Modeling Labor Market Policy in Developing Countries: A Selective Review of the Literature and Needs for the Future

Gary S. Fields

May, 2007
Modeling Labor Market Policy in Developing Countries: A Selective Review of the Literature and Needs for the Future

Gary S. Fields
Cornell University
May, 2007

Outline

Introduction
The Welfare Economics of Policy Analysis in the Labor Market
  The Pareto criterion
  First order dominance
  Abbreviated labor market well-being functions
  Minimizing a low-earnings index
  Social cost-benefit analysis
  Conclusions regarding welfare economics

Theoretical Models
  The overall framework
  The need for multisector models
  Modeling the formal sector labor market
  Modeling the informal sector labor market
  Modeling intersectoral linkages in the labor market
  Models with skill differentials
  Modeling social protection
  Conclusions regarding theoretical models

Empirical Evidence
  What is asked versus what is needed for policy purposes
  Causal versus correlative analysis
  From cross-section regressions to market-level empirical analysis: benefits
  From cross-section regressions to market-level empirical analysis: costs
  Conclusions regarding empirical evidence

Conclusion: Putting the Three Components Together
“The progressive decline of cost-benefit analysis . . . may be explained by the formidable demands it places on data availability, on the exhaustive understanding required of economic mechanisms in a non-competitive model of the economy, and on the knowledge demanded regarding institutions and governance parameters. But as our discussions have revealed, these issues still loom large . . . – we just are not dealing with them in a comprehensive framework.

François Bourguignon, Chief Economist, The World Bank (Bourguignon, 2006)

“The purpose of cost-benefit analysis is to provide a consistent procedure for evaluating decisions in terms of their consequences. This might appear as an obvious and sensible way to proceed, but it is by no means the only one (examples of alternative procedures are majority voting, collective bargaining, the exercise of power, or the assertion of rights).”


“Around the world, many governments find it necessary, and worthwhile to subsidize artistic enterprise such as opera and theater, and some, including France and Morocco, provide subsidies for cinema production as well. . . What I find so striking about this example is the social cost-benefit analysis. . . There is little chance that French-language films, subsidized or not, will make a major dent in Hollywood profits. Whether subsidizing them is a good way for the French government to spend its money should be a matter for the French people to decide. If they spend it well, not only those in France but filmgoers around the world will benefit.”


Introduction

Sound labor market policy requires sound labor market models. Sound models have three characteristics. First, from a welfare economic point of view, the policy judgments are explicit, mutually consistent, and thoroughly worked out. Second, from a theoretical point of view, the models are sufficiently detailed and suitably rigorous. And third, from an empirical point of view, the models guide and are guided by solid quantitative evidence.

This paper reviews models of labor markets in developing countries from both a positive and a normative point of view. The survey is selective in that it exposit only some of the more significant contributions to the literature and highlights the most important gaps and analytical needs.
Why study labor markets and labor market policies? Throughout the world, there are fundamentally two, and only two, ways that people can escape from poverty. One is by earning their way out of poverty. The other is by receiving goods and services that lift them out of poverty. Even with multilateral and bilateral assistance, developing country governments are too poor to be able to make a significant dent in poverty by the social services route alone, and families are too poor to make large enough transfers to the less-advantaged among them. This means that creating more and better earning opportunities for the poor is the only other option available.

In labor economics, the “labor market” is defined as the place where labor services are bought and sold. Two types of employment are distinguished: (1) wage and salaried employment, in which workers sell their labor services to others, and (2) self-employment, in which workers sell their labor services to themselves. The poorer the country, the higher the proportion of workers in self-employment and the smaller the proportion in wage and salaried employment (Squire, 1981, Table 17). Provided that the idea of “labor market” is understood as comprising both wage and salaried workers and the self-employed alike, it is as relevant to analyze labor markets in the low-income countries of Africa and Asia as it is to analyze them in the middle-income countries of Latin America, the Middle East and North Africa, and Eastern and Central Europe or in the high-income countries of Europe and North America.

Just as “labor market” is defined broadly in labor economics, so too are “labor market policies.” The core model of a labor market involves labor demand, labor supply, and determination of wages and other conditions of employment. It is useful to think of those policies that have a bearing on the operation of a single labor market and those that involve the connection between labor markets. Accordingly, in any given labor market, there are those policies that have a bearing on the buying and selling of labor and that are usually called labor market policies (wage-setting mechanisms, industrial relations systems, and the like) as well as those policies that impinge upon the labor market
(macroeconomic policies and investment policies which affect the position of the labor demand curve, education and training policies which affect the position of the labor supply curve for different skill levels, fiscal policies which affect the size of the tax wedge, etc.). And then there are those policies that affect the movement of workers and firms between labor markets (hiring and firing regulations, labor market information systems, and so on).

Some units within the World Bank use the MILES framework:

<table>
<thead>
<tr>
<th>Policy issues</th>
<th>Conditions for growth</th>
<th>Macroeconomic stability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Macroeconomic conditions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Investment climate</strong></td>
<td>Regulatory environment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Government transparency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Taxes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Financing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infrastructure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Legal environment</td>
<td></td>
</tr>
<tr>
<td><strong>Labor market policies and institutions</strong></td>
<td>Labor market regulation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wage-setting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-wage costs</td>
<td></td>
</tr>
<tr>
<td><strong>Education and skills</strong></td>
<td>Basic education</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Higher education</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Training and lifelong learning</td>
<td></td>
</tr>
<tr>
<td><strong>Safety net for workers</strong></td>
<td>Income support</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Active labor market policies</td>
<td></td>
</tr>
</tbody>
</table>

All of these aspects fit into the purview of labor market policies as defined in this review.

The status of labor market analysis and labor market policy in the development economics community now is similar to the status of poverty analysis and anti-poverty policy in the development economics community two or three decades ago. At the time, the profession knew that it wanted to take on poverty more fully but most in it didn't know how. Knowledgeable insiders in development institutions joined with outsiders in universities and think tanks to set the direction for poverty analysis and anti-poverty policy and show those who were not poverty specialists how to do good poverty work. Today, anyone who is not already an expert in poverty analysis and anti-poverty policy
can acquire such expertise via a rich array of books, documents, articles, and training courses.

A parallel task is still to be performed for labor market work today. Imagine that you are a country economist formulating a labor market policy proposal which you propose to take to a client country’s government. A well-formulated proposal would answer the following questions.

- What specific labor market objective or objectives are you trying to achieve and by what welfare economic criterion or criteria will you decide if your objective(s) is/are being achieved?
- What is the labor market model you are using to analyze the effects of the proposed policy?
- What is the empirical evidence favoring one view of labor market functioning over another?

If I were your manager, it is only after these questions are satisfactorily answered that I would feel comfortable authorizing you to make policy proposals to a client country.

Typically, labor market policies have been proposed without such questions being answered. Let me give three examples, each of which is discussed at greater length later in this paper.

The first comes from Chapter 7 of the World Development Report 2005, which calls for creating more and better jobs by 1) fostering a skilled and healthy workforce that can contribute to a productive and prosperous society, 2) crafting labor market interventions to benefit all workers, and 3) helping workers cope with change in a more dynamic economy. These recommendations rely on an implicit model in which more and better employment is constrained by the poor quality of the labor supply. Others, myself included, work with a different model: that in a typical developing country, the principal constraint is the deficiency of labor demand. Both views have a long and distinguished history, being associated respectively with two economists who shared the Nobel Prize in
Economics, T.W. Schultz (1961) and Sir W. Arthur Lewis (Lewis, 1954). Of course, policy-makers do not have the luxury of being able to wait for more comprehensive modeling and policy analyses to be undertaken, and so development institutions are (properly) playing a role in advising on these issues based on what can be said from what is known at the time that policy decisions are made. Nonetheless, research can and should be undertaken to determine whether the principal constraint to employment creation in a given country lies on the supply side or the demand side of the labor market. With such research findings in hand, policy-makers and the development institutions that advise them will be better able to decide where to allocate their scarce development resources in the future.

Here is another example. Why do so many working people in low-income countries sell their labor to themselves and not to others, and what does the answer imply for policy? One answer to the analytic question is that a) the wage and salaried jobs are the better ones, but b) employers do not want to buy all of the labor services that are being supplied, and so c) workers who cannot afford to remain unemployed while holding out for wage and salaried jobs enter self-employment so that they can earn cash right away. Based on such a model, some policy analysts conclude that the appropriate policy response is to create more wage and salaried jobs. This may be the correct policy conclusion, but then again it may not be. Maybe more development could be achieved at the same cost by helping the self-employed earn more in the sectors where they already are. Neither policy option – create more wage and salaried jobs or help the self-employed earn more in self-employment- is obviously correct. A comparison of the development effects of each policy alternative is needed.

