilled unemployment rose in nearly all European countries and that skilled unemployment rose in some. Nickell & Bell (1996) show relative unemployment rates evolving generally in the same manner in the US and Europe. They present patterns of unskilled relative to skilled unemployment that rose in the 1970s and 1980s but fell in the 1990s. Moore & Ranjan (2005) present evidence from the US, UK, Italy, West Germany and France that unemployment rates rose for both skilled and unskilled workers. Thus far, the evidence on balance suggests no difference in relative unemployment performance across countries. It is also not obvious that the unskilled experienced worse unemployment outcomes than the skilled. However, inspection of evidence in Moore & Ranjan suggests that the proportional rise in skilled unemployment was slower than the rise in unskilled unemployment, except for the US, where it was roughly the same. Furthermore, Acemoglu (2003) reports that, at least for some European countries, relative skilled employment rose faster than in the US. This evidence suggests that the unskilled may have suffered more in terms of unemployment outside of the US. On balance, however, the evidence is equivocal.

6.2.2 Alternative explanations

The three broad categories of explanation for variations in wage and perhaps employment outcomes between the US and Europe are: i) differences in relative demand shifts, ii) differences in the rise of skill supply, and iii) differences in labour market institutions. Nickell & Layard (1999) examine relative demand shifts. They find econometric evidence that assigns such changes a big responsibility for the differential shifts in the skill premium across countries, but conclude that they had no effect on relative unemployment performances.

Moore & Ranjan (2005) show theoretically that globalisation, which raises the relative demand for skilled workers in their model, can lead to a rise in the ratio of unskilled to skilled unemployment and/or the skill premium. While their search unemployment model generates a rise in the skill premium after exogenous SBTC, whether the relative unemployment ratio rises or falls depends on the degree of complementarity between skilled and unskilled labour.[38]

Nickell & Layard (1999) argue that wage compression in some countries can be attributed to skill compression: wage inequality in Europe is lower than in the US because educational inequality is lower. This supply-side argument is consistent with the argument in Acemoglu (2003a) that European skill supply rose faster than American skill supply.

Moore & Ranjan (2005) formalise the Krugman (1994) argument that European labour market rigidities helped mitigate the rise in wage inequality observed in the US: rigidities result in relative quantity adjustment rather than relative price adjustment. While intuitive, the comparisons of relative unemployment performance across continents do not yet provide sufficient support for this view. Furthermore, the econometric findings in Nickell & Layard (1999) assign no role to labour market institutions for explaining wage/unemployment differences by skill: *"Supply and demand does fine"* (pg 3078).

The discussion of the role of rigidities and other labour market features suggests that models analysing SBTC and wage inequality do not become less relevant in settings with high unemployment. Section 8.4 discusses work by Sanders (2005), which links directed SBTC and labour market

[38]They use the term complementarity to describe the cross partial derivative of the production function with respect to skilled and unskilled labour, which is the basis for the Hicks Elasticity of Complementarity.

rigidities.

6.3 Why might technological change favour skilled workers?

Although there is agreement that SBTC has contributed to relative labour demand shifts (Berman, Bound & Griliches, 1994; Katz & Autor, 1999), the causes and nature of SBTC are still debated. One source of disagreement is whether technological change is skill-biased by nature or whether it is skill-biased by design (Sanders, 2005). We consider each of these possibilities in turn.

6.3.1 Skill-bias by nature

One view holds that technological change by its very nature favours skilled workers. Griliches (1969) argues skills and the technology embodied in machines are relative complements in the production function such that the elasticity of substitution between capital and unskilled workers is higher than that between capital and skilled workers. He labels this capital skill complementarity (CSC). Building on CSC, Krusell, Ohanian, Rios-Rull & Violante (2000) posit a dramatic fall in the price of capital and the resulting capital adoption as the main cause of relative labour demand shifts. In some developing countries, an exogenous rise in the quantity of capital - through the lifting of sanctions or the opening up of domestic markets to foreign investors for example - could favour the wages of skilled workers.[39]

We have mixed evidence of CSC in South Africa from earlier work in this thesis. Based on Allen elasticities of substitution in chapter 4, we do not support the hypothesis. Based on Hicks elasticities of complementarity in chapter 5, we do. Berman & Machin (2000) argue that, even if CSC holds, developing countries have not had nearly the extent of capital deepening required to generate the observed relative labour demand shifts. By similar reasoning, the downward trend in investment in physical capital over the last three decades (Fedderke, Kayemba, Henderson, Mariotti & Vaze, 2001) suggest capital deepening could not explain SBTC in South Africa, even if CSC does exist.

Nelson & Phelps (1966) refer to skilled workers as intrinsically complementary to new technologies precisely because they are new; skilled workers have the ability to understand and implement new machines and/or processes. An exogenous boom in new technologies would temporarily favour skilled workers because they are needed to apply them. Sanders (2005) adapts Grossman & Helpman's (1991ab) North-South product lifecycle model to a single closed economy. Technologies for

[39] When there are more than two factors, it might be better to measure these effects using the HEC rather than traditional elasticities of substitution.

new products are developed and produced in a skilled sector but eventually imitated and produced by the unskilled sector. The introduction of a major general purpose technology (GPT) like information and communications technology (ICT) is modelled as a fall in the cost of innovation in the skilled sector and leads to a rise in the skill premium.

The view that new technologies are intrinsically skill-biased relies on an exogenous boost in technology growth to explain relative labour demand shifts. There is support for this in the form of ICT adoption (Katz & Autor, 1999) but Acemoglu (2002a) argues total factor productivity has not accelerated sufficiently to support the claim that there has been a pronounced revolution.

6.3.2 Skill-bias by design

The argument which follows is that technologies are purposefully created/adopted in pursuit of profits. Markets for technologies and costs of developing them are important determinants of the pattern of factor-bias in technological change. In the discussion below (and in the formal treatment, which follows), it is helpful to consider two alternative intermediate means of production. One is skill intensive and the other is not. The skill intensive intermediate process uses skilled labour and machines that complement skilled labour ("skilled machines"). The other process uses unskilled labour and "unskilled machines".

Markets for technologies: the role of relative skill supply Acemoglu (2002a) argues that technologies have historically favoured a particular factor because of profit motives. Drawing on evidence from the industrial revolution and elsewhere, he argues technological change was not always skill-biased, nor did it always lead to a rise in the skill premium. Hicks (1963) introduces factor saving inventions as those which increase the marginal product of the other factor relative to that factor. Such inventions seek to economise on the use of the more expensive factor, are thus spurred by changes in relative factor prices and are called "induced innovations" (pg 125). A skilled intermediate producer will, ceteris paribus, demand more skilled machines (capital) if skilled wages are higher and an unskilled intermediate producer would do likewise for unskilled wages. Thus a given rise in the skill premium would induce innovations in favour of skilled machines.

Alternatively, Schmookler (1966) introduces the market-size effect later advanced by Kiley (1999): a higher number of skilled workers raises the marginal product of skilled machines and hence increases demand for skilled machines. A lower number of unskilled workers reduces the marginal product of unskilled machines and decreases their demand. Thus, a rise in the proportion of skilled workers makes skilled machines more attractive to develop/licence. For a further discussion of the developments of these ideas, see Sanders (2005).

While both market size and induced innovation arguments are consistent with profit-driven innovation, it appears that they conflict: if we were to follow Habakkuk (1962), we would say a rise in the relative scarcity of unskilled labour would make it more expensive and spur the development of unskilled machines. Alternatively, we can try to combine both these arguments by defining the attractiveness of a market for a particular technology in terms of the share of the factor using it, which captures both the quantity of a factor employed and its price. If the share of skilled labour in output is higher, the market for technologies that complement skilled labour is more attractive. We will do this in chapter 8.

Acemoglu (1998, 2002a), Kiley (1999) and others build their models of directed technological change on the endogenous growth literature[40]. Firms can produce final goods using skilled labour together with machines that complement skilled workers and unskilled labour together with machines than complement unskilled workers. Depending on the value of the elasticity of substitution between the skilled and unskilled processes, the demand for machines that complement skilled labour is positively related to the availability of skilled labour. The same applies for unskilled labour. Thus the market for skill-complementary technologies is relatively more attractive if there are more skilled workers. Because researchers are driven by profits, they will find it relatively more attractive to produce skill-biased technologies if the proportion of skilled workers is higher. Thus an exogenous rise in the supply of skilled workers can lead to skill-biased technological progress. Depending on the elasticity of substitution between skilled and unskilled workers, this can lead to a rise in wage inequality. We will present formal treatments of this in section 6.4.

Costs of technologies: the role of global technology patterns The skill-bias of technological change in developing countries is affected by technology patterns in developed countries. Berman, Bound and Machin (1998) argue that SBTC is pervasive across the globe, affecting both OECD and developing countries. Furthermore, most of the developing countries in their sample experienced increases in the skill premium and in the skilled share of employment.

Savvides & Zachariadis (2005) use manufacturing data to find that developing countries undertake no own R&D but rely on foreign technology transfer. Furthermore, they find that foreign R&D generally has a bigger impact on domestic productivity and value added growth than capital goods imports and foreign direct investment. Berman & Machin (2000) also argue that technology

[40]For example Romer (1990), Grossman & Helpman (1991b) or Aghion & Howitt (1992). These are models of endogenous growth because technology adoption and hence growth are determined by profit maximizing research firms. See Barro & Sala-i-Martin (2004) or Aghion & Howitt (1998).

adoption in the South is driven by that in the North.[41] Using skill-upgrading as an indicator of skill-biased technology transfer, they find that the same industries experiencing SBTC in the South in the 1980s were those experiencing it in the North in prior decades. They conclude the choices of technologies available to developing countries are becoming increasingly skill-biased.

Appropriate technology for a developing country? Caselli & Coleman (2000) develop a model in which countries have a choice of technology options subject to a technology frontier; the degree of skill-bias in the technologies chosen depends on relative skill endowments. From a sample of a wide variety of rich and poor countries, they present a strong correlation between relative factor shares and relative factor productivities (pg 23). To the extent that factor productivity is an indicator of the technologies available to that factor and factor shares indicate the attractiveness of the market, this evidence is consistent with the view that the skill bias of technology is directed in both rich and poor countries. We can therefore also argue that, just like in the North, the relative attractiveness of the market for skill-biased technologies will influence the skill-bias of technology adoption and could lead to a rise in the skill premium.

We therefore have evidence that skill bias in developing countries is influenced both by external skill-biased technological progress and by domestic conditions like relative skill supply. Put differently, technology adoption can be "appropriate" to local skill supply, in the sense given by Atkinson & Stiglitz (1969) and Caselli & Coleman (2000), but also "inappropriate" because it is driven by international factors.

6.4 Directed SBTC and wage inequality

We briefly review two contributions in a class which models SBTC as the result of deliberate decisions of agents in a profit seeking technology sector. The purpose of the section is to convey the main ideas, so we present only brief sketches of the models. We start with the approach of Kiley (1999), which views skilled and unskilled labour as perfect substitutes. We continue with the approach of Acemoglu (2002a), which allows for imperfect substitutability.

6.4.1 Perfect substitutes

Kiley (1999) builds on a variety expansion model (Romer, 1990) to model the production technology for final goods as:

$$Y_{it} = (L_{it}^{s})^{1-\alpha} \sum_{j=1}^{N_t} (X_{ijt})^{\alpha} + (L_{it}^{u})^{1-\alpha} \sum_{j=1}^{M_t} (Z_{ijt})^{\alpha} \tag{106}$$

[41] This finding is limited to middle income countries, not the least developed countries in their sample.

The model is in discrete time. Y_{it} is output for perfectly competitive firm i at time t. L^s_{it} and L^u_{it} are skilled and unskilled labour. X_{ijt} is machine input of type j used by firm i at t. It is the quantity of each of N machines (capital) and complements skilled labour. Similarly, Z_{ijt} is the quantity of each of M machines (capital) complementing unskilled labour. Capital depreciates fully in each period. We will refer to N and M as the numbers of skilled machines and unskilled machines. The production technology has diminishing returns to the quantity of each machine employed, but has constant returns to the variety of machines used.

Technological progress takes the form of increases in M or N. This production technology is additive in the skilled and unskilled components, which implies skilled and unskilled labour are perfect substitutes. As we will see in more detail in chapter 7, the perfect substitutes assumption allows us to write demand for X and Z in terms of q and hence write the production function for the economy in a simple manner. Dropping the time subscripts:

$$Y = \alpha^{\frac{2\alpha}{1-\alpha}} \left[Nq + M(1-q)\right] \tag{107}$$

q and $1-q$ are the proportions of skilled and unskilled labour in the labour force, which has been normalised to unity. The proportion of skilled labour is exogenous.

Firms hire labour such that marginal product equals the wage. For economy-wide equilibrium, wages are set endogenously by competitive labour markets. As will be shown in chapter 7, by differentiating (106) and writing X and Z in terms of q, we can write skilled and unskilled wages as:

$$W^s = (1-\alpha)\alpha^{\frac{2\alpha}{1-\alpha}} N \tag{108a}$$

$$W^u = (1-\alpha)\alpha^{\frac{2\alpha}{1-\alpha}} M \tag{108b}$$

The skilled wage is directly proportional to the number of skilled technologies in the economy and the unskilled wage is directly proportional to the number of unskilled technologies. Having established this relationship, we proceed to sketch how Kiley determines technology adoption.

Firms in a technology production sector incur a cost to develop the next available technology, for which they will have a patent that yields perpetual monopoly profits. A slightly simplified version of his equations for the costs are given by:

$$C^s = \beta^s \left(\frac{N_t}{R_t}\right) \tag{109a}$$

$$C^u = \beta^u \left(\frac{M_t}{R_t}\right) \tag{109b}$$

$\beta^s \neq \beta^u$ are parameters and R_t is the stock of basic research available at time t. Thus, Kiley distinguishes between "applied technologies" like N, M and "basic research", which can be used to produce either applied technology. If the stock of skilled technologies is high relative to basic research, it is relatively expensive to research and develop a new skilled technology. The same applies to unskilled technologies. Basic research is assumed to evolve exogenously with a growth rate of $\gamma > 1$. Having described the costs of technology development, we now consider the benefits.

The value of a patent for the next technology is the discounted present value of monopoly profits. The values for skilled technologies, unskilled technologies and their ratio are:

$$V^s = \Omega\frac{q}{r} \tag{110a}$$

$$V^u = \Omega\frac{1-q}{r} \tag{110b}$$

$$V \equiv \frac{V^s}{V^u} = \frac{q}{1-q} \tag{110c}$$

r is the exogenous discount rate and $\Omega \equiv (1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}$. There is free entry into the technology sector such that, in equilibrium, the value of adopting a technology must equal the costs. Setting (109a) equal to (110a) and equating (109b) with (110b) yields the following expression for the ratio of skilled to unskilled machines:

$$T \equiv \frac{N}{M} = \frac{\beta^u}{\beta^s}\frac{q}{1-q} \tag{111}$$

(111) is the key result in Kiley's paper: the ratio of skilled to unskilled technologies is increasing in the relative value of skilled technologies, which is represented by the relative proportion of skilled labour. This captures the market size argument of Schmookler (1966) discussed in section 6.3.2. If there are relatively more skilled people to use a technology, it is more attractive to produce the technology to complement skilled workers. The ratio of skilled to unskilled technologies is decreasing in the relative cost of developing a skilled technology.

If the skill composition stays constant, the value of technologies must stay constant. Therefore, in equilibrium, the cost of developing technologies must stay constant. Because basic research evolves and acts to reduce the costs, both skilled and unskilled machines are adopted to preserve equilibrium. In other words, because R is rising at rate γ, both N and M must grow at rate γ. Along the balanced growth path, N and M grow at the same rate and the ratio of technologies T stays constant.

Kiley does not model changes in relative costs, preferring to focus on the effects of a rise in skill supply on the relative value of skilled machines and hence on the ratio given by (111). Abstracting from the dynamics involved in adjustment to a new equilibrium, it is clear that a rise in $\frac{q}{1-q}$ will lead to a rise in T. This can be interpreted as SBTC. By (108), it is clear that $\frac{W^s}{W^u} = \frac{N_t}{M_t}$ and the skill premium is:

$$W \equiv \frac{W^s}{W^u} = \frac{\beta^u}{\beta^s} \frac{q}{1-q} \tag{112}$$

A rise in skill supply leads to a rise in wage inequality. The relationships between wages and technology on the one hand, and between technology and the skill proportion on the other hand, are very simple because of the assumption that skilled and unskilled labour are perfect substitutes.

6.4.2 Imperfect substitutes

In a series of papers, Acemoglu (1998, 2002ab, 2003ab) develops variations of models of directed technological change assuming imperfect substitution between factor inputs. They are in continuous time but we adapt to a discrete time setting.[42] We draw mostly on the introductory treatment in the (2002a) paper. Capturing the imperfect substitution between skilled and unskilled labour, economy-wide output is now given by the Constant Elasticity of Substitution production function:[43]

$$Y = \alpha^{\frac{2\alpha}{1-\alpha}} \left[(Nq)^\rho + (M(1-q))^\rho\right]^{\frac{1}{\rho}} \tag{113}$$

where the elasticity of substitution[44] $\sigma \equiv \frac{1}{1-\rho}$. It is clear that $\lim_{\rho \to 1} \sigma = \infty$ yields the perfect substitutes example just discussed. (113) permits various interpretations. We choose one where Y is final output, Nq produces a skilled intermediate y^s and $M(1-q)$ produces an unskilled intermediate y^u. An alternative is to view y^s and y^u as final goods in a utility function for the economy (see Acemoglu, 2002b). These intermediates are purchased by final goods producers in the

[42] We do this because, in this section, we consider comparative statics only.

[43] We present the model as one of expanding technology varieties based on Romer (1990) to keep it consistent with our discussion of the Kiley (1999) model. This is done by Acemoglu (2002b). An alternative is one based on increasing product quality (eg Acemoglu, 1998). The exposition in the 2002a paper is not particular about what form technological progress takes. The terms N, M are replaced by coefficients which augment the productivity of the factor. Furthermore, the exposition subsumes our constant $\alpha^{\frac{2\alpha}{1-\alpha}}$ in each of these terms.

[44] In chapter 2, we considered various elasticity measures. While the models are not explicit about it, this elasticity refers to the Robinson (1933) concept; that is $-\frac{d low \frac{L^s}{L^i}}{d \log \frac{w^s}{w^u}}$. In the two-input context, $\sigma > 0$ and the elasticity concepts associated with constant output factor demands are equivalent. They also for example correspond to the inverse of the Hicks Elasticity of Complementarity.

production of final output. The imperfect substitution between y^s and y^u is what makes skilled and unskilled labour imperfect substitutes. Competitive labour markets ensure wage equals marginal revenue product and the skill premium is:

$$W = (T)^{\frac{\sigma-1}{\sigma}} \left(\frac{q}{1-q}\right)^{\frac{-1}{\sigma}} \tag{114}$$

Because $\sigma < \infty$, we have a conventional substitution effect: the premium is negatively related to the relative skill supply, holding the ratio of technologies constant. This effect was absent from the Kiley (1999) model.

The elasticity of the skill premium with respect to relative technologies is now less than unity, unlike Kiley (1999). The reasoning for this is as follows: If there is a rise in N relative to M, ceteris paribus, there will be a rise in y^s relative to y^u. This means the price of y^s relative to y^u must fall to entice final output producers to use more skilled intermediates relative to unskilled intermediates.[45] From (113), equilibrium in the economy gives the price of the skilled intermediate relative to that of the unskilled intermediate as:

$$p = \left(\frac{y^s}{y^u}\right)^{\rho-1} = \left(\frac{y^s}{y^u}\right)^{\frac{-1}{\sigma}} \tag{115}$$

So, because $\frac{y^s}{y^u}$ rises, p falls. In turn, this reduces the marginal revenue product of skilled labour relative to that of unskilled labour, which has a negative effect on the skill premium. This negative influence counters the original positive effect on the relative productivity of skilled labour and hence its relative wage. If $\sigma > 1$, the relative price of y^s does not need to fall by a lot to accommodate the increased supply, so the negative effect on the relative marginal revenue product of skilled labour is small. Therefore, the positive physical productivity effect dominates. If $\sigma < 1$, the negative effect dominates and the skill premium falls. Acemoglu (2002a) concludes skill-augmenting technological change (a rise in T) is only skill-biased if $\sigma > 1$.

The analysis of the preceding paragraph was one of a given rise in T, but the directed technological change literature has at its core arguments about what may cause T to rise. Like Kiley (1999), potential patent holders compare the relative costs and relative values of adopting skilled or unskilled technologies. The relative value of developing a skilled machine depends on the market size effect, and is thus positively related to the relative proportion of skilled workers as in Kiley (1999), but also depends on the price effect. This is given by (115), which implies a negative relationship

[45] The treatment in Acemoglu (2002a) is based on the utility function interpretation.

between the relative skill supply and the relative attractiveness of a skilled technology.

To understand the overall effect, we rely on a similar intuition to before. In equation (114), we had the effects of relative technological variety on relative marginal products of labour. Now we have the effects of relative labour quantities on the relative returns to machines, which in turn affect relative profitability. If we have a rise in the relative supply of skilled labour, this increases the relative physical productivity of each machine (this is shown explicitly in the perfect substitutes set up; cf. (106)). This is the market size effect. However, relatively more skilled labour translates to relatively more skilled intermediates, which through (115) translates to a relative fall in the price of skilled intermediates. In turn, this reduces the relative demand for skilled technologies.

How big this price effect is depends on the size of σ. The relative value of a skilled technology is proportional to $\left(\frac{q}{1-q}\right)^{\frac{\sigma-1}{\sigma}}$. If σ is big, price need only fall a little to accommodate a rise in $\frac{y^s}{y^u}$, so the price effect is smaller than the market size effect and there is a positive relationship between the relative skill supply and the relative attractiveness of skilled technologies. Furthermore, through another price effect, the relative value of a skilled machine is proportional to $(T)^{\frac{-1}{\sigma}}$. This reflects the fact that a higher relative quantity of skilled machine varieties will reduce p and hence negatively affect relative demand. Thus we have (Acemoglu, 2002b):

$$V = (T)^{\frac{-1}{\sigma}} \left(\frac{q}{1-q}\right)^{\frac{\sigma-1}{\sigma}} \tag{116}$$

For the case of perfect substitutability, this corresponds to (110c) from the Kiley (1999) model.

Turning to the costs of developing technologies, various modelling options have been employed in the literature. They share in common the characteristic that R&D firms incur costs in the form of units of final output to research either skilled or unskilled technologies. The return to spending may differ between skilled and unskilled machines. In a stochastic setting, the probability of developing a new technology is proportional to R&D effort. Costs are linear in effort and there are decreasing or constant returns to effort (Acemoglu, 1998). A deterministic specification is given in Acemoglu (2002b), where the rate of advance is a constant proportion of R&D spending. This corresponds to the "lab equipment" model of R&D (Rivera-Batiz & Romer, 1991) and implies a constant deterministic cost of developing a new machine.[46] Acemoglu (2002b) also allows for knowledge-based R&D (Rivera-Batiz & Romer, ibid.), where spillovers from past research to current R&D activity are necessary to sustain growth. This would imply a rise in T today would make future

[46] In the continuous time setting, this is formally written as $\dot{N} = \beta^s E^s$ and $\dot{M} = \beta^u E^u$. Given the constant returns to effort, we can interpret this as a constant marginal cost of development.

research into skilled machines relatively cheaper.

The simplest specification, which is sufficient for this chapter, is that given in Acemoglu (2002a), where it simply costs β units to develop a new technology of either sort deterministically. Under this set up, the relative cost of producing a skilled technology is unity. By free entry, this requires $\frac{V^s}{V^u} = 1$. Then, by (116),

$$T = \left(\frac{q}{1-q}\right)^{\sigma-1} \tag{117}$$

This equation summarises the trade-off between the price and market size effects. If $\sigma > 1$ the price effect is small and the market size effect dominates. In (114), we showed the relationship between the skill premium and skill supply, holding technology constant. Substituting from (117) allows us to account for the directed technology effect:

$$W = \left(\frac{q}{1-q}\right)^{\sigma-2} \tag{118}$$

This shows that a rise in the skill supply will lead to a rise in the skill premium if $\sigma > 2$. In other words, σ must be large enough for the positive effect on the premium arising from SBTC to outweigh the conventional substitution effect. In the Acemoglu (2002b) model with knowledge-based R&D, σ need not be as high.[47]

Although $\sigma > 2$ is needed for a rise in relative skill supply to lead to a rise in the skill premium, any value of $\sigma \neq 1$ will lead to SBTC. This was neglected by Acemoglu (1998) but shown in Acemoglu (2002b). If $\sigma > 1$, we see from (117) that T will rise. Acemoglu (2002b) calls this skill augmenting technological change; change which affects the marginal physical product of skilled labour. We also see in (114) that, with $\sigma > 1$, skill augmenting technological change is also skill-biased. For $\sigma < 1$, T falls, but this fall leads to a rise in the skill premium because $\sigma < 1$. The intuition for this comes from the discussion of the price effect under (114): a fall in the relative number of skilled machines raises their price. Because $\sigma < 1$, this rise is large, so there is a positive effect on the skill premium. Only for $\sigma = 1$ is there no effect.

6.5 Our models of SBTC and wage inequality for developing countries

The models developed in chapters 7 and 8 allow for a number of departures from the existing literature. First, instead of firms in a technology sector paying to develop technologies, we have firms in an import sector that acquire the licence for a new product in exchange for units of

[47] The simplified intuition for this is that a rise in skill supply, if it leads to a relative rise in the development of skilled machine varieties, will also lead to further rises in skilled machines varieties such that the SBTC effect is bigger.

output exported. Second, we assume that skill-biased technologies advance faster than technologies complementing unskilled labour in the technologically advanced countries. We call this Northern skill-biased technological change (NSBTC) and model this by building on the framework laid by (109). NSBTC will feed directly into the pace at which technologies are adopted by the South so that developing countries can have skill-biased technological progress and wage inequality because the North does. Besides this substantive result, our modelling is novel in that it allows for changes in the relative costs of adopting skilled technologies. Existing work only considers changes in the relative values of these technologies, which are driven by relative skill supply.

Section 6.3.2 motivated why relative costs might differ or evolve. However, the central aim of this thesis is the role of skill supply, which allows for the third departure. In addition to one-off changes in skill supply, we allow for a gradual change in the skill composition. To capture the effects of education reforms on the entire population, we model changes in the proportion of cohorts being educated as they enter the labour force and replace those cohorts that die. This generates periods of growth in the skill composition of the population.

6.6 Conclusion

We have presented evidence that SBTC is a potential contributor to documented rises in wage inequality. We have considered how technological change can be directed. This has been formally shown in models by Kiley and Acemoglu. In particular, a rise in the skill supply can lead to SBTC and a rise in wage inequality. This has implications in the context of this thesis, which investigates the interactions between skilled and unskilled labour. We will therefore adapt such models to a developing country setting in a number of ways in chapters 7 and 8.

We have also presented evidence that the skill-bias of technological change in developing countries can be directed by the skill-bias of change in technological leaders. We will build this into the formal models in chapters 7 and 8 by showing how technological developments in the North can affect relative costs of development in the South.

7 A model of SBTC and wage inequality in a developing country: expanded educational access and global technology developments

7.1 Introduction

This thesis examines what effects a rise in the supply of skilled labour will have on demand for unskilled labour. We have so far done so assuming a constant technology. Here, we concentrate on the role of changes in technology that may favour skilled workers. In particular, our model falls within the literature on directed technological change discussed in chapter 6. Directed technological change offers a possible explanation for why the rise in supply of skilled workers coincided with a rise in the skill premium in some countries.[48] A useful feature for our purposes is the possible predictive value of expanded education access.

We use this class of model to see the possible effects of expanded education access on SBTC in developing countries. To our knowledge, the only other paper with an emphasis on developing countries is work by Rahman (2005). His work combines directed technological change with endogenous demographic change. The emphasis there is on growth equilibria and convergence patterns, so the focus, although complementary, is rather different.

In the empirical chapters, we modelled the interactions between more than two factors. This allowed for them to be complements or substitutes. Our models in this chapter and the next will be set up such that skilled and unskilled labour are necessarily p-substitutes (or, equivalently in the two factor case, q-complements). This chapter further assumes skilled and unskilled labour are perfect (p-) substitutes, which means that there will be no standard substitution effects from changes in the skill composition. This will allow us to focus exclusively on the role of technology in a relatively straightforward manner. Chapter 8 will relax this assumption.

As argued in chapter 6, there is evidence that the skill-bias of technology adoption in developing countries is influenced by the relative attractiveness of skilled technologies. Caselli & Coleman (2000) present a correlation between the relative factor shares of skilled and unskilled labour and a measure of the skill-bias of technology. Chapter 6 also presented evidence from Berman, Bound and Machin (1998) that the skill-bias of technological change in developing countries is affected by technology patterns in developed countries. Savvides & Zachariadis (2005) use manufacturing data to find that developing countries undertake no own R&D but rely on foreign technology transfer.

