We have all visited several stores to check prices and/or to find the right item or the right size. Similarly, it can take time and effort for a worker to find a suitable job with suitable pay, and for employers to receive and evaluate applications for job openings. Search theory explores the workings of markets once facts such as these are incorporated into the analysis. Adequate analysis of market frictions needs to consider how reactions to frictions change the overall economic environment: not only do frictions change incentives for buyers and sellers, but the responses to the changed incentives also alter the economic environment for all the participants in the market. Because of these feedback effects, seemingly small frictions can have large effects on outcomes.

Equilibrium search theory is the development of basic models to permit analysis of economic outcomes when specific frictions are incorporated into simpler market models. The primary friction addressed by search theory is the need to spend time and effort to learn about opportunities—opportunities to buy or to sell, to hire or to be hired. There are many aspects of a job and of a worker that matter when deciding whether a particular match is worthwhile. Such frictions are naturally analyzed in models that consider a process over time—of workers seeking jobs, firms seeking employees, borrowers seeking lenders, and shoppers buying items that are not part of frequent shopping. Search theory models have altered the way we think about markets, how we interpret market data, and how we think about government policies.

The complexity of the economy calls for the use of multiple models that address different aspects of the determinants of unemployment (and other) outcomes. This view was captured so well by Alfred Marshall (1890: 1948 edition, p. 366) that I have quoted this passage repeatedly since coming upon it while doing research for the Churchill Lectures (Diamond 1994b).

The element of time is a chief cause of those difficulties in economic investigations which make it necessary for man with his limited powers to go step by step; breaking up a complex question, studying one bit at a time, and at last combining his partial solutions into a more or less complete solution of the whole riddle. … The more the issue is thus narrowed, the more exactly can it be handled: but also the less closely does it correspond to real life. Each exact and firm handling of a narrow issue, however, helps towards treating broader issues, in which that narrow issue is contained.

---

† This article is a revised version of the lecture Peter Diamond delivered in Stockholm, Sweden, on December 8, 2010, when he received the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel. This article is copyright © The Nobel Foundation 2010 and is published here with the permission of the Nobel Foundation.

‡ Diamond: Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142 (e-mail: pdiamond@mit.edu). I am grateful for help with interpretation of the literature and comments on drafts from George Akerlof, Jim Albrecht, Nick Barr, Olivier Blanchard, Jesse Bump, Ken Burdett, Steve Davis, Bill Dickens, Glenn Ellison, Laura Feiveson, Ben Friedman, Bob Hall, John Haltiwanger, Bengt Holmstrom, Frank Levy, Matt Mitchell, Dale Mortensen, Giuseppe Moscarini, Mike Piore, Mike Rothschild, Jenny Simon, Kathy Swartz, Peter Temin, Susan Vroman, Ivan Werning, Janet Yellen, and Joel Yellin.
more exactly than would otherwise have been possible. With each step…

exact discussions can be made less abstract, realistic discussions can be
made less inexact than was possible at an earlier stage.

This passage is particularly suitable here since the heart of the difference between
standard demand-supply analysis and search theory lies in the treatment of time—the
use of time as well as resources to learn about opportunities. I am concerned that, in
contrast with Marshall’s view, too many economists take the findings of individual
studies literally as a basis for policy thinking, rather than drawing inferences from an
individual study, and combining them with inferences from other studies that con-
sider other aspects of a policy question, as well as with intuitions about aspects of
policy that have not been formally modeled. Assumptions that are satisfactory for
basic research, for clarifying an issue by isolating it from other effects, should not
play a central role in policy recommendations if those assumptions do not apply
to the world. To me, taking a model literally is not taking a model seriously. It is
worth remembering that models are incomplete—indeed, that is what it means to be
a model.

This essay is not a survey of uses of search theory in analyzing many labor market
questions or questions in other markets.1 Instead, it explores a few of the contrib-
utions of search theory, focusing primarily on early work that marked the break
from analyzing equilibrium in markets to analyzing equilibrium in a search setting.
Although the focus is on the labor market, the discussion starts with a retail market
example, followed by discussions of labor market flows, aggregate demand, and the
Beveridge curve, which shows unemployment and vacancy rates for the economy.
Considerable attention has been given to possible policy implications of the recent
evolution of the Beveridge curve, during what has been called the Great Recession
and the Long Slump. I add to this ongoing debate from a search perspective.