But maybe the analytical model presented in the previous paragraph is wrong. What if, as Maloney (2003, other) has argued in the case of Mexico and Webb (2005) in the case of Peru, a majority of the workers prefer the combination of earnings and job conditions in self-employment to those offered in wage employment and so are self-
employed for that reason? This is a very different model of the labor market, one that leads to very different policy judgments from the first view.

As a final example, consider education policy. The World Development Report 2006 addressed equity, which it defined as a normative concept including equality of opportunities and aversion to absolute consumption deprivation. It called too for policies to level the economic and political playing fields. One example of inequity cited in the report is that school enrollment rates in India vary by income quintile: the higher the income, the higher the enrollment rate. Consider now a policy of creating additional educational opportunities for children in low-income-quintile families. Assuredly, each newly-educated person will have improved opportunities for competing for whatever good jobs there are. But will the economy offer more good jobs because more people are educated, or will a given number of good jobs simply be reallocated to different individuals? Different labor market models make different assumptions about the answer, but the issue has been all but ignored empirically. The efficacy of education policy as an approach to development depends critically on which model in fact holds in a given setting.

In this introduction, I have tried to illustrate how sound labor market policies require explicit welfare economic analysis, well-specified theoretical models, and appropriate empirical evidence. In the balance of this review, I elaborate in turn on each of these three components.

**The Welfare Economics of Policy Analysis in the Labor Market**

Economics is often defined as the science of the allocation of scarce resources among alternative uses. In the current context, the scarce resources are the funds available for development purposes, for example, in a World Bank budget for a country. The alternative uses are many: labor market interventions, social sector programs, infrastructure activities, and the like.
Implicit in the preceding definition of economics, and the point that should be made explicit here, is that a rational allocation of scarce resources among alternative uses is possible only once an objective or a number of objectives has/have been specified. It is important that this be the ultimate objective or objectives of policy, not intermediate objectives.

**Specifying the ultimate objective or objectives**

Think of familiar maximization problems in economics. An individual maximizes utility by choosing how much to spend on each of a number of alternative products or by allocating one’s time between work and leisure. A firm maximizes profits by choosing how much output to produce, what production technology to use, and how much of each input to employ. A government or international agency seeks to minimize poverty by allocating an anti-poverty budget among the populace. In these and other decisions like them, what is being maximized or minimized - technically speaking, the maximand or minimand - is made explicit.

The great majority of studies purporting to offer “policy implications” from labor market research do not have a well-formulated maximand or minimand. Typically, we analyze inputs to or components of the maximand or minimand but not the maximand or minimand itself. Minimum wage policy is sometimes analyzed in terms of the disemployment caused, active labor market policy by the penetration rate of the national employment service, pension policy by the social security replacement rate, and unemployment insurance by the percentage of workers who are eligible for benefits in the event of job loss.

Let us consider some more comprehensive alternatives. The interested reader is referred to Fields (2001, Chapter 8) for further details.

**The Pareto criterion**

The Pareto criterion maintains that a policy is beneficial if it makes some people better off and nobody worse off. “Better off” and “worse off” are defined in utility terms.
A policy that makes some people better off and none worse off is said to be “Pareto-improving.”

Seeking Pareto-improvements and implementing them when they are found is indeed a sound criterion for social decision-making. However, implementing the Pareto criterion raises three practical problems.

First, the analyst must know the utility functions of the persons involved. Hardly ever is this the case.

Second, the analyst must accept the utility functions of the persons involved, even if they are nasty, malevolent, or weird. In practice, we may well find such judgments objectionable and might therefore want to “launder” them (Harsanyi, 1977, Goodin, 1986).

And third, we must have data on the actual or prospective changes in utility for each specific individual; rarely are prospective models or panel data on actual outcomes available.

Suppose though that these practical problems could be overcome. There remains the fact that it would be an unusual policy indeed in which nobody is made worse off. Typically, there are losers and rarely are the losers compensated fully for their losses.

Although the Pareto criterion is appealing conceptually, the practical issue is the general absence of Pareto-improving policy interventions. In the great majority of circumstances, other more applicable policy criteria must be sought.

**First order dominance and rank dominance**

Two of the problems with the Pareto criterion – its reliance on utility information and its need to compare the same individuals before and after a proposed policy intervention – can be overcome by switching to first order dominance as a policy criterion. The problem of not knowing utilities is overcome by working instead with incomes. The problem of not being able to compare the same individuals before and after
a policy intervention is overcome by looking instead at anonymous individuals – for example, those in the lowest income quintile, those in the second income quintile, etc.

Specifically, first order dominance is the following idea. One income distribution is said to first-order-dominate another if, at each income level, there are fewer people below the specified income amount in the first distribution compared with the second. First order dominance therefore overcomes the two aforementioned problems by looking at the incomes (not utilities) of anonymous (not identified) individuals.

Practically speaking, what needs to be done to implement first order dominance is quite simple. Draw two cumulative distribution functions, which virtually any statistical software can do. If one cdf is below the other everyplace and strictly below it someplace, then the first distribution first-order-dominates the second.

First order dominance is equivalent to rank dominance. Rank dominance looks at the income received at each rank in the income distribution. If the income is everywhere at least as high in one distribution compared with another and strictly higher someplace, then the first distribution rank-dominates the second. Practically speaking, to implement rank dominance, draw exactly the same graphs as with first order dominance but look at them differently. For rank dominance, draw the two cdfs, and if one cdf is found to lie to the right of the other one everyplace and strictly to the right someplace, then the first distribution first-order-dominates the second. Of course, if one curve lies to the right of another, then it necessarily lies below the other one also, which is why rank dominance and first order dominance are equivalent.

Implementing first order dominance requires that the analyst compare the entire income distribution vectors before and after a policy is put into effect. Careful modeling is required to derive the counterfactual vector of incomes under the proposed policy. This is demanding informationally, but no more demanding than the other policy evaluation methods that follow.
Finally, it bears mention that when the pre- and post-policy cdfs are drawn, they may cross. In such a case, the post-policy distribution neither dominates nor is dominated by the pre-policy distribution. The analyst must then rely on some other criterion for evaluating the policy.

**Abbreviated labor market well-being functions**

Lambert (1989) coined the term “abbreviated social welfare function” to indicate that social welfare (or, equivalently, “economic well-being”) is a function of statistics which are themselves calculated from the before and after income distribution vectors. Examples of such statistics are GDP per capita, inequality, poverty, total wage bill, and the unemployment rate. Each of these components must itself be carefully defined and measured.

In addition to specifying the arguments of the abbreviated social welfare function, the analyst must also specify the direction of the effect: for example, that an increase in GDP per capita is judged favorably and an increase in inequality negatively. Social welfare is judged to have increased for all functions belonging to a broad class of abbreviated social welfare functions if and only if all of the components that have been deemed to be good are found to have increased or remained unchanged and all of those deemed bad are found to have decreased or remained unchanged. If even one component goes in the “wrong” direction from the others, that broad class of abbreviated social welfare functions produces an ambiguous evaluation. Note that the evaluations we reach are contingent: all evaluators who adopt the particular class of social welfare functions in question would reach the same judgment, but evaluators who adopt different social welfare functions might reach different judgments (Atkinson and Stiglitz, 1980; Atkinson, 1983).

An example of such an abbreviated social welfare function is one that I used in Fields (2005b) to evaluate alternative labor market policies in one of the models reviewed below, the Harris-Todaro labor market model. The welfare function chosen was
(1) \( SW = f(\text{Total labor earnings, unem, ineq, pov}) \), \( f_1 > 0, f_2 < 0, f_3 < 0, f_4 < 0 \) – that is, social welfare is a function of total labor earnings, unemployment, inequality, and poverty, with the judgments that other things equal an increase in total labor earnings increases social welfare and increases in unemployment, inequality, or poverty lower social welfare. Three policies were evaluated: modern sector enlargement, which increases the number of modern sector jobs holding wages in the two sectors constant; modern sector enrichment, which increases wages in the modern sector holding the number employed in the modern sector and wages in the traditional sector constant; and traditional sector enrichment, which increases wages in the traditional sector holding wages and employment in the modern sector constant. For the class of social welfare functions given by (1), a policy of traditional sector enrichment was shown to be unambiguously welfare-improving, whereas policies of modern sector enrichment and modern sector enlargement were shown to be welfare-ambiguous.