[48] For a critique of these models, see Sanders (2005).

Berman & Machin (2000) find that the same industries experiencing SBTC in the South in the 1980s were those experiencing it in the North in prior decades. We therefore have evidence that skill bias in developing countries is influenced both by external skill-biased technological progress and by domestic conditions like relative skill supply.

This chapter presents a model that captures both internal and external influences on the skill bias of technology adoption. We build on existing work by allowing for a gradual change in the skill composition rather than simply one-off changes. To capture the effects of education reforms on the entire population, we model changes in the proportion of cohorts being educated as they enter the labour force and replace those cohorts that die. Modelling cohort effects is particularly relevant to South Africa, where the change of government has coincided with attempts to broaden access to education.[49] Using a Markov model, we see expanded education generates periods of growth in the skill composition of the population. The model will show this leads to potentially long periods of accelerated wage inequality. We also show that the unskilled lose not only relative to skilled workers, but that they receive lower wages than if there had been no rise in skill supply.

We adapt our model to a developing country setting. Instead of firms in a technology sector paying to develop technologies, firms in an import sector acquire the licence for a new product in exchange for units of output exported. This allows us to model the influence of global technological patterns. To our knowledge, this has not been done in the directed technological change framework. We assume that, in the technologically advanced countries, skill-biased technologies advance faster than technologies complementing unskilled labour. We call this Northern skill-biased technological change (NSBTC).

Northern skill-biased technological change will directly affect the costs of acquiring skilled rather than unskilled technologies from abroad and hence influence the pace at which technologies are adopted by the South. A key result is that developing countries can have skill-biased technological progress because the North does. In our model, this means there can be a rise in wage inequality simply because of international technology patterns.

Section 7.2 describes the model, establishing the key relationships between wages, technology adoption, skill supply and global technology patterns. Section 7.3 presents the steady-state evolution of the economy, showing how NSBTC leads to rising wage inequality along the steady-state growth path. To be clear, by "rising" wage inequality, we mean the first derivative of the ratio of

[49] Van der Berg (2001) shows that, after 1994, major shifts of fiscal resources from traditonally white to traditonally black schools took place. This was not accompanied by major changes in teacher:pupil ratios nor outcomes. Chisolm (2004) provides a summary of the intended measures taken. These include fiscal distribution and a largely unsuccessful teacher redeployment programme.

skilled to unskilled wages with respect to time is positive. Section 7.4 uses the model to examine the effects of a one-off rise in skill supply, showing this is ultimately a levels effect on wage inequality. Section 7.5 models the demographic effects of expanded educational access with a Markov model. It can lead to temporary but potentially long periods of accelerated wage inequality. "Accelerated" wage inequality means the first derivative is higher than along the steady state growth path. Section 7.6 concludes the chapter with a brief discussion of potential links between domestic and international effects on wage inequality.

7.2 Model description

In this section, we describe the model assuming a constant skill composition. Producers use skilled labour (and a variety of machines that complement skilled labour) as well as unskilled labour (and machines that complement unskilled labour) to produce final output. We describe the forces driving the adoption of skilled and unskilled technologies. Any agent can decide to acquire from a technologically advanced nation a licence. The licence carries the exclusive right to import and distribute a particular technology, which works with either skilled or unskilled labour. The decision to acquire a licence will depend on the domestic skill composition, which affects the potential revenues from a licence and will depend on how advanced technologies are abroad, which affects the cost of a licence. We will show how the skilled (unskilled) wage is directly proportional to the number of skilled (unskilled) machine varieties available.

7.2.1 The population and labour force

The economy has a constant population $L = 1$ consisting of portion q skilled workers and $1 - q$ unskilled workers. Consumer i, skilled or unskilled, has utility function

$$U_{it} = \sum_{h=t}^{\infty} G_{ih} (1+r)^{-h+t}, \tag{119}$$

where G is output consumed. It is linear[50] and pins down the interest rate at r for all t. Consumers earn wages and the profits from any licences they may hold.

[50] The role of consumers in models of this class is to describe the evolution of consumption in terms of the subjective rate of time preference and the interest rate. CRRA utility functions are commonly used. In those models, the interest rate is determined by an arbitrage equation between the value of developing new technologies and the bond market. This is generally possible because the marginal cost of adopting technologies is exogenously given (see Barro & Sala-i-Martin (2004) for this class of model). Here, in contrast, the marginal cost will be a function of the number of technologies adopted so there would be a co-determinacy between interest rates, marginal cost and the value of intermediates that cannot be resolved with only two equations. It is for this reason that we use (119), where the interest rate equals the subjective rate of time preference. While we could assume the interest rate is determined by the world interest rate, the utility function assumed also removes any need to use the capital market for risk sharing (Aghion & Howitt, 1992).

7.2.2 Production

The economy has perfectly competitive producers of final output and monopolist suppliers of a variety of machines. There is one supplier for each type of machine. Machine suppliers are monopolists because they had to pay for the licence to import and sell that machine variety. In aggregate, final output can be used for consumption of final goods, to import machines for further production, or to acquire licences for new types of machine.

Using a variety expansion model (Romer, 1990) and building on the presentation in Barro & Sala-i-Martin (2004), the linearly homogeneous production technology for final goods is:

$$Y_{it} = (L^s_{it})^{1-\alpha} \sum_{j=1}^{N_t} (X_{ijt})^\alpha + (AL^u_{it})^{1-\alpha} \sum_{j=1}^{M_t} (Z_{ijt})^\alpha \tag{120}$$

Y_{it} is output for firm i at time t. L^s_{it} and L^u_{it} are skilled and unskilled labour. X_{ijt} is machine input of type j used by firm i at t. It is the quantity of each of N machines (capital), which complements skilled labour. Similarly, Z_{ijt} is the quantity of each of M machines complementing unskilled labour. Capital depreciates fully in each period.[51] $A < 1$ for unskilled labour makes production a function of effective units of labour, with the coefficient for skilled labour normalised to one. We will refer to N and M as the number of skilled machines and unskilled machines.

The production technology, which is employed in Kiley (1999), implies the elasticity of substitution between the skilled and unskilled processes is infinite. As argued in Rahman (2005), this functional form reflects the phenomenon that there are discrete different ways to produce a good and the choice of method depends on a region's endowments.

The price of final output is unity. Firms are profit maximisers and the quantity of each skilled machine demanded by each firm is such that the marginal product of the machine equals its price and $\frac{X_{ijt}}{L^s_{it}} = \left(\frac{\alpha}{P^X_{jt}}\right)^{\frac{1}{1-\alpha}}$. For the economy as a whole, we can condition demand for skilled machines on the quantity of skilled labour. Because final goods are produced using a constant returns to scale technology, we know that, in equilibrium, economy-wide demand for each skill-biased intermediate j must be:

$$X_{jt} = q\left(\frac{\alpha}{P^X_{jt}}\right)^{\frac{1}{1-\alpha}} \tag{121}$$

q is the quantity of skilled labour available to the economy. P^X_{jt}, the price of each skilled machine, is set by the firm holding the licence for that type of machine. Firms acquire this licence by importing

[51] This assumption is made for convenience. In endogenous growth models, this assumption is usually not important for steady-state equilibria, although it can affect the path of adjustment. See Barro & Sala-i-Martin (2004).

it from abroad in exchange for exports of Y at a cost described below.[52] Firms in the technology import sector must receive ex post profits to persuade them to incur the ex ante licence cost. Once the fixed cost of acquiring the licence from abroad has been incurred, it costs 1 exported unit of Y, which has a price of 1, to import each machine. Using (121), the own-price elasticity of demand is $\frac{1}{1-\alpha}$ for all machines of any type. Therefore each monopolist sets a profit maximizing price of $\frac{1}{\alpha}$ for all j, t. Thus, demand for each and every skilled intermediate good in the economy is equal and given by:

$$X = \alpha^{\frac{2}{1-\alpha}} q \tag{122}$$

Similarly, demand for intermediates that use unskilled labour is:

$$Z = A\alpha^{\frac{2}{1-\alpha}}(1-q) \tag{123}$$

Output for the economy is given by:

$$Y_t = \alpha^{\frac{2\alpha}{1-\alpha}} \left[N_t q + AM_t(1-q)\right] \tag{124}$$

From (120), rises in X and/or Z (a rise in the quantity of every machine used) will encounter diminishing returns. However, the fact that N and M enter additively ensures constant returns to increases in the variety of inputs. As long as N or M rise, Y will rise. This is the basis for endogenous growth. The model literally considers N and M as the number of different types of machines a firm can use, but can also be thought of as the technological complexity of the firms' production processes (Barro & Sala-i-Martin, 2004). This latter interpretation is why, as described by the production technology, new machine types will always be employed.[53]

7.2.3 Technology adoption

We are interested in establishing the equilibrium number of technologies N and M at time t. We start by describing the decision of a potential licence holder whether or not to acquire a licence for a technology from abroad. At any time t, the agent considers if the value of the licence exceeds the cost. The agent would incur the cost at t and start receiving profits at $t+1$. The value is the

[52] Because utility is linear, we can either think of our economy as having frictionless credit markets and/or the consumers being shareholders in the firms (Aghion & Howitt, 1992). In either case, consumers are the final claimants on the profits.

[53] As an alternative to variety expansion models, we can also have technological progress in the form of increases in the quality of the machines used (Aghion & Howitt, 1992; Grossman & Helpman, 1991c). Such models argue why it is always in the interest of final output producers to employ the highest quality of machine available. Grossman & Helpman (ibid) and Barro & Sala-i-Martin (2004) compare machine quality and machine variety models, finding many of the positive conclusions are the same.

discounted present value of all future profits. Because profits for all skilled machine types are equal, the value is the same for all skilled machine types. Thus the value of a skilled licence at time t is $V_t^s = \sum_i^{\infty} \left(P_{t+i}^X - 1\right) X_{t+i} \left(1+r\right)^{-i}$. Recalling $P^X = \frac{1}{\alpha}$, using (122) and defining $\Omega \equiv (1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}$, the per period profit from a licence for a skilled machine is Ωq_t. Summing over infinite periods when the skill composition is constant, the present value of a skilled licence is:

$$V^s = \Omega \frac{q}{r} \tag{125}$$

Similarly:

$$V^u = A\Omega \frac{1-q}{r} \tag{126}$$

Equations (125) and (126) capture the market size arguments of Schmookler (1966) introduced in chapter 6: the value of acquiring a particular type of machine technology is directly proportional to the size of the market, which is determined by the quantity of workers available to work with the machine.

The specification of costs is an important feature of the model. The cost of acquiring a licence for a new technology variety depends on how many technology varieties of a particular type - skilled or unskilled - exist in the economy relative to the stock of internationally developed technologies of that type. If an economy is relatively far from the technology frontier, then it can acquire a licence for a machine that is relatively old and hence relatively cheap (but equally productive). We adapt the modelling of costs in Kiley (1999) so that the cost of acquiring a licence to import a skilled machine depends on the number of machines already in use (N) relative to the number available in the world (R^s):

$$C_t^s = \left(\beta^s \frac{N_t}{R_t^s}\right)^{\kappa} \text{ if } \frac{N_t}{R_t^s} < 1, \text{ where } \kappa > 0 \tag{127a}$$

$$C_t^s = \infty \text{ if } \frac{N_t}{R_t^s} \geq 1 \tag{127b}$$

The cost of acquiring from abroad a licence for an unskilled machine is given by:

$$C_t^u = \left(\beta^u \frac{M_t}{R_t^u}\right)^{\kappa} \text{ if } \frac{M_t}{R_t^u} < 1, \text{ where } \kappa > 0 \tag{128a}$$

$$C_t^u = \infty \text{ if } \frac{M_t}{R_t^u} \geq 1 \tag{128b}$$

R_t^s is the number of skilled machine varieties available in the world and R_t^u is the number of unskilled varieties. κ is the elasticity of cost of acquiring a technology from abroad with respect to the proportional distance from the frontier. A higher κ denotes costs rise by more as the ratio of the number of machines already in use relative to those available in the world increases. For example, holding N constant, a rise in R^s will lead to a large fall in C^s if κ is high. Should a country be on the frontier for a type of technology, then it is of course infinitely expensive to copy a newer technology, but we expect developing countries to be well within this frontier. β^s and β^u allow for possible differences in the ability to adopt technologies across developing countries and between skilled and unskilled technologies. Factors affecting these parameters might be the regulatory environment or proximity to technological leaders.

The stock of skilled and unskilled technologies available worldwide is assumed to evolve exogenously according to:

$$R^s_{t+1} = \gamma^s R^s_t \tag{129}$$

$$R^u_{t+1} = \gamma^u R^u_t \tag{130}$$

The treatment of basic research as exogenous is appropriate here. A developing country is unable to influence the decisions by first world producers to develop new technologies. $\gamma^s > \gamma^u > 1$ denotes exogenous skill-biased technological change in the North (NSBTC). We assume this in the model, as justified in chapter 6. Differences in growth rates of R^s and R^u are an important component of the model in this chapter and are a key difference with the specification in Kiley (1999).

Any firm is free to acquire a licence. A firm will decide to acquire a licence if the value of the licence, given by (125) or (126), exceeds the cost, given by (127a) or (128a). If the value does not exceed the cost, it does not acquire a licence. By free entry, the value of acquiring a licence can never exceed the cost, so

$$V^s_t \leq C^s_t \tag{131}$$

and

$$V^u_t \leq C^u_t \tag{132}$$

In equilibrium, (131) and (132) hold with equality. Define $V = \frac{V^s}{V^u}$ as the relative value of skilled technologies, $C_t \equiv \frac{C^s}{C^u}$ as the relative cost, $\frac{N_t}{M_t} \equiv T_t$ as the ratio of skilled to unskilled technologies in the developing country, $\frac{R^s_t}{R^u_t} \equiv R_t$ as the skill bias of the world technology frontier, $\beta \equiv \frac{\beta^s}{\beta^u}$ and $\left(\frac{q}{A(1-q)}\right) \equiv Q$ as the effective ratio of skilled to unskilled labour. We can summarise the analysis so far as follows:

Proposition 2 *The ratio of skilled to unskilled technologies is increasing in the relative values of skilled technologies and decreasing in the relative costs of skilled technologies:*

$$T_t = \frac{R_t}{\beta} (V_t)^{1/\kappa} \tag{133}$$

Proof. The result follows when (131) and (132) hold with equality and when we substitute from (127a) and (128a). ■

(133) expresses the ratio of the technologies in terms of relative costs and relative values. This form is general and will carry over to the imperfect substitution and non-constant population cases. For our production technology, in equilibrium,

$$N_t = \frac{R_t^s}{\beta^s} \left(\frac{\Omega q}{r} \right)^{1/\kappa} \tag{134}$$

$$M_t = \frac{R_t^u}{\beta^u} \left(\frac{A\Omega(1-q)}{r} \right)^{1/\kappa} \tag{135}$$

We can use (125) and (126) to show the relative values are given by the relative effective ratio of skilled to unskilled labour:

$$V = Q \tag{136}$$

Therefore, applying (133) to the production technology (120) and a constant population, we can make the following statement:

Corollary 2 *When skilled and unskilled labour are perfect substitutes, the ratio of skilled to unskilled technologies is proportional to the relative effective skill supply (Q) and increasing in the skill-bias of the world technology frontier (R):*

$$T_t = \frac{R_t}{\beta} (Q)^{1/\kappa} \tag{137}$$

Relative values would change over time if the skill composition of the labour force changed, but for now only relative costs change as given by (129) and (130). This will be discussed further in section (7.3).

7.2.4 Wages

This section establishes the link between wages, the world technology frontier and the domestic skill supply. Labour is exogenously supplied to the economy. Individual firms hire each labour unit such that its marginal product equals the wage. For equilibrium in the economy, wages are such that, for a given level of technology and intermediates,

$$W_t^s = (1-\alpha) N_t \left(\frac{X}{q}\right)^{\alpha} \tag{138}$$

$$W_t^u = A^{1-\alpha} (1-\alpha) M_t \left(\frac{Z}{1-q}\right)^{\alpha} \tag{139}$$

Using (122) and (123), we find,

$$W_t^s = (1-\alpha) \alpha^{\frac{2\alpha}{1-\alpha}} N_t \tag{140}$$

$$W_t^u = A (1-\alpha) \alpha^{\frac{2\alpha}{1-\alpha}} M_t \tag{141}$$

(140) and (141) expose one of the simplifications induced by the production function. The terms in labour and intermediates cancel, so we see that wages are proportional to the number of technology varieties. By (134) and (135), we see this translates into a positive relationship between wages and labour supply and between wages and the research frontier:

$$W_t^s = \theta^s \frac{R_t^s}{\beta^s} \left(\frac{q}{r}\right)^{1/\kappa} \tag{142}$$

$$W_t^u = \theta^u \frac{R_t^u}{\beta^u} \left(\frac{1-q}{r}\right)^{1/\kappa} \tag{143}$$

where $\theta^s \equiv \left[(1-\alpha)^{1+\kappa} \alpha^{\left(\frac{1+\alpha+2\alpha\kappa}{1-\alpha}\right)}\right]^{1/\kappa}$ and $\theta^u \equiv \left[A^{1+\kappa} (1-\alpha)^{1+\kappa} \alpha^{\left(\frac{1+\alpha+2\alpha k}{1-\alpha}\right)}\right]^{1/\kappa}$. Let $W_t \equiv \frac{W_t^s}{W_t^u}$ be the skill premium or degree of wage inequality. We can summarize the discussion in the following proposition.

Proposition 3 *When skilled and unskilled labour are perfect substitutes, the skill premium is proportional to the effective relative skill supply and increasing in the skill-bias of the world technology frontier.*

$$W_t = \frac{R_t}{\beta} \frac{(Q)^{1/\kappa}}{A} \tag{144}$$

Proof. Dividing (142) by (143) yields the result. ■

This nests the corollary to Proposition 1 in Kiley (1999).

In order to interpret our setting as one for a developing economy, we need to assume that i) it consists mostly of unskilled technologies - $M_t > N_t$ - and that ii) it still has a positive skill premium

$(W > 1)$ for any skill composition.[54] We impose the restrictions that:

$$A\left(\frac{1-q_t}{q_t}\right) > \left(\frac{\beta^u R_t^s}{\beta^s R_t^u}\right)^{\kappa} \quad (145)$$

$$A\left(\frac{1-q_t}{q_t}\right) < \left(\frac{1}{A}\right)^{1/\kappa}\left(\frac{\beta^u R_t^s}{\beta^s R_t^u}\right)^{\kappa} \quad (146)$$

The first inequality implies that the expense of acquiring unskilled machines relative to skilled machines is sufficiently low to overcome the lower productivity of unskilled workers, so there are more unskilled technologies. The second inequality implies that the greater variety of unskilled machines used by unskilled workers, which makes them more productive, is not sufficient to overcome their inherently lower productivity relative to skilled workers, so there is a positive skill premium. The implied additional criterion $\left(\frac{1}{A}\right)^{1/\kappa} > 1$ is naturally met when $A < 1$. These conditions are not needed for the main results of this chapter, only those on the impact on growth.

7.3 Steady-state evolution of the economy

7.3.1 Technology and wages

By assumption, we have a constant rate of skill-biased technological change in the North (NSBTC). $\gamma^s > \gamma^u$ such that $\frac{R_{t+1}}{R_t} > 1$. We will say the economy is in steady state whenever the population is constant. This allows us to present a key result:

Theorem 3 *When skilled and unskilled labour are perfect substitutes, skill-biased technological change in the North is transmitted directly to SBTC in the South such that, on the steady state growth path:*

$$\frac{T_{t+1}}{T_t} = \frac{\gamma^s}{\gamma^u} \quad (147)$$

Proof. From (134) and (129):

$$\frac{N_{t+1}}{N_t} = \gamma^s \quad (148)$$

From (135) and (130):

$$\frac{M_{t+1}}{M_t} = \gamma^u \quad (149)$$

Recalling $T_t \equiv \frac{N_t}{M_t}$ yields (147). ■

The simple expression (147) reflects the persistent SBTC observed in developing countries and delivers one of the key messages of the chapter: holding skill composition constant, product variety

[54] $0 < q < 1$

expansion can be skill-biased simply because technological change research is skill-biased in the North ($\gamma^s > \gamma^u$). It captures the observation that:

> "*...developing countries must be choosing from a menu of best practices that includes an ever-increasing proportion of skill-biased technologies.*" - (Berman & Machin, 2000:3)

This phenomenon occurs even if the population is largely unskilled. Rather than complementing their abundance of unskilled labour, such countries acquire technologies that complement their skilled workers, giving rise to the phenomenon of "inappropriate" technology transfer (eg Atkinson & Stiglitz, 1969).

Inspection of (147) would suggest that T gets very large over time and that output would eventually consist solely of skilled processes. However, we note that the developing country restrictions set up in (145) and (146) would no longer hold at some finite value of t. Our model for technology adoption would cease to apply. Furthermore, we do not literally believe the phenomenon of NSBTC is a permanent one. This would be inconsistent with the directed technological change literature, which argues technological change in the North is only skill-biased in response to changing profit incentives. We abstract from this important issue for modelling purposes only.

It is easy to confirm that skilled wages grow at γ^s and unskilled wages grow at γ^u, which allows us to make the following statement about the link between global technological patterns and wage inequality in a developing country.

Corollary 3 *When skilled and unskilled labour are perfect substitutes, skill-biased technological change in the North is transmitted directly to wage inequality in the South such that, on the steady state growth path,*

$$\frac{W_{t+1}}{W_t} = \frac{\gamma^s}{\gamma^u} \tag{150}$$

7.3.2 Consumption and GDP growth

We briefly consider the evolution of output and consumption. GDP growth is given by:

$$\frac{Y_{t+1}}{Y_t} \equiv \gamma^y = \frac{N_{t+1}q + AM_{t+1}(1-q)}{N_t q + AM_t(1-q)} \tag{151}$$

We can show[55] that, if $\gamma^s > \gamma^u$, then $\gamma^s > \gamma^y > \gamma^u$ and if $\gamma^s = \gamma^u = \gamma$, then $\gamma^y = \gamma$. This last result coincides with standard growth models, which have only one type of technology (see for example Romer, 2001): GDP (and hence GDP per capita) grows at a pace matching the rate of technological progress. Furthermore, the lower is q, the closer is the growth rate of output to the growth rate of unskilled rather than skilled technology. We will discuss the effects of changes in the skill composition on growth in the next section but complete the description of the model by analysing how GDP is allocated.

We still assume the skill composition is constant but allow $\gamma^s \neq \gamma^u$. Consumers earn income in the form of wages and returns on any claims on licences for intermediates they hold. Aggregate wages are $W_t^s q + W_t^u(1-q) = (1-\alpha)Y_t$. Returns on licences are $r(N_t V^s + MV^u)$. By (125) and (126), returns are $\Omega\left[N_t q + AM_t(1-q)\right] = (1-\alpha)\alpha Y_t$. Thus consumer income Y^C is:[56]

$$Y_t^C = Y_t - \alpha^2 Y_t \tag{152}$$

However, using (122) and (123), $\alpha^2 Y_t = N_t X + M_t Z$, so

$$Y_t^C = Y_t - N_t X - M_t Z \tag{153}$$

Let all assets (licences) owned in the economy $B_t \equiv N_t V^s + M_t V^u$. Analogous to Barro & Sala-i-Martin (2004), consumption in the economy is thus given by,

$$\begin{aligned} G_t &= Y_t^C - (B_{t+1} - B_t) && (154)\\ &= (1-a^2)Y_t - (B_{t+1} - B_t), && (155) \end{aligned}$$

where $(B_{t+1} - B_t)$ is the consumption foregone in the form of exports in exchange for new licences (for both types of intermediate). Using the fact that $(B_{t+1} - B_t) = \frac{\Omega}{r}\left[qN_t(\gamma^s - 1) + A(1-q)M_t(\gamma^u - 1)\right]$, algebra shows

$$\frac{G_{t+1}}{G_t} = \frac{r(1-\alpha^2)Y_{t+1} - \frac{Y_{t+2}-Y_{t+1}}{\alpha(1-\alpha)}}{r(1-\alpha^2)Y_t - \frac{Y_{t+1}-Y_t}{\alpha(1-\alpha)}} = \gamma^y, \tag{156}$$

so consumption grows at the same rate as output. If $\gamma^s = \gamma^u = \gamma$, consumption, GDP, wages and income from licences all grow at γ.

[55] Write $\gamma^y = \frac{\gamma^s N_t q + \gamma^u M_t A(1-q)}{N_t q + M_t A(1-q)}$. If $\gamma^s = \gamma^u = \gamma$, $\gamma^y = \frac{\gamma[N_t q + M_t A(1-q)]}{N_t q + M_t A(1-q)} = \gamma$. If $\gamma^s = \lambda\gamma^u$ where $\lambda > 1, \gamma^y = \frac{\gamma^u[\lambda N_t q + M_t A(1-q)]}{N_t q + M_t A(1-q)} > \gamma^u$. Also, $\gamma^y = \frac{\gamma^s[N_t q + M_t A(1-q)/\lambda]}{N_t q + M_t A(1-q)} < \gamma^s$.

[56] As expected from the production technology, the wage share of output is $(1-\alpha)$ and the capital share is α, of which α^2 is lost in foregone/depreciated inputs and the remaining $\alpha - \alpha^2$ is earned as monopoly profits.

7.4 A one-off rise in the skill composition

As a prelude to the modelling of improved education access in the next section, this section compares the state and evolution of the economy after an instant change in the steady-state population. Besides providing a useful indicator of the final effects of a rise in the skill composition discussed in the next section, it can be appropriate to model one-off rises. An economy can experience a relatively fast rise in the skill composition if race or ethnicity barriers to highly skilled jobs are removed. In South Africa for example, people who may have been skilled were effectively barred from participating in the skilled labour force and were in general forced to take unskilled jobs. Once this discrimination in the labour market is removed, there can be a relatively fast rise in the skill composition of the effective labour force. We therefore analyse an unexpected rise in q from q_0 to q_1 in time period $t = 0$ to see what impact this has on the varieties of technologies.

7.4.1 Technology

Inspection of (125) shows $\frac{V^s_{jt}|q_1}{V_{jt}|q_0} = \frac{q_1}{q_0}$. A rise in the number of skilled workers raises demand for each available skilled machine and therefore raises the present value of profit received for holding a skilled licence. Therefore, initially, $V^s > C^s$. The number of varieties at $t = 0$ is not affected as it takes one period to import a new technology. However, the economy can jump to the equilibrium quantity the following period. (134) shows the ratio is:[57]

$$\frac{N_t|q_1}{N_t|q_0} = \left(\frac{q_1}{q_0}\right)^{1/\kappa}, t > 0 \tag{157}$$

The ratio holds for all t as variety expansion continues at rate γ^s. Similarly, $\frac{V^u_{jt}|(1-q_1)}{V^u_{jt}|(1-q_0)} = \frac{1-q_1}{1-q_0}$ and the *equilibrium* value of unskilled technologies falls:

$$\frac{M_t|(1-q_1)}{M_t|(1-q_0)} = \left(\frac{1-q_1}{1-q_0}\right)^{1/\kappa}, t \geq t^* \tag{158}$$

Initially, $V^u < C^u$ but the number of varieties available cannot fall; the technologies have already been acquired. The *actual* number of unskilled intermediates remains constant until the research frontier (R^u) has advanced sufficiently. In other words, because the value of unskilled intermediates has fallen, the cost C^u must fall sufficiently before technology adoption can resume. This occurs

[57] The negative effect of κ can be explained as follows: to restore equilibrium, the rise in value generated by higher q must be matched by a rise in cost. The higher the value of κ, the faster cost rises as N_t moves closer to the technology frontier and thus the smaller the rise in N_t needed to restore the equivalence.

when $\frac{M_{t^*}|(1-q_1)}{M_0|(1-q_0)} > 1$. By (130) and (135), $(\gamma^u)^{t^*} > \left(\frac{1-q_1}{1-q_0}\right)^{1/\kappa}$ so that:

$$t^* > \frac{\ln \frac{1-q_0}{1-q_1}}{\kappa \ln \gamma^u} \tag{159}$$

The time required is shorter if the population change is smaller and the unskilled research frontier advances faster. A large value of κ means the cost of a new technology decreases fast as the economy falls further behind the advancing frontier. At t^*, technology imports resume at rate γ^u.

7.4.2 Wages

At $t = 0$, the quantity of machines and variety of technologies is unaffected by the change in skill composition. Holding the quantity of machines demanded and technology adoption constant, (138) shows skilled wages are lower than they would have been, in accordance with downward sloping labour demand. Similarly, (139) shows unskilled wages are higher than they would have been. We thus see a fall in wage inequality at t_0. However, machine demand (X_j and Z_j) can adjust to the new skill composition for $t > 0$. (140) and (141) show that, still holding the number of varieties constant, the effect of the change in machine demand on wages exactly cancels the effect of the change in labour supply. Wage behaviour is thus driven solely by the nature of technology adoption in the perfect substitutes case.