I. A Retail Market Example

Economics education starts with the abstraction of an ideal market—demand,
supply, and a price that clears the market. That is an extremely valuable starting
place—it shows some of the effects that are present in pretty much all markets. And,
since such ideal markets can achieve efficient outcomes, it helps us understand the
sources of inefficiencies that can occur when market structures differ from the ideal-
ized version. In the simple market abstraction, each buyer knows how to purchase at
the lowest price available in the market. As a result, each firm faces a discontinuity
in sales when the price varies from the lowest price elsewhere and the “law of one
price” holds—all transactions happen at the same price, which is the competitive
price, equal to the marginal cost of providing the good.

To analyze a search setting, the first step is to consider how individuals approach
both the search process and the ensuing purchase or sale decision. And that is where
the literature started. But a full analysis needs to combine individual decision making

1 For a survey, see Richard Rogerson, Robert Shimer, and Randall Wright (2005). My citations to the literature
primarily reflect my memory of papers that particularly influenced my research, as well as recent articles that have
stimulated a response from me. Having stopped work on search 15 years ago, having shifted my focus to pensions,
my awareness of the current literature is limited, although I have enjoyed catching up somewhat.
with an analysis of how the interactions of buyers and sellers determine the economic environment in which these decisions are made.\(^2\)

Consider the simplest standard model, a retail market for a standardized good with a large number of identical buyers and a large number of identical sellers. My 1971 paper alters this model by assuming that the only way to find a price is by visiting a store, and that stores are visited randomly. Then, visiting another store to get a second price quote has a cost. That implies that the first store you happen to visit has a little bit of market power over the other stores. Surprisingly, in this uniform setting, these little bits of market power interact so that all firms want to charge the price a monopolist would set. If the price were lower and the same in every store, then each store would want to make a small price increase to take advantage of its little bit of market power.\(^3\) So the only way to have equilibrium, where no store has an incentive to change price, is when the monopoly price is set. Notice that it does not matter how small is the cost of visiting another store—any positive cost gives the same result. Technically, there is a discontinuity in the equilibrium price as the cost of search hits zero.

My description of this example reflects an equilibrium approach similar to the approach of a standard competitive equilibrium—equilibrium in Arrow-Debreu theory is a set of conditions on an allocation of resources such that no price-taker would want to change behavior from what is consistent with that allocation. However, this description does not reflect how I came to my 1971 paper, nor how I presented the analysis. The process that led to this, my first search paper, began in 1969, while I was visiting Hebrew University, Jerusalem, and then Nuffield College, Oxford (rounding out 15 months of leave that started at University College, Nairobi).

Arrow-Debreu theory does not contain a mechanism or process for an economy to achieve its equilibrium allocation. In the 1960s there was ongoing work to find a hypothetical process that would converge to this equilibrium, with a focus on equations for price adjustment based on excess demands and supplies at tentative prices (referred to as tâtonnement). It struck me that the wrong question was being asked. Rather than asking whether a process could be found that would converge to a standard competitive equilibrium, I chose to work on the question of finding the allocation to which a plausible process would converge. The research approach was a dynamic overlapping generations model, which presented conditions under which the (uniform) prices of identical firms converged to the monopoly price in finite time. Paying attention to a dynamic process, and not just conditions for a stationary equilibrium, was a central part of my research strategy. To my surprise, I found the monopoly pricing equilibrium, about as far from a competitive equilibrium as one could be.\(^4\) With monopoly prices, equilibrium is not efficient, even with the search frictions taken into account.

---

\(^2\)The need for equilibrium analyses was stressed in Michael Rothschild (1973).

\(^3\)Formal modeling is readily done assuming shoppers know the distribution of prices, but not which price is charged at which store. While convenient, this is an unrealistic picture of what people know. People do not know the distribution of prices and often have been given some partial information about relative prices, from others and from store reputations. Conventional modeling of both random search and directed search straddle plausible information flows.

\(^4\)Earlier, my dissatisfaction with the completeness of markets assumed in Arrow-Debreu theory led me to explore alternative incomplete allocation mechanisms. My search over mechanisms did not last long, as I settled quickly on a model of a stock market, with a limit on allowable trades from the set of existing stocks, which were
This model was not meant to be a realistic model of price determination, but an exploration of how important search frictions could be—and they were found to be very important, even when the search costs were small. This result highlights one of the central aspects of economics. Economists study the nature of equilibrium, reflecting the interactions among buyers and sellers. Small search costs can have a large impact because price setters respond to the prices set by others, so there is a feedback process that greatly expands the impact of search costs—as each firm reacts both to the presence of the search costs of potential customers and to the responses of other suppliers to the same search costs. That a small amount of friction could create a large change, even in such an unrealistic setting, served as a marker of the importance of the study of equilibrium with frictions.

