Is Monopsony the Right Way to Model Labor Markets?
A Review of Alan Manning’s *Monopsony in Motion*

PETER KUHN

ABSTRACT Manning proposes that the ‘traditional’ monopsony model, once regarded as an analytical curiosity, be adopted as a widely-applicable description of firms’ behavior in labor markets. In Manning’s view, search frictions in the labor market generate upward-sloping labor supply curves to individual firms even when firms are small relative to the labor market. Thus a model of ‘monopsonistic competition’ best characterizes labor markets as a whole. Manning’s book applies this new perspective to a wide range of ‘traditional’ topics in labor economics, ranging from labor supply, to gender discrimination, to the effects of trade unions on wages and employment, generating refreshing new insights in each case. Ultimately, however, this reader was left unconvinced that monopsony is the ‘right’ model of most firms’ labor market behavior in the long run.

Key Words: Monopsony; Search; Labor Markets; Wages; Employment.

JEL Classification: J00, J42, J63.

1. Introduction

Alan Manning has written an impressive new book on labor markets. The central claim of *Monopsony in Motion* is a strong one: that a particular theoretical model—‘monopsony’—is ‘the best simple model to describe the decision problem facing an individual employer’ (2003:3), at least when the issue is the determination of wages and employment. Further, to describe the labor market as a whole, Manning argues that ‘models of oligopsony or monopsonistic competition are what is needed’ (p.3). In fact, throughout much of this book, Manning argues that a particular version of such a model—that of Burdett and Mortensen (1998), where monopsonistic behavior at the firm level is generated by a job search process with posted wages—provides an empirically-relevant picture of labor markets.

Clearly, *Monopsony in Motion* is an ambitious piece of work. Aside from the strength of the claims it makes concerning the ‘right’ way to model labor markets,
this book covers virtually every ‘traditional’ topic in labor economics, including, for example, labor supply, human capital, gender discrimination, the evolution of wages over a worker’s career, and the effects of trade unions. In considering each of these subjects Manning argues two points: (a) that the profession’s understanding of that issue can be improved by adopting a monopsonistic perspective, and (b) that the positive and normative implications of a wide range of policy interventions (such as, for example, minimum wages or mandated benefits) can be very different when the perfectly competitive model is replaced by a monopsonistic one. I can think of only two other labor economics books appearing in the last three decades that aim even remotely as high as this book does. Devine and Kiefer (1991) make a similar argument for search theory; in practice, however, their book reads more like an overview of existing results than Manning’s more synthetic approach. Like Manning, Teulings and Hartog (1998) combine theory and evidence in thoughtful and original ways to advance a particular theoretical perspective on labor markets. While their goal—understanding international differences in wage formation and labor market performance—is different from Manning’s, these authors share his goal of replacing the competitive neoclassical paradigm with another whose assumptions correspond more closely to the real world.

Manning’s book is, therefore, both unique and impressive. Indeed, while little new theory is developed in this book, the use of theory and the discussion of its relation to the facts is uncommonly thoughtful. And while no new econometrics is developed in the book, its command of multiple U.S. and U.K. data sets ranging from the familiar (Current Population Survey, Panel Study of Income Dynamics, National Longitudinal Survey of Youth, Labour Force Survey, and British Household Panel Survey) to the arcane (the UK’s Employers’ Manpower and Skills Practices Surveys), and its careful consideration of identification problems and possible solutions is impressive. Indeed, it is an unfortunate fact that the style of analysis in this book, which goes back and forth between deriving empirically refutable implications of relatively simple theoretical models and confronting them with data, has become almost extinct in a subfield currently obsessed with atheoretical identification of ‘treatment effects’, and in a profession that has become hyper-specialized. Since I would be very pleased to see the profession produce more work of this kind, I am naturally reluctant to criticize this book. However, criticize I must, for that is the reviewer’s role, and I can only hope that my criticisms will encourage, rather than discourage, more research of this style and quality.