Because the variety of unskilled intermediates remains constant, unskilled wages also remain constant until t^*. Thereafter, they rise at rate γ^u. Therefore, relative to what they would have earned had the change in skill supply not taken place, unskilled workers suffer a loss of wages of $\left(\frac{1-q_1}{1-q_0}\right)^{1/\kappa}$ in all time periods ($t > 0$). However, because the number of varieties does not fall, unskilled wages never fall below the level in the period immediately preceding the rise in skill supply. We summarise this with the following statement:

Corollary 4 *If skilled and unskilled labour are perfect substitutes, a one-off rise in skill supply will cause unskilled workers to earn less than they would have throughout the steady-state growth path, but they do not experience an actual drop in wages.*

Similarly, skilled wages jump when the population rises before resuming their normal rate of increase γ^s. Wage inequality jumps by $\left(\frac{q_1}{q_0}\right)^{1/\kappa}$ at $t = 1$ and continues to grow at γ^s until $t < t^*$. Once adoption of unskilled technologies resumes, wage inequality grows at $\frac{\gamma^s}{\gamma^u}$ as before. The net effect of skill supply on wage inequality is a levels effect:

$$\frac{W|q_1}{W|q_0} = \left(\frac{q_1(1-q_0)}{q_0(1-q_1)}\right)^{1/\kappa}, t \geq t^* \tag{160}$$

The results presented are so far consistent with Proposition 2 and its corollary in Kiley (1999), as summarised in chapter 6. It is important to stress that the one-off change in population causes annual rises in wage inequality only as long as M_t is stagnant. The main effect is in levels. If one believes the skill-biased technological change observed in developing countries is more than simply a transition, the one-off changes in the skill composition produced by models in this class are an inadequate explanation for developing economies. Steady state SBTC is generated only because we believe $\gamma^s > \gamma^u$.

7.4.3 Output effects

Before turning to gradual population changes, we consider the effects of the rise in the skill composition on GDP and GDP growth for a developing country, as defined by restrictions (145) and (146).

Proposition 4 *For a developing country, a rise in skill supply will increase GDP. If there is NSBTC, a rise in skill supply will increase GDP growth in a developing country.*

Proof. Partially differentiating (124), holding the levels of technology constant, $\frac{\partial Y_t}{\partial q} = N_t - AM_t$. Allowing for technology to change, $\frac{dY_t}{dq} = \frac{1+k}{k}(N_t - AM_t)$. It is clear that GDP rises iff $N_t > AM_t$, but (146) ensures this condition holds. Also, let $\gamma^s = \lambda\gamma^u$. Using (151),

$$\frac{d\frac{Y_{t+1}}{Y_t}}{dq} = \left(\frac{\gamma^u}{(Y_t)^2}\right)[(\lambda N_t - AM_t)(N_t q + (1-q)AM_t) - (N_t - AM_t)(\lambda N_t q + (1-q)AM_t)], \tag{161}$$

which is strictly positive iff $\lambda > 1$; that is, iff $\gamma^s > \gamma^u$. ■

Hence, in aggregate, a rise in the skill composition of the economy allows the economy to exploit the skill-biased research being done in the North and grow faster. The finding that a rise in skill composition can boost growth is also consistent with theoretical models and cross country empirical growth studies, which document a positive correlation between skills acquisition and growth (e.g. Barro & Sala-i-Martin, 2004). The reasons are different to those proposed by Nelson & Phelps (1966), as discussed in chapter 6. Furthermore, this model suggests that, even if unskilled workers lose out in wage terms, the overall gain should be large enough for them to be compensated. One caveat is that we have not accounted for any resource or opportunity costs.

7.5 An improvement in educational access

7.5.1 Demography

Any improvement in the quantity and quality of education will raise the skill composition of school-age cohorts rather than an economy as a whole. Rather than experiencing an instantaneous shift, an economy's skill composition can rise only gradually as cohorts of people dying are replaced by those born and going to school.

To model this, we assume people can exist in one of three states. They can either be skilled, unskilled or deceased. The labour force (people that are not deceased) is fully employed and normalised to one such that proportion q_t of the labour force is skilled and $1 - q_t$ is unskilled. In each period, both skilled and unskilled workers have probability δ of entering the deceased state (dying). In turn, proportion Ψ of those in the deceased state are reborn as skilled people and $1 - \Psi$ are reborn as unskilled people. One alternative interpretation is that, upon dying at the end of one period, people are immediately reborn and exist as children at the start of the next period. Proportion Ψ of children receive schooling to a given level before entering the workforce. The dynamics are captured by the following Markov process:

$$\begin{pmatrix} q_{t+1} \\ 1 - q_{t+1} \\ D_{t+1} \end{pmatrix} = \begin{pmatrix} 1-\delta & 0 & \Psi \\ 0 & 1-\delta & 1-\Psi \\ \delta & \delta & 0 \end{pmatrix} \begin{pmatrix} q_t \\ 1 - q_t \\ D_t \end{pmatrix} \tag{162}$$

The solution to the system is:

$$\begin{pmatrix} q_{t+1} \\ 1 - q_{t+1} \\ D_{t+1} \end{pmatrix} = \begin{pmatrix} \Psi \\ 1-\Psi \\ \delta \end{pmatrix} + (\Psi - q_0) \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} (1-\delta)^t \tag{163}$$

q_0 is the proportion of skilled workers at the initial date $t = 0$. At any given time, the number of deceased people is constant at δ; all δ people are reborn to replace the δ who die, maintaining a constant living population.

If the proportion of the new cohort being educated Ψ is the same as the proportion of the labour force that is skilled, then the proportion of skilled workers remains constant at Ψ; that is, $q = \Psi$. If not, we see that the proportion of skilled workers approaches its steady state Ψ as t gets large.

In particular:

$$\frac{q_{t+1}}{q_t} = 1+\left(\frac{\Psi - q_t}{q_t}\right)\delta \tag{164a}$$

$$\frac{1-q_{t+1}}{1-q_t} = 1-\left(\frac{\Psi - q_t}{1-q_t}\right)\delta \tag{164b}$$

This is likely to be a slow transition. A birth/death rate of 2% per annum - corresponding incidentally to an expected working life of 50 years - implies that the half-life of $\Psi - q_t$ is approximately 34 years.

Because each consumer's lifespan is uncertain, agents may leave unexpected and unintended bequests or debts. Therefore, building on Blanchard (1985) and Yaari (1965), each consumer takes a bet offered by insurance companies. If the consumer dies during the time period, she gives up all her assets. If she remains alive, she receives a certain portion of her assets. Insurance companies offer this risklessly and without profit. The actuarially fair portion they pay out contingent on the consumer staying alive is $\frac{\delta}{1-\delta}$. In other words, insurance companies collect the assets from the δ people dying every year and turn proportion $\frac{\delta}{1-\delta}$ of these assets over to the $1-\delta$ people who stay alive.

We now model the effects of a new education policy that gradually raises the proportion of skilled workers in the economy. We implicitly assume the main driver of educational attainment is ease of access through the supply of education, not shifts in demand by individuals. This assumption is appropriate in the context of widening access to formerly barred segments of the population.[58]

We assume the economy initially has skill proportion q_0, with the number of people being educated $\Psi = \Psi_0$ so that the proportion is constant. At $t = 0$ (t_0), better access to education raises Ψ to $\Psi_1 > \Psi_0$. This is credibly announced at t_0 but only starts to take effect one period later as given by (163). Over time, the economy's proportion of skilled workers will move towards Ψ_1. These demographics are known to all agents in the economy.

7.5.2 Technology adoption

When both N and M are at their equilibrium values, but the skill composition is changing, we refer to this as the equilibrium transition path. However, it is possible for the equilibrium level of M to be below its actual level, such that the acquisition of unskilled machines ceases. This is a period when the economy is not on the equilibrium transition path. We start by describing the paths of N and M on the equilibrium transition path. Thereafter, we present a fuller description of the paths

[58] Acemoglu (1998) and Rahman (2005) allow for endogeneity of the skill supply.

of the values and costs of skilled and unskilled licences, including the equilibrium transition path and periods when unskilled technology acquisition is suspended.

Technology along the equilibrium transition path The cost of acquiring a licence is as in (127a) and (128a). Similarly, the value of a licence is the discounted present value of the profits from selling machines. Recall $P^s = P^u = \frac{1}{\alpha}$ and $\Omega \equiv (1-\alpha)\alpha^{\frac{1+\alpha}{1-\alpha}}$ such that the per period profit from a licence is Ωq_t for a skilled machine and $A\Omega(1-q_t)$ for an unskilled machine. The value of a skilled licence is therefore given by:

$$V_t^s = \Omega \left[\sum_{i=1}^{\infty} \frac{q_{t+i}}{(1+r)^i} \right] \tag{165}$$

Because q_t is changing, we use (163) to find, for a skilled licence,

$$V_t^s = \Omega \left[\frac{\Psi_1}{r} - \frac{(\Psi_1 - \Psi_0)(1-\delta)^{t+1}}{r+\delta} \right] \tag{166}$$

Similarly, for an unskilled licence,

$$V_t^u = A\Omega \left[\frac{1-\Psi_1}{r} + \frac{(\Psi_1 - \Psi_0)(1-\delta)^{t+1}}{r+\delta} \right] \tag{167}$$

The equilibrium varieties of technologies are those that equate their value and cost:

$$N_t = \frac{R_t^s}{\beta^s}\Omega^{1/\kappa} \left[\frac{\Psi_1}{r} - \frac{(\Psi_1 - \Psi_0)(1-\delta)^t}{r+\delta} \right]^{1/\kappa} \tag{168}$$

$$M_t = \frac{R_t^u}{\beta^u}(A\Omega)^{1/\kappa} \left[\frac{1-\Psi_1}{r} + \frac{(\Psi_1 - \Psi_0)(1-\delta)^t}{r+\delta} \right]^{1/\kappa} \tag{169}$$

Recall $\Psi_0 = q_0$ is the original proportion of skilled workers and Ψ_1 is the new proportion of people being educated and the eventual new steady state proportion of skilled workers. The terms inside the square brackets capture the present value of the number of skilled or unskilled workers. As t gets large, the terms converge on the present value of a constant population, as in (125) and (126). Dividing (168) by (169) would give an expression for T_t that fits the general form in Proposition 2.

As is discussed shortly, it is possible for the equilibrium value of M to be lower than the actual value, such that the number of unskilled varieties is stuck. When the population is changing and the actual values of M (and N) are at their equilibrium values, we refer to this as the equilibrium transition path. We can make the following statement about the growth rates of the varieties on the equilibrium transition path.

Theorem 4 *If skilled and unskilled labour are perfect substitutes, improved educational access in a developing country will result in accelerated SBTC such that, on the equilibrium transition path,*

$$\frac{T_{t+1}}{T_t} > \frac{\gamma^s}{\gamma^u} \tag{170}$$

Proof. The growth rate of N is:

$$\begin{aligned} \frac{N_{t+1}}{N_t} &= \gamma^s \left[1 + \frac{\Psi - q_t}{q_t}\delta\right]^{1/\kappa} \text{ (by (134) and (164a))} && (171) \\ &= \gamma^s \left[1 + \phi_t^s\right]^{1/\kappa} > \gamma^s \text{ for finite } t \text{ (by (168))} && (172) \end{aligned}$$

Here, $\phi_t^s = \frac{r\delta(\Psi_1 - \Psi_0)(1-\delta)^t}{\Psi_1(r+\delta) - r(\Psi_1 - \Psi_0)(1-\delta)^t} > 0$ for finite values of t. Also,

$$\begin{aligned} \frac{M_{t+1}}{M_t} &= \gamma^u \left[1 + \frac{\Psi - q_t}{q_t}\delta\right]^{1/\kappa} \text{ (by (135) and (164b))} && (173) \\ &= \gamma^u \left[1 - \phi_t^u\right]^{1/\kappa} < \gamma^u \text{ for finite } t \text{ (by (169))} && (174) \end{aligned}$$

Here, $\phi_t^u = \frac{r\delta(\Psi_1 - \Psi_0)(1-\delta)^t}{(1-\Psi_1)(r+\delta) - r(\Psi_1 - \Psi_0)(1-\delta)^t} > 0$ for finite values of t. (170) follows easily from combining (172) and (174). ■

Corollary 5 *On the equilibrium transition path, the rate of SBTC will fall back towards the steady-state rate of* $\frac{\gamma^s}{\gamma^u}$.

$\lim_{t \to \infty} \phi_t^s = 0$ and $\lim_{t \to \infty} \phi_t^u = 0$ and hence $\lim_{t \to \infty} \frac{T_{t+1}}{T_t} = \frac{\gamma^s}{\gamma^u}$.

It is clear from (172) that growth in skilled technologies is accelerated by the rising proportion of skilled workers. This combines the effects of a fall in cost from an advancing research frontier with the increasing attractiveness of a rising skill composition. As t gets large and the population approaches its steady state, technology growth falls to the steady state rate γ^s. Similarly, (174) shows that growth is less than the steady state value: falling costs are counteracted by the decreasing attractiveness of the market for unskilled technologies. As t gets large, the growth rate rises back to γ^u.

The path of V_t^s and C_t^s (166) can be used to compare the values of a skilled licence with and without the change in educational access. Noting that $V_t^s|\Psi_0 = V_0^s|\Psi_0$ for all t because the skill proportion would have remained constant,

$$\frac{V_t^s|\Psi_1}{V_t^s|\Psi_0} = \frac{\Psi_1}{\Psi_0} - \frac{r}{r+\delta}\left(\frac{\Psi_1}{\Psi_0} - 1\right)(1-\delta)^t \tag{175}$$

$\lim_{t\to\infty} \frac{V_t^s|\Psi_1}{V_t^s|\Psi_0} = \frac{\Psi_1}{\Psi_0}$ so the value eventually changes by the same amount as reported for the one-off change.[59] To characterise what happens initially, we set $t = 0$. In this case $\frac{V_0^s|\Psi_1}{V_0^s|\Psi_0} = \frac{\Psi_1}{\Psi_0} + \frac{r}{r+\delta}\left(1 - \frac{\Psi_1}{\Psi_0}\right)$: the initial jump in value is bigger if δ/r is bigger. If we set $\delta = r$, the initial change is approximately half of the eventual change. After the initial jump, the value continues to rise in small increments towards its steady state value. Figure 1 shows the approximate initial jump and path of $V^s = C^s$ as well as their new steady state values. Using (168), the level of skilled licences in period one relative to what it would have been is given by $\frac{N_1|\Psi_1}{N_1|\Psi_0} = \left[\frac{\Psi_1}{\Psi_0} - \frac{r}{r+\delta}\left(\frac{\Psi_1}{\Psi_0} - 1\right)(1-\delta)\right]^{1/\kappa} > 1$. Dropping all terms in $r\delta$ because they are second order relative to the magnitude of the initial jump and assuming $\delta = r$, $\frac{N_1}{N_0} \approx \left[\frac{\Psi_0+\Psi_1}{2\Psi_0}\right]^{1/\kappa}$. After the initial jump, N rises along the equilibrium transition path as given in (171). This rate is gradually decreasing over time and approaches γ^s.

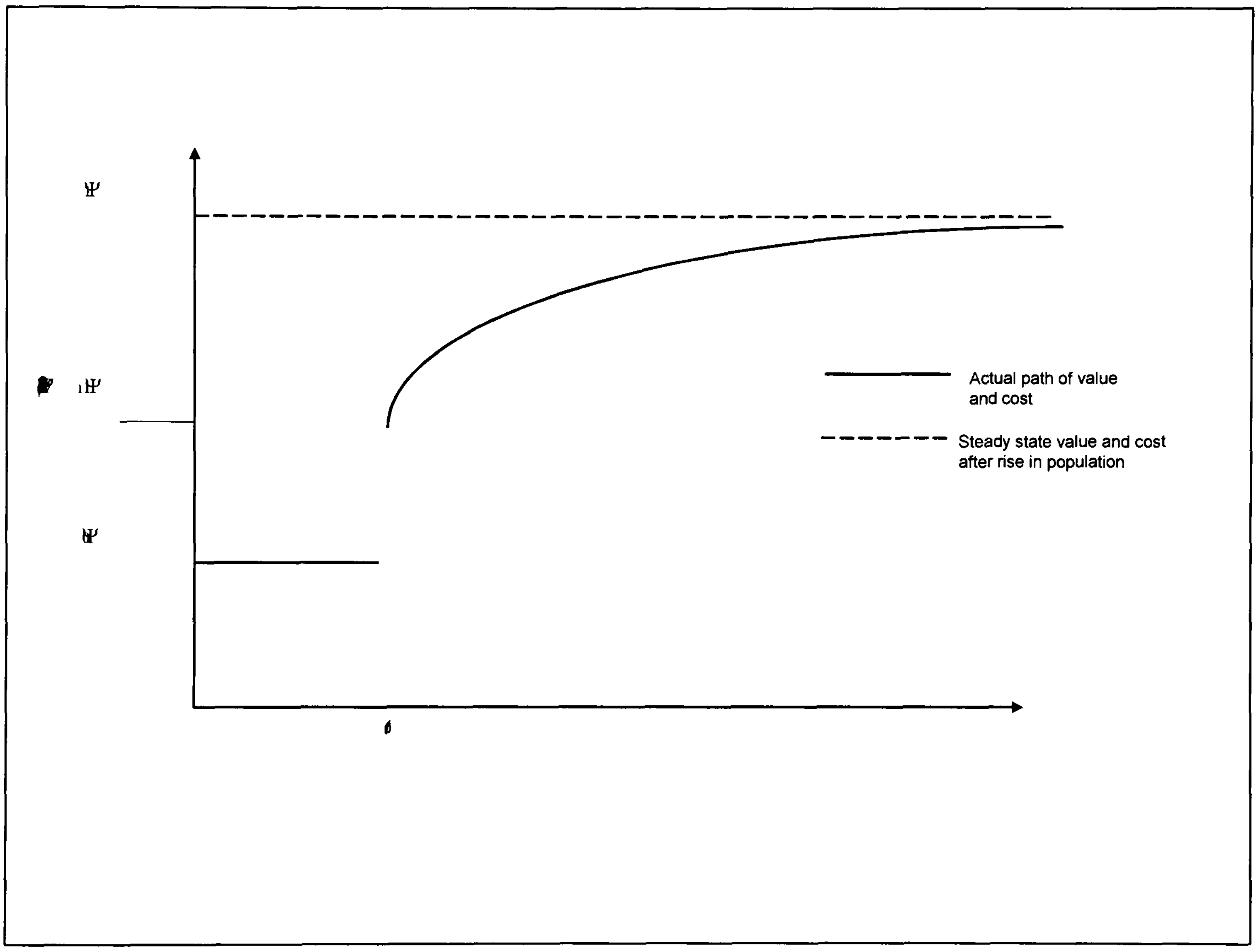


Figure 1: The evolution of V^s and C^s after expanded educational access. Graph scale divided by Ω/r.

[59] The change in value is lower for large r because future changes in the population are heavily discounted. A large value of δ means more people are being educated in absolute terms in a given period. This speeds up the rise in skill composition, which means the change in value is bigger.

The path of V_t^u and C_t^u The change in the value of unskilled licences is similarly:

$$\frac{V_t^u|\Psi_1}{V_t^u|\Psi_0} = \frac{1-\Psi_1}{1-\Psi_0} + \frac{r}{r+\delta}\left(1 - \frac{1-\Psi_1}{1-\Psi_0}\right)(1-\delta)^t \tag{176}$$

The initial jump at $t = 0$ is $\frac{V_0^u|\Psi_1}{V_0^u|\Psi_0} = \left(\frac{1-\Psi_1}{1-\Psi_0}\right) + \frac{r}{r+\delta}\left(\frac{1-\Psi_1}{1-\Psi_0}\right)$. If $\delta = r$, the downward jump is half the size of the steady state change. Therefore, it is likely that the equilibrium value of M - the value of M such that $V_t^u = C_t^u$ - will be less than the actual value. M therefore remains constant until t^+, when $(\gamma^u)^{\kappa t^+} > \left(\frac{V_{t^+}^u|\Psi_1}{V_0|\Psi_0}\right)$.

To understand the significance of t^+ note that the research frontier advances to reduce the cost of acquiring a new technology. However, the value of a licence also declines as the population falls. At first, whether the cost or value fall quicker depends on the parameters. Over time, the rate of fall in value decelerates to 0 while the research frontier keeps advancing and cost keeps falling. Eventually, cost will fall faster. Technology adoption resumes once the fall in cost has caught up with the fall in value. t^+ denotes this period:

$$t^+ : t > \frac{\ln(1-\Psi_0)(r+\delta) - \ln\left[(1-\Psi_1)(r+\delta) + r(\Psi_1 - \Psi_0)(1-\delta)^t\right]}{\kappa \ln \gamma^u} \tag{177}$$

t appears on both sides of the inequality. To characterise the solution for t^+, note that the total fall in value is given by the steady state fall, such that t^* given by (159) is an upper bound. Similarly, the initial jump provides an approximate lower bound for t^+. So, for example, for a rise in the skill proportion from 20% to 40% with $\kappa = 1$ and $\gamma^u = 1.02$, the upper bound is 15 years and the approximate ($\delta = r$) lower bound is 7 years. Figure 2 illustrates the actual (separate) paths of V^u and C^u, the steady state values, t^+ (when imports of unskilled licences resume) and t^* (the upper bound for t^+).

Once technology imports resume, the rate of technology adoption along the equilibrium transition path is given by (174). That expression shows that the rate of technology adoption is below γ^u but accelerates until it reaches the rate of frontier growth. These results echo Proposition 2 in Kiley (1999) (and section 7.4) but the stagnation is now driven by a continually falling value of unskilled technologies rather than a one-off fall.

7.5.3 Wages

Growth in wages has exactly the same pattern as growth in technologies.

Corollary 6 *While the skill composition of the labour force is rising, wage inequality is accelerated*

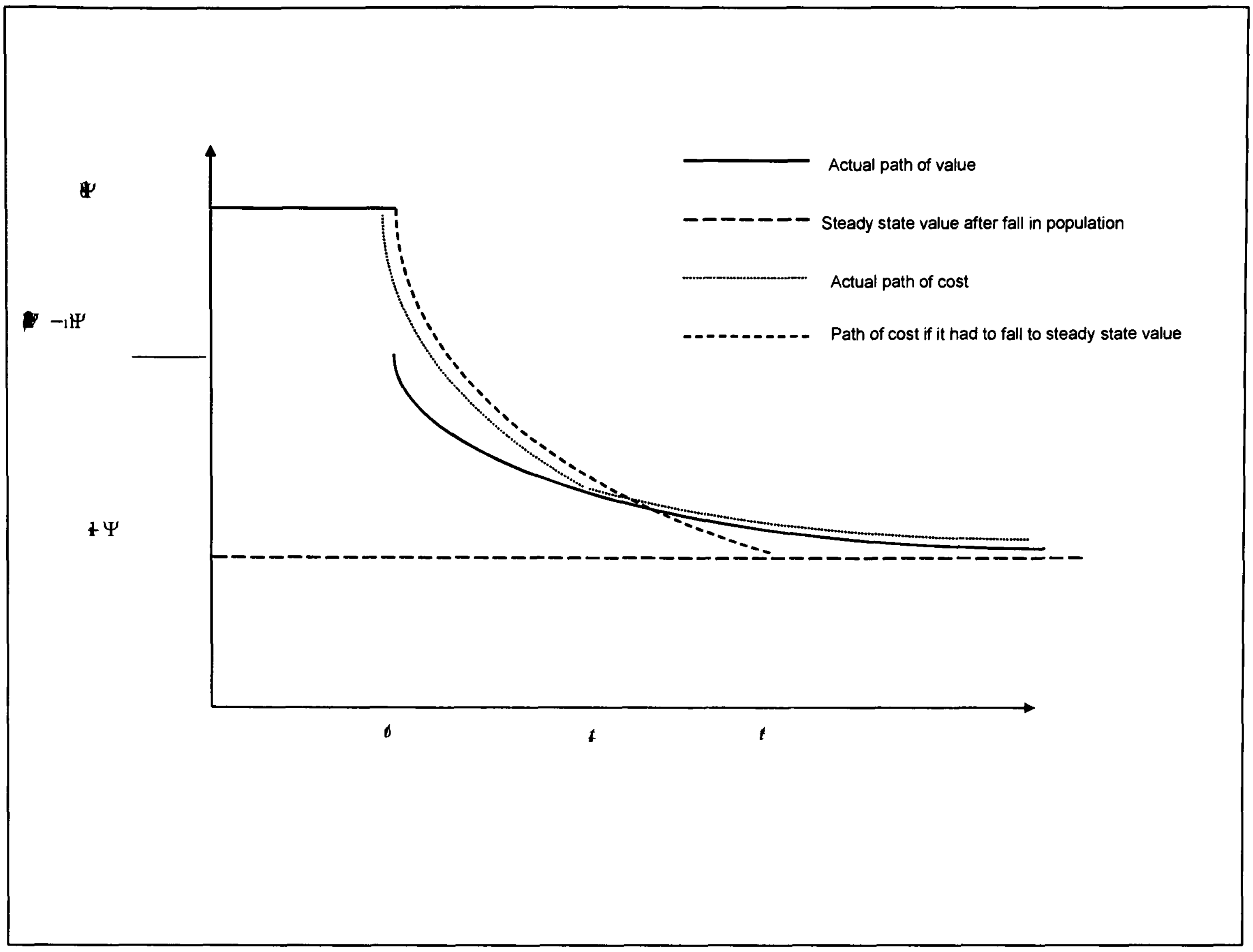


Figure 2: Evolution of V^u and C^u after a rise in Ψ from Ψ_0 to Ψ_1. Graph scale divided by $\frac{A\Omega}{r}$. t^+ denotes when marginal cost has caught up to value and imports of unskilled patents resume.

along the equilibrium transition path. Once the skill composition settles at its new steady-state, wage inequality grows at the steady-state rate $\frac{\gamma^s}{\gamma^u}$.

Recall that the one-off change in skill composition modelled in the previous section generated a one-period fall in wage inequality. This does not happen here. At t_0, the level of skill supply has not changed yet. By the time it does start to change, demand for intermediates can adjust as the population change was anticipated. Thus, holding the level of technology constant, there is no initial adjustment in wage inequality. The effect of the initial jump in N is to raise skilled wages at t_1 by $\frac{W_1^s|\Psi_1}{W_1^s|\Psi_0} = \left[\frac{\Psi_1}{\Psi_0} - \frac{r}{r+\delta}\left(\frac{\Psi_1}{\Psi_0} - 1\right)(1-\delta)\right]^{1/\kappa}$. For example, dropping terms in δr, setting $\kappa = 1$ and making $\delta = r$, the proportional jump in wages is approximately half the proportional change in the steady state skill composition. Skilled wages continue to grow at $\frac{W_{t+1}^s}{W_t^s} = \gamma^s \left(1 + \phi_t^s\right)^{1/\kappa}$ and decelerate to γ^s as $q_t \to \Psi_1$.

While skilled wages jump and grow fast, unskilled wages are stuck at W_0^u. They remain there until t^+, when unskilled technology adoption resumes. From t^+, unskilled wages grow at $\frac{W_{t+1}^u}{W_t^u} =$

$\gamma^u (1-\phi_t^u)^{1/k} < \gamma^u$. This is a period of accelerated wage inequality; wage inequality grows at:

$$\frac{W_{t+1}}{W_t} = \frac{\gamma^s}{\gamma^u}\left(\frac{1+\phi^s}{1-\phi^u}\right)^{1/\kappa} > \frac{\gamma^s}{\gamma^u}$$

We therefore have potentially long periods of accelerated wage inequality. The continual rise in wage inequality driven by differential advances in the research frontiers is compounded as the skill composition of the workforce gradually rises. Furthermore, at t_1, there is a one-off jump in relative wages. The comparative statics section showed that, once the skill composition settles at its new level, the effect of skill supply is a levels effect. The fact that the unskilled labour force falls for a long time makes it possible for unskilled wages to remain stagnant for long periods while skilled wages advance at a higher (although decelerating) rate.

Furthermore, it is possible to have very long periods of accelerated / transitional SBTC and wage inequality that are hard to distinguish from steady state changes. In the extreme case of $\gamma^u = 1$ - if all new technologies are skill-biased - then adoption of technologies complementing unskilled labour never resumes, the economy does not reach the optimal transition path and unskilled wages remain stagnant forever.

7.6 Concluding comment

This chapter adds to the directed technological change literature by allowing for external technology patterns to affect those of a developing country and by analysing the effects of expanded educational access. Firms in the technology sector pay technologically advanced countries for exclusive licences to import and distribute a new technology, which can complement either skilled or unskilled workers.