In similar fashion, in analyzing labor market policies, we can move to what might be called an “abbreviated labor market well-being function.” What should be included in such a function?

Following traditional practice, an abbreviated labor market well-being function would presumably include the unemployment rate. Even though the unemployment rate should probably receive some weight, how much weight it should receive is open to discussion (Peek, 2006).

The abbreviated labor market well-being function would also recognize that in addition to how many workers have jobs, what also matters is how much workers earn in those jobs (Fields, 1980b; Squire, 1981; Turnham, 1971, 1993; Somavía, 1999). Thus, a second argument might be a measure of labor market earnings such as the average earnings of those employed, the incidence of low earnings, or the extent of earnings deprivation.
A third component might be social protection. For example, Bourguignon and Dethier (2006) have written, “The term ‘good job’ has two connotations: higher wages and more protection against dismissals, earnings variation and several risks linked to health, old age and unemployment.”

Taking these three components into account, we arrive at an abbreviated labor well-being function of the form:

\[
(2) \quad \text{Labor market well-being} = g(\text{level of employment, earnings of those employed, extent of social protection}), g_1, g_2, g_3 > 0.
\]

In the same way that the abbreviated social welfare function given by (1) requires careful specification of the total earnings, unemployment, inequality, and poverty arguments, so too does the abbreviated labor market well-being function given by (2) require careful specification of the employment, earnings, and social protection arguments.

**Minimizing a low-earnings index**

An attractive way of combining the level of employment with the earnings of those employed is to draw a parallel with poverty measurement. A poverty index widely used at the World Bank and elsewhere is the index developed by Foster, Greer, and Thorbecke (1984):

\[
(3) \quad P_\alpha = \sum_{i=1}^{q} \left( \frac{z - y_i}{z} \right)^\alpha,
\]

where \( q \) is the number poor, \( z \) is the poverty line, and \( y_i \) is the income of recipient \( i \). To allow for diminishing marginal utility of income, the parameter \( \alpha \) should be chosen to be greater than one. The value \( \alpha=2 \) is conceptually appealing and commonly utilized. The \( P_2 \) index is sometimes called the “severity of poverty” (Ravallion, 1994). In similar fashion, an index of the severity of low-earnings might be defined as

\[
(4) \quad P_2 = \sum_{i=1}^{q} \left( \frac{z - y_i}{z} \right)^2.
\]
where now $q$ is the number of low earners, $z$ is the low-earnings line, and $y_i$ is the labor market earnings of worker $i$. Observe that the index of severity of low-earnings given by (4) treats the unemployed as low-earners: the earnings deficit of an employed worker with earnings 10% below the low-earnings line is counted with a weight of 10%, the deficit of one with earnings 50% below the low-earnings line is counted with a weight of 50%, and the deficit of a potential earner who is unemployed and who therefore has zero earnings is counted with a weight of 100%.

The approach used in (4) is almost utilitarian but not quite. As described in Foster and Sen (1997), simple utilitarianism can be factored into three components: (i) consequentialism (the well-being of an action or institution is judged only by the goodness of the consequent state of affairs); (ii) welfarism (judging the goodness of states of affairs only by utility information); and (iii) sum-ranking (judging utility information for a given population simply by summing utilities). Let $u_i(.)$ denote the actual utility function of individual $i$ as a function of his/her labor market earnings and let $v_i(.)$ denote how the evaluator evaluates $i$'s labor market earnings. Denote by (ii') what we get if we use $v_i(.)$ in place of $u_i(.)$ in (ii). The combination of (i), (ii'), and (iii) might be called “quasi-utilitarianism.” Quasi-utilitarianism is similar in spirit to the approach taken by Atkinson (1970) in his famous paper on the meaning and measurement of inequality. As he later wrote (Atkinson, 1983, p. 5), “It should be stressed that the [welfare] function $W = \int_0^\infty U(y)f(y)dy$ is defined in terms of individual incomes, and not individual utilities. . . In retrospect, it would have been wiser to use a letter other than $U$.” [Emphasis in the original.] As Foster and Sen (1997, p. 116) later put it, “Taking $U_i$ to be individual utility . . . would be one possible interpretation, but in general $U_i$ would stand for the individual component of social welfare that is associated with person $i$.” [Emphasis in the original.]
In my view, the case for choosing labor market policies according to the criterion of maximizing \( W = \int_0^\gamma U(y) f(y) dy \) by minimizing \( P_2 = \sum_{i=1}^q \left( \frac{z - y_i}{z} \right)^2 \) is a strong one, deserving serious consideration in policy discussions.

**Social cost-benefit analysis**

In order to assess the returns from public expenditures, economists and other social scientists have long relied on social cost-benefit analysis. In a comprehensive survey article, Prest and Turvey (1965) described social cost-benefit analysis thus: “Cost-benefit analysis is a practical way of assessing the desirability of projects, where it is important to take a long view (in the sense of looking at repercussions in the further, as well as the nearer, future) and a wide view (in the sense of allowing for side-effects of many kinds on many persons, industries, regions, etc.), i.e., it implies the enumeration and evaluation of all the relevant costs and benefits.” See also Atkinson and Stiglitz (1980), Gramlich (1981), and Rosen (1992), among others, for further elaborations of the social cost-benefit analysis approach.

The cost-benefit criterion may be stated as follows: A project is socially profitable if the marginal social benefits (broadly defined and appropriately discounted) exceed the marginal social costs. The larger the difference between discounted benefits and costs, the more worthwhile the project.

Four aspects of the cost-benefit criterion for social decision-making are worth highlighting. The first is the use of social benefits and costs rather than private benefits and costs. Often, the distinction between private and social rates of return seems to be little more than a difference of a few percentage points after adjusting for mortality, taxes, and public subsidies. Second, when analysts have done cost-benefit analysis, they typically have calculated average rates of return while generally neglecting to point out that they are in fact averages and not marginals. Taken together, the calculation of average private benefits rather than marginal social benefits may be seriously flawed: in
some circumstances, the average private returns may exceed the marginal social returns by an order of magnitude or more (Fields, 1995).

Third, while early social cost-benefit analysis was formulated in terms of total benefits and total costs, we social scientists have become much more sensitive in the last twenty or thirty years to distributional aspects. We now ask: Who (in terms of position in the income distribution) receives the benefits? Who (in terms of position in the income distribution) pays the costs? This concern with distributional issues reflects the reformulation of the very notion of economic development as having an important anti-poverty component rather than being just economic growth itself.

Finally, the role of non-quantifiable benefits and costs should be addressed. There are some who maintain that if something cannot be quantified, it does not exist. I could not disagree more: if that were the case, there would be no such thing as love, pain, or friendship, and humanity would be much the worse for it. Revealed preference is often helpful in setting a bound on the non-quantifiable benefits or costs. When I buy my wife a dozen roses, I can be understood to have judged the benefits to our relationship to be worth at least the price paid. The same can be said for accommodations made to provide the disabled with equal access in the workplace, to provide special education programs for students with special needs, or to provide pension benefits to those who have no legal entitlement to them - provided, that is, that the benefits and costs have been carefully considered and weighed.

Cost-benefit analysis may well lead to different decisions from what other criteria might. Here are two examples.

In the education area, decisions are sometimes made on the basis of human resource planning (what used to be called “manpower planning”). It is right to ask, as the human resource planner does, “What is the absorptive capacity of the economy? Will the newly-educated person be employed in an occupation for which he or she is trained?” However, it is also right to ask “What is the nature of the work the graduate will perform
and what benefits will this work confer on society?” and also to ask “How much will it cost society to educate another graduate? Do the benefits outweigh the costs?” The human resource planner does not ask these latter questions; his/her goal is to equate the supply of persons with different skills and educational attainments to the demand for them. This is the wrong criterion, one that is likely to lead to an overproduction of education. Why overproduction? Because the demanders consider only the private costs to them of hiring workers with certain qualifications but not the costs to society of producing these qualifications.

As another example of how the social cost-benefit approach differs from other approaches, consider the merit goods argument originated by Musgrave (1959): that a project (be it education, health care, infrastructure, or whatever) is inherently meritorious and therefore deserving of additional public expenditures. The problem with the merit goods argument is that it fails to recognize opportunity costs: given limited budgets, spending more money in any one of these areas means spending less in another. The social cost-benefit analyst would be justified in concluding that we have too much public spending on X, because we are spending too little on Y, given the costs and benefits of X compared with Y.