In what follows, I begin by providing some background on exactly what Manning means by ‘monopsony’, and how this compares with the general usage of this term in the profession. I then discuss some of the evidence Manning presents on the empirical relevance of monopsony in U.S. and U.K. labor markets. I close by discussing two features of the theoretical models that form the heart of this book which in my opinion severely constrain their empirical relevance relative to what is claimed in the book.

2. Semantics: What is Monopsony?

Given the focus of their field on the effects of market structure on outcomes, it may come as a surprise to industrial organization economists that the attention given by labor economists to market forms other than the perfectly competitive
one has been, to date, very limited. Some research has modeled unions as monopolistic sellers of labor (see Kuhn, 1998 for an overview); other than that the only variant on perfect competition that has made it into most labor economics texts is the ‘traditional’ or ‘textbook’ monopsony model.

The textbook monopsony model is just the mirror image of the more familiar textbook monopoly model, applied to factor markets rather than product markets. Thus, there is a single buyer in a labor market (the example offered most often is the case of nurses in a town with only one hospital). Because there is only one buyer, that buyer faces the entire market’s labor supply curve, which—in contrast to the horizontal one faced by a competitive firm—is likely to be upward-sloping. Diagrammatically, let the supply curve be given by the function \( w(L) \) in Figure 1.

With a production function \( F(L) \) and an output price \( p \), the monopsonistic firm’s profits are given by \( pF(L) - w(L)L \); first-order conditions for a maximum can be written \( pF'(L) = w(L) + Lw'(L) \). In other words, profits are maximized at an employment level where the value marginal product of labor, \( pF'(L) \), equals the marginal factor cost of labor, \( w(L) + Lw'(L) \). Since the latter exceeds the wage, the monopsonist hires \( L^m \) units of labor, which is less than the socially efficient amount \( L^* \). The monopsonist chooses to pay a wage of \( w^m \), which is below the efficient, market-clearing level, \( w^* \).

In this book, Manning argues that the situation depicted in Figure 1, rather than being an analytical curiosity, is a good representation of the situation facing the vast majority of firms in the U.S. or U.K. today. Thus, Manning does not intend to restrict the term ‘monopsony’ to its literal and traditional meaning of markets with a single buyer. Instead, Manning argues that even in labor markets with an arbitrarily large number of employers, search-related frictions will generate an upward-sloping labor supply curve to each employer like that shown in Figure 1. Thus, a descriptor focusing on ‘search’ or ‘frictions’ instead of monopsony would be a more accurate representation of the types of models Manning advocates in this book. Manning is, of course, correct in drawing a distinction between the models he is interested in and search models in general—which sometimes ignore wages completely or assume they are bargained over after a match is made—preferring to reserve the term ‘monopsony’ for search models where wages are posted in advance by employers. But search models of this type.

![Figure 1](image_url)
have been around for some time (e.g. Peters 1991, or even Burdett 1978) without the ‘monopsony’ label. The key point seems to be that the title Search Models with Ex-Ante Posted Wages in Motion, while considerably more accurate than Manning’s, is certainly less catchy.

Continuing on the subject of semantics, although Manning insists the term derives from early uses by Hicks and Pigou and is meant to be value-free, one can’t help thinking that he is having some fun by using the term ‘exploitation’ rather than something less dramatic like ‘mark-up’ to describe the difference between the wage and value of workers’ marginal product.

3. Evidence for Monopsony

Turning now to substantive critiques, what is the evidence that search frictions (or for that matter, any other factor) generate an upward-sloping labor supply curve for a typical firm in the U.S. or U.K? At the very start of his book—indeed on the inside front jacket—Manning makes a prima facie case for this which runs as follows: ‘What happens if an employer cuts the wage it pays its workers by one cent? Much of labor economics is built on the assumption that all existing workers immediately leave the firm….’ (p. 3). Since this assumption seems patently absurd, Manning argues that an upward-sloping supply curve of the form shown in Figure 1 simply must be a better representation of the situation facing a typical firm today.