The cost of a skilled (unskilled) licence depends on how advanced the skilled (unskilled) frontier is - how many technologies have been developed by the technological leaders. Adopting the label "Northern skill-biased technological change" (NSBTC), we assume that skilled technologies advance faster than technologies complementing unskilled labour in the technologically advanced countries. NSBTC will directly affect the costs of acquiring skilled rather than unskilled technologies and hence influence the pace at which technologies are adopted by the South. Therefore, the South can experience skill-biased technological progress (SBTC) simply because the North does. Assuming skilled and unskilled labour are perfect substitutes, this means there can be a rise in wage inequality simply because of international technology patterns.

In our model, there is permanent exogenous NSBTC and hence a rise in wage inequality every period for ever. If we are consistent and believe NSBTC is itself directed, then it will not be permanent and we should restate our finding: There will be periods of SBTC and increased wage inequality in the South as long as SBTC takes place in the North.

To capture the effects of education reforms, we model changes in the proportion of cohorts being educated as they enter the labour force and replace those cohorts that die. This generates periods of growth in the skill composition of the population, during which wage inequality growth is higher: as the economy adjusts slowly to its new steady-state skill composition, we have potentially long periods of accelerated SBTC and wage inequality. During these periods, the SBTC driven by external technology patterns is amplified by domestic events.

We have modelled education policy as exogenous. In South Africa it is to a large extent a response to its history and hence appropriately modelled as such. For developing countries in general, education policy may itself be a response to global requirements. The Global Competitiveness Index views an increase in the skills base through expansion of both primary and higher education as a key ingredient for competitiveness in the light of observed shifts in demand away from unskilled labour (Lopez-Claros et. al., 2006). Thus expanded education may be a response to global labour demand shifts.

Ignoring the resource costs of expanded access, our brief discussion on the output effects of a one-off rise in skill supply suggest, under a scenario of NSBTC, GDP and GDP growth would rise. However, we have seen higher skill supply will itself lead to further skill-biased labour demand shifts. If expanded educational access is a response to global effects on wage inequality, it could be self-defeating.

8 Evaluating NSBTC and skill supply as contributors to SBTC in a developing country

8.1 Introduction

In the previous chapter, we modelled final output as the sum of two intermediate processes - skilled and unskilled. An alternative approach is for final output to be a CES aggregate of the two intermediate processes, which implies skilled and unskilled labour are not perfect substitutes. Existing literature suggests the assumption of perfect substitution is important for the conclusion that a rise in skill supply leads overall to a rise in wage inequality (Acemoglu, 2002ab; Sanders, 2005). Section 6.4.2 presented a model in this class where the elasticity must exceed 2.

The ominous conclusion in chapter 7 is that developing countries can experience increasing wage inequality because of external technological events, and that increasing the skill supply as a response may accelerate the process. The objective of this chapter is to evaluate to what extent the elasticity of substitution matters, both for the effects of skill supply and of Northern skill-biased technological change (NSBTC).

Introducing imperfect substitutability complicates the analysis, making it necessary to distinguish between skill-augmenting and skill-biased technological change. As explained in section 6.4.2, the former increases the relative marginal physical product of skilled labour while the latter increases the relative marginal revenue product. The former does not imply the latter. Finite substitutability necessitates price adjustment between the two intermediate inputs. It is the endogenous response of relative prices to changes in relative skill supplies and/or technological varieties that is the main source of complexity.

The model set-up in section 8.2 shows that the skill premium is now negatively related to the skill supply, holding technology constant, in accordance with standard labour demand theory. The model set-up shows that the equilibrium technology ratio now depends on the value of the elasticity of substitution between skilled and unskilled labour. Instead of defining the attractiveness of a market for a new licence in terms of the quantity of skilled or unskilled labour, as done in section 6.4.2, this chapter does so in terms of factor shares.

Section 8.2 concludes with a summary of a steady state equilibrium in which the skill supply (Q), the ratio of skilled to unskilled technologies available world-wide (R) and the ratio of skilled to unskilled technologies adopted by the developing country (T) are constant. We do this to bring out clearly the effects of imperfect substitution between skilled and unskilled labour.

A comparative static analysis is conducted in section 8.3. We model Northern skill-biased technological change (NSBTC) as a one-off rise in R (the ratio of skilled to unskilled technologies available in the world) and a rise in skill supply as a one-off rise in Q (the ratio of skilled to unskilled workers in the economy). Consistent with chapter 7, we find that a rise in skill supply leads to SBTC (except if the elasticity of substitution is unity). The most important qualification introduced by this chapter is that, for a rise in skill supply to lead to a rise in wage inequality, a high elasticity of substitution is needed ($\sigma > 2 + k$, where $k > 0$). The parameter values required for NSBTC to cause SBTC and raise domestic wage inequality are not onerous ($\sigma > 1$), making global SBTC a more likely cause of increased wage inequality.

The models here and in chapter 7 assume perfectly clearing labour markets, but section 8.4 draws on existing work to discuss how imperfectly clearing labour markets may affect the model predictions. In section 8.5, we use existing empirical estimates of σ applicable to developing countries to predict the likely effects of skill supply and global technology patterns. We conclude that, despite leading to SBTC, a rise in skill supply is unlikely to lead to a rise in wage inequality. Section 8.6 offers brief closing comments.

8.2 Model set-up

8.2.1 Production

Total output of final goods is a CES aggregate of two types of intermediate, as described by the linearly homogeneous technology:

$$Y_t = \left[(y_t^s)^{\frac{\epsilon-1}{\epsilon}} + (y_t^u)^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}} \tag{178}$$

Final output is produced by perfect competitors using two intermediate inputs purchased from intermediates producers. The price of final output is unity. $\epsilon > 0$ is the finite elasticity of substitution between intermediate inputs.[60] Individual producers take the price of final output and intermediate input prices as given before choosing their optimal quantities of intermediates. For the economy as a whole, a rise in supply of one of the inputs relative to the other would necessitate a relative price adjustment. Each final output producer l chooses its ratio of intermediates such that $\frac{y_{lt}^s}{y_{lt}^u} = \left(\frac{p_t^s}{p_t^u}\right)^{-\epsilon}$. For the economy to be in equilibrium, intermediates must have prices p^s and p^u such that:

[60] As $\epsilon \rightarrow 1$, the CES function approaches a Cobb Douglas Production Function with the corresponding unit elasticity of substitution.

$$\frac{p_t^s}{p_t^u} = \left(\frac{y_t^s}{y_t^u}\right)^{-\frac{1}{\epsilon}} \tag{179}$$

Intermediates are produced by i perfectly competitive firms, where y^s uses skilled labour and N different machines while y^u uses unskilled labour and M different machines. Specifically, $y_{it}^s = (L_{it}^s)^{1-\alpha} \sum_{j=1}^{N_t} X_{ijt}^\alpha$ and $y_{it}^u = A(L_{it}^u)^{1-\alpha} \sum_{j=1}^{M_t} Z_{ijt}^\alpha$.

The labour force is as described in and around equation (119) in chapter 7. While we will ultimately describe the equilibrium with constant skill proportions, we initially allow for time subscripts on the proportions of skilled and unskilled labour. Firms making y^s take p^s and factor prices as given and maximize profits. Demand for each skilled machine is $X_{ijt} = \left(\alpha \frac{p_t^s}{p_{jt}^X}\right)^{\frac{1}{1-\alpha}} L_{it}^s$. As in chapter 7, the price set by the monopolist is $\frac{1}{\alpha}$ and the quantity demanded of every skilled machine is equal. Demand for the whole economy given by

$$X_t = \alpha^{\frac{2}{1-\alpha}} (p_t^s)^{\frac{1}{1-\alpha}} q_t \tag{180}$$

and economy-wide demand for each unskilled machine is:

$$Z_t = A\alpha^{\frac{2}{1-\alpha}} (p_t^u)^{\frac{1}{1-\alpha}} (1 - q_t) \tag{181}$$

Economy-wide output of skilled and unskilled intermediates is:

$$y_t^s = N_t q_t^{1-\alpha} X_t^\alpha \tag{182a}$$

$$y_t^u = M_t A^{1-\alpha} (1 - q_t)^{1-\alpha} Z_t^\alpha \tag{182b}$$

As in chapter 7, there are constant returns to increases in the number of machine varieties.

8.2.2 Prices and wages

An exogenous change in the relative skill supply would lead to a rise in the relative quantity of y^s produced. By (179), this would necessitate a relative price adjustment. If the ratio of N to M were to change, this too would necessitate a relative price adjustment. Using (179) and (182),

$$p \equiv \left(\frac{p^s}{p^u}\right)^{\frac{1}{1-\alpha}} = \left(\frac{Nq}{MA(1-q)}\right)^{\frac{-1}{\sigma}} = (TQ)^{\frac{-1}{\sigma}}, \tag{183}$$

where $\sigma = \epsilon + \alpha - \epsilon\alpha$ is the elasticity of substitution between skilled and unskilled labour. It captures the percentage change in relative demand for the two factors due to the change in relative factor prices at constant output; that is[61] $-\frac{d\log\frac{L^s}{L^u}}{d\log\frac{w^s}{w^u}}$, the Robinson (1933) interpretation. Although we have more than two inputs in the production technology, it consists of two separable aggregates in y^s and y^u such that $\sigma > 0$.[62]

In (183), we see a negative relationship between the relative price of the skill intensive good on the one hand and the relative number of skilled technologies on the other. We can show that $\epsilon = 1 \Leftrightarrow \sigma = 1$ and that $d\sigma/d\epsilon > 0$. Also, $\epsilon \to \infty \Leftrightarrow \sigma \to \infty$. Therefore, as $\epsilon \to \infty$, changes in relative labour quantities or relative numbers of machine varieties would have no effect on p. The ratio would be unity and we could without loss of generality normalise $p^s = p^u$ to unity. While this would take us back to the model in chapter 7, (183) has a major effect on the decision to acquire a licence, as will be explained in section 8.2.3.

Producers of intermediate goods hire labour such that wage equals marginal revenue product. For equilibrium in the economy, relative wages are given by:

$$W \equiv \frac{W^s}{W^u} = \frac{N}{AM}p = \left(\frac{N}{AM}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{q}{1-q}\right)^{\frac{-1}{\sigma}} \tag{184}$$

(184) mirrors the findings of Acemoglu (2002a) discussed in chapter 6. The far right of the equation shows relative wages are affected by two things. First, the standard substitution effect, where a rise in the relative quantity of skilled labour reduces the relative skilled wage. Second, relative technologies. The effect of relative technologies can be positive or negative. As will be explained in section 8.3, a rise in N/M raises the relative physical productivity of skilled labour but, via its effect on relative product prices, can result in a net positive or negative effect on the relative marginal revenue products. In the simpler perfect substitutes setup, the only impact is on marginal physical product so the effect is always positive. As $\sigma \to \infty$, we see there is no substitution effect, $W = \left(\frac{N}{AM}\right)$ and a rise in N/M necessarily has a positive effect on wage inequality, as in chapter 7.

Equation (184) describes the important direct relationship between wages and the skill supply. It also describes the relationship between wages and technology. The next section describes how technology adoption is determined by the skill supply and by international technology.

[61] Put differently, it can be shown that (relative) derived demand $\frac{L^s}{L^u} = \left(\frac{w^u}{w^s}\right)^{\epsilon+\alpha-\epsilon\alpha} \left(\frac{M}{N}\right)^{1-2\alpha-\epsilon+\epsilon\alpha}$

[62] In the context of the earlier chapters in the thesis, notably the discussion of various elasticity concepts in chapter 2, it may seem odd to use the Robinson elasticity σ, where factor prices are exogenous and factor quantities are endogenous, in a model where factor quantities determine factor prices. When we have two separable inputs, we saw in chapter 2 that $\sigma = \frac{1}{H}$. We could reproduce these results and discussion in terms of the Hicks elasticity. Indeed, it some places it is a more comfortable and efficient explanation. However, we choose to employ the elasticity used in the literature for easier comparison. It would also be more accessible to audiences not familiar with the Hicks elasticity.

8.2.3 Technology adoption

The marginal cost for a machine of any type is unity and the cost of acquiring a licence is as given in (127a) and (128a) in chapter 7, although we set $\beta^s = \beta^u = 1$ for simplicity:[63]

$$C_t^s = \left(\frac{N_t}{R_t^s}\right)^{\kappa} \text{ if } \frac{N_t}{R_t^s} < 1, \text{ where } \kappa > 0 \tag{185a}$$

$$C_t^s = \infty \text{ if } \frac{N_t}{R_t^s} \geq 1 \tag{185b}$$

and

$$C_t^u = \left(\frac{M_t}{R_t^u}\right)^{\kappa} \text{ if } \frac{M_t}{R_t^u} < 1, \text{ where } \kappa > 0 \tag{186a}$$

$$C_t^u = \infty \text{ if } \frac{M_t}{R_t^u} \geq 1 \tag{186b}$$

Much of the discussion will set the cost elasticity $\kappa = 1$, indicating clearly when this is the case. For $\frac{N_t}{R_t^s} < 1$ and $\frac{M_t}{R_t^u} < 1$, the relative cost is:

$$C_t = \left(\frac{T_t}{R_t}\right)^{\kappa} \tag{187}$$

Profit from the sales of skilled and unskilled machines in a given period is:

$$\pi^x = \Omega \left(p_t^s\right)^{\frac{1}{1-\alpha}} q_t \tag{188a}$$

$$\pi^z = \Omega \left(p_t^u\right)^{\frac{1}{1-\alpha}} A(1-q_t) \tag{188b}$$

We can use (188a) to show that the value of a skill-augmenting licence is:

$$V_t^s = \Omega \left[\sum_{i=1}^{\infty} \frac{q_{t+i}\left(p_{t+i}^s\right)^{\frac{1}{1-\alpha}}}{(1+r)^i}\right] \tag{189}$$

The value of acquiring a licence today is affected by the future path of prices, which in turn is affected by the future path of T (cf. (183)). Analogously:

$$V_t^u = A\Omega \left[\sum_{i=1}^{\infty} \frac{(1-q_{t+i})\left(p_{t+i}^u\right)^{\frac{1}{1-\alpha}}}{(1+r)^i}\right] \tag{190}$$

[63] As explained in section 7.2.3, these parameters might represent environmental factors affecting the relative costs of acquiring skilled and unskilled technologies. They affect the level of the ratio of skilled to unskilled technologies but do not qualitatively affect the comparative static impact of a change in R or in Q.

The ratio of values is:

$$V_t = \frac{\left[\sum_{i=1}^{\infty} \frac{q_{t+i}\left(p_{t+i}^s\right)^{\frac{1}{1-\alpha}}}{(1+r)^i}\right]}{A\left[\sum_{i=1}^{\infty} \frac{(1-q_{t+i})\left(p_{t+i}^u\right)^{\frac{1}{1-\alpha}}}{(1+r)^i}\right]} \tag{191}$$

While prospective agents consider the values of skilled and unskilled licences independently, they rely on the paths of prices p_t^s and p_t^u. To bring out the effects of relaxing the assumption of perfect substitution between skilled and unskilled labour as simply as possible, we will from now on assume constant values of $Q \equiv \frac{q}{A(1-q)}$ and $R \equiv \frac{R^s}{R^u}$. This will allow us to find an equilibrium in which p, V and $T \equiv \frac{N}{M}$ are constant. Under these assumptions, equation (191) can be simplified to:

$$V = Q^{\frac{\sigma-1}{\sigma}} T^{\frac{-1}{\sigma}} \tag{192}$$

To gain intuition into the effects of limited substitutability, note a rise in relative skill supply raises the relative marginal (physical) product of skilled machines. For finite elasticities of substitution, there is also a negative effect, because the rise in Q raises y^s/y^u and thus lowers p. As explained in chapter 6, this lowers the relative marginal revenue product of each machine. σ determines the size of the second effect and hence whether the overall effect is positive ($\sigma > 1$) or negative ($\sigma < 1$), as shown in (192). Equation (192) also shows that the relative supply of skilled machines has a negative effect. This is (also) because of an effect on p and hence the relative marginal revenue products of X and Z. Note the perfect substitutes case has $V = Q$ as in equation (136), where the relative value of skilled machines depends only on the relative skill supply. Here, it depends on the relative skill supply *and* the state of technology.

By free entry, we have the equilibrium conditions (131) and (132) from chapter 7, which implies the relative cost of acquiring a skilled licence must equal the relative value:

$$V_t = C_t \tag{193}$$

We can substitute from (187) to confirm proposition 2 in chapter 7 still holds: with $\beta = 1$, $T_t = R_t \left(V_t\right)^{1/\kappa}$. We can also adapt our corollary from the perfect substitutes case for a constant T:

Corollary 7 *When skilled and unskilled labour are imperfect substitutes and R, Q are constant over time, the ratio of skilled to unskilled technologies (T) is increasing in the relative effective skill supply (Q) if and only if $\sigma > 1$ but is always increasing in the skill-bias of the world technology* *frontier (R):*

$$T = Q^{\frac{\sigma-1}{\sigma k+1}} R^{\frac{\sigma k}{\sigma k+1}}, k > 0 \quad (194)$$

$$T = Q^{\frac{\sigma-1}{\sigma+1}} R^{\frac{\sigma}{\sigma+1}}, k = 1 \quad (195)$$

This is shown by substitution from (187) and (192). We will use this result as well as (184) to conduct comparative statics on wages in section 8.3. Before we do that, we generalise the treatment in chapter 7 and rephrase the discussion of (191) and (192), which is in terms of factor quantities and product prices, in terms of factor shares.

Proposition 5 *When skilled and unskilled labour are imperfect substitutes and R, Q are constant over time, the value of a licence for a skilled machine relative to that for an unskilled machine is directly proportional to the share of skilled labour in output relative to that of unskilled labour.*

Proof. Using the fact that wage equals marginal revenue product and that T, Q are constant, we can write (191) as $V = \frac{qw^s}{(1-q)w^u}\frac{1}{T}$. But $qw^s/(1-q)w^u \equiv \zeta$ is an expression for the relative factor shares of skilled and unskilled labour, so,

$$V = \frac{\zeta}{T} \quad (196)$$

■

Equation (196) expresses the relative value of skilled and unskilled technologies in terms of relative factor shares and relative technology availability. This captures both the induced innovations (Hicks, 1963) and market size (Schmookler, 1966) arguments for factor-biased technological change. The treatment in Acemoglu (1998, 2002ab) is based on (183), where relative product prices (and hence wages) act to counter the market size effect. Alternatively, we can combine the two arguments into "market attractiveness" based on relative shares. Holding technology constant, a rise in the supply of skilled labour will raise (lower) its factor share if the elasticity of substitution is more (less) than unity. Thus, if $\sigma > 1$, a rise in relative skill supply increases the skill share and overall makes the value of adopting a skilled machine licence relatively more attractive. By free entry and using (196) and (187), this translates into a relationship between the ratio of technologies and relative factor shares:

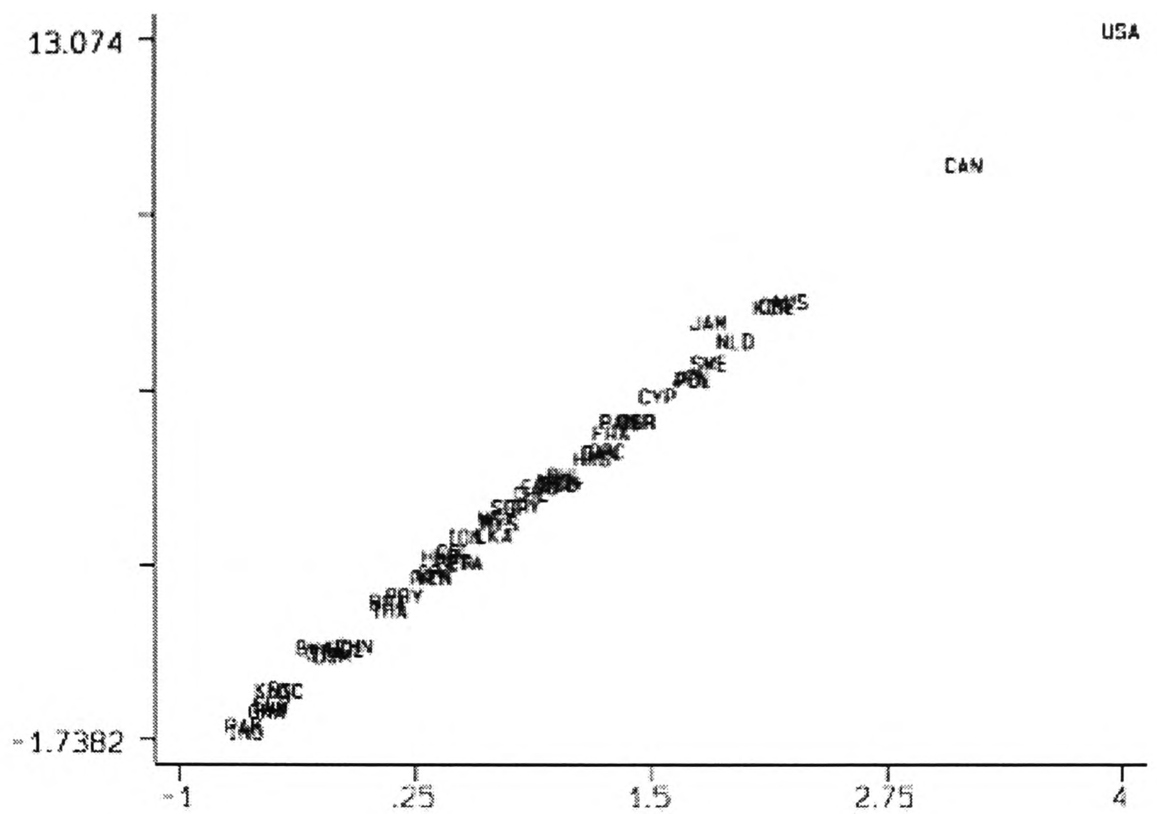


Figure 3: Scatterplot revealing relationship between relative labour efficiency (y-axis) and relative factor shares (x-axis). Source: Caselli & Coleman (2000:23, figure 9). Using their interpretation and our notation, relative labour efficiency is $\log N - \log AM$. Relative factor shares are $\log \varsigma_s - \log \varsigma_u$. Each observation is a country.

Corollary 8 *When skilled and unskilled labour are imperfect substitutes and R, Q are constant, the ratio of skilled to unskilled technologies is positively related to the relative factor share of skilled labour and positively related to the skill-bias of the world technology frontier:*

$$T = R^{\frac{k}{1+k}} \zeta^{\frac{1}{1+k}}, k > 0 \quad (197)$$

$$T = \sqrt{R\zeta}, k = 1 \quad (198)$$

The relative factor shares of skilled and unskilled labour represent the relative attractiveness of skilled machines while R^{-1} represents the relative cost of skilled machines. (197) is consistent with empirical work. Caselli & Coleman (2000:23) use cross country data spanning both developed and developing nations, based on the dataset of Hall & Jones (1999). As shown in figure 3, they note a strong positive correlation between relative factor shares and the productivities of skilled workers relative to unskilled workers, which they interpret as the relative numbers of machines skilled workers have. They propose this is evidence of appropriate technology adoption driven by factor endowments.

Corollary 8 is useful for combining the market size and induced innovations arguments as well as for comparison with the findings of Caselli & Coleman (2000). However, for effective comparative statics, we solve explicitly for T to obtain (194) or (195).

8.2.4 Equilibrium in steady state

Before conducting comparative statics, we describe our equilibrium in the economy to summarise the discussion so far. Equilibrium in the final goods (Y) market has all goods sold at price unity for consumption, to acquire licences or to import machines. Equilibrium in the intermediates (y^s, y^u) market is established by relative price p such that (cf. equation (183))

$$p = (TQ)^{\frac{-1}{\sigma}} \tag{199}$$

Supply equals demand for each of $N + M$ machine varieties sold at the monopoly price of $\frac{1}{\alpha}$. The labour market clears at the relative wage given by equation (184):

$$W = \left(\frac{N}{AM}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{q}{1-q}\right)^{\frac{-1}{\sigma}} \tag{200}$$

By free entry, the relative value of acquiring a skilled technology equals the relative cost: $V_t = C_t$. From proposition 2, $T_t = R_t (V_t)^{\frac{1}{\kappa}}$. (191) describes the relative value of a skilled technology when relative skill supply and relative prices, which are influenced by skill supply and by the state of technology T, are allowed to vary.

Equation (191) offers no easy solution for the path of a non-constant T. Analogously, for the economy modelled in Acemoglu & Zilibotti (2001), skilled and unskilled machines must grow at the same rate along the balanced growth path. While the varieties of skilled (N) and unskilled (M) machines can grow and there is growth in R^s and R^u, we restrict ourselves to analyses of the steady state *ratio* T. We therefore have $V = Q^{\frac{\sigma-1}{\sigma}} T^{\frac{-1}{\sigma}}$ (cf. 192) and the equilibrium condition:

$$T = Q^{\frac{\sigma-1}{\sigma k+1}} R^{\frac{\sigma k}{\sigma k+1}}, k > 0 \tag{201a}$$

$$T = Q^{\frac{\sigma-1}{\sigma+1}} R^{\frac{\sigma}{\sigma+1}}, k = 1 \tag{201b}$$

We can therefore perform comparative statics to see the effects of skill supply and NSBTC on technology adoption (from (201)) and wages (from (200)).

8.3 Comparative statics

We model NSBTC as a one-off rise in R and a rise in skill supply as a one-off rise in Q. Specifically, the comparative statics shows what elasticity parameter values are needed for NSBTC and a rise in skill supply to generate SBTC and increased wage inequality.

As introduced in corollary (7), inspection of equation (201b) shows that a rise in relative skill supply would lead to a rise in T if $\sigma > 1$. The intuition is clear with the help of (198). If $\sigma > 1$, a rise in the supply of skilled labour will lead to a rise in its factor share. This is because the factor inputs are relatively easy to substitute and the wage of skilled labour, *ceteris paribus,* would not need to fall by much to accommodate the rise in skill supply. If $\sigma < 1$, the number of skilled machines relative to unskilled ones would fall after a rise in skill supply. If $\sigma = 1$, factor shares are constant, so there is no effect on T. In the case of a perfect substitutes technology, the elasticity of T with respect to Q is unity, which is evident from (201b).[64]

Importantly, we can also see from equation (201b) that a rise in R has a positive effect on T, *regardless* of σ. We now turn to the comparative static effects on wages.

Proposition 6 *When skilled and unskilled labour are imperfect substitutes, a rise in skill supply will lead to a rise in wage inequality iff* $\sigma > 2 + \kappa$.

Proof. The derivative of (200) with respect to Q is $\frac{d\log W}{d\log Q} = \frac{\partial \log W}{\partial \log T}\frac{d\log T}{d\log Q} - \frac{1}{\sigma}$, but we know from (200) that $\frac{\partial \log W}{\partial \log T} = \frac{\sigma-1}{\sigma}$. We also know from (201a) that $\frac{d\log T}{d\log Q} = \frac{\sigma-1}{\sigma\kappa+1}$. Therefore $\frac{d\log W}{d\log Q} = \frac{\sigma-2-\kappa}{\sigma\kappa+1}$.
■

A rise in N relative to M has two effects on wages. First, it increases the relative (physical) productivity of skilled labour. Second, however, it increases y^s/y^u, which necessitates a fall in p and therefore has a negative effect on the relative marginal revenue product of skilled labour. The net effect on relative wages depends on σ. If $\sigma > 1$, a rise in T will have a positive effect on the skill premium, because the second effect is relatively small. Conversely, $\sigma < 1$ means relative wages will have to adjust a lot so that the overall effect is negative. Building on this, we see that:

$$\frac{d\log W}{d\log Q} = \frac{\partial \log W}{\partial \log T}\frac{d\log T}{d\log Q} - \frac{1}{\sigma} \tag{202}$$

$$= \underbrace{\frac{\sigma-1}{\sigma}\frac{\sigma-1}{\sigma k+1}}_{\text{directed technology effect}} - \underbrace{\frac{1}{\sigma}}_{\text{substitution effect}} \tag{203}$$

$$= \frac{\sigma-2-k}{\sigma k+1} \tag{204}$$

[64] To reconcile this with (197), note that, in this model, $\frac{d\log\zeta}{d\log Q} = 2$ because of the positive effect of technology on wages.

The overall effect of skill supply on wages is divided into two parts. The substitution effect comes from conventional labour demand theory. The directed technology effect, as labelled by Acemoglu (2002a), shows the effect of relative labour supply changes on wages through the adoption of skilled and unskilled technologies. The directed technology effect of skill supply on wages is unequivocally (weakly) positive. This is consistent with a key advance of Acemoglu (2002a) over Acemoglu (1998). From corollary 7, T falls after a rise in Q if $\sigma < 1$. However, if $\sigma < 1$, a fall in T would have a *positive* effect on the skill premium because of the rise in p it would create (cf.(183)).