As the quotations at the beginning of this paper indicate, Bourguignon, Stern, and Stiglitz all believe in the importance of social cost-benefit analysis. I quite agree: in my view, it is much better to ask the right questions and answer them imprecisely than to answer the wrong questions precisely. Social cost-benefit analysis asks the right questions. It is a shame that is used less now than it used to be.

**Conclusions regarding welfare economics**

The following are the main points reached in this section:

In labor market policy work, what is being maximized or minimized needs to be made explicit, but seldom is.
The Pareto criterion is conceptually appealing but of limited practical applicability.

Some of the practical difficulties with the Pareto criterion may be overcome by using first order dominance or an abbreviated labor market well-being function, although these criteria may produce ambiguous results and therefore be indecisive. For those analysts who wish to focus on the level of employment and the earnings of those employed, it may be appealing to minimize a low-earnings index.

Finally, social cost-benefit analysis merits more attention than it currently gets.

**Theoretical Models**

**The overall framework**

Economics is usually divided into two branches, macroeconomics and microeconomics. For present purposes, though, neither of these is the relevant level of analysis. Rather, the focus here will be the market level, specifically the labor market.

As noted above, the labor market is defined as the place where labor services are bought and sold. All of the models reviewed here maintain that the aggregate labor market consists of a small number of labor market sectors (or segments) linked to one another by actual or potential mobility of workers and of firms. Conditions in one labor market affect and/or are affected by conditions in other labor markets. Thus, all of these models maintain that labor market outcomes are best understood by analyzing conditions in markets, not conditions in individual firms.

“Dualistic labor market models” maintain that there are just two labor market sectors, which I shall term “formal” and “informal.” “Segmented labor market models” add other sectors, for example, distinguishing between “traditional agriculture” and “urban informal sector.”

For any given labor market sector, market conditions are determined by market labor demand, market labor supply, and a process determining wages and conditions of
employment. Market labor demand and market labor supply underlie all of the models reviewed in this paper. However, the models differ depending on whether earnings levels and conditions of employment (hereafter “wages”) within a sector are determined by supply and demand, by institutional forces, by efficiency wage considerations, or by something else.

Different intersectoral linkages can be specified as alternative stylizations of different countries’ conditions. Among those considered are models featuring non-competing groups, full market-clearing, crowding, expected wage equalization, and restrictions on entry due to capital market constraints.

Many of the models reviewed in this section are elaborated at greater length in Fields (2005a), to which the reader is referred for additional details.

**The need for multisector models**

Following a long and distinguished tradition in development economics, I view developing countries’ economies as being divided into a number of distinct sectors or segments. Leading the way were the pathbreaking works of two Nobel Prize-winning economists, W. Arthur Lewis (Lewis, 1954) and Simon Kuznets (Kuznets, 1955).

Though Lewis and Kuznets differed in many respects, they agreed on one fundamental one: that the feature distinguishing one economic sector from another is the fact that workers earn different wages depending on the sector of the economy in which they are able to find work. Lewis wrote (p. 150): “Earnings in the subsistence sector set a floor to wages in the capitalist sector, but in practice wages have to be higher than this, and there is usually a gap of 30 per cent or more between capitalist wages and subsistence earnings.” Part of the gap was “illusory,” Lewis wrote, because of the higher cost of living in the capitalist sector. Yet, there remained a real wage gap due to a) the “psychological cost of transferring from the easy going way of life of the subsistence sector to the more regimented and urbanized environment of the capitalist sector,” b) the payoff to experience in the capitalist sector, and c) “workers in the capitalist sector
acquiring tastes and a social prestige which have conventionally to be recognized by higher real wages.” He then went on to analyze the dynamics of economic growth as profits earned in the high-income sector were reinvested, leading to capital formation, an increased demand for labor, and continued intersectoral shifts.

At the same time, Kuznets (1955) explored how various measures of income inequality would change as the high-income sector comes to employ an increasing share of the population. All of the inequality measures used by Kuznets exhibited an inverted-U pattern, which later came to be known as the “Kuznets Curve.”

Research on labor market dualism has been developed considerably in the intervening years. Some of the subsequent writings on labor market dualism were grounded in the then-emergent theory of human capital, which also earned its developers the Nobel Prize (Schultz, 1961; Becker, 1964). Dual labor market theory received a boost through the work of Doeringer and Piore (1970). As summarized by Wachter (1974, p. 639), the dual labor market model advanced four hypotheses:

First, it is useful to dichotomize the economy into a primary and a secondary sector. Second, the wage and employment mechanisms in the secondary sector are distinct from those in the primary sector. Third, economic mobility between these two sectors is sharply limited, and hence workers in the secondary sector are essentially trapped there. Finally, the secondary sector is marked by pervasive underemployment because workers who could be trained for skilled jobs at no more than the usual cost are confined to unskilled jobs.

Wachter and also Cain (1976) stressed that for dualism to exist, different wages must be paid in different sectors to comparable workers. Empirical evidence has been compiled demonstrating that observationally equivalent workers earned different amounts in different parts of the economy (Fields, 1980a), but skeptics (e.g., Rosenzweig, 1988) remain unconvinced that dualism or segmentation in fact exists. Nonetheless, the good
jobs/bad jobs dichotomy remains widespread in the current literature (e.g., Inter-American Development Bank, 2003; Bourguignon, 2005).

The alternative to a segmented labor market model is a unitary labor market model in which all workers with given skills receive the same wage regardless of the sector in which they work – a view which I find to be at odds with developing countries’ reality. When possible, Occam’s Razor suggests limiting the analysis to two sectors. Models with two sectors – formal and informal – are elaborated below. But when two sectors are not enough, three-sector or n-sector models can prove insightful. This point is developed further in the sequel.

**Modeling the formal sector labor market**

In this subsection, I review three groups of models of labor markets for formal sector workers of a given skill level. They are the supply-demand model with market wage determination, the supply-demand model with institutional wage determination, and efficiency wage models.

The supply-demand model with full market-clearing (e.g., Ehrenberg and Smith, 2006) is depicted in Figure 1. The model displays three essential features. First, the amount of labor demanded in the market is taken as a decreasing function of the wage, other things equal. The market labor demand curve slopes downward because of diminishing marginal revenue product of labor and the associated substitution and scale effects of a wage change. Second, the amount of labor supplied to the market is taken as an increasing function of the wage, other things equal. The market labor supply curve slopes upward because a higher wage induces workers to enter this labor market from other labor markets and induces non-workers to enter the labor force. And third, the wage is set by supply and demand in order to clear the market.

According to the market-clearing model, three equilibrating forces operate: behavior of firms, behavior of workers, and behavior of wages. In the model, firms are free to hire workers or not depending on what is in their profit-maximizing interest to do.
If market conditions change, what is in their profit-maximizing interest to do will change accordingly, and firms are free to act on these changes. Similarly, workers are free to supply their labor in any given labor market or not depending on what is in their utility-maximizing interest to do. For them too, if market conditions change, what is in their utility-maximizing interest to do will change accordingly, and they (workers) are free to act on these changes. And finally, if supply and/or demand conditions change, real wages are free to rise or fall accordingly. (In this paper, wages should always be thought of in real terms.)

A second group of models retains the demand and supply curves but not the market-clearing feature of Figure 1. Rather, in these models, wages are set by forces different from supply and demand (Fields, 2005a). A number of institutional forces are identified. Minimum wages aim to assure covered workers an adequate standard of living. Trade unions are often encouraged by government policy as a means of entitling workers to a “just” share of the fruits of their labor. Public sector pay policies often result in substantially higher wages being paid to government workers than to their private sector counterparts. Multinational corporations frequently pay above-market wages. Labor codes in some countries regulate hiring and firing, impose payroll taxes on firms, and mandate that employers provide certain benefits to their workers. What unites all of these models is that the usual kind of labor market equilibration does not take place. One of the standard equilibrating forces is that if the wage is above the market-clearing level and consequently workers are unemployed, the unemployed would offer to work for lower wages rather than remain unemployed. However, the existence of the institutions just mentioned – minimum wages, trade unions, public sector pay policies, multinational corporations, and labor codes - prevents the wage from falling. The equilibrium is characterized by a higher-than-market-clearing wage with consequent unemployment.