While appealing, this literal reading of the neoclassical model is, in my opinion, beside the point. After all, even within the context of the monopsony model, Manning himself acknowledges the distinction between a short-run and a long-run labor supply elasticity (pp. 32–3) with the latter exceeding the former. Further, Manning then decides—explicitly, and in my view correctly—to focus on long-run elasticities throughout this book (p. 34). Thus, the observation that short-run responses of employment to wages are surely not infinite does not shed much light on the central question at hand: what we really care about is the long run, and whether firm-level labor supply curves are effectively horizontal in the long run remains an issue that needs to be settled by data rather than by introspection.

Focusing now on the magnitude of long-run firm-level labor supply elasticities, what does the evidence say? In my view, the most direct evidence Manning brings to bear on this question is his discussion of the literature on the employer-size wage effect in chapter 4. While it is well established that large firms pay higher wages, much uncertainty remains regarding the explanations for this relationship, including whether it actually reflects the upward-sloping firm-level labor supply curve required by the monopsony model. Manning’s discussion of the issues here is excellent, and I cannot help but agree with his ultimate assessment:

This is all rather depressing: a good estimate of the elasticity of the labor supply curve facing the firm seems very elusive so perhaps there is a good reason for the lack of research into this area. Progress seems to be dependent on finding a good firm-level instrument (p. 96).

Thus, when all is said and done, there is not much hard evidence of the kind of upward-sloping long run labor supply curve at the firm level that is required by a monopsony model.³ In the remainder of this review, I would like to argue
that this lack of evidence is not surprising, because of two theoretical features of
the monopsony model that, in my opinion, make the question—as posed in this
book—not sufficiently precisely defined for empirical testing. They are, in turn,
(a) that heterogeneity in workers’ abilities does not play enough of a role in exist-
ing models of search or monopsony, while (b) issues of scale (specifically, the
absolute number of workers in a firm) play too large a role.

4. Introducing Heterogeneous Worker Ability

In the simplest general-equilibrium search model presented by both Burdett-
Mortensen and Manning, all workers are assumed to be equally productive in all
firms. Rather than being a limitation, this is a brilliant assumption, because it
allows the authors to make a dramatic point: in posted-wage search models of
this kind, equilibrium wage dispersion must exist even when there is no underly-
ing heterogeneity on either the firm- or the worker side. Assumptions that help us
make important theoretical points, however, are not always appropriate for
bringing a model to the data; thus it is often important to consider various exten-
sions before assessing the empirical realism of a particular modeling approach.

To that end, as Burdett-Mortensen have shown, extending this basic model to
incorporate heterogeneity in workers’ value of leisure or in the productivity of firms
is both interesting and relatively straightforward. Unfortunately, incorporating
heterogeneity in worker productivity of a form that is portable across firms (i.e.
‘ability’) is not so simple, at least if done in a nontrivial way. Indeed, Burdett-
Mortensen do not attempt it at all, and the only theoretical analysis in Manning’s
entire book that does not assume identical worker ability is in Section 11, on work-
ers’ education decisions and firms’ training decisions. Unsurprisingly given the
challenges involved, this analysis is quite mechanical, with completely segmented
labor markets for educated versus other workers, and exogenous wage distribu-
tions in both.

The assumption of identical worker ability is important because it has
profound effects on how we must relate the predictions of models such as these to
the real world. To see this, consider the following alternative model of labor
markets. A worker of ability \( \alpha \) has a productivity of \( \alpha \) in every firm where he
could be employed. There is a continuous distribution of worker productivities,
\( f(\alpha) \), and perfect competition ensures that a worker of ability \( \alpha \) must be paid
exactly \( \alpha \) in any firm that employs him. Since we wish to consider models in
which firms post wages, note that in this model a firm that posted a single wage \( w \)
will attract only workers of ability exactly \( w \), and will just break even by doing so.
Although firm-level employment in such a model is arbitrary, assume for simplic-
ity and for the sake of argument that every firm has a fixed number of jobs, \( N \), and
fills them all at its posted wage.