Also, if $\sigma > 1$, T would rise and, although there would be a downward effect on p, it would not be big, so the overall effect is also *positive*. As discussed in section 6.4.2, Acemoglu (2002b) distinguishes between skill *augmenting* technological change (a rise in T) and skill-*biased* technological change, which overall raises the marginal revenue product of skilled labour. Thus, a rise in Q will lead to SBTC for any value of $\sigma \neq 1$.

The overall effect of skill supply on wages is positive if $\sigma > 2 + \kappa$. Our introduction to this class of model in section 6.4.2 presented thresholds for σ that were up to and including 2. Our model produces a threshold that exceeds 2. This stems from differences in how we model the costs of importing/developing a new technology. In chapter 6, costs were constant or falling with every technology developed. The extent to which the threshold was below 2 depended on the degree of increasing returns, if any. Our model has costs which increase in the acquisition of licences. As more licences are acquired, it becomes more expensive to purchase technologies closer to the frontier. If we set $\kappa = 1$ such that the elasticity of cost with respect to the number of technologies imported is unity, the threshold is $\sigma = 3$.

When assuming perfect elasticities, this criterion is automatically met: the responsiveness of T to Q is high and the ordinary substitution effect would be zero, so the effect of a rise in Q on the skill premium would always be positive. Formally, $\lim_{\sigma \to \infty} \frac{\sigma - 2 - \kappa}{\sigma + 1} = 1$ and the elasticity of wages with respect to skill supply would be $+1$.

While we distinguish between skill-augmenting and skill-biased technological change in the technology importing country, we will not make this distinction when referring to the technology frontiers. In other words, we continue to make a rise in R synonymous with NSBTC.

Proposition 7 *When skilled and unskilled labour are imperfect substitutes, NSBTC will lead to a rise in wage inequality iff* $\sigma > 1$.

Proof. Differentiating (184) with respect to R, using (194):

$$\frac{d\log W}{d\log R} = \frac{\partial \log W}{\partial \log T}\frac{d\log T}{d\log R} = \frac{\sigma - 1}{\sigma k + 1} \tag{205}$$

■

Here, there is no standard substitution effect as the quantity of labour is unchanged. We simply recall NSBTC leads to an unequivocal rise in T. This translates to SBTC in the South, which has a net positive effect on the relative marginal revenue product of skilled labour, if $\sigma > 1$. We therefore see that the conditions required for global SBTC to generate a rise in the skill premium are less onerous than the conditions for a rise in skill supply to do so.

8.4 Directed technological change and labour market institutions

The previous section performed a comparative static analysis of effects on relative wages. Chapter 6 introduced the potential for labour market rigidities to affect the relationship between SBTC and relative wages or employment. Here, we briefly discuss some of the interactions between rigidities and *directed* SBTC before considering whether such imperfections might have a material effect on the results presented thus far.

Sanders (2005) develops a simple framework that allows for various forms of unemployment. The main impact is to convert potential labour supply into effective labour supply. Using a similar CES setup to that in this chapter, Sanders allows for a minimum unskilled wage defined in terms of the skilled wage such that the relative wage is fixed. Casually, one might rewrite the equation for the endogenous skill premium (200) in terms of relative firm-level labour demand as $\frac{\bar{L}^s}{L^u} = (\bar{w})^\sigma \left(\frac{N}{AM}\right)^{\sigma-1}$. Thus he interprets an exogenous rise in the minimum wage (fall in $\bar{w}$) as a rise in relative firm-level demand for skills $\left(\frac{\bar{L}^s}{L^u}\right)$. From a macroeconomic perspective, this is a rise in the effective skill supply, which has the impacts on the relative value of adopting skilled machines discussed in this chapter.

Alternatively, we might look directly at the share of unskilled labour. For the macroeconomy, a rise in the minimum wage will reduce its share if $\sigma > 1$, make unskilled machines relatively unattractive and lead to a rise in the relative number of skilled machines. Sanders formally confirms the intuition that the (induced) shift in relative labour demand leads to relative quantity adjustment (a relative rise in unskilled unemployment) rather than relative wage adjustment. The implication is that, depending on the value of σ, a rise in the minimum wage could through directed technological change have a larger or smaller effect than predicted by standard labour demand theory.

In chapter 6, we introduced evidence that SBTC influences relative wage outcomes despite potential labour market rigidities. We now summarise how Sanders (2005) considers these issues in a directed technological change context: for explaining overall unemployment differences across countries, Nickell & Layard (1999) argue that bargaining institutions can contribute to higher total unemployment but that minimum wages (with the exception of France) and other candidate institutional features play no role. Drawing on this as well as the evidence in Nickell & Bell (1996) cited in chapter 6, Sanders concludes that minimum wages did not have an important effect on relative wages and that they do not explain why Europe's skill premium responded to relative demand shifts differently to that of the US.

Instead, Sanders considers the potential role of European bargaining institutions in determining relative unemployment outcomes. He performs numerical simulations to see whether bargaining institutions i) generate higher European than US unemployment and ii) produce meaningful differences in the wage and employment responses to a rise in skill supply.

Sanders finds the different bargaining setup can explain higher European total unemployment, but not relative unemployment. More importantly, the simulations suggest the predicted wage outcomes are virtually no different to the baseline models without the bargaining setup. This indicates that, should we care only about modelling relative wage behaviour in developing countries, the models in the previous chapters do not suffer materially from such an omission. One caveat in drawing inference from his results is that a feature of the bargaining set-up is that people of all skill types engage in bargaining. For a South African example, 40% of employees (excluding domestic work and agriculture) belong to a trade union in South Africa, but membership tends to be predominantly blue collar workers (Makgetla, 2006). However, section 3.3.1 suggested as many of 70% of (manufacturing) firms are affected by bargaining agreements and that there appeared to be no material variation across the occupations. With this potential issue in mind, we claim that the model is useful for predicting the effects of skill supply and NSBTC on wages, even if labour markets don't work perfectly.

8.5 Model predictions

Section 8.3 showed the values for the elasticity of substitution[65] between skilled and unskilled labour needed for rises in the skill composition or NSBTC in the North to cause SBTC and a rise in wage inequality in the South. Table 22 summarises the results. We compare this with estimates of the

[65] This is the Robinson (1933) measure, which captures the change in relative demand for two factors due to the change in relative factor prices at constant output.

Table 22: Summary of model predictions

	SBTC	Rise in W
Rise in Q	$\sigma \neq 1$	$\sigma > 2 + k$
Rise in R	$\sigma > 1$	$\sigma > 1$

Table 23: International Elasticity of Substitution Estimates

Source	**Region**	**Elasticity**
Caselli & Coleman (2000)	Large cross section of countries	1.32
Teal (2000)	Ghana	0.28;2.2
Fajnzylber & Maloney (2001)	Chile	0.33
Fajnzylber & Maloney (2001)	Mexico	0.26
Fajnzylber & Maloney (2001)	Colombia	0.42
Roberts & Skoufias (1997)	Colombia	0.38
Behar (chapter 4)	South Africa	-1.71
Edwards & Behar (2006)	South Africa	0.27

elasticity of substitution, shown in table 23.

Using estimates of the technological parameters reported in Caselli & Coleman (2000), we calculate an elasticity of 1.32, as shown in the first row of table 23. Teal (2000) reports values of 0.28 (fixed effects) and 2.2 (no fixed effects). Depending on how concerned one is about attenuation bias in fixed effects estimates, one might think the true elasticity is closer to 0.28 than 2.2 or alternatively take an average of about unity.

The South American labour demand studies did not have readily available measures of σ. However, we were able to calculate them using their estimates of labour demand elasticities. The Morishima elasticity of substitution[66] between two inputs can be calculated using the following formula (Blackorby & Russell, 1989):

$$M_{ij} = \bar{\lambda}_{ji} - \bar{\lambda}_{ii} \tag{206}$$

Blackorby & Russell (1981) present various conditions under which the Morishima and Allen/Uzawa elasticities are equivalent, including an implicit CES structure and the case of two inputs. All the South American studies have two inputs Therefore, *if* such estimates based on two inputs are legitimate, we can interpret the values in table (23) as σ. Estimates are well below unity, like the fixed effects estimate for Ghana.

Chapter 4 produced estimates of -1.71 for South Africa, which means the inputs are p-complements and not p-substitutes. The production technology (178) does not permit $\sigma < 0$.

[66] This was introduced in section 2.5.

It is therefore difficult to attach a valid interpretation to what a coefficient of -1.71 might mean.[67] With (178) being a CES function, which makes skilled and unskilled labour separable inputs, it may seem more appropriate for the purposes of this particular exercise to refer to estimates that also make this assumption. Edwards & Behar (2006) do so using the data discussed in chapter 3. For aggregates of skilled and unskilled labour, the elasticity is 0.27, which is in line with other studies for particular economies in the table.

Comparing the elasticity estimates with the model predictions, we see the elasticities of less than unity for South Africa, Ghana (fixed effects) and the South American countries suggest NSBTC will not lead to a rise in wage inequality. The large international study by Caselli & Coleman (2000) suggests NSBTC could be an explanation for rising wage inequality in developing countries. We conclude the evidence is mixed. While it is unclear to what extent NSBTC will affect wage inequality in the South, it is a far more likely contributor than a rise in skill supply, which we discuss next.

To examine the potential effects of skill supply on wage inequality, a natural threshold to consider has $\kappa = 1$ such that $\sigma > 3$. Acemoglu (1998) cites empirical examples of high elasticities of substitution (some exceeding his threshold of 2), but these do not appear to occur for South Africa and a broad range of other developing countries. All the values are *well* below the threshold of 3 for a rise in wage inequality to result. Only if we believe the estimates of Teal (2000) with no fixed effects and that $\kappa < 0.2$ can we find an example where a rise in skill supply will lead to a rise in wage inequality. This is bad news for those advocating a rise in skill supply as a direct cause of wage inequality in developing countries, but is good news for policy-makers aiming to reduce wage inequality. This result stands in sharp contrast to the ominous warnings at the end of chapter 7.

Therefore, even if we allow for (unequivocal) SBTC, a rise in skill supply will tend to reduce the skill premium and benefit the unskilled.

8.6 Conclusion

In chapter 7, where we assume skilled and unskilled labour are perfect substitutes, there is an unequivocally positive relationship between the proportion of skilled workers and the ratio of skilled to unskilled technologies. We also saw technological change is biased in the South simply because it is in the North. This chapter allows for imperfect substitution between skilled and unskilled labour.

[67] This raises both the practical and conceptual difficulty associated with using elasticity estimates from a multiple input technology to infer the elasticity between only two inputs.

Our model shows a rise in skill supply will lead to SBTC for all elasticities of substitution except $\sigma = 1$. However, for a rise in skill supply to raise wage inequality, the elasticity must be higher than a threshold which exceeds 2. In other words $\sigma > 2 + k$ where $k > 0$. This contrasts with the model of Acemoglu (2002b), where the threshold is below 2 ($k < 0$). Our developing country model has a higher threshold because it becomes more expensive to acquire a technology as the country gets closer to the technological frontier. The high threshold makes it very unlikely that a rise in skill supply will increase the skill premium in developing countries, as evidenced by available empirical estimates of σ. A comparative static change in global technology possibilities leads to skill-biased technological change and a rise in the skill premium if $\sigma > 1$, a requirement much less onerous than that for a rise in skill supply. Individual country estimates are below unity but those across a wide sample of countries in Caselli & Coleman (2000) exceed unity.

In our model, the source of skill-biased technological change in the North is exogenous. It could be due to a number of reasons. For complete consistency, we would have to draw on a rise in the skill supply to lead to NSBTC. According to the directed technological change literature (see chapter 6), a rise in the skill supply in the US, UK and other technological leaders unequivocally leads to NSBTC. This, according to our model, potentially leads to SBTC and increased wage inequality in middle income countries.

Therefore, while parameter estimates of σ (see Hamermesh, 1993) suggest a rise in skill supply will not lead to higher wage premia in the technological leaders, it may do so in the technological followers. Nonetheless, a developing country that expands educational access will overall see a fall in the skill premium. Skilled and unskilled labour are complements.

9 Synthesis and conclusion

This thesis has investigated whether skilled and unskilled labour are complements or substitutes, doing so within the South African context of mass unemployment. The role of relative factor prices may be important: have rises in the cost of unskilled labour relative to capital hampered the employment performance of the economy? Part of the unemployment problem is due to South Africa's apartheid history, which effectively barred much of its population from a decent education. As more people get access to education and training, they are expected to enjoy better job market outcomes, but what will happen to those who remain unskilled? Will having more skilled workers in the economy have a positive or negative effect on the demand for unskilled labour? Worsening labour market prospects for the unskilled seem to be attributable to technologically driven shifts in global demand away from unskilled workers in favour of skilled workers. In this context, we asked how these global patterns affect developing economies.

We also investigated what role increases in the skill composition may have in causing directed technological change. Tinbergen (1975:79) warned of a *"race between education and technology."* This thesis asked if it's the case that, the faster education runs, the further ahead it propels (skill-biased) technology. In the South African context, will attempts to redress historical inequality and confront global skill-biased global labour demand shifts be self defeating? Will having more skilled workers decrease demand for unskilled workers, possibly via skill-biased technological change? Will increases in the relative cost of labour compound the problem?

Chapter 2 argued that we should employ the Allen elasticity of substitution (σ) and associated elasticities of factor demand to address the role of relative factor prices. It also argued that we should use the Hicks elasticity of complementarity (H) to ask what effects a rise in the supply of skilled labour will have on the productivity and demand for unskilled labour, holding technology constant.

By generalising the Uzawa (1962) result, we confirmed it is legitimate to use cost functions to measure σ when returns to scale are not constant. We also built on Berndt & Christensen (1973a) to show one can employ functional separability and use disaggregated data to calculate elasticities of labour demand for aggregated labour inputs.

Chapter 3 explained in detail the data used for the empirical estimations. This included a description of the procedure used to supplement firm-level technology data with wage data from household surveys as well as an analytical discussion of the importance of accounting for firm-size effects on wages in this procedure.

Table 24: Comparison of estimated and implied elasticities of substitution

Input pair	σ_{cost}	$\sigma_{\text{production}}$
Capital/More	+	-
Capital/Less	+	+
More/Less	-	-

Addressing the role of relative factor prices, we saw in chapter 4 that capital and labour are p-substitutes. A fall in the price of capital relative to labour will lead to substitution away from labour in favour of capital. Unlike other work, we are able to take account of potential output effects and still reach this conclusion. The result suggests recent labour legislation, which has increased the bargaining power of workers, has worsened their employment prospects. We also find that skilled and unskilled labour are p-complements. A fall in the price of more skilled labour relative to less skilled labour will lead to a rise in demand for less skilled labour.

Chapter 5 suggests a rise in supply of skilled/artisanal workers will benefit the unskilled while a rise in supply of semi-skilled workers will harm the unskilled. Unlike existing work, we are able to adjust for finite elasticities of product demand. The finding reveals the value of a disaggregated study. It also finds more and less skilled labour are q-complements overall.

One might argue chapter 5 shows a rise in skill supply leads to a rise in demand for unskilled labour while chapter 4 shows a fall in the price of skilled labour, which may have been caused by a rise in skill supply, will also lead to a rise in demand for unskilled labour. This intuitive consistency is comforting, although it is not entirely appropriate because of the differences in assumptions employed for each elasticity concept. Furthermore, Hamermesh (1993) argues how it is possible for a pair of factors to be p-complements and for that same pair to be q-substitutes (or vice versa). Field (1988) finds that, in some cases, free and slave labour are Hicks (q) complements but Allen (p) substitutes. Especially when we consider that we don't use the exact same data for both sets of estimation, the finding that more and less skilled labour are both p- and q-complements is not trivial.

We saw in section 2.3 that there is a formula for the relationship between H and σ. Table 24 presents two sets of signs.[68] The first column is based on the aggregate elasticity of substitution as estimated using the cost data in chapter 4. The second set uses estimates of H from chapter 5 and equation (25) from section 2.3 to calculate the values and hence signs of σ implied by those estimates. We see consistency in the signs for two out of three rows, which could be random. The

[68] The numbers are not presented because there are some difficulties associated with what value of returns to scale to attach. (They are constant in chapter 5 but non-homogenous in chapter 4.) The objective in any case is to compare the qualitative effects.

fact that we see differences in the signs in the first row leads one to consider a potentially fruitful area for further research. Using a data set that is of relatively high quality, it would be informative to find estimates for σ from both production and cost functions. (A two factor example need only test whether $H = \frac{1}{\sigma}$.) If similar, they would provide grounds for confidence in the results and simultaneously find support for the theoretical links. If not similar, this would raise the theoretical and econometric question of why not. It would also raise the issue of which estimation method is more reliable. This is an important methodological question with broad applicability.

In the second row, the consistency in the signs leads us to re-iterate we believe a fall in the cost of capital relative to that of less skilled labour will have negative effects on demand for less skilled labour. In row three we find that, were we to use our production function to calculate the elasticity of substitution between more and less skilled labour, possibly as an alternative to constructing wage data, we would still conclude that they are p-complements.

Having held technology constant in the first half of the thesis, we considered the interactions between skilled labour, technology, and unskilled labour in the second half. In chapter 6, we presented support for the claims that i) skill-biased technological change (SBTC) has impacted negatively on the demand for less skilled labour, ii) SBTC in developing countries can be heavily influenced by technology patterns in the technologically advanced nations and iii) a rise in skill supply can stimulate SBTC.

Chapter 7 used a model where skilled and unskilled labour are perfect substitutes. It showed how SBTC by technologically advanced countries (NSBTC) lowers the relative costs of adopting machines designed to work with skilled labour, which leads to SBTC in developing countries and persistent rises in wage inequality. This model captures the idea that firms in developing countries are over time finding it less attractive to use technologies appropriate to their skill endowments.

To investigate the effects of expanded educational access, we model cohort educational effects such that the skill composition rises. While the skill composition rises, SBTC is greater and wage inequality is accelerated. Ultimately, once the skill composition settles, it is only NSBTC that leads to persistent rises in wage inequality. However, given that the skill composition can take a long time to reach its steady state, the effect of increased skill supply can have a long duration. Sustained periods of observed SBTC can be equally attributable to either cause.

In chapter 8, we relax the assumption that skilled and unskilled labour are perfect substitutes. NSBTC only leads to a rise in wage inequality if $\sigma > 1$. Available empirical estimates of σ yield mixed predictions on whether this is likely to occur. A rise in skill supply unequivocally leads to SBTC (except for $\sigma = 1$). However, for a rise in skill supply to lead to increased wage inequality in our baseline model, $\sigma > 3$. Our threshold exceeds 2 because of increasing costs of adopting more advanced technologies as the developing country gets closer to the frontier. Existing work (Acemoglu, 2002ab), has constant or decreasing costs, leading to a threshold below 2. In any case, available elasticity estimates are nowhere near high enough to suggest a rise in skill supply will increase the skill premium. Thus, as a response to potential global SBTC, a rise in skill supply will not have unintended wage inequality consequences.

Alternative explanations suggest SBTC can be induced by trade. In an argument similar to Acemoglu (2003b), if technological leaders open to trade with developing countries, this stimulates demand for skilled goods and leads to the development of skilled technologies. Developing countries may encounter increased demand-side incentives for unskilled technologies. However, if economic globalisation makes it easier for NSBTC to drive SBTC, then the overall effect can be a rise in wage inequality in the South. Similarly, middle income countries trading with low income countries may experience demand shifts in favour of relatively skill intensive products such that technological change is directed towards skilled technologies.[69] This coincides with what Wood (1994) terms "defensive innovation".

While trade induced SBTC and capital skill complementarity, which was discussed in chapter 6, are potential explanations for increases in wage inequality, providing a comprehensive account of observed wage inequality patterns is not the aim of this thesis. The central role is to predict what role skill supply might have. Our conclusion for South Africa and developing countries in general is clear: despite SBTC, skilled and unskilled labour are complements.

Thus, whether we hold technology constant or not, whether we use the Hicks or Allen elasticity, and whether we use production function estimates or cost function estimates, skilled and unskilled labour are complements.

[69] This is the subject of work in progress.

References

[1] Acemoglu, D (1998); "Why do new technologies complement skills? Directed technological change and wage inequality"; Quarterly Journal of Economics 113

[2] Acemoglu, D (2002a); "technological Change, Inequality and the Labor Market"; Journal of Economic Literature 40

[3] Acemoglu, D (2002b); "Directed technological Change"; Review of Economic Studies 69

[4] Acemoglu, D (2003a); "Cross-country inequality trends"; Economic Journal 113

[5] Acemoglu, D (2003b); "Patterns of Skill Premia"; Review of Economic Studies 70

[6] Acemoglu, D & D Zilibotti (2001); "Productivity Differences"; Quarterly Journal of Economics 116

[7] African National Congress (2005); Discussion Document prepared for National General Council on 29 June – 3 July; Tshwane

[8] Aghion, P & P Howitt (1992); "A Model of Growth Through Creative Destruction"; Econometrica 60

[9] Aghion, P & P Howitt (1998); "Endogenous Growth Theory"; MIT Press, London

[10] Allen, R (1938); "Mathematical Analysis for Economists"; Macmillian & Co Ltd, London

[11] Anderson, G & J Thursby (1986). "Confidence Intervals for Elasticity Estimators in Translog Models"; Review of Economics and Statistics 68

[12] Appelbaum, E & U Kohli (1997); "Import Price Uncertainty and the Distribution of Income"; Review of Economics and Statistics 79

[13] Atkinson, A & J Stiglitz (1969); "A New View of Technological Change"; Economic Journal 79

[14] Barro, R & X Sala-i-Martin (2004); "Economic Growth second edition"; MIT Press

[15] Behar, A (2006); "ASGISA and JIPSA: Will the unskilled also get to share?"; TIPS/DPRU Conference Paper

[16] Behar, A & L Edwards (2004); "Estimating Elasticities of Demand and Supply for South African Manufactured Exports Using a Vector Error Correction Model"; Centre for Study of African Economies Working paper WPS/2004-04

[17] Bergström, V & E Panas (1992); "How robust is the capital-skill complementarity hypothesis?"; Review of Economics and Statistics 74

[18] Berman, E & S Machin (2000); "Skill-biased technology transfer: evidence of factor biased technological change in developing countries"; NBER Working Paper Series

[19] Berman, E, J Bound & Z Grliches (1994); "Changes in the Demand for Skilled Labor within US Manufacturing: Evidence from the Annual Survey of Manufacturers"; Quarterly Journal of Economics 109

[20] Berman, E, J Bound & S Machin (1998); "Implications of skill-biased technological change: international evidence"; Quarterly Journal of Economics 113

[21] Berndt, E (1991); "The Practice of Econometrics, Classic and Contemporary"; Addison-Wesley, Reading

[22] Berndt, E & L Christensen (1973a); "The Internal Structure of Functional Relationships: Separability, Substitution and Aggregation"; Review of Economic Studies 40

[23] Berndt, E & L Christensen (1973b); "The Translog Function and the Substitution of Equipment, Structures and Labour in U.S. Manufacturing 1929-68"; Journal of Econometrics 1

[24] Berndt, E & L Christensen (1974); "Testing for the Existence of a Consistent Aggregate Index of Labor Inputs"; American Economic Review 64

[25] Berndt, E & M Khaled (1979); "Parametric Productivity Measurement and Choice Among Flexible Functional Forms"; Journal of Political Economy 87

[26] Bertoletti, P (2003); "Elasticities of substitution and complementarity: A synthesis"; mimeo

[27] Bhorat, H & J Hodge (1999); "Decomposing shifts in labour demand in South Africa"; South African Journal of Economics 67

[28] Bhorat, H, M Leibbrandt, M Maziya, S van der Berg & I Woolard (2001); "Fighting Poverty: Labour Markets and Inequality in South Africa"; UCT Press, Cape Town

[29] Bhorat, H & P Lundall (2002); "Employment, wages and skills development: firm specific effects - evidence from two firm surveys in South Africa"; Development Policy Research Unit Working Paper 68

[30] Bils, M & P Klenow (2000); "Does schooling cause growth?"; American Economic Review 90

[31] Binswanger, H (1974a); "A Cost Function Approach to the Measurement of Elasticities of Factor Demand and Elasticities of Substitution"; American Journal of Agricultural Economics 56

[32] Binswanger, H (1974b); "The Measurement of Technological Change Biases with Many Factors of Production"; American Economic Review 64

[33] Blackorby, C & R Russel (1989); "Will the Real Elasticity of Substitution Please Stand Up? (A Comparison of the Allen/Uzawa and Morishima Elasticities)"; American Economic Review 79

[34] Blackorby, C, D Primont & R Russel (1977); "On Testing Separability Restrictions with Flexible Functional Forms"; Journal of Econometrics 5

[35] Blanchard, O (1985); "Debts, deficits and finite horizons"; Journal of Political Economy 93

[36] Booth, A (1995); "The Economics of the Trade Union"; Cambridge University Press

[37] Borjas, G (1986); "The Demographic Determinants of the Demand for Black Labor"; in R Freeman & H Holzer (eds.); "The Black Youth Employment Crisis"; University of Chicago Press

[38] Cahuc, P & A Zylberberg (2004); "Labor Economics"; MIT Press; London

[39] Caselli, F & W Coleman (2000); "The world technology frontier"; NBER working paper series

[40] Chisholm, L (2004); "The Quality of Primary Education in South Africa"; at http://portal.unesco.org/education/en/file_download.php/6284aa8e1c97b3a9da28a7448344933fChisholm,+South+Africa,+13+April.doc; last accessed 5 December 2006

[41] Christensen, L, D Jorgenson & L Lau (1973); "Transcendental Logarithmic Production Frontiers"; Review of Economics and Statistics 55

[42] Chung, J (1987); "On The Estimation of Factor Substitution in the Translog Model"; Review of Economics and Statistics 69

[43] Chung, J (1994); "Utility and Production Functions"; Blackwell, Oxford

[44] Clague, C (1969); "Capital-labor Substitution in Manufacturing in Underdeveloped Countries" Econometrica 37

[45] Congress of South African Trade Unions (1998); "COSATU's Approach to Budget Hearings on the Role of the South African Reserve Bank and Monetary Policy"; Memorandum to Parliamentary Finance Committee

[46] Congress of South African Trade Unions (2005); "COSATU takes jobs campaign to the streets"; COSATU Weekly News, 27 May

[47] Crotty, A (2006); "Sustaining SA's hopes through education is the best way to go"; Business Report 18 October

[48] Deaton, A (1997); "The Analysis of Household Surveys: A Microeconometric Approach to Development Policy"; World Bank / John Hopkins University Press, Baltimore

[49] Denny, M & M Fuss (1977); "The Use of Approximation Analysis to test for Separability and the Existence of Consistent Aggregates"; American Economic Review 67

[50] Department of Labour (1997); "Skills Development Strategy for Economic and Employment Growth in South Africa"; Green Paper

[51] Diewert, W (1971); "An Application of the Shephard Duality Theorem: a Generalized Leontief Production Function"; Journal of Political Economy 79

[52] Doms, M, T Dunne & K Troske (1997); "Workers, Wages and Technology"; Quarterly Journal of Economics 112

[53] Du Toit, C & R Koekemoer (2003); "A labour model for South Africa", South African Journal of Economics 71

[54] Edwards, L (2001); "Globalisation and the Occupational Structure of Employment in South Africa"; South African Journal of Economics 69

[55] Edwards, L (2003); "A Firm Level Analysis of Trade Technology and Employment in South Africa"; Journal of International Development 15

[56] Edwards, L & A Behar (2006); "Trade liberalization and labour demand within South African Manufacturing firms"; Studies in Economics and Econometrics 20

[57] Estrin, S & D Laidler (1995); "Introduction to Microeconomics"; Harvester Wheatsheaf New York

[58] Fajnzylber, P & W Malony (2001). "How Comparable are Labor Demand Elasticities Across Countries?"; Policy Research Working Paper 2658; World Bank Latin America and the Caribbean Region Poverty Reduction and Economic Management Sector Unit

[59] Fallon, P & P Layard (1975); "Capital Skill Complementarity, Income Distribution, and Output Accounting"; Journal of Political Economy 83

[60] Fallon, P & R Lucas (1998); "South Africa: Labor Markets: Adjustment and Inequalities"; World Bank Southern Africa Department Discussion Paper 12

[61] Fallon, P & D Verry (1988); "The Economics of Labour Markets"; Philip Alan Publishers Limited, Oxford

[62] Fedderke, J, Y Shin & P Vaze (2003); "Trade and Labour Usage: An Examination of the South African Manufacturing Industry; Econometric Research Southern African Working Paper 15

[63] Fedderke, J, J Kayemba, S Henderson, M Mariotti & P Vaze (2001); "Changing factor market conditions in South Africa: the capital market - a sectoral description", Development Southern Africa 8

[64] Field, E (1998); "Free and Slave Labor in the Antebellum South: Perfect Substitutes or Different Inputs?"; Review of Economics and Statistics 70

[65] Fuss, M (1977); "The Demand for Energy in Canadian Manufacturing"; Journal of Econometrics 5