A different type of institutional force may act to maintain wages above market-clearing levels for reasons having to do with the supply side of the labor market (e.g.,
Bardhan and Rudra, 1981; Solow, 1990). Suppose we have a casual labor market in which fresh hiring takes place each day and suppose further that the demand for labor is inelastic. The fact that labor demand is inelastic means that the total wage bill paid to labor is higher the higher is the daily wage. The fact that the labor market is casual means that each worker will receive his or her share of the higher total wage bill is s/he does not undercut the established wage. Knowing this, each worker has an incentive to leave the wage where it is. A wage above the market-clearing level therefore persists in equilibrium.

A third group of models has firms in the formal sector setting wages above the market-clearing level for efficiency wage reasons; see Katz (1986), Akerlof and Yellen (1986), and Weiss (1990) for reviews. The basic postulate of efficiency wage theory is that firms will pay higher-than-market-clearing wages if and only if the gains in productivity from doing so outweigh the costs, so that profits are increased. Two types of mechanisms for realizing productivity gains are specified. One set of explanations is that higher wages enable firms to hire better-quality workers from a heterogeneous labor pool – for example, workers who have more education or who perform better on tests of potential job performance. The other set of explanations is that higher wages induce workers of a given skill level to perform in a more productive manner through better nutrition, improved morale, lower absenteeism, reduced shirking, diminished labor turnover, and greater discretionary effort.

Let us turn now to models of the informal sector.

**Modeling the informal sector labor market**

As described above, the essential feature of labor market dualism is that one sector offers relatively attractive wages and other terms and conditions of employment while another sector offers relatively unattractive ones. In accordance with the most common way of viewing these sectors, the more attractive sector is called here the “formal sector” while the less attractive sector is called the “informal sector.” Thinking of
the labor market in this way leads working in the informal sector to be thought of as a type of employment of last resort. More recently, though, a different view has been proposed: that workers choose to work in the informal sector in preference to formal sector work. A third view is that the informal sector has its own internal dualism, combining the first two characterizations. Finally, the notion of informal work has broadened from thinking about workers in the informal sector to also include workers engaged in informal jobs outside the informal sector. These four views are taken up in turn.

Developing country labor markets exhibit low unemployment and substantial working poverty (ILO, 2006). The usual explanation offered for this is that workers in poor countries cannot afford to remain unemployed for very long, so unless they have a better option, in order earn cash quickly, the majority of them take up wage-employment or self-employment in the informal sector. I believe that this explanation, proposed long ago in ILO (1972) and Hart (1973), is essentially correct. The informal sector thus plays the role of a free-entry or fallback sector for those who cannot obtain formal sector jobs. As Reynolds (1969, p. 91) put it, “The urban trade-service sector . . . is a relatively open sector in the sense that, with little skill and little initial capital, a newcomer can crowd his way into employment. It is thus a natural entry point to the urban economy for migrants from the countryside.”

The earliest dualistic labor market models (e.g., Lewis, 1954; Harris and Todaro, 1970) had a single informal sector. Later models built in two informal sectors - an urban informal sector and a rural agricultural sector – to go along with an urban formal sector and unemployment (Fields, 1975). In these models, the best jobs were assumed to be located in the urban formal sector, in view of which job-seekers had an incentive to locate in the urban areas to improve their chances of being hired for the better jobs. Those choosing to work in either of the informal sectors were enabled to eke out a meager existence. This is not to say that they were happy about what they could earn when
informally employed, but working and earning in the informal sector was better for most of them than being openly unemployed. Once this was recognized, the view of the employment challenge in developing countries changed from a concern with employing the unemployed to the additional concern of raising the earnings of those employed (Turnham, 1971, 1993; Squire, 1981).

Workers entering the labor market were assumed to face a limited number of job search strategies. In Fields (1975), these were: 1) Be unemployed and search for a formal sector job full time. 2) Take up an urban informal sector job and search for a formal sector job part time. 3) Take up a rural agricultural sector job and search for a formal sector job part time. Because the probability of being hired for a formal sector job falls as we move from 1) to 2) to 3), the wages vary inversely – that is, the urban unemployed earn nothing, the urban informally employed earn something, and the rural agriculturally employed earn something more.

Informal sector wage determination can be modeled in a variety of ways. One is to assume that there is a fixed amount of income to be earned in the informal sector regardless of the number of people working in that sector - that is, the marginal product of labor is literally zero. A second approach is to regard a part of the informal sector as facing, instead of zero marginal product, constant marginal product. This assumption was adopted by many in what has come to be called the simplified Harris-Todaro model (Fields, 1975; Anand and Joshi, 1979; Heady, 1981; Stiglitz, 1982; Sah and Stiglitz, 1985; Bell, 1991). A third approach is intermediate between the first two: a positive but diminishing marginal product. Harberger (1971, pp. 574-5) put it thus:

[This] variant associates disguised unemployment not just with low wages but with situations in which the marginal productivity of labour lies below the actual wages earned. . . There are a variety of activities to which this argument applies. A classic example is that of fishermen on a lake. The addition of more fishermen increases the total catch, but not
proportionately, yet the last fisherman has an equal chance of making a
given catch as the first. The expected catch is the same for all, and is equal
to their average productivity. But, owing to the fact that the total catch
does not increase in proportion to the number of fishermen, the marginal
productivity of a fisherman is less than what he earns.

A fourth approach is to model a full demand system for agricultural and non-agricultural
products and workers (Bourguignon, 1990).

More recently, one more need has become apparent. This is to give due
recognition to the fundamental duality to be found within the informal sector. On the one
hand, the informal sector has free-entry activities of the sort just described. On the other
hand, it also has restricted entry activities that people who could be working formally
choose to work in instead. In Fields (1990), I highlighted this distinction and labeled
these two sectors the “free entry” part of the informal sector and the “upper tier” of the
informal sector. Recently, in a series of papers, William Maloney has maintained that, in
Mexico at least, self-employment in the informal sector provides a package of wages,
non-wage benefits, and working conditions that is at least as attractive for many people as
what they could receive as wage employees in the formal sector (e.g., Maloney, 2003,
2004; see also Webb, 2005). Still, though, there is “no consensus” (Maloney’s term) on
how many enterprises and individuals are to be found in each tier.

Finally, it bears mention that a broader notion of “informal economy” is on the
ascendancy (ILO, 2002; Jhabvala et al, 2003; Chen et al, 2005). The ILO now defines the
informal economy as comprising: informal employment (without secure contracts,
worker benefits, or social protection) of two kinds: a) self-employment in informal
enterprises (small unregistered or unincorporated enterprises) including employers, own-
account operators, and unpaid family workers in informal enterprises and b) paid
employment in informal jobs (for informal enterprises, formal enterprises, households, or
no fixed employer) including: casual or day laborers, industrial outworkers, unregistered
or undeclared workers, contract workers, and unprotected temporary and part-time workers.

**Modeling intersectoral linkages in the labor market**

In this section, I review five ways in which the various sectors are linked to one another. They are the non-competing groups model, the integrated labor market model, the crowding model, the Harris-Todaro model, and the Banerjee-Newman model.

The non-competing groups model (Cairnes, 1874) is one I mention because it is there, not because I believe it is a particularly useful stylization. This model maintains that the various labor market segments operate entirely separately from one another. Individuals belong either to labor market segment A or to labor market segment B, and they cannot switch from one to the other regardless of differences in pay, prestige, or anything else. It is conceivable that such a model might apply to labor markets separated by caste in India, tribe in Africa, or gender in more traditional societies. On the other hand, it is not very likely to apply even in these contexts, because even if workers cannot move to the more attractive labor market segments, employers can switch to the workers who are more efficient relative to their cost. The outsourcing of jobs from developed to developing economies that is taking place at an accelerated rate today can be understood in precisely this way.

A second multisector model is the integrated labor market model. This model starts with two or more sectors but assumes that all of the equilibrating forces that apply to a single labor market with market-clearing also apply to a labor market with a multiplicity of sectors. In particular, workers are free to move from one sector to another, firms are free to move from one sector to another, and the wage in each sector is free to rise or fall. Given such equilibrating forces, any wage differential between one sector and another would quickly be eroded. Moreover, in the integrated labor market model, a policy leading to economic growth in one sector results in higher wages in all sectors. Such a model provides a good stylization of Taiwan and other East Asian economies.
A third model, the crowding model, makes a different set of assumptions. The crowding model starts with the assumption of a real wage differential between sectors. Also by assumption, any worker who is not employed in the high wage part of the economy takes up employment in the low wage part of the economy. The larger the number of people who “crowd” into the low wage sectors, the lower the wage will be there. A development economist will recognize that the model of Lewis (1954) is a crowding model. A labor economist is more likely to associate crowding with the labor market discrimination model formulated by Bergmann (1971). In the crowding model, as in the integrated labor market model, a policy that brings about economic growth in one sector of the economy results in higher wages in the rest of the economy – in this case, because economic growth and the consequent increase in employment in the high-wage sector of the economy leads to less crowding and hence higher wages in the low-wage sector of the economy.