In the world described above, what happens when a firm paying \( w \) cuts its
wage by an infinitesimal amount, to \( w - \epsilon \)? In one sense, the estimated labor
supply curve it faces will be not just upward sloping but vertical: if we simply
count the number of workers employed (as is the practice in the firm-size wage
literature), employment will remain unchanged at \( N \) while all the workers of abil-
ity \( \alpha \) are replaced by workers of ability \( \alpha - \epsilon \). In another sense, the firm’s labor
supply curve is upward-sloping: measuring the amount of labor in ‘efficiency units’
(i.e. the number of workers times their quality), total labor input falls when the
firm cuts its wage. Just as in Manning’s search model, a rising firm-level labor
supply curve ensures that firms in this model are indifferent among offered wages: in the long run, higher wages elicit just enough extra labor (in quality-adjusted units) to keep profits unchanged. In a third sense, the firm under consideration can be said to have an infinitely-elastic supply curve, in the sense that an infinitesimal wage cut causes all its existing workers to quit (and be replaced by lower-quality workers). This third sense seems to correspond most closely to Manning’s caricature of the perfectly competitive model, but note its fragility: on the one hand, it depends on measuring gross, not net flows of labor out of the firm (something that is rarely done in empirical work); on the other, with a slight modification to the model this worker ‘quality’ change need not involve any worker turnover at all: marginal changes in the quality or intensity of work among existing employees can accomplish the same purpose. Thus, depending on how one measures labor input (number of workers vs. quality-adjusted units), and depending on whether one measures gross versus net flows of labor to the firm, the same model of wage and employment determination can characterized as one with zero, positive, or infinite labor supply elasticity to an individual firm. Further, note that—in contrast to oligopoly—employment in this alternative model is always fully efficient. Thus, depending on the details of measurement, all three notions of firm-level labor demand elasticity described above are consistent with Pareto-optimality.

In sum, measuring the elasticity of the firm-level labor supply curve in the real world—where there surely exists a continuum of worker abilities—requires careful attention to definitions as well as some measure of the quality of a firm’s entire workforce. More to the point, until we can derive the predictions of search models in worlds where workers with a continuum of abilities compete with each other for jobs (and understand how these predictions differ from the purely neoclassical model sketched in the preceding paragraph), it is not clear that we even know how to assess the empirical relevance of these ‘new monopsony’ models of the labor market.

5. The Excessive Importance of Scale

The discussion of worker quality in the previous section highlights a surprising absence from the theoretical discussion in this otherwise remarkably comprehensive book: the analysis of any dimension of labor input other than the absolute number of workers employed at a firm. While there is a brief discussion of hours per worker on pp. 227–34, the possibility that average worker quality or effort might respond to the offered wage is, as far as I can tell, never analyzed at all. This draws attention to the central, and in my view excessive role played by a firm’s scale of operations (as measured by the number of workers in a firm) in both the theory and empirical work here.

Imagine asking a production- or human resource manager at a ‘typical’ U.S. or U.K. firm the following question: ‘Aside from the obvious increase in compensation costs, what do you expect would be the main consequences of raising all your production workers’ wages by, say, 5 percent?’ My guesses (based on little more than extrapolation from the fascinating interview results in Bewley (1999)) is that answers like ‘better morale’, ‘a reduction in quits’, and ‘would attract a more qualified group of workers’ would be reasonably frequent. So, perhaps, would ‘make it easier to hire new workers’, which could indicate some short-run monopsony power. On the other hand, I find it much harder to imagine a
manager rating ‘it would make it easier for us to expand our scale of operations in the long-run’ as an important consequence of offering a higher wage. Yet it is this latter response that lies at the heart of the general-equilibrium models of monopsony in this book.