[66] Grant, J & D Hamermesh (1981); "Labour Market Competition Among Youths, White Women and Others"; Review of Economics and Statistics 63

[67] Greene, W (2003); "Econometric Analysis 5th edition"; Pearson Education International, Upper Saddle River

[68] Griffin, J & P Gregory (1976); "An Intercountry Translog Model of Energy Substitution Responses"; American Economic Review 66

[69] Griliches, Z (1969); "Capital-skill complementarity"; Review of Economics and Statistics 51

[70] Griliches, Z & J Mairesse (1995); "Production Functions: The Search for Indentification"; NBER Working Paper Series

[71] Grossman, G & E Helpman (1991a); "Endogenous Product Cycles"; Economic Journal 101

[72] Grossman, G & E Helpman (1991b); "Quality Ladders and Product Cycles"; Quarterly Journal of Economics 106

[73] Grossman, G & E Helpman (1991c); "Quality Ladders in the Theory of Growth"; Review of Economic Studies 58

[74] Guarda, P (2000); "Luxembourg's Cross-Border Workers: Estimating a System of Factor Demands"; CREA Working Paper

[75] Gujarati, D (1995); "Basic Econometrics"; McGraw-Hill Inc, London

[76] Habbakuk, H (1962); "American and British technology in the Nineteenth Century: Search for Labor Saving Inventions"; Cambridge University Press, Cambridge

[77] Hall, R & C Jones (1999); "Why do some Countries Produce So Much More Output Per Worker Than Others?"; Quarterly Journal of Economics 114

[78] Hamermesh, D (1993); "Labor Demand"; Princeton University Press

[79] Heathfield, D & S Wibe (1987); "An Introduction to Cost and Production Functions"; Macmillan Education Ltd, Hong Kong

[80] Hicks, J (1932); "The theory of wages first edition"; Macmillan & Co Ltd

[81] Hicks, J (1963); "The theory of wages second edition"; Macmillan & Co Ltd

[82] Hicks, J (1970); "Elasticities of Substitution Again: Substitutes and Complements"; Oxford Economic Papers

[83] Humphrey, D & B Wolkowitz (1976); "Substituting intermediates for capital and labour with alternate functional forms: an aggregate study"; Applied Economics 8

[84] Johnson, G (1980); "The Theory of Labor Market Intervention"; Economica 47

[85] Jorgenson, D (1963); "Capital Theory and Investment Behaviour"; American Economic Review 53

[86] Kaplinsky, R (1995); "Capital Intensity in South African Manufacturing and Unemployment, 1972-90"; World Development 23

[87] Katz, L & D Autor (1999); "Changes in the wage structure and earnings inequality"; in O Ashenfelter & D Card (eds.); "Handbook of Labour Economics Volume 3A"; Elsevier North Holland, Amsterdam

[88] Kiley, M (1999); "The supply of skilled labour and skill-biased technological progress"; Economic Journal 109

[89] Kim, H (1997); "Functional Separability and Elasticities of Complementarity"; American Journal of Agricultural Economics 79

[90] Kim, H (2000); "The Antonelli versus Hicks Elasticitiy of Complementarity and Inverse Input Demand Systems"; Australian Economic Papers 39

[91] Krugman, P (1994); "Past and prospective causes of unemployment"; Federal Reserve Bank of Kansas City Economic Review

[92] Krusell, P, L Ohanian, J Rios-Rull & G Violante (2000); "Capital-skill complementarity and inequality: a macroeconomic analysis"; Econometrica 68

[93] Lau, L (1969); "Duality and the Structure of Utility Functions"; Journal of Economic Theory 1

[94] Lau, L (2000); "Research on the Cost of Capital; Past, Present, and Future"; in Lau, L (ed.); "Econometrics Volume 2: Econometrics and the Cost of Capital: Essays in Honor of Dale W Jorgenson"; MIT Press

[95] Layard, P & A Walters (1987); "Microeconomic Theory"; McGraw-Hill, Singapore

[96] Levinsohn, J & A Petrin (2003); "Estimating Production Functions Using Inputs to Control for Unobservables"; Review of Economic Studies 70

[97] Lopez-Claros, A, L Altinger, J Blanke, M Drzeniek & I Mia (2006); "The Global Competitiveness Index: Identifying the Key Elements of Sustainable Growth"; in World Economic Forum; "Global Competitiveness Report 2006-2007"

[98] Lucas, R (1988); "On the Mechanics of Economic Development"; Journal of Monetary Economics 22

[99] Mak, K (2000); "The contribution of Canadian education to industrial production"; Education Economics 8

[100] Makgetla, N (2006); "In praise of labour movement's overlooked virtues"; Business Day 25 August

[101] Maki, D & L Meredith (1987); "A note on Unionization and the Elasticity of Substitution"; Canadian Journal of Economics 20

[102] Manuel, T (2006); "National Budget Speech"

[103] Marshall, A (1920); "Principles of Economics 8th edition"; Macmillan & Co Ltd, London

[104] Mas-colell, A, M Whinston & J Green (1995); "Microeconomic Theory"; Oxford University Press

[105] May (2006); "Persistent Poverty, Asset Accumulation and Shocks in South Africa: Evidence from KwaZulu-Natal" in Bhorat, H & R Kanbur; Poverty and Policy in Post-Apartheid South Africa; HSRC Press

[106] Mdladlana, 2004; "Address to the Sector Education Training Authorities"; Ministry of Labour, Pretoria

[107] Mlambo-Ngcuka, P (2006); "Address delivered by the Deputy President at the launch of the Joint Initiative for Priority Skills Acquisition (JIPSA)"

[108] Moll, P (1996); "Compulsory Centralisation of Collective Bargaining in South Africa", American Economic Review: Papers and Proceedings 86

[109] Moore, M & P Ranjan (2005); "Globalisation vs Skill-biased Technological Change: Implications for Unemployment and Wage inequality"; Economic Journal 115

[110] Nattrass, N (2000); "Inequality, Unemployment and Wage-setting in South Africa"; Studies in Economics and Econometrics 24

[111] Nattrass, N (2004); "Unemployment and AIDS, the Social-democratic challenge for South Africa"; Development Southern Africa 21

[112] Negash, M (1999); "Corporate tax and capital structure: some evidence and implications"; The Investment Analysts Journal 56

[113] Nelson, R & E Phelps (1966); "Investment in humans, technological diffusion and economic growth"; American Economic Review 56

[114] Nickell, S & B Bell (1996); "Changes in the distribution of wages and unemployment in OECD countries"; American Economic Review 86

[115] Nickell, S & R Layard (1999); "Labour market institutions and economic performance" in O Ashenfelter & D Card (eds.); "Handbook of Labor Economics Vol 3C"; Elsevier North Holland, Amsterdam

[116] Ntuli, D (2006); "SA has eyes trained on skills shortage"; Sunday Times 30 April

[117] Oi, W & T Idson (1999); "Firm size and wages" in O Ashenfelter & D Card (eds.); "Handbook of Labour Economics Volume 3C"; Elsevier North Holland, Amsterdam

[118] Paton, C (2003); "Artisan Alert"; Financial Mail 20 June

[119] Pollin, R, G Epstein, J Heintz & L Ndikumana (2006); "An employment-targeted economics program for South Africa"; United Nations Development Programme, New York

[120] Pritchett, L (1999); "Where has all the education gone?"; World Bank Manuscript

[121] Rahman, A (2005); "The convergence implications of directed technological change"; unpublished

[122] Republic of South Africa (1995); "Basic Conditions of Employment Act"

[123] Republic of South Africa (1998); "Budget Review 1998"

[124] Rivera-Batiz, A & P Romer (1991); "Economic Integration and Endogenous Growth"; Quarterly Journal of Economics 106

[125] Robinson, J (1933); "The Economics of Imperfect Competition"; Macmillan & Co Ltd, London

[126] Roberts, M & E Skoufias (1997); "The Long-Run Demand for Skilled and Unskilled Labor in Colombian Manufacturing Industries"; Review of Economics and Statistics 79

[127] Romer, P (1986); "Increasing Returns and Long Run Growth"; Journal of Political Economy 94

[128] Romer, P (1990); "Endogenous Technological Change"; Journal of Political Economy 98, part II

[129] Sanders, M (2005); "Technology and the decline in demand for unskilled labour"; Edward Elgar

[130] Sato, K (1975); "Production Functions and Aggregation"; North Holland Publishing Company, Oxford

[131] Sato, K & T Koizumi (1973); "On the Elasticities of Substitution and Complementarity"; Oxford Economic Papers 25

[132] Savvides, A & M Zachariadis (2005); "International Technology Diffusion and the growth of TFP in the manufacturing sector of developing economies"; Review of Development Economics 9

[133] Schmookler, J (1966); "Invention and economic growth"; Harvard University Press

[134] Smith, H (2003); "Note on the changing structure of the manufacturing sector"; South African Reserve Bank Quarterly Bulletin March 2003

[135] Soderbom, M & F Teal (2004); "Size and efficiency in African manufacturing firms: evidence from firm-level panel data"; Journal of Development Economics 73

[136] South African Qualifications Authority (2004); "Trends in Public Higher Education in South Africa: 1992-2001"

[137] Statacorp.(2003); "Stata: Reference Manual: Release 8."; Texas

[138] Statistics South Africa (1998); "Metadata report for 1997 October Household Survey"

[139] Statistics South Africa (2000); "South African Statistics"

[140] Statistics South Africa (2005); "Labour Force Survey September 2000 to March 2005, Historical Series of Revised Estimates"; Statistical Release P0210 September

[141] Stern, D (2004); "Elasticities of Substitution and Complementarity"; Rensselaer Working Papers in Economics no. 0403

[142] Syrquin, M & G Hollander (1982); "Elasticities of Substitution and Complementarity: The General Case"; Oxford Economic Papers 34

[143] Teal, F (2000); "Real wages and the demand for skilled and unskilled male labour in Ghana's manufacturing sector: 1991-1995"; Journal of Development Economics 61

[144] Temple, J (2001); "Generalizations that aren't? Evidence on education and growth"; European Economic Review 45

[145] The Economist (2004); "Running to stand still"; 12 February

[146] Tinbergen, J (1975); "Income Distribution: Analysis and Policies"; North-Holland Publishing Co, Amsterdam

[147] Topel, R (1999); "Labor markets and economic growth" in O Ashenfelter & D Card (eds.); "Handbook of Labor Economics Vol 3C"; Elsevier North Holland, Amsterdam

[148] Troske, (1991); "Evidence on the employer size-skill premium from worker-establishment matched data"; Review of Economics and Statistics 81

[149] UNESCO (2000); "Education For All: Meeting Our Collective Commitments: Text adopted by the World Education Forum Dakar, Senegal, 26-28 April"

[150] Uzawa, H (1962); "Production Functions with Constant Elasticities of Substitution"; Review of Economic Studies 29

[151] Van den Honert, R (1997); "Intermediate Statistical Methods for Business and Economics"; UCT Press, Cape Town

[152] Van der Berg, S (2001); "Resource shifts in South African schools after the political transition"; Development Southern Africa 18

[153] Varian, H (1992); "Microeconomic Analysis 3rd edition"; WW Norton & Company, London

[154] Vere, J (2001); "Education, Technology and the Wage Structure in Taiwan: 1979:1998"; Research Programme in Development Studies Working Paper 201, Woodrow Wilson School of Public and International Afairs

[155] Wood, A (1994); "North South Trade, Employment and Inequality"; Oxford University Press

[156] Wood, E (1995); "Essays on Trade and Manufacturing in South Africa"; University of Cambridge PhD thesis

[157] Wooldridge, J (2002); "Econometric Analysis of Cross Sectional and Survey Data"; MIT Press

[158] Yaari, M (1965); "Uncertain lifetime, life insurance, and the theory of the consumer"; Review of Economic Studies 32

[159] Zellner, A, J Kmenta & J Dreze (1966); "Specification and Estimation of Cobb-Douglas Production Function Models"; Econometrica 34

Appendices

Appendix 1: Full output regression used to test value added separability

Dependent variable: Turnover

Variable	coef	p value	Variable	coef	p value
Constant	-0.92	0.30	ind2	-0.08	0.44
Capital	0.10	0.54	ind3	-0.13	0.21
Man/Prof	0.95	0.00	ind4	0.19	0.12
Sale/Cler	0.02	0.94	ind5	-0.17	0.11
Skil/Art	0.39	0.01	ind6	0.05	0.70
Semi	-0.21	0.30	ind7	-0.01	0.91
Un	0.23	0.09	ind8	-0.10	0.31
Raw	-0.04	0.89	ind9	0.07	0.55
0.5*Capital^2	0.00	0.91	loc2	0.30	0.05
Capital*Man/Prof	0.05	0.16	loc3	0.01	0.93
Capital*Sale/Cler	0.00	0.98	loc4	0.02	0.78
Capital*Skil/Art	0.01	0.58	loc5	0.05	0.78
Capital*Semi	-0.03	0.30	loc6	-0.02	0.92
Capital*Un	0.00	0.90	loc7	-0.28	0.19
0.5*Man/Prof^2	0.01	0.92	loc8	-0.08	0.75
Man/Prof*Sale/Cler	0.03	0.65	loc9	0.08	0.18
Man/Prof*Skil/Art	0.00	0.96	Exports as % sales	0.26	0.04
Man/Prof*Semi	0.03	0.56	Raw materials as % costs	0.00	0.21
Man/Prof*Un	0.01	0.74	Recruitment ease ManProf	0.00	0.92
0.5*Sale/Cle^2	0.02	0.80	Recruitment ease SaleCle	0.02	0.57
Sale/Cler*Skil/Art	0.04	0.27	Recruitment ease Skilart	-0.02	0.60
Sale/Cler*Semi	0.00	0.94	Recruitment ease Semi	0.07	0.14
Sale/Cler*Un	-0.01	0.64	Recruitment ease Un	-0.06	0.30
0.5*Skil/Art^2	-0.07	0.05	Productivity disatisfaction	0.01	0.73
Skil/Art*Semi	0.02	0.34	Training Expenditure	0.00	0.06
Skil/Art*Un	0.03	0.13	Market Conditions	0.00	0.80
0.5*Semi^2	0.03	0.42	Collective Bargaining	-0.05	0.07
Semi*Un	-0.04	0.14	Firm size > 50 employees	0.27	0.00
0.5*Un^2	-0.02	0.45	Ownermanaged	0.47	0.00
Capital*Raw	-0.02	0.64	Computer Investment as % Assets	0.00	0.87
Man/Prof*Raw	-0.16	0.00	Observations	210	
Sale/Cler*Raw	-0.01	0.89	R^2	0.98	
Skil/Art*Raw	-0.07	0.01	test of strong separability	0.00	
Semi*Raw	0.03	0.50			
Un*Raw	-0.01	0.74	(p values based on robust standard errors)		
0.5*Raw^2	0.15	0.02			

Appendix 1 (continued): Tests of weak separability of raw materials

Restriction number	F statistic	p value
1	0.01	0.918
2	0.01	0.918
3	0.01	0.931
4	0	0.992
5	0.01	0.943
6	0	0.956
7	0	0.999
8	0.59	0.444
9	0.03	0.854
10	0.11	0.741
11	0.07	0.790
12	0.02	0.879
13	0	0.962
14	0.25	0.621
15	0.03	0.854
16	0.01	0.943
17	0	0.944
18	0	0.957
19	0.01	0.938
20	4.71	0.032
21	0.07	0.797
22	0.84	0.361
23	0.85	0.358
24	0.07	0.790
25	0.31	0.581
26	2.46	0.119
27	0.07	0.793
28	0.07	0.797
29	0.64	0.424
30	0.06	0.804
31	0	0.947
32	0.01	0.915
33	0.04	0.846
34	0.04	0.847
35	0.05	0.816
36	0.02	0.894
37	0.11	0.742
38	0.97	0.326
39	0.1	0.750
40	3.76	0.055
41	2.38	0.125
42	0.07	0.788
43	0.03	0.867
44	0.74	0.392
45	0.05	0.827

Note, unadjusted p values reported

Appendix 2: List of variables

VARIABLE NAME	SYMBOL	OTHER
VARIABLES IN PRODUCTION FUNCTION (in logarithms unless otherwise stated)		
Sales	Q	R million turnover
Value Added	Y	Turnover*(1 - raw materials percentage)
Capital	x1	R million fixed assets - adjusted for shift capacity utilization
Managerial/Professional Labour	x2	Number of fulltime employees + 0.5*(number of part time employees)
Sales/Clerical Labour	x3	Number of fulltime employees + 0.5*(number of part time employees)
Skilled/Artisan Labour	x4	Number of fulltime employees + 0.5*(number of part time employees)
Semiskilled Labour	x5	Number of fulltime employees + 0.5*(number of part time employees)
Unskilled Labour	x6	Number of fulltime employees + 0.5*(number of part time employees)
Raw Materials	x7	(Raw materials percentage)*(turnover)
These variables are used to calculate the higher order and interaction terms		
VARIABLES IN COST FUNCTION AND COST SHARE EQUATIONS (in logarithms unless indicated otherwise)		
All variables based on wages are calculated using each of the four wage variables; wages are annualised for the regressions		
Cost of Capital	w1	Real interest rate + depreciation rate + effective tax rate + risk premium
Managerial/Professional Labour	w2	4 wages predicted using household survey data
Sales/Clerical Labour	w3	4 wages predicted using household survey data
Skilled/Artisan Labour	w4	4 wages predicted using household survey data
Semiskilled Labour	w5	4 wages predicted using household survey data
Unskilled Labour	w6	4 wages predicted using household survey data
Total Cost	C	Sum of factor costs; each factor cost (i) = w(i)x(i)
Value Added	Y	Turnover*(1 - raw materials percentage)
Factor share Man/Prof	ShareMan/Prof	w2*x2/C (not in log form)
Factor share Sale/Cle	ShareSale/Cle	w3*x3/C (not in log form)
Factor share Skil/Art	ShareSkil/Art	w4*x4/C (not in log form)
Factor share Semiskilled	ShareSemi	w5*x5/C (not in log form)
Factor shareUnskilled	Shareun	w6*x6/C (not in log form)
Managerial/Professional Labour - Cost of Capital	d2	w2-w1
Sales/Clerical Labour - Cost of Capital	d3	w3-w1
Skilled/Artisan Labour - Cost of Capital	d4	w4-w1
Semi-skilled Labour - Cost of Capital	d5	w5-w1
Unskilled Labour - Cost of Capital	d6	w6-w1
These variables are used to calculate the higher order and interaction terms		

Appendix 2 (continued)

VARIABLE NAME	SYMBOL	OTHER
CANDIDATES FOR CONTROLS/INSTRUMENTS		
8 Industry Dummies	Ind	
8 Province Dummies	Loc	
Exports as a % of sales	q10q9	
Raw Materials as % of Total Costs	q18c	
% Raw Materials Imported	q18d	
Difficulty Recruiting Man/Prof workers	q43_a	index; higher value means the firm has less difficulty
Difficulty Recruiting Sale/Cle workers	q43_b	index; higher value means the firm has less difficulty
Difficulty Recruiting Skil/Art workers	q43_c	index; higher value means the firm has less difficulty
Difficulty Recruiting Semi-skilled workers	q43_d	index; higher value means the firm has less difficulty
Difficulty Recruiting Unskilled workers	q43_e	index; higher value means the firm has less difficulty
Productivity Dissatisfaction	q45a	index
Training Expenditure	q47	R million
Market Effects	q50atot	Aggregation of six indices of effects macro-factors have on hiring decisions; index ranging from serious obstacle to major benefit
Unionised	q52q53	Index combining firm's exposure to firm- or plant-level collective bargaining and bargaining council agreements
Size dummy	nf2a	1 if there are more than 50 employees
Owner Managed	ownermanaged	1 if there is only one managerial/professional worker
Computer Investment in past year as percentage of assets	q19b1q11	
Average Equipment Age	q24_e	Constructed using information on what percentage of equipment is in a certain age range
Capital:labour ratio	Klratio	Capacity-adjusted capital stock divided by total number of employees (weighted to account for part time workers)

Appendix 3: Basic descriptive statistics of key variables

Variable	Mean	Std. Dev.	p25	Median	p75
Turnover (R million)	90	331	3	9	37
Capacity adjusted fixed capital stock (R million)	42	203	0.8	3	13
Managerial/Professional employees	18	67	2	4	8
Sales/Clerical employees	33	128	2	4	15
Skilled/Artisanal employees	23	73	1	4	12
Semi-skilled employees	99	537	4	13	50
Unskilled employees	72	335	2	10	38

Employee numbers include part time workers with a weighting of 0.5; p25 and p75 refer to percentiles

Appendix 4: Gross monthly wages according to $wage_{ind}$ definition

Category by Occupation and Industry	Mean	Std. Err.	Obs
Managerial/Professional Food & Beverages	3475	468	38
Managerial/Professional Wood, Pulp & paper	2844	565	16
Managerial/Professional Chemicals, Rubber & plastic	3427	523	17
Managerial/Professional Auto	2603	584	9
Managerial/Professional Textiles and Clothing	2562	326	19
Managerial/Professional Fabricated Metal	3527	947	5
Managerial/Professional Furniture	3699	833	5
Managerial/Professional Electrical, Electronic and other Machinery	3048	431	11
Managerial/Professional Printing & Publishing	3051	398	4
Sales/Clerical Food & Beverages	2479	431	58
Sales/Clerical Wood, Pulp & paper	1803	242	18
Sales/Clerical Chemicals, Rubber & plastic	2157	291	20
Sales/Clerical Auto	2977	629	8
Sales/Clerical Textiles and Clothing	1535	163	29
Sales/Clerical Fabricated Metal	2600	0	1
Sales/Clerical Furniture	2755	442	9
Sales/Clerical Electrical, Electronic and other Machinery	2159	417	19
Sales/Clerical Printing & Publishing	1571	420	8
Skilled/Artisans Food & Beverages	1562	161	88
Skilled/Artisans Wood, Pulp & paper	1689	139	50
Skilled/Artisans Chemicals, Rubber & plastic	2134	255	45
Skilled/Artisans Auto	2510	303	34
Skilled/Artisans Textiles and Clothing	1337	66	187
Skilled/Artisans Fabricated Metal	2047	437	17
Skilled/Artisans Furniture	1310	143	18
Skilled/Artisans Electrical, Electronic and other Machinery	2408	265	38
Skilled/Artisans Printing & Publishing	2462	317	31
Semi-skilled Food & Beverages	1720	88	131
Semi-skilled Wood, Pulp & paper	1604	119	38
Semi-skilled Chemicals, Rubber & plastic	1706	120	49
Semi-skilled Auto	1908	124	35
Semi-skilled Textiles and Clothing	1533	69	120
Semi-skilled Fabricated Metal	1756	229	17
Semi-skilled Furniture	1387	197	17
Semi-skilled Electrical, Electronic and other Machinery	1313	143	20
Semi-skilled Printing & Publishing	1817	260	10
Unskilled Food & Beverages	1200	83	177
Unskilled Wood, Pulp & paper	1047	70	39
Unskilled Chemicals, Rubber & plastic	1898	223	35
Unskilled Auto	1925	268	21
Unskilled Textiles and Clothing	1351	73	86
Unskilled Fabricated Metal	1401	140	18
Unskilled Furniture	955	114	16
Unskilled Electrical, Electronic and other Machinery	1780	232	22
Unskilled Printing & Publishing	1171	237	12

Appendix 5: Comparison of $wage_{some}$ and $wage_{all}$

		ManProf	Salecle	Skilart	Semi	Un
$Wage_{all}$	Categories	31	32	34	34	32
	Observations	3.9	5.3	14.6	11.8	13.2
$Wage_{some}$	Categories	7	10	15	10	14
	Observations	17.7	16.9	33.9	43.7	30.2

Number of categories each wage series is divided into, and the average number of observations in the household survey used to calculate the mean wage for each wage category.

Appendix 6: Costs of capital

Cost of Capital %	Frequency (# firms)
40	1
41	8
42	28
43	62
44	7
45	70
46	89
47	78
48	97
49	20
50	74
51	60
52	14
53	17

Appendix 7: Factor shares constructed from factor price data

Statistic	ManProf	Salecle	Skilart	Semi	Un	Capital
mean	8.7%	6.9%	6.6%	15.3%	10.8%	50.9%
median	6.1%	4.4%	2.8%	9.8%	6.7%	52.2%
p5	0.9%	0.0%	0.0%	0.0%	0.0%	12.8%
p25	3.2%	2.2%	0.1%	4.5%	2.4%	32.1%
p75	11.5%	8.6%	9.2%	22.5%	14.9%	69.2%
p95	24.8%	22.3%	25.1%	44.6%	36.0%	86.1%