A fourth way in which the various sectors may be linked is the model put forth by Harris and Todaro (1970). This model has both intersectoral wage differentials (as the crowding model does) and unemployment (which the crowding model does not). The Harris-Todaro model also has a spatial dimension: the high-wage jobs are located in urban areas while the low-wage jobs are to be found in rural areas. The reason that unemployment arises is that workers are attracted to the urban areas by the relatively high wages there. The lucky ones become employed, while the unlucky ones are unemployed. A Harris-Todaro equilibrium arises when the urban search strategy (obtain a high wage with probability p, become unemployed and earn nothing with probability 1-p) has the same expected value as the rural search strategy (obtain a low wage with probability one).

The Harris-Todaro model helps explain a puzzle that arose in East Africa: why a policy of urban job creation did not reduce urban unemployment and seemed to exacerbate it. It also led to a counter-intuitive policy conclusion: that the solution to
urban unemployment is rural development. In response to the model, the government of Kenya put into place an integrated rural development program, which indeed had the effect of reducing urban unemployment.

All of the models reviewed thus far have just two employment sectors. Subsequent modeling efforts added in more sectors. Fields (1975) extended the Harris-Todaro model to have three employment sectors – formal urban unemployment, informal urban employment, and agricultural employment – plus unemployment. Moene (1988, 1992) built duality into the rural sector while maintaining a single urban sector. Others extended the Harris-Todaro model to allow for on-the-job search, mobile capital, endogenous urban wage setting, risk-aversion, a system of demand for goods, and many other factors; see Fields (2005a) and Khan (forthcoming) for citations.

A final multisector model is the occupational choice model of Banerjee and Newman (1993). Earlier individual-level models of occupational choice (e.g., Boskin, 1974) posited that an individual would choose that occupation that offered the highest present discounted value of lifetime earnings. These micro-level models, however, did not ask where the present discounted values in the different occupations came from. One answer was provided by Banerjee and Newman, who modeled a labor market consisting of four possible occupations: subsistence labor, wage labor, self-employment, and entrepreneurship. They assumed a large number of workers who were identical in their preferences and abilities but who differed in terms of their initial wealth. Because of imperfect capital markets, occupations that required high levels of investment were out of the reach of poor people. At any given time, individuals could invest their time and money to increase their wealth, which they then bequeathed to their children, thereby improving the children’s occupational potential. Over time, the path of economic development and occupational attainment would evolve accordingly.

Note that the Banerjee-Newman model is a market-level model of occupational choice, not merely a micro model. The important feature differentiating this model from
micro models such as Boskin’s is that the time paths of economic development and the
distribution of individuals across occupations are endogenous - that is, the returns to
choosing a given occupation depend on how many others in the economy are choosing
that and other occupations. In this respect, the Banerjee-Newman model is like the
Harris-Todaro model, in which the returns to choosing a given job-search strategy
depends on the number of others choosing that and other job-search strategies.

The models just reviewed are distinguished by the existence of multiple
employment sectors. Another important group of models focuses on skill differentials
among workers. It is to these that we now turn.

Models with skill differentials

One of the earliest models of skill differentials was formulated by Roy (1951).
The Roy model had two types of skills: hunting and fishing. Some workers were
relatively better at one or the other. Workers were assumed to sort themselves between
hunting and fishing according to where the payoff to their particular skills was highest.
The Roy model had limited influence among economists for quite some time, although its
influence has been growing. More important was a later line of work: human capital
theory.

As developed in the Nobel Prize-winning work of Schultz (1961, 1962), Becker
(1962, 1964), and Mincer (1962, 1974), human capital theory posits that individuals
differ from one another in terms of the amount of skill they possess. Unlike the Roy
model, skill is essentially unidimensional in the human capital model and can be
augmented in two ways: through schooling and through training.

The human capital model couples the preceding assumption that education and
training create market-useful skills with another assumption: that the returns to skill are
determined by supply and demand in what are essentially non-competing occupations. A
policy of educational expansion increases the supply of relatively educated labor and
decreases the supply of relatively uneducated labor. Given the human capital model’s
assumption that wages are set by supply and demand for each skill category, two consequences follow. The first is that all workers are employed in occupations befitting their skills. The second is that educational expansion lowers the market wage of educated labor and raises the market wage of uneducated labor.

As alternatives to the human capital model, other models of education have been formulated. **Signaling models** (Spence, 1973) maintain that workers get educated in order to signal to employers that they (the educated workers) are inherently more productive than other workers. **Screening models** and their close cousin **bumping models** examine what happens when the educational system certifies which workers are more productive than others and analyze linked labor markets in which educated workers use their education to move to the front of the queue and be hired preferentially for jobs for which education is advantageous but not required (Fields, 1972, 1974; Stiglitz, 1975). What all of these models have in common is that the social returns to education might differ substantially from the private returns, which in turn has important implications for countries’ decisions about the quantity of resources to invest in education (Fields, 1972; Stiglitz, 2002). This issue is so important that I return to it later in this paper.

**Modeling social protection**

Theoretical models of social protection abound. For overviews, the interested reader is referred to textbooks in public finance (e.g., Rosen, 1992; Stiglitz, 2000) and in labor economics (e.g., Cahuc and Zylberberg, 2004; Ehrenberg and Smith, 2006) as well as other sources (e.g., Ahmad et al, 1991; Bardhan and Udry, 1999; Holzmann and Jorgensen, 2001).

From my reading of the literature, three features of existing social protection models stand out. First, the great bulk of the literature models social protection at the level of the individual firm or the individual worker; models of social protection at the level of the market are much more limited. Second, when social protection is modeled at the level of the market, it is usually assumed that all firms face the same institutions;
seldom are covered and non-covered sectors modeled. And third, most of the literature models social protection in a deterministic environment; only infrequently is the environment a stochastic one.

To model social protection in a manner consistent with the multisector labor market models discussed above, two features are essential: the model must be market-level, and it must provide for covered and non-covered sectors. As for the deterministic/stochastic distinction, for some types of protection, such as providing old-age pensions or banning child labor, a deterministic approach may be satisfactory. For other types of protection though, such as employment security regulations or unemployment insurance benefits, only a stochastic approach will do.

From my admittedly less-than-encyclopedic reading of the literature, I cannot find a single instance of a model of social protection in the labor market with all three features (is formulated at the market-level, allows for covered and non-covered sectors, and incorporates a stochastic element). NOTE TO READERS OF THE REVIEW DRAFT: I WOULD BE VERY GRATEFUL IF ANY OF YOU COULD POINT ME TOWARD A MODEL THAT HAS THESE FEATURES. Absent such a literature, it would be best to formulate policies cautiously.

**Conclusions regarding theoretical models**

Good policy work requires sound theoretical foundations. Developing countries’ labor markets are marked by distinct labor market sectors that work in different ways from one another and by complicated interrelationships among the sectors. Unfortunately, few existing labor market models begin to capture the rich empirical reality of developing countries’ labor market conditions, and no existing model captures them all.

In order for improved labor market policy analysis to be undertaken, policy-makers and advisors need to work with models that contain enough sectors, provide realistic stylizations of the workings of the labor markets in each sector, and contain reasonable specifications of the linkages between sectors in the country in question. This
is a demanding set of requirements, requiring more time than is available for many policy
decisions. When time is of the essence, as it often is, I would say that it would be better to
base policies on a less rigorous model with approximately the right features than on an
explicit model with demonstrably wrong features.

Let us turn our attention now to empirical evidence.

**Empirical Evidence**

**What is asked versus what is needed for policy purposes**

There exists a huge statistical/econometric literature on labor markets in
developing countries. Highlights are surveyed in various articles in the three-volume of
the Handbook of Labor Economics (Ashenfelter and Card, 1999).