Manning clearly recognizes this weakness of search-based monopsony models, and does his best to address it in his discussion of ‘random’ vs. ‘balanced’ matching on pages 284–96. Manning’s basic general-equilibrium monopsony model, set out in chapter 2, assumes ‘random matching’, which means that, regardless of its size, every firm—from the local bakery to Microsoft—receives the same absolute number of job applications per period. The only way for a firm to expand its scale of operations in this model is to offer a higher wage, thereby ensuring that a higher share of its fixed flow of job applicants accept the firm’s wage offer, and that a smaller share of its own workers are bid away to work for other firms. Of course, as Burdett and Vishnawath (1988) point out, this means that it would be in every firm’s interests to split itself into as many pieces as possible, but this is simply not allowed in Manning’s basic model.

An intuitive alternative to this troubling assumption is ‘balanced matching’, in which a firm’s contact arrival rate (at a given wage) is proportional to its level of employment. This would be plausible, for example, if job contacts were made through networks of existing workers. However, as Manning himself acknowledges, if matching is balanced (which effectively amounts to constant returns to scale in the technology for recruiting new workers), all elements of monopsony disappear from the model and the neoclassical equilibrium again prevails: in the long run firms can expand without limit without needing to raise their wages. Thus it is absolutely critical to the search-based monopsony model at the core of this book that there be diminishing returns to scale in the technology for recruiting new workers. In other words, for the theory to apply, firms must find it harder to recruit a single new worker the larger the absolute number of workers they currently employ.

Needless to say, the above strikes me as a very thin nail on which to hang an entire book. And unfortunately, Manning’s attempt to support it with evidence (on pp. 286–92) is creative but ultimately unconvincing. In fact, I can think of several reasons why there might be increasing returns to scale in recruitment (there are fixed setup costs for a hiring system or personnel department; it is no costlier to post a newspaper advertisement announcing five vacancies than one). Indeed, any economist who has been on the buying side of the academic labor market would, I suspect, readily agree that—given setup costs and the cost of sorting through candidates—it is almost as much work filling one junior position in a year as three!

Aside from the gains to splitting firms and empirical realism, the diminishing-returns-in-recruitment assumption raises two additional difficulties in relating the predictions of search/monopsony models to the real world. One of these is that—in contrast to the constant-returns case—the thorny issue of what exactly constitutes a firm matters for the predictions of the model. For example, if diminishing returns are assumed to operate at the firm level, then for these models to make sense it must be the case that the forcible division of Microsoft into (say) 10 component firms would make it significantly easier for each of those new, smaller firms to recruit employees than when they were part of Microsoft. This seems unlikely (if for no other reason than they probably already do most of their hiring at a lower level anyway). Alternatively, if diminishing returns operate at the plant or establishment level (which seems more plausible), then firms always have the
The option of expanding at constant marginal cost by replicating existing plants in a new geographic location, and diminishing returns effectively disappear from the model.

The mention of geographic location raises a second difficulty. In the industrial organization literature, assessments of a firm’s product market power tend to pay close attention to (a) defining the market in which the firm operates, and (b) assessing the firm’s size relative to this market, in part by calculating concentration ratios, Herfindahl indexes and the like. Strikingly, discussion of both these issues is essentially absent from this book, whose central thesis is that firms have power in labor markets. Empirically, it seems likely that any given firm operates in a considerable number of labor markets at once, defined by worker skill level, skill type, and geography. Just as clearly, attempting to measure market power in any one of these markets without reference to the firm’s size relative to the market seems, on reflection, somewhat bizarre. That is, however, the consequence of working with search models of the type considered in this book, in which every firm is infinitesimal in size relative to the market as a whole. In sum, careful consideration of how one might actually measure a firm’s labor market power, perhaps surprisingly, leads one back towards the concerns of the ‘traditional’ monopsony model—how large is a firm relative to its labor market—and away from the concerns of the search-based ‘new monopsony’ literature—where all that matters is the number of existing workers in the firm.

In sum, despite (and in some cases because of) the evidence presented in this book, it strikes me as highly unlikely that any individual firm in a labor market as large as the U.S. or U.K. faces an upward-sloping supply curve for labor of a given quality in the long run. Contrary to what Manning sometimes implies in this book, this is not inconsistent with the notion that, in the real world, firms choose what wage to pay. Clearly, they do, and human resource management textbooks are full of discussions of the relative merits of choosing high versus low wages. The main consequences of a higher wage, in my opinion, however are not the possibility of expanding the firm’s long-run scale of operations (as the search models in this book assert), but an increase in the average ability of workers a firm ends up hiring (as in the alternative, neoclassical model sketched in Section 4 of this review).