Appendix 8: Distributions of factor shares based on factor costs

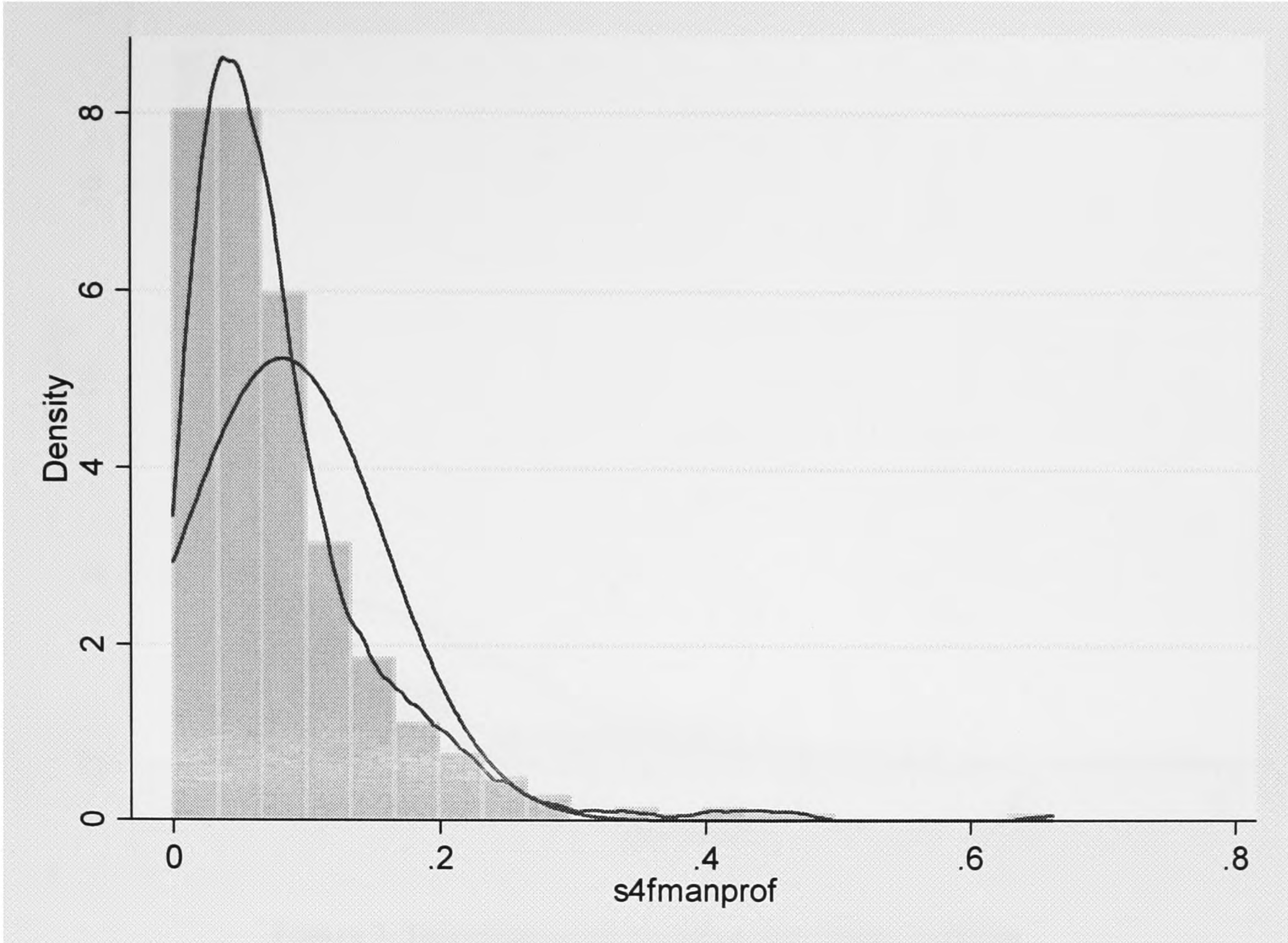


Figure 1: Distribution of cost share for Managerial/Professional

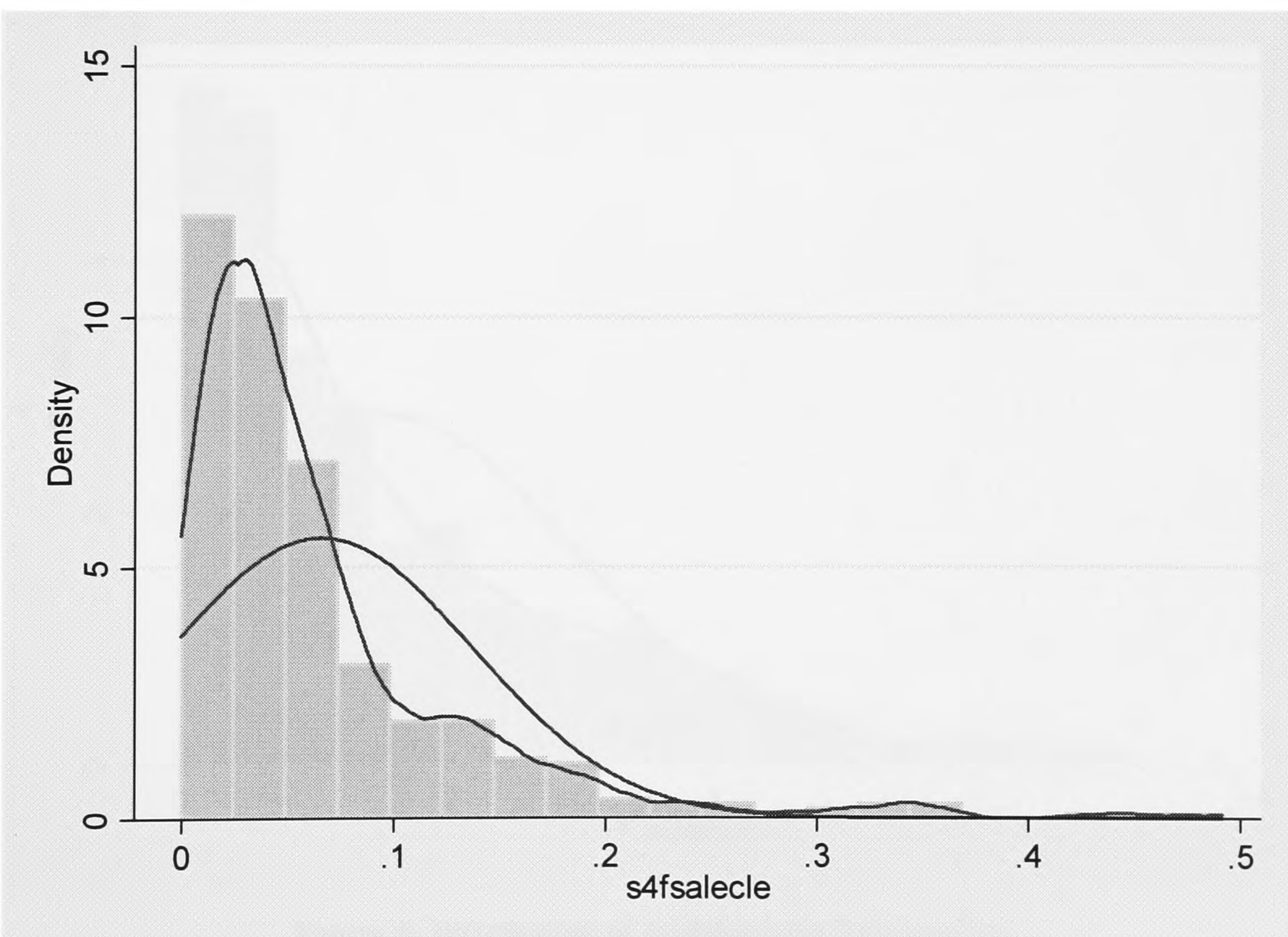


Figure 2: Distribution of cost share for Sales/Clerical

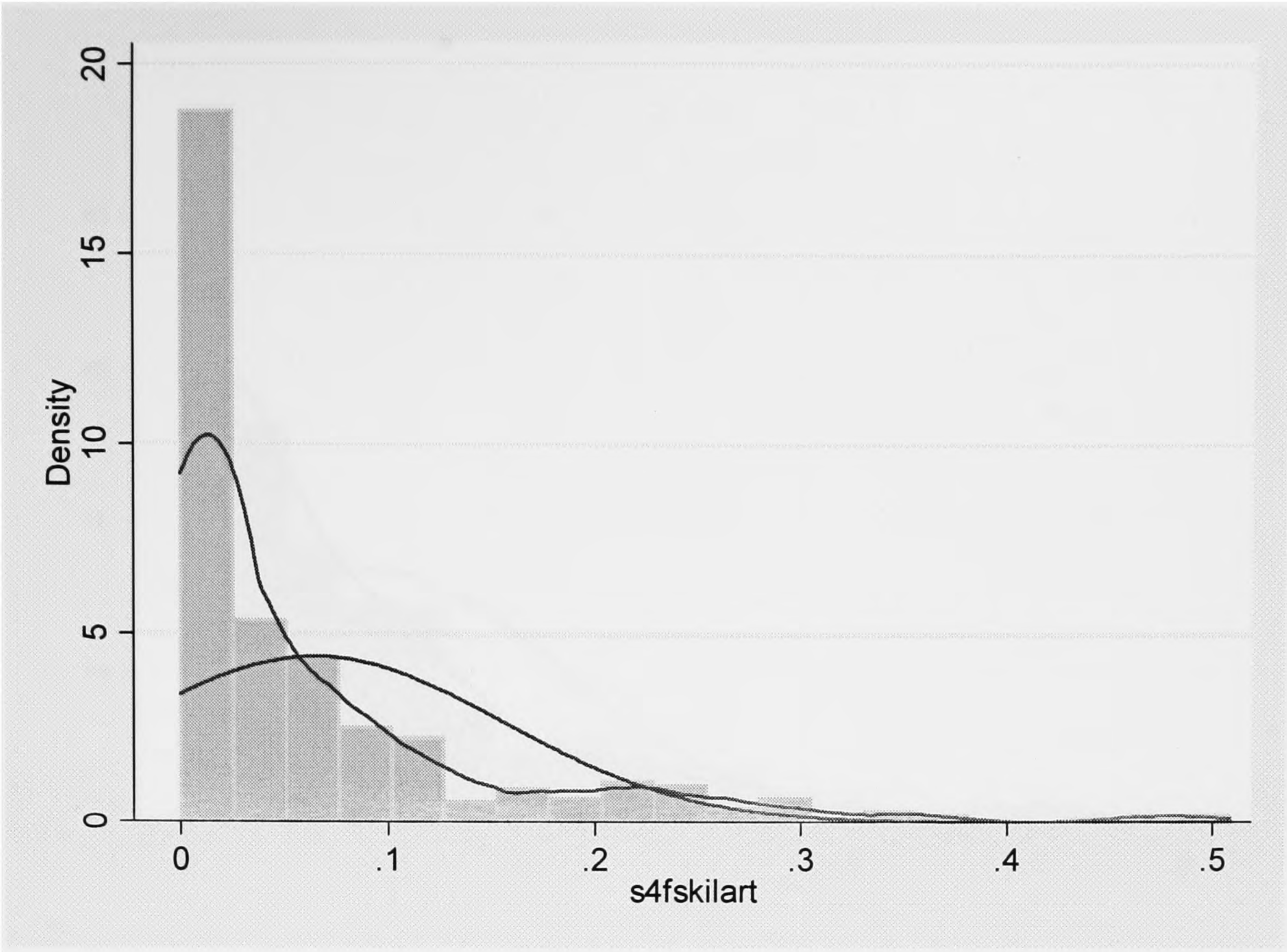


Figure 3: Distribution of cost share for Skilled/Artisan

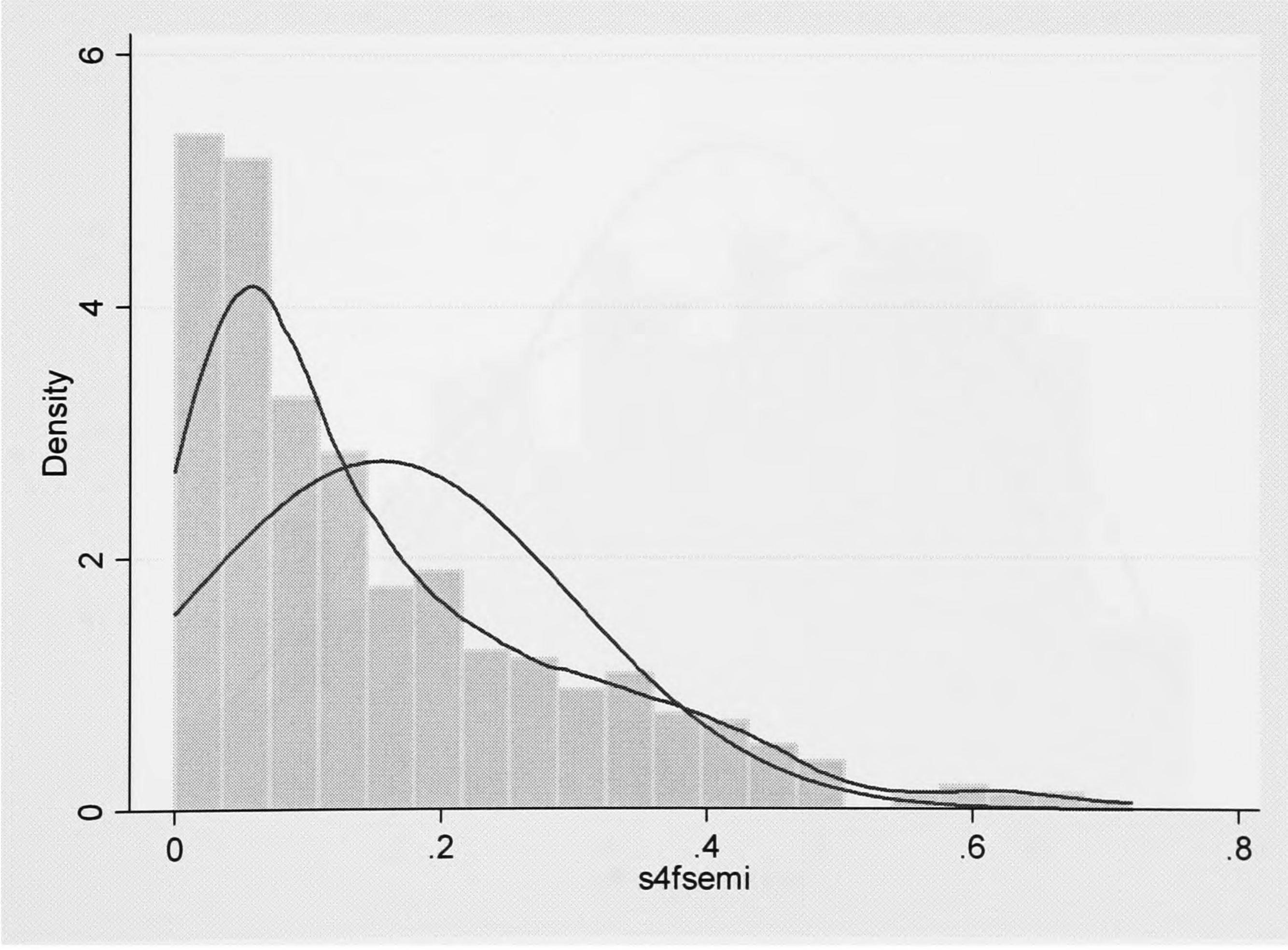


Figure 4: Distribution of cost share for Semi-skilled

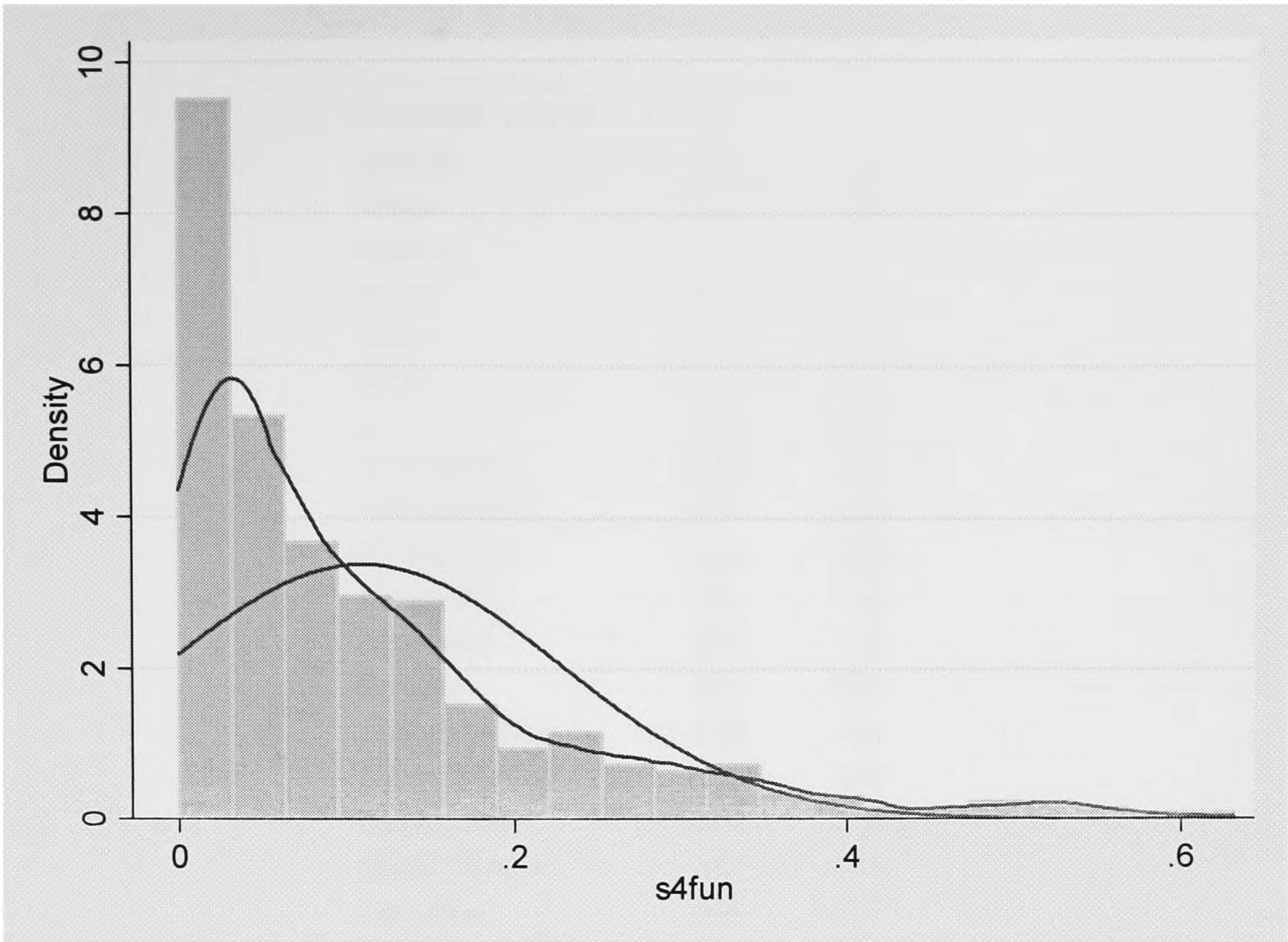


Figure 5: Distribution of cost share for Unskilled

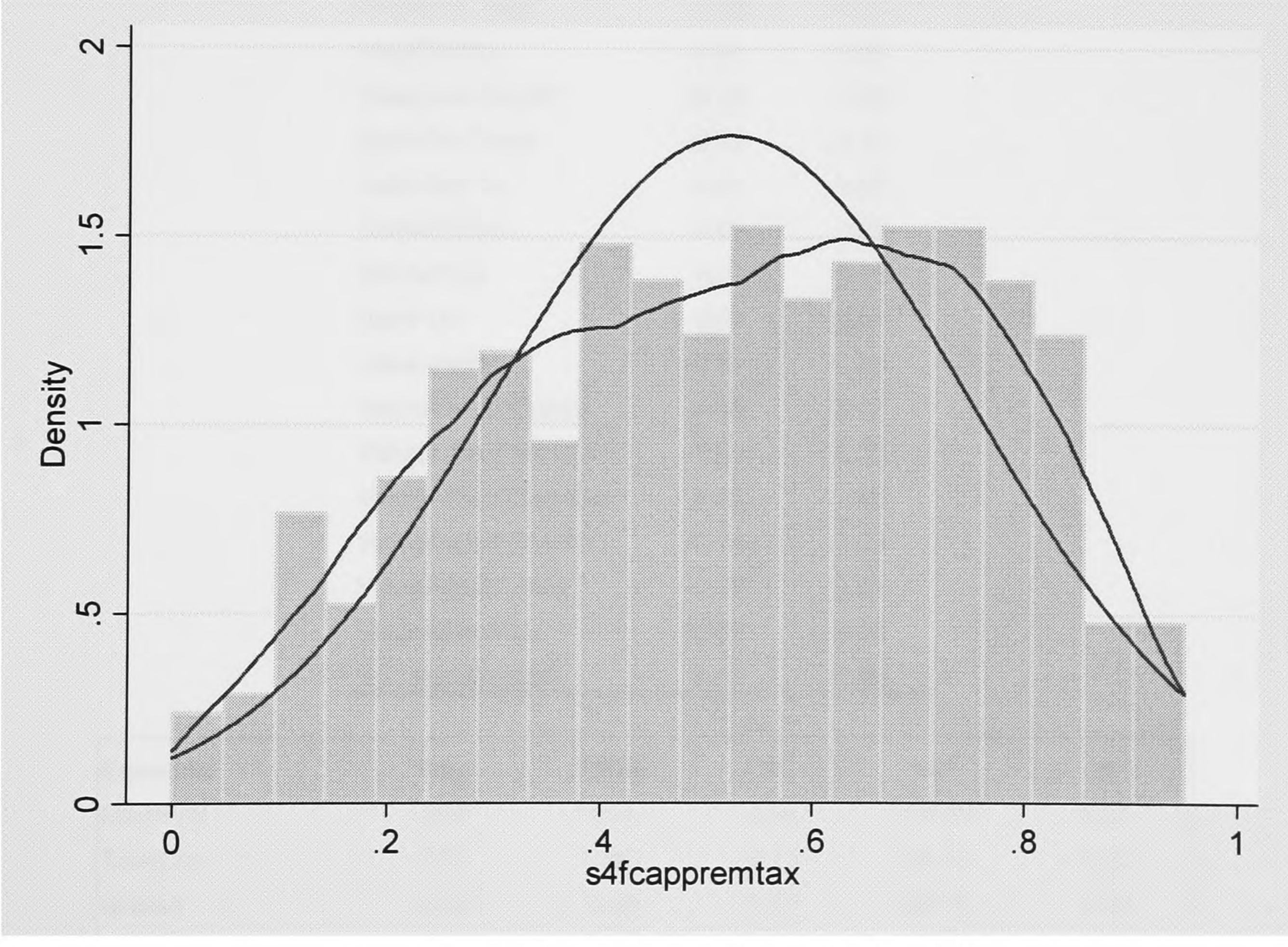


Figure 6: Distribution of cost share for Capital

Appendix 9: Regression using $wage_{some}$

Dependent Variable: Total cost		
Variable	coef	p>t
Capital	0.14	0.53
ManProf	0.20	0.00
Salecle	0.10	0.10
Skilart	-0.02	0.78
Semi	0.33	0.03
Un	0.25	0.03
0.5*Capital^2	0.06	0.43
0.5*ManProf^2	0.07	0.00
0.5*Salecle^2	0.05	0.00
0.5*Skilart^2	0.06	0.00
0.5*Semi^2	0.15	0.00
0.5*Un^2	0.05	0.07
Capital*Man/Prof	-0.02	0.39
Capital*Sale/Cler	0.00	0.82
Capital*Skil/Art	0.03	0.30
Capital*Semi	-0.03	0.51
Capital*Un	-0.03	0.37
Man/Prof*Sale/Cler	-0.02	0.02
Man/Prof*Skil/Art	0.00	0.83
Man/Prof*Semi	-0.04	0.01
Man/Prof*Un	0.00	0.86
Sale/Cler*Skil/Art	-0.02	0.08
Sale/Cler*Semi	-0.02	0.17
Sale/Cler*Un	0.01	0.57
Skil/Art*Semi	-0.06	0.00
Skil/Art*Un	-0.02	0.16
Semi*Un	-0.01	0.81
Value Added	0.30	0.00
ValueAdded*Capital	0.08	0.00
ValueAdded*Man/Prof	-0.04	0.00
ValueAdded*Sale/Cler	0.00	0.62
ValueAdded*Skil/Art	-0.01	0.03
ValueAdded*Semi	-0.02	0.08
ValueAdded*Un	-0.01	0.25
0.5* $ValueAdded^2$	0.13	0.00

Equation	**Obs**	**RMSE**	**"R^2"**	**chi^2**	**p**
Man/Prof	309	0.07	0.38	234.23	0.00
Sale/Cler	309	0.07	0.17	63.39	0.00
Skil/Art	309	0.08	0.21	95.78	0.00
Semi	309	0.14	0.19	77.51	0.00
Unskilled	309	0.10	0.10	34.34	0.23
Cost	309	0.65	0.75	2024.68	0.00

Note: controls not presented

Appendix 10: Compensated own-price elasticities using $wage_{some}$

statistic	$\lambda_{Capital}$	$\lambda_{ManProf}$	$\lambda_{Salecle}$	$\lambda_{Skilart}$	λ_{Semi}	λ_{Un}
p25	-0.46	-0.29	-0.36	-0.34	-0.53	-0.51
p50	-0.33	0.18	0.07	0.25	-0.41	-0.32
p75	-0.21	1.19	0.93	2.16	-0.08	0.40

Appendix 11: Variance inflation factors from single equation estimate

Variable	VIF
0.5*Salecle^2	211473
Sale/Cler*Semi	182653
Sale/Cler*Skil/Art	168828
Semi*Un	162363
Man/Prof*Skil/Art	101204
0.5*Semi^2	95128
0.5*Un^2	68038
Skil/Art*Semi	56352
Man/Prof*Skil/Art	47955
Skilart	38510
Mean	*18702*

Appendix 12: Regression using firm-size adjusted wages including the Sales/Clerical occupation

Dependent Variable: Total Cost

Variable	coef	p value
Capital	0.69	0.31
ManProf	-0.03	0.93
Salecle	-0.60	0.11
Skilart	0.26	0.20
Semi	0.57	0.27
Un	0.12	0.66
0.5*Capital^2	-0.38	0.04
0.5*ManProf^2	0.02	0.70
0.5*Salecle^2	0.07	0.11
0.5*Skilart^2	0.01	0.58
0.5*Semi^2	0.06	0.58
0.5*Un^2	0.03	0.52
Capital*Man/Prof	0.09	0.12
Capital*Sale/Cler	0.02	0.60
Capital*Skil/Art	0.11	0.04
Capital*Semi	0.06	0.56
Capital*Un	0.10	0.17
Man/Prof*Sale/Cler	-0.02	0.49
Man/Prof*Skil/Art	-0.04	0.04
Man/Prof*Semi	-0.03	0.54
Man/Prof*Un	-0.02	0.48
Sale/Cler*Skil/Art	-0.01	0.32
Sale/Cler*Semi	0.01	0.77
Sale/Cler*Un	-0.06	0.02
Skil/Art*Semi	-0.06	0.08
Skil/Art*Un	-0.01	0.78
Semi*Un	-0.04	0.48
Value Added	1.38	0.00
0.5*ValueAdded^2	0.03	0.40
ValueAdded*Capital	-0.34	0.00
ValueAdded*Man/Prof	0.10	0.29
ValueAdded*Sale/Cler	0.08	0.32
ValueAdded*Skil/Art	-0.09	0.12
ValueAdded*Semi	0.25	0.01
ValueAdded*Un	0.01	0.93

Equation	Obs	RMSE	"R^2"	chi^2	p
Man/Prof	296	0.07	0.26	103.57	0.00
Sale/Cler	296	0.06	0.10	30.74	0.02
Skil/Art	296	0.08	0.17	64.79	0.00
Semi	296	0.14	0.11	41.48	0.02
Unskilled	296	0.10	0.11	42.36	0.02
Cost	296	0.58	0.82	1563.57	0.00

Note: Controls not reported

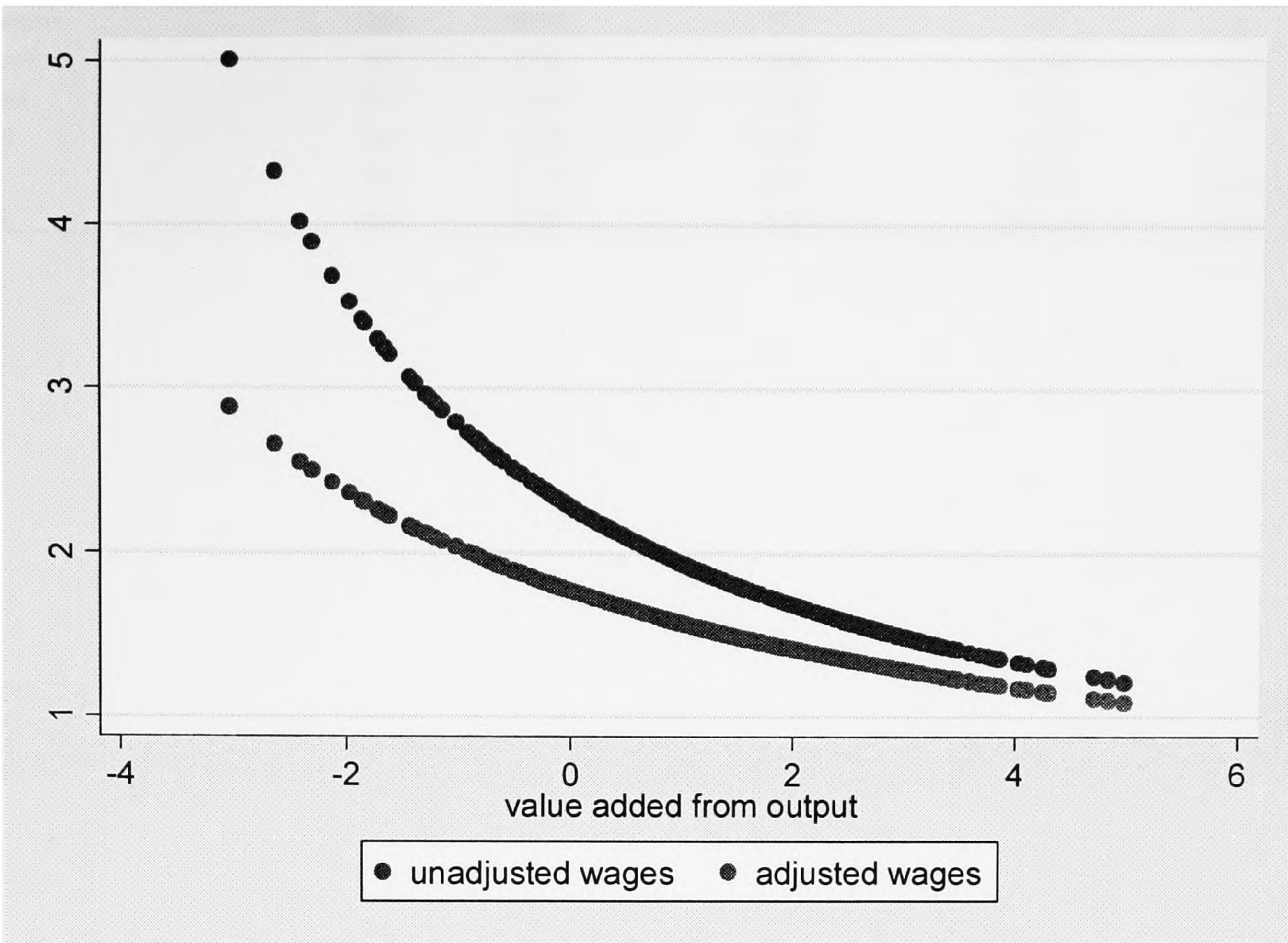
Appendix 13: Compensated own-price elasticities using firm-size adjusted wages including the Sales/Clerical occupation

statistic	$\lambda_{Capital}$	$\lambda_{ManProf}$	$\lambda_{Salecle}$	$\lambda_{Skilart}$	λ_{Semi}	λ_{Un}
p25	-1.75	-0.73	-0.16	-0.73	-0.49	-0.62
p50	-1.15	-0.68	0.41	-0.65	-0.38	-0.51
p75	-0.84	-0.48	1.63	-0.27	0.08	-0.09

Appendix 14: Returns to scale using firm-size adjusted wages and unadjusted wages

statistic	adjusted	unadjusted
mean	1.59	2.01
p5	1.25	1.44
p25	1.39	1.65
p50	1.57	1.93
p75	1.73	2.21
p95	2.09	2.91

Appendix 15: Graphical comparison of calculated returns to scale



Estimates of returns to scale using adjusted ($wage_{size}$) and unadjusted ($wage_{some}$) wages converge with each other and approach a value of unity at high output (Value Added in the Log of Millions of Rands).