On the empirical side, five kinds of data analysis are potentially useful. Analysis
of aggregate cross-sectional quantitative data is a useful and often-neglected starting
point. Such data analysis permits answers to questions such as how do labor market
outcomes now compare with what they were earlier before growth took place (if in fact
growth took place), how do the outcomes in one country compare with those in another,
or how do labor market outcomes for one population sub-group compare with those of
another? Second, analysis of micro data from comparable cross sections enables
researchers to find out what makes a difference at the level of the individual worker or
the individual household. Such data analysis forms the basis for most contemporary
research studies on labor markets and will be highlighted below. Third, panel data
analysis is a new and promising approach in developing countries. Panel studies follow
the same individuals or families over time. Such studies enable development processes to
be analyzed in ways that analysis of comparable cross-sections cannot -- for example, in
the area of earnings mobility and labor market transitions. Fourth, time series data allow
for the determinants of changes over time to be modeled explicitly. For example, the
wage elasticity of formal sector labor demand in a country can be ascertained using such
data. Fifth and finally, also useful is qualitative data analysis, an excellent example of which is the World Bank’s Voices of the Poor study. In my judgment, qualitative data analysis can supplement but in no way should supplant quantitative data analysis. In view of what is now known, panel data studies could contribute the most new knowledge and therefore merit the highest priority in the coming years.

The bulk of empirical labor market studies use micro data from comparable cross sections. The question these studies typically answer is, which individuals or which firms have higher levels of the dependent variable Y? More specifically, the regression coefficients can be understood thus: when we compare those individuals or firms for which an independent variable X is one unit higher, on average how much higher is Y?

As an example, take the related literatures on earnings functions and returns to education. As reviewed by Willis (1986) and World Bank (1995), the earnings function studies show similar patterns of earnings differentials in various countries. Other things equal, earnings generally are higher for those workers with more schooling and more labor market experience, for men compared with women, for formal sector workers compared with informal sector workers, etc. Focusing on the schooling variable, George Psacharopoulos is best known for a long series of studies summarizing studies of returns to education. His most recent compilation of evidence (Psacharopoulos and Patrinos, 2002) reaffirms three earlier findings about returns to education as conventionally measured: they are higher in developed than in developing countries; they are highest for primary education, next highest for secondary education, and lowest for higher education; and the private rates of return exceed social rates of return.

These findings are not what we need for policy analysis. A fundamental postulate of public economics is that policy decisions should be made by comparing the marginal social benefits of a policy action with the marginal social costs. The social benefits include the gains to all members of society from the proposed action; the social costs
include the costs paid by all members of society. As with all economic decisions, resources should be allocated on the basis of marginal benefits and marginal costs.

For policy purposes, two types of knowledge are needed. First, how would market conditions change if a particular action were to be undertaken? Second, what are the marginal social benefits of the proposed action, what are the marginal social costs, and how do the marginal social benefits and marginal social costs compare?

For purposes of answering such questions, the usual regressions are helpful but insufficient in a number of respects. The next subsections explain why, continuing with the illustrative case of the social returns to education.

**Causal versus correlative analysis**

The first limitation of regressions across individuals and across firms is that while they are correlative, they are not necessarily causal. I plead guilty to having asserted, but not convincingly demonstrated, that the coefficients obtained from earnings functions can be interpreted causally and not merely correlatively (Fields, 1980a). More recently, leading economists including David Card, François Bourguignon, and James Heckman have built sophisticated econometric models of the causal effects of education; see respectively Card (1999), Bourguignon and Ferreira (2003), and Cunha, Heckman, and Navarro (2006). The question asked in each of these papers is, what would be the earnings gains from acquiring additional schooling? In this literature, the earnings gain is called the treatment effect. The results may be summarized by such measures as the average treatment effect, the average treatment effect for those treated, and the distribution of treatment effects for different individuals.

There have been three generations of results in this literature. In the first generation, earnings were related to education using ordinary least squares. In the second, the estimates were adjusted for omitted factors such as ability. In the third, the estimates were further adjusted for variations in returns to schooling, assuming that those
individuals who chose to get additional education were the ones that had the most to gain by doing so.

Even the third generation results, as sophisticated as they are, do not necessarily provide the needed information. What they answer is a micro question: if individual i were to get additional education, by how much would i's earnings increase? Suppose that i has the same chance of being hired for an available job as do others with the same level of education. In that case, i's benefit from education can be approximated by the average benefit from education for others with the same level of ability as i. This is i’s private benefit from education.

How closely does the private benefit from education approximate the social benefit from education? The answer depends on how the labor market responds to having additional educated persons. Let us now turn to this issue.

**From cross-section regressions to market-level empirical analysis: benefits**

A second limitation of standard empirical studies is that they **assume** that the social benefits of education can be approximated by the private benefits of education but they do not **test** whether this is the case. As I shall now demonstrate, the standard assumption may be problematical.

From the empirical findings throughout the world that have shown conclusively that individuals with more education earn more in the labor market, it is very frequently argued that society should invest in more education. As I shall now argue, while there is no question that additional education produces benefits for the individuals who receive the education, the case for additional education producing social benefits for all members of society including those who do not get the education is much less compelling. In this subsection, I explain these doubts by considering two alternative underlying models that yield very different conclusions about the social benefits of education.

Take first the familiar human capital model (Becker, 1964; Mincer, 1974). The model makes a number of assumptions, usually implicitly: that educated workers are
more productive than less educated workers; that the difference in productivity can be approximated by the difference in earnings; that better-educated workers earn more because of the additional education they have received; that there is full employment; and that if one more worker is educated, there will be one more educated worker employed in the educated workers’ labor market and one less less-educated worker employed in the less-educated workers’ labor market.

An alternative to the human capital model is the screening model (Fields, 1972, 1974; Stiglitz, 1975). In screening models, the assumptions are different from those in the human capital model: the wage is set according to the job, employers seek to fill each job with the most productive worker available, and workers with more education are on average more productive than workers with less education. In such circumstances, employers use education as a screening device, because they end up with workers who are more productive on average as a result.

A special version of the screening model is the job assignment model which Sattinger (1993) graphically termed “the dog-bone economy.” In his particular version of the screening model, jobs differ in terms of their quality according to a fixed system of rewards. Thus, the best job goes to the most highly-qualified applicant, the second-best job to the second-best applicant, and so on – by analogy, the biggest juiciest bone goes to the fiercest, most powerful dog, the second-best bone to the second-strongest dog, and so on. Just as the weakest dogs may end up with no bone, in a dog-bone economy, the least-qualified workers may end up with no job.

Screening models assume that educated workers are more productive but take no strong position as to why educated workers are more productive. This could be for any of a number of reasons or a combination of them: human capital is formed in schools; schools selectively admit the best students, who are likely to be the best workers; and workers get additional education to signal their superior ability. (This last is the foundation for the Nobel Prize-winning model of Spence, 1973.)
Consider now how the human capital model and the screening model differ in terms of the social return to additional education. The marginal social returns to education must be evaluated at the market level, not the individual level. (In both models, the individual who gets additional education will enjoy a private benefit.) Under the assumptions of the human capital model, if individual i gets more education, the effects are the following. Society will have one more employed better-educated person and one less employed less-educated person. Individual i will be more productive because of the additional education received. The amount of the productivity gain to society is the difference between the earnings of those with and without the education level in question. Given the preceding, the marginal social return to education is closely approximated by the average social return.

Contrast this with the effects of additional education under the screening model. In the screening model, the availability of one additional educated person results in one more job being filled by a better-educated person rather than a less-educated person, the reason being that employers judge that the better-educated individuals will be more productive than the less-educated ones and so hire the better-educated preferentially. However, there is no assumption now that the difference in productivity between better-educated and less-educated workers can be approximated by the difference in average earnings or that the difference in productivity is caused by human capital produced in schools rather than by pre-labor-market differences in ability. The assignment model goes one step further: in that version of the screening model, there is no productivity gain whatsoever: the now-stronger worker gets the job which somebody else does not now get. In a dog-bone economy, the distribution of bones does not change; all that changes is the distribution of dogs among bones.

Which model is right for China, Chile, or Chad? The answer cannot be found by running earnings functions across individuals. Instead, the answer must be found by market-level tests of the assumptions. The question then is, which set of assumptions –
those of the human capital model or those of the screening model – best approximate conditions in the country in question?