6. Summary

This book is a tour de force by a creative and talented economist. It combines theory and evidence on a wide array of labor market issues in thoughtful and creative ways. I wish the profession had several books a year like this, rather than the one per decade or so that currently appear. Ultimately, however, I am left unconvinced that the class of models advocated in this book is the ‘right’ way to think about labor markets today. Unless one focuses on workers with very specific skill types in very defined geographical areas, upward-sloping labor supply curves—whether induced by search or other factors—seem unlikely to me to be a serious constraint for most firms. This seems even more likely to be the case in the near future, as barriers to goods and factor flows across regions and countries continue to fall, and as information technology has the potential to reduce search frictions (Kuhn 2003).

That said, this book provides a refreshingly different perspective on almost every one of the ‘traditional’ topics in labor economics, from labor supply to
discrimination. As such, it makes fascinating reading for labor economists, and would make a nice addition to many graduate reading lists. Finally, how many economics books come with a downloadable theme song and video? Get your copy, written and performed by John DiNardo, at http://econ.lse.ac.uk/staff/amanning/work/book.html (Alan Manning’s web site). It’s cool! (at least to a middle-aged economist like me).

Notes
1. Manning expands on this theme in a more recent article, ‘Monopsony and the Efficiency of Labour Market Interventions’ (2004).
2. In fact, Burdett and Mortensen (1998), whose model forms the basis of much of Manning’s analysis, reserve the term ‘monopsony’ for the polar case of their search model in which only unemployed workers receive job offers and the wage distribution collapses to the value of unemployment.
3. As Manning notes, the evidence that a firm’s quit rate responds to the wage it pays with an elasticity less than infinity in absolute value is both stronger and more convincing than the firm-size and wages literature (see pp. 100–104). However, this is not in itself evidence of the kind of monopsony studied in this book. It is, for example, consistent with a labor market such as that modeled by Hashimoto (1981) that is perfectly competitive \textit{ex ante}, but in which workers acquire firm-specific skills after being hired.
4. It is trivial to replicate markets of the sort described an arbitrary number of times, with one market for each ‘type’ or quality of labor. Since this does not allow workers of different (even adjacent) ability levels to compete with each other for the same jobs, it has little hope of providing a realistic depiction of labor markets.
5. If work intensity is chosen optimally given the wage, the envelope theorem implies workers will be indifferent at the margin between the original situation and a small cut in total earnings accompanied by a proportionate reduction in work intensity.
6. Capelli and Chauvin (1991) provide convincing evidence that worker effort does respond to offered wages.
7. In fact, one can even argue that filling a single position is more costly since the possibilities for compromise between competing demands within a department are much more limited, thus adding to (within-department) bargaining costs.
8. Another way in which the diminishing-returns-in-recruitment assumption stretches the bounds of empirical plausibility concerns the simple fact that firms of one worker coexist with firms of 50,000 workers in today’s economy. One suspects that calibrating an equilibrium search model in a way that permits this to exist would require a very small degree of diminishing returns (and hence a very small amount of monopsony power). Relatedly, the diminishing-returns-to-recruitment assumption would seem to contradict a well-known empirical regularity in the economics of firms, namely Gibrat’s law that firm growth rates seem to be independent of firm size.
9. In practice, the empirical literature on ‘firm’ size and wages actually tends to use \textit{plants} or \textit{establishments} as the unit of analysis more often than firms. Alternatively, some theoretical search models adopt a convenient convention of equating ‘firms’ with ‘jobs’. In this case firm size can be recast as the fraction of time the job is occupied. But that requires one to model the fact that workers who ‘bump into’ filled vacancies in the search process now must remain unemployed until they encounter a vacant position, which significantly complicates the analysis.

References