Appendix 16: Diagnostics for system of final cost equation

Equation	Obs	RMSE	"R^2"	chi^2	p
Man/Prof	307	0.06	0.43	232.62	0.00
Skilart	307	0.08	0.18	71.78	0.00
Semi	307	0.13	0.16	61.78	0.00
Un	307	0.11	0.11	38.65	0.02
Cost	307	0.54	0.85	2021.29	0.00

Appendix 17: Normality tests for predicted factor shares and residuals

Variable	**p(skewness)**	**p(kurtosis)**	**adjusted $Chi^2(2)$**	**p**
Un	0	0	32.01	0
Un (res)	0	0	.	0
Semi	0	0.922	12.74	0.002
Semi (res)	0	0.022	39.93	0
Skil/Art	0.032	0.389	5.37	0.068
Skil/Art (res)	0	0	.	0
Man/Prof	0	0	.	0
Man/Prof (res)	0	0	.	0

Appendix 18: Uncompensated elasticities of labour demand for clothing, furniture and exports

λ_{ij} **for clothing**		j				
		Capital	Man/Prof	Skil/Art	Semi	Un
	Capital	-1.28*	0.15*	0.16*	0.41*	0.14*
i	Man/Prof	1.12*	-0.56*	-0.43*	-0.27*	-0.30*
	Skil/Art	2.24*	-0.79*	-0.34	-1.72*	0.19*
	Semi	0.91*	-0.08*	-0.26*	--0.83*	-0.16*
	Un	0.85*	-0.24*	0.08*	-0.47*	-0.65*

λ_{ij} **for furniture**		j				
		Capital	Man/Prof	Skil/Art	Semi	Un
	Capital	-1.61*	0.12*	0.15*	0.29*	0.10*
i	Man/Prof	0.57*	-0.71*	-0.30*	-0.27*	-0.23*
	Skil/Art	0.78*	-0.31*	--0.75*	-0.73*	0.06*
	Semi	0.91	-0.17	-0.42*	-0.95*	-0.32*
	Un	0.45*	-0.21*	0.05*	-0.47*	-0.76*

λ_{ij} **for exports**		j				
		Capital	Man/Prof	Skil/Art	Semi	Un
	Capital	-4.48*	-0.28*	-0.14	-0.60*	-0.46*
i	Man/Prof	-2.13*	-1.01*	-0.66*	-1.25*	-0.84*
	Skil/Art	-1.77	-0.9	-0.94	-2.05	-0.46*
	Semi	-1.93*	-0.58*	-0.70*	-1.82*	-0.85*
	Un	-2.43*	-0.65*	-0.17	-1.40*	-1.29*

* denotes sign consistent for at least 95% of firms

Appendix 19: Diagnostics for system with separability imposed

Equation	Obs	RMSE	"R^2"	chi^2	p
Man/Prof	307	0.06	0.43	236.06	0
Skilart	307	0.08	0.17	71.06	0
Semi	307	0.13	0.16	60.85	0
Un	307	0.11	0.11	41	0.0082
Cost	307	0.54	0.85	2018.55	0

Appendix 20: OLS vs 2SLS

Dependent variable: Value Added

	coef	p>t	[80% Conf	Interval]
Capital	**0.164**	**0.000**	**0.118**	**0.210**
Man/Prof	**0.277**	**0.000**	**0.186**	**0.368**
Sale/Cle	**0.296**	**0.000**	**0.208**	**0.384**
Skil/Art	**0.156**	**0.000**	**0.099**	**0.213**
Semi	0.037	0.453	-0.026	0.100
Un	0.051	0.148	0.006	0.096
Ind dummy 1	0.007	0.957	-0.159	0.173
Ind dummy 2	**0.443**	**0.012**	**0.220**	**0.667**
Ind dummy 3	0.055	0.648	-0.101	0.212
Ind dummy 4	0.208	0.193	0.003	0.412
Ind dummy 5	-0.089	0.543	-0.276	0.098
Loc dummy 1	0.055	0.642	-0.097	0.207
Loc dummy 2	0.042	0.717	-0.108	0.192
Loc dummy 3	0.048	0.659	-0.092	0.188
Exports %	0.360	0.047	0.128	0.591
Recruitment MP	-0.001	0.982	-0.067	0.064
Recruitment SC	**0.137**	**0.044**	**0.050**	**0.224**
Recruitment Skilart	0.008	0.905	-0.079	0.096
Recruitment Semi	0.014	0.866	-0.092	0.120
Recruitment Un	-0.121	0.216	-0.247	0.004
Productivity dissat.	-0.034	0.390	-0.084	0.017
Training expenditure	**0.000**	**0.000**	**0.000**	**0.000**
Market conditions	0.000	0.998	-0.014	0.014
Size dummy	0.450	0.001	0.275	0.626
Ownermanaged	**0.424**	**0.019**	**0.193**	**0.656**
_cons	-0.176	0.664	-0.698	0.345

OLS for value-added with robust standard errors. The R^2 is 0.87, the root mean square error is 0.59, and the F-statistic is 111. Significant variables are in bold. The sum of the input coefficients is 0.98.

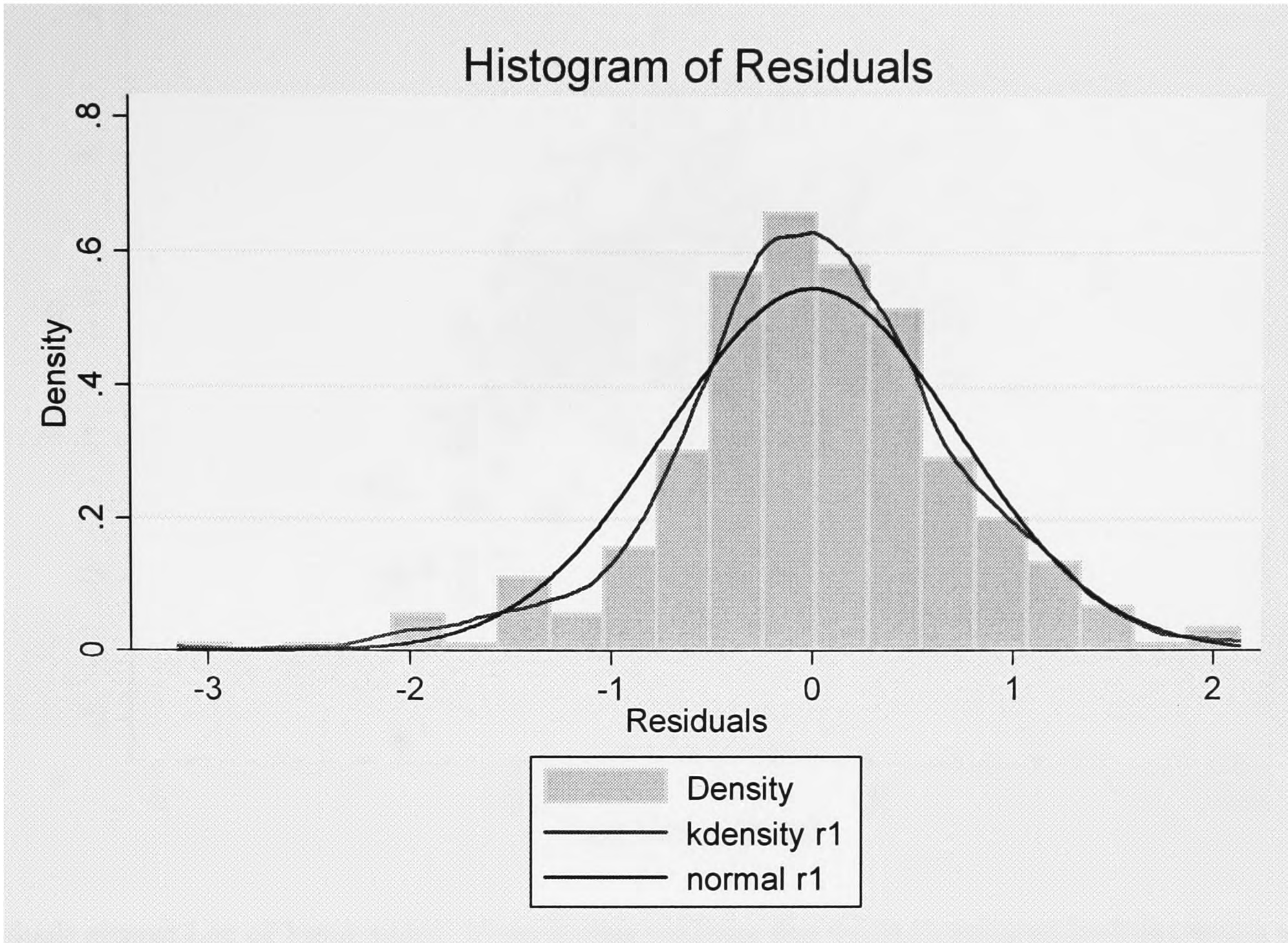
Dependent Variable: Value Added (second stage)

	coef	p>t	[80% Conf	Interval]
Capital	0.245	0.214	-0.008	0.497
Man/Prof	0.545	0.242	-0.052	1.142
Sale/Cle	0.364	0.379	-0.167	0.895
Skil/Art	-0.055	0.792	-0.323	0.212
Semi	0.029	0.854	-0.172	0.230
Un	0.019	0.898	-0.170	0.208
Ind dummy 1	-0.169	0.622	-0.610	0.272
Ind dummy 2	0.275	0.392	-0.137	0.687
Ind dummy 3	0.130	0.403	-0.070	0.330
Ind dummy 4	**0.377**	**0.088**	**0.094**	**0.661**
Ind dummy 5	-0.106	0.625	-0.386	0.173
Loc dummy 1	-0.069	0.662	-0.273	0.134
Loc dummy 2	0.091	0.544	-0.101	0.282
Loc dummy 3	0.063	0.613	-0.097	0.222
Exports %	0.233	0.556	-0.275	0.741
Size dummy	0.145	0.648	-0.263	0.554
Ownermanaged	**0.804**	**0.102**	**0.175**	**1.432**
_cons	-0.289	0.471	-0.803	0.225

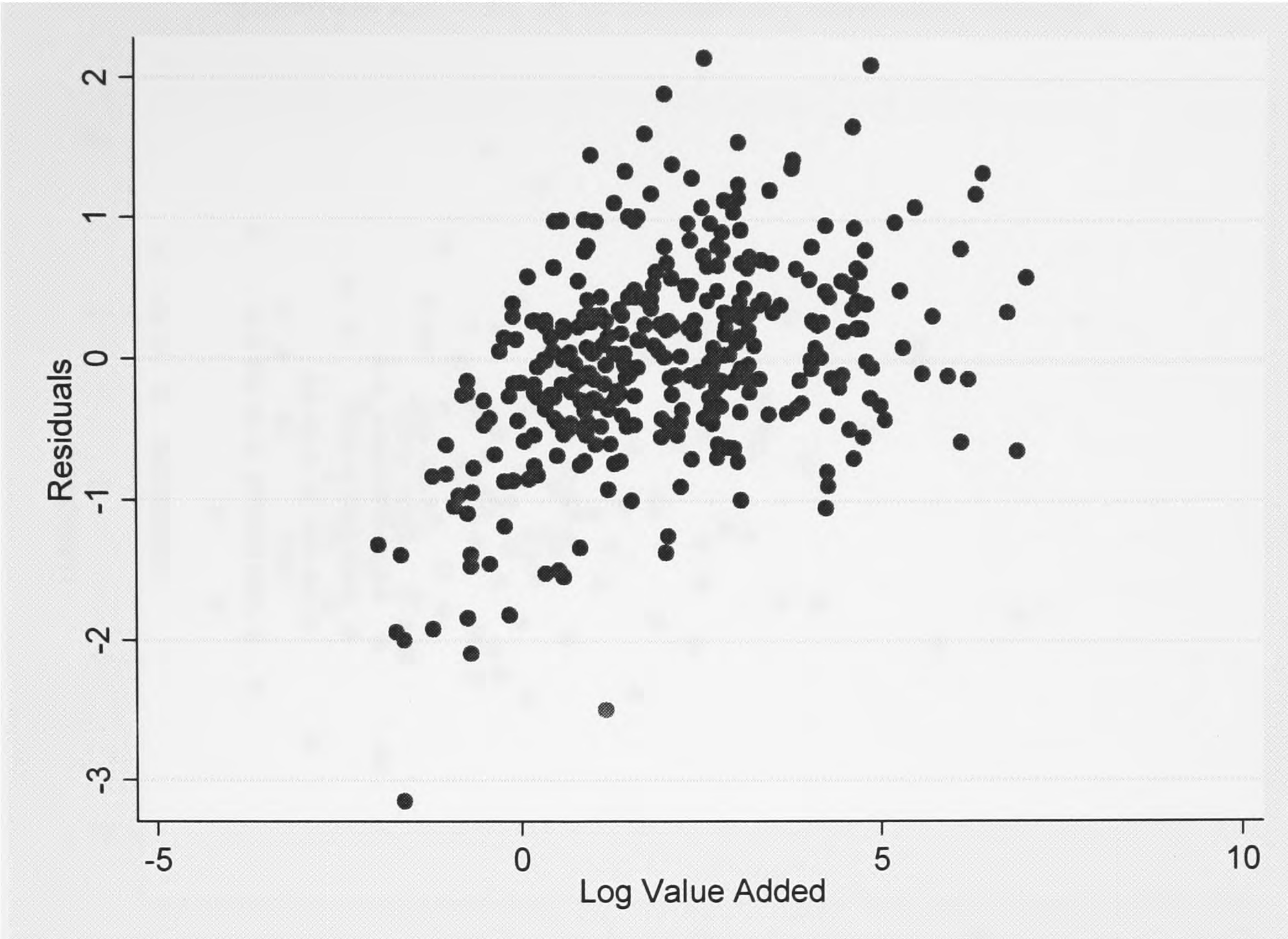
Second stage of 2SLS regression for value added with all 6 factors instrumented, with robust standard errors; additional instruments are factor prices, recruitment difficulties, productivity dissatisfaction, training expenditure and market conditions. The R^2 is 0.83, the root mean square error is 0.66 and the F-statistic is 69. Significant variables are in bold. The sum of the input coefficients is 1.15

Hausman specification test for endogeneity of OLS: $Chi^2(17)=10.08$; $p>0.90$
Test for endogeneity of instruments: $Chi^2(8)=8.435$; critical value for $p<0.20$ is 11.03

Appendix 21: Residuals from OLS value added regression

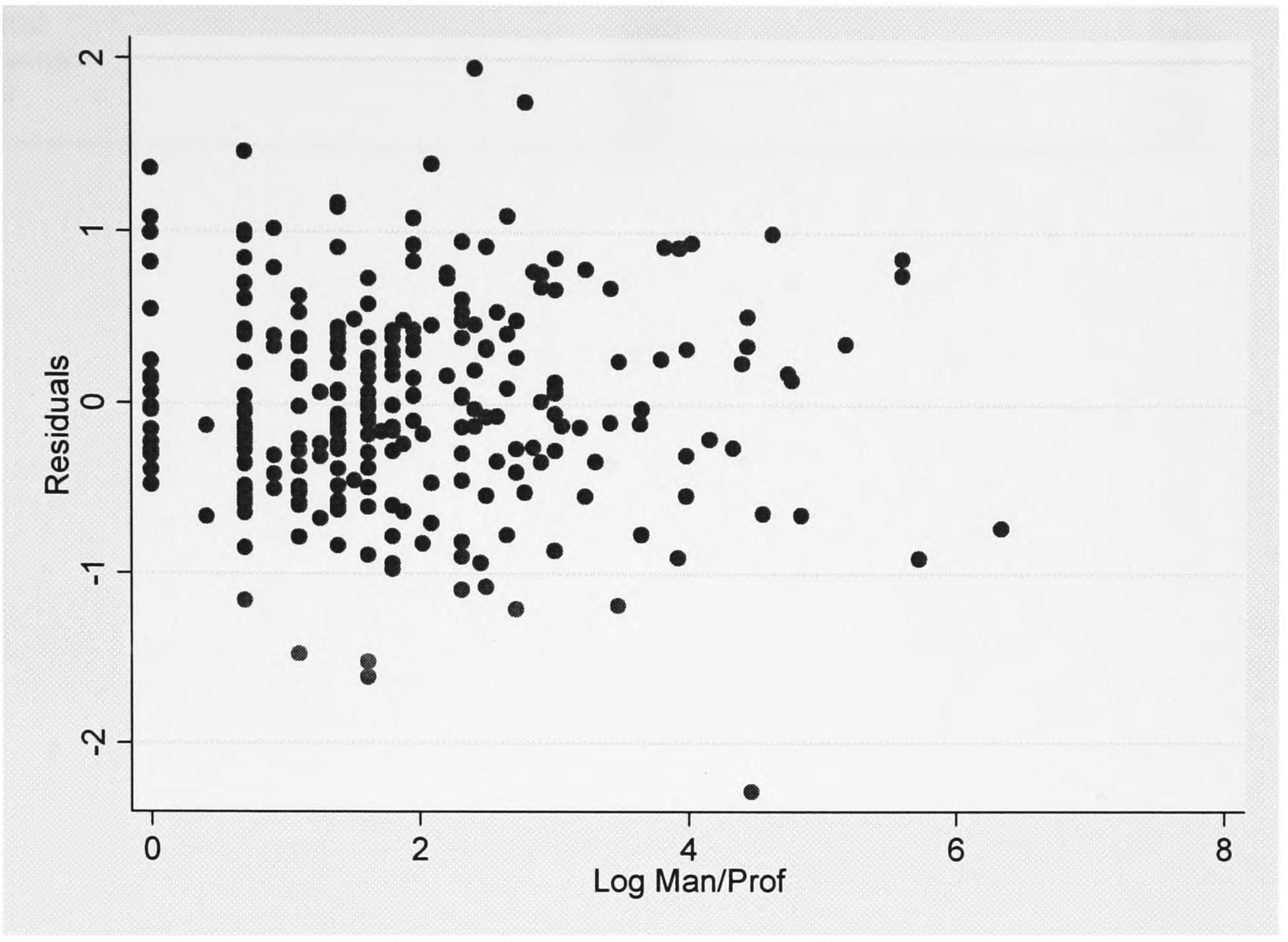


Appendix 22: Plot of residuals



Plot of residuals against Log of Value added. There is clear evidence that the fit is not good for firms with a value added of less than 1 million Rand (Log Value Added = 0). The measured returns to scale are 1.13, significantly different from unity at 5%.

Appendix 23: Plot of residuals on restricted sample



After removing small firms, the residuals suggest firms with 1 manager (Log=0) tend to be more efficient.

Appendix 24: Predicted factor shares

Factor	Predicted share	p value
Capital	0.15	0.00
Managerial/Professional	0.22	0.00
Sales/Clerical	0.40	0.00
Skilled/Artisanal	0.10	0.01
Semiskilled	0.09	0.03
Unskilled	0.10	0.00

Appendix 25: Manual two-stage procedure for predicting endogenous unskilled labour quantity

Dependent variable: Unskilled workers (first stage)	coef	p value
Unskilled wages	0.00	0.02
0.5*Capital²	-0.01	0.91
Capital*Man/Prof	0.03	0.79
Capital*Sale/Cler	0.02	0.81
Capital*Skil/Art	0.05	0.50
Capital*Semi	-0.01	0.87
0.5*Man/Prof²	0.08	0.69
Man/Prof*Sale/Cler	-0.11	0.48
Man/Prof*Skil/Art	0.00	1.00
Man/Prof*Semi	0.01	0.89
0.5*Sale/Cle²	-0.09	0.58
Sale/Cler*Skil/Art	0.20	0.06
Sale/Cler*Semi	0.04	0.66
0.5*Skil/Art²	-0.26	0.03
Skil/Art*Semi	0.04	0.57
0.5*Semi²	-0.07	0.34
ind2	0.07	0.90
ind3	0.31	0.50
ind4	0.50	0.13
ind5	0.45	0.21
ind6	-0.21	0.62
ind7	-0.08	0.83
ind8	0.20	0.70
ind9	0.52	0.10
loc2	0.41	0.46
loc3	-0.07	0.80
loc4	0.45	0.12
loc5	-0.17	0.75
loc6	-0.22	0.79
loc7	1.32	0.14
loc8	0.42	0.55
loc9	-0.13	0.63
Raw materials as % costs	0.00	0.24
Training expenditure as % turnover	0.00	0.29
Firm size > 50 employees	1.24	0.00
Computer Investment as % Assets	-1.61	0.44
Recruitment ease: unskilled	0.14	0.49
_cons	2.74	0.01
Number of observations	246	
R^2	0.64	

Appendix 25 (continued)

Dependent variable: Value Added (second stage)

Variable	coef	p value
constant	-0.65	0.05
Capital	0.26	0.00
Man/Prof	0.01	0.91
Sale/Cler	0.34	0.00
Skil/Art	-0.05	0.53
Semi	0.16	0.04
Un	0.28	0.01
0.5*Capital²	0.12	0.00
Capital*Man/Prof	-0.04	0.43
Capital*Sale/Cler	-0.01	0.85
Capital*Skil/Art	0.01	0.86
Capital*Semi	-0.05	0.13
Capital*Un	-0.04	0.62
0.5*Man/Prof²	-0.04	0.66
Man/Prof*Sale/Cler	-0.04	0.49
Man/Prof*Skil/Art	0.04	0.44
Man/Prof*Semi	0.11	0.21
Man/Prof*Un	-0.04	0.48
0.5*Sale/Cle²	-0.03	0.59
Sale/Cler*Skil/Art	-0.05	0.44
Sale/Cler*Semi	0.07	0.08
Sale/Cler*Un	0.11	0.03
0.5*Skil/Art²	-0.04	0.37
Skil/Art*Semi	0.11	0.14
Skil/Art*Un	0.00	0.98
0.5*Semi²	0.05	0.29
Semi*Un	-0.15	0.01
0.5*Un²	0.02	0.16

Note: All terms involving unskilled labour include predicted value for unskilled labour from first stage; p values do not account for this; controls omitted

Appendix 26: Manual two-stage procedure for predicting endogenous less skilled labour quantity

Dependent variable: less skilled workers (first stage)		
Variable	coef	p value
_cons	1.54	0.07
Capital	0.12	0.13
more	-0.05	0.77
0.5*Capital²	-0.07	0.05
Capital*More	0.09	0.05
0.5*More²	0.04	0.58
ind2	0.79	0.01
ind3	0.90	0.00
ind4	0.92	0.00
ind5	1.13	0.00
ind6	1.06	0.00
ind7	0.83	0.01
ind8	0.82	0.02
ind9	0.68	0.00
loc2	0.62	0.07
loc3	-0.11	0.54
loc4	0.11	0.52
loc5	0.30	0.38
loc6	1.16	0.05
loc7	1.28	0.01
loc8	0.80	0.08
loc9	-0.01	0.98
Market Conditions	-0.01	0.30
Raw materials as % costs	0.01	0.01
Training expenditure as % turnover	0.00	0.13
Firm size > 50 employees	1.10	0.00
Owner managed	-0.64	0.00
Computer Investment as % Assets	0.96	0.08
Unskilled wages	0.00	0.16
Semiskilled wages	0.00	0.13
Recruitment ease ManProf	-0.01	0.89
Recruitment ease SaleCle	0.11	0.17
Recruitment ease Skilart	0.02	0.78
Recruitment ease Semi	-0.03	0.67
Recruitment ease Un	-0.08	0.54
Observations	319	
R^2	0.79	

Appendix 26 (continued)

Dependent variable: Value added (second stage)	
Variable	coef
Constant	-1.43
Capital	0.36
More	-0.07
Less	0.71
0.5*Capital²	0.22
Capital*More	-0.08
Capital*Less	-0.14
0.5*More²	-0.01
More*Less	0.10
0.5*Less²	0.04

Note: All terms involving less skilled labour include predicted value for less skilled labour from first stage; p values do not account for this; controls omitted

Less skilled labour is predicted using other factor types and their higher order interactions (not those involving less skilled labour), unskilled and semi-skilled wages as well as various controls. Predicted values are interacted with other factor inputs to generate predicted higher order terms. p values are not valid given this procedure. As shown in Appendix 27, the values for the own elasticities are implausibly high and more skilled and less skilled labour are found to be substitutes.

Appendix 27: Predicted (aggregated) elasticities of factor price based on manual two-stage procedure

i	j	coef
ε^{ρ}_{ij}		
Capital	Capital	9.8
	More	-9.8
More	Capital	-3.17
	More	3.17
ρ		
Less	Capital	-1.35
	More	-2.35

Appendix 28: 2SLS for predicted less skilled labour quantity

Unskilled wages, semi-skilled wages and five measures of recruitment difficulty are used as instruments for less skilled labour and its three higher order terms. The first stage regression for less skilled labour shows plausible signs including significantly negative unskilled wages (but insignificantly positive semi-skilled wages). The results in Appendix 29 find evidence for complementarity between skilled and unskilled labour, but the evidence is very weak given the high p values. Furthermore, the elasticities, especially the own-elasticities, are implausibly high.

Equation	Obs	RMSE	R-sq	F-Stat	P
Value added	318	0.74	0.80	9.27	0.00
Less	318	1.46	0.09	4.29	0.00
Capital*Less	318	11.47	0.09	4.53	0.00
More*Less	318	9.57	0.12	5.99	0.00
0.5*Less2	318	5.97	0.10	4.82	0.00

Endogenous variables: Value added, Less, Capital*Less, More*Less, 0.5*Less^2. Exogenous variables: Capital, More, 0.5*Capital^2, Capital*More, 0.5*More^2, ind2, ind3, ind4, ind5, ind6, ind7, ind8, ind9, loc2, loc3, loc4, loc5, loc6, loc7, loc8, loc9, Raw materials as % costs, Training expenditure as % turnover, Firm size > 50 employees, Market conditions, Computer Investment as % Assets, Owner managed, Unskilled wages, Semi-skilled wages, Recruitment ease Man/Prof, Recruitment ease Sale/Cle, Recruitment ease Skil/Art, Recruitment ease Semi, Recruitment ease Unskilled

Appendix 28 (continued)

Dependent Variable: Less skilled workers		
Unskilled wage	-0.00	0.04
Semi-skilled wage	0.00	0.16
Recruitment ease ManProf	-0.40	0.01
Recruitment ease SaleCle	0.11	0.45
Recruitment ease Skilart	-0.27	0.05
Recruitment ease Semi	0.34	0.02
Recruitment ease Un	0.39	0.01
Constant	2.78	0.00
Dependent variable: Capital*Less		
Unskilled wage	-0.00	0.17
Semi-skilled wage	0.00	0.06
Recruitment ease ManProf	-3.58	0.00
Recruitment ease SaleCle	1.08	0.36
Recruitment ease Skilart	-2.10	0.05
Recruitment ease Semi	2.01	0.08
Recruitment ease Un	3.05	0.10
Constant	-0.80	0.90
Dependent variable: 0.5*Less^2		
Unskilled wage	-0.00	0.03
Semi-skilled wage	0.00	0.09
Recruitment ease ManProf	-1.85	0.00
Recruitment ease SaleCle	0.44	0.47
Recruitment ease Skilart	-0.92	0.10
Recruitment ease Semi	1.49	0.01
Recruitment ease Un	1.54	0.11
Constant	3.76	0.26
Dependent variable: Less*More		
Unskilled wage	-0.00	0.34
Semi-skilled wage	0.01	0.01
Recruitment ease ManProf	-3.32	0.00
Recruitment ease SaleCle	0.49	0.62
Recruitment ease Skilart	-1.63	0.07
Recruitment ease Semi	2.40	0.01
Recruitment ease Un	2.08	0.18
Constant	0.30	0.96

Dependent variable: Value Added (second stage)	
Capital	0.79
More	0.13
Less	0.09
0.5*Capital^2	0.27
Capital*More	0.11
Capital*Less	-0.39
0.5*More^2	-0.32
More*Less	0.20
0.5*Less^2	0.18
ind2	-0.13
ind3	-0.07
ind4	-0.11
ind5	-0.29
ind6	-0.29
ind7	-0.12
ind8	-0.52
ind9	-0.04
loc2	-0.67
loc3	-0.01
loc4	-0.25
loc5	-0.04
loc6	-1.44
loc7	-1.72
loc8	-0.46
loc9	-0.07
Raw materials as % costs	-0.02
Training expenditure as % turnover	0.00
Market Conditions	0.03
Firm size > 50 employees	0.18
Computer Investment as % Assets	-0.61
Ownermanaged	0.75
Constant	-0.24

Appendix 29: Predicted (aggregated) elasticities of factor price based on 2SLS

i	j	coef	p
ε^{ρ}_{ij}			
Capital	Capital	17.83	0.87
	More	-17.83	0.87
More	Capital	-6.5	0.85
	More	6.5	0.85
ρ			
Less	Capital	-5.51	0.87
	More	6.5	0.85