In my view, the conditions posited by the screening model appear at least as relevant for developing economies as those posited by the human capital model, if not more so. So called “crowding out” of less-educated workers by the better-educated has been demonstrated for Africa (Knight, Sabot, and Hovey, 1992; Bennell, 1996). Bennell’s evidence, discussed further by Pritchett (2001), shows that wage employment grew by nearly the same number of people as the change in school enrollments in just two countries (Botswana and Zimbabwe). In other countries studied, the number of newly-educated individuals entering the labor force was four times as large as the growth of wage employment in Senegal, Kenya, and Malawi, and between ten and twenty-nine times larger in Lesotho, Burkina Faso, Ghana, Uganda, and Sierra Leone. Consequently, I feel that the marginal social benefits of education may be greatly overstated by the standard methods, at least in these particular African countries.

This analysis can be carried beyond the field of education. The general issue is that there may well be a fallacy of composition. Individual i or firm j may be able to get ahead by doing more of something, but if that very action causes individual m or firm n to fall behind, the social benefits may be very much smaller than the private benefits. This is what Nobel Laureate George Akerlof (1976) called “The Economics of the Rat Race.” Nothing in individual-level analysis rules it out.

So far, we have considered the benefits of additional education. It is time to turn now to turn to costs.

**From cross-section regressions to market-level empirical analysis: costs**

The third limitation of standard empirical studies is that social costs are rarely given the same attention that social benefits are. The social costs of any project include the direct costs of the project plus the opportunity costs. In the case of educational projects, the direct social costs are the costs to society of providing the education, which
in turn equals the direct costs paid by the student plus the subsidies provided by the government or others (e.g., the university or its alumni). The social opportunity costs consist of the value of the output that is not produced because some individuals are removed from the labor force while they are in school. Properly speaking, the social rate of return to education is found by equating the marginal social benefits of education to the marginal social costs of education and solving for r:

\[ PV_{benefits} = PV_{costs} \]

or

\[ (5) \quad B_0 + \frac{B_1}{(1+r)} + \frac{B_2}{(1+r)^2} + \ldots = C_0 + \frac{C_1}{(1+r)} + \frac{C_2}{(1+r)^2} + \ldots, \]

where \( B_t \) and \( C_t \) are the social benefits and social costs in year \( t \). To avoid confusion, let us call the value of \( r \) that solves (5) the “internal rate of return.”

Following upon a long tradition originated by Mincer (1974) and continued by Psacharopoulos (1973, other), it is common to use the term “rate of return” to refer to the coefficient on schooling in a Mincer-type earnings function of the form

\[ (6) \ln Y = a + bS + c\text{EXP} + d\text{EXP}^2, \]

where \( S \) is the individual’s schooling and \( \text{EXP} \) is the post-schooling experience. Let us call the regression coefficient \( b \) in (6) the “Mincer rate of return.”

Social decisions about resource allocation should be based on the internal rate of return. As noted by Mincer himself, the Mincer rate of return equals the internal rate of return if and only if the only costs of education are opportunity costs. Direct costs are neglected in the Mincer rate of return calculation. Of course, there are always direct social costs of education, even if education is entirely “free” to the student. Given that these costs are not included in the Mincer rate of return, the Mincer rate of return is not a suitable basis for social decision-making. Unfortunately, analysts often think they are estimating an internal rate of return when in fact what they have gotten is a Mincer rate of return.
In sum, the social costs of investments in education and other public projects need to be factored into social decision-making. Often, though, they are not, which means that the results so derived are less than believable.

**Conclusions regarding empirical evidence**

Despite the huge number of statistical and econometric studies of labor markets in developing countries, the literature is far from conclusive on policy implications. This is because these studies do not provide the information that is needed for public policy: knowledge of how market conditions would change if a particular action were to be undertaken along with knowledge of how the marginal social benefits compare with the marginal social costs.

Specifically, three limitations have been highlighted in this section. First, regressions across individuals and firms are correlative but not necessarily causal. Second, standard empirical studies assume (but do not test) whether the social benefits can be approximated by the private benefits. And third, social costs often do not receive the same attention that social benefits do.

**Conclusion: Putting the Three Components Together**

In this paper, I have argued that three components of a work program on labor market policy in developing countries merit the highest priority. One is to specify explicitly the welfare economic criteria by which labor market policy judgments are being made. The second is for fuller, more comprehensive theoretical models of how developing countries’ labor markets actually function. And the third is for statistical and econometric studies that a) are guided by and b) guide market-level models.

Four pitfalls are common enough in the literature on labor markets in developing countries that I would like to caution readers specifically against them. The first is to think of labor markets in terms of productivity rather than in terms of supply and demand. A simple example illustrates the point. Consider a two person economy in which workers
dig holes. Suppose that together, two workers working with one shovel can dig 10 holes a day, i.e., 5 holes each. Suppose one worker is replaced by a power shovel, which costs the same as the worker displaced, and that the one worker who remains employed working with a power shovel is able to dig ten holes a day. Suppose the other worker is unemployed. Productivity has doubled. “Should” the wage of the employed worker double? No – supply and demand analysis tells us that if anything the employed worker’s wage would fall, not rise.

A second pitfall is to rely on the wrong kinds of empirical studies. What we need most for policy purposes are empirical studies that give guidance on what the right theoretical labor market model should be or what the empirical magnitudes included in existing models actually are. Standard household and firm surveys can be used to perform market-level as opposed to individual-level analysis. For example, if we want to know what the consequences of future educational expansion might be, we can derive useful policy insights by asking what were the market-level effects of past educational expansion. If, in the recent past, the supply of educated people has been increased by 100 university graduates, are 100 more people now working in “university-level” occupations? Sometimes, though, non-standard data are needed. In South Africa, the broad unemployment rate is 41%, whereas in most developing economies, a typical unemployment rate is in the range of 5% to 10%. Why is South Africa so different? An interesting field study in South Africa (Cichello et al., 2006) asked the unemployed why they have not entered self-employment. The principal result of this field study – that the number one factor impeding entry of South Africans into informal self-employment is the fear of crime if the enterprise is successful – is more informative than any regression would have been.

Third, many policy conclusions are offered with no supporting cost-benefit analysis. As an example, consider a country in which a poverty profile shows (as most if not all poverty profiles do) that those who work in agriculture are disproportionately
poor. An unwarranted policy conclusion would be to decide to invest the available development resources in the sectors where the poor are. Equally unwarranted would be the policy conclusion to invest the available resources in the sectors where the poor are not. The policy conclusion that is warranted is to invest development resources in whichever sector, agriculture or non-agriculture, produces the highest marginal social benefit compared with the marginal social cost – gauged, for example, in terms of poverty reduction.

A fourth policy pitfall is to limit one’s analysis to only a limited subset of the relevant goods or bads. Two examples are common in the literature on the economic effects of labor unions. One is to say that strong unions are good because those who remain employed will earn higher wages. The other is to say that strong unions are bad because the higher wages negotiated by unions are likely to cause higher unemployment.

As an example of how these three components might be put together and these four pitfalls avoided, let us look more carefully at the literature on the labor market effects of minimum wages. The familiar starting point is the basic supply-demand model in which a higher minimum wage in a sector reduces the quantity of labor demanded and induces in-migration of labor, thereby increasing unemployment in that sector. Also well-known is the monopsony model in which a higher minimum wage in a sector can lead to increased employment in that sector provided that the wage increase is not too large (Stigler, 1946). Moving from one sector to two, minimum wages in labor market models in economies with covered and non-covered sectors have been analyzed by Harris and Todaro (1970), Harberger (1971), Mincer (1976, 1984), Gramlich (1976), and Fields (1975, 1997). We now know from this literature that a higher minimum wage in the covered sector may but need not result in more unemployment. But unemployment should probably not be the only criterion for evaluating a minimum wage. When total labor earnings, inequality, and poverty are also included as components of the welfare judgment, Fields (2005b) showed that a higher minimum wage in a Harris-Todaro model
could improve social welfare or not depending on parameter values. Finally, taking poverty as the welfare criterion, Fields and Kanbur (2006) showed in a single-sector model with income-sharing that a higher minimum wage could lower poverty or that it could raise poverty depending on parameter values which might then be estimated empirically. Despite the progress that has been made, more work remains to be done in modeling minimum wages – in particular, enriching the model to include enough heterogeneity, most importantly, placing some minimum wage workers in poor families and others in non-poor ones.

I would conclude where I started: sound labor market policy requires sound labor market models. Let us be both bold enough to challenge ourselves and those whom we advise to state our policy evaluation criteria, specify our theoretical models, and present our empirical evidence and humble enough to know when the best policy conclusion to draw is no policy conclusion at all. The stakes are too high for carelessness.
References


Figure 1.
The Standard Market-Clearing Labor Market Model.