Of course, this extreme result is dependent on a number of special and unrealistic assumptions. But the point of the analysis was to open up a field of study, not to zero in on a particular market. Later work, by a number of economists, explored models with more realistic assumptions. Two findings of this literature are to expect a variety of prices for identical commodities and that pretty much all the prices will exceed the marginal cost of providing the good. Models show these results when they incorporate variation in the demand curves of different buyers, variation in the costs of different sellers, mixed strategies being followed by homogeneous sellers, or more complex methods of becoming informed about prices. With richer modeling, the level of prices does depend on the size of search costs.

A price per unit set by a seller has two roles: affecting the number of customers who buy and affecting the average quantity purchased by those who do buy—the extensive and intensive margins. In the simple setting originally analyzed, search frictions imply that the extensive margin (affecting the number of purchasing customers) plays no role in equilibrium price setting, resulting in the equilibrium price being the one that optimizes relative to the intensive margin (average sales per buying customer). In contrast, thought of as a price-setting (Bertrand) equilibrium, standard competitive market theory gives full weight to the extensive margin and none to the intensive margin.

After my return from 15 months abroad, I learned that search ideas were very much in the air from the publication of the Phelps volume (1970), written while I was away. It is no surprise when many researchers are thinking along parallel lines, although I came to search from general equilibrium theory, while the Phelps volume had an unemployment focus. The process of going from my start on search assumed not to be rich enough to span the space of possible outcomes. To analyze efficiency I took the bases of the limits on market outcomes to also limit how the government can affect resource allocation, an approach that has also been taken in search theory. While the simple, basically one-period model I constructed, 1967, showed efficiency of the market, I was aware that having a single period was special. Later work, incorporating more periods, showed that generically the private allocation is inefficient, quite a contrast to the efficiency with complete markets (Oliver D. Hart 1975; John D. Geanakoplos and Heraklis M. Polemarchakis 1986). For broader discussion of my research motivations and the diversity of ways I came to work on different topics, see Guiseppe Moscarini and Wright (2007) and Diamond (forthcoming).

5Kenneth Burdett and Kenneth L. Judd (1983) clarified the link between information patterns and the type of equilibrium that occurs. Wage dispersion for similar workers has been a significant topic (Dale T. Mortensen 2003). Presumably similar underlying mathematical structures are involved, as in the retail price literature, although the citations across literature appears limited.

6Not surprisingly, studies of the impact of the low search costs when using the Internet have been an interesting part of the search literature.
equilibrium to my writings on the labor market (Diamond 1981, 1982a) and on the entire economy (Diamond 1982b, 1984) had two phases. The first phase was an exercise in brain rewiring. As an undergraduate, in the spring of 1960, I studied general equilibrium theory with Gerard Debreu, when his Theory of Value (1959) was new. Gerard was an outstanding teacher, and I became well grounded in the theory, so much so that it became the starting place for my thinking about allocation questions. But it was not a good starting place for thinking about search issues. Nevertheless, I found my mind running down that path as I attempted to find a different way to approach questions. In order to free up my thinking process, I pressed myself to explore a question that could not possibly be answered within the Arrow-Debreu framework. This resulted in a little noticed, read, or cited 1978 paper and greater ease in creating models.

The second phase was writing a pair of papers with Eric Maskin (1979, 1981), considering search equilibria in a setting of pairwise contracting, with a focus on a law and economics question—the role of damage payments for breach of contract. For a contracting paper, it was natural to consider a bargained price rather than one set by one party, as in the retail market paper. Thus the focus had shifted from the price outcome to the efficiency of the search-and-matching process. We assumed equal splitting of the surplus from signing a contract. These papers drew on a contracting paper by Mortensen (1978) that helped me realize the power of Poisson processes as modeling tools. After this work, I was ready to work on models focused directly on the labor market and on the entire economy. In this work, I continued with the assumption of a negotiated wage rather than one set on a take-it-or-leave-it basis as in the retail model above. To set the framework for later discussion, I begin with a review of some data on market flows.

II. Labor Market Flows

At any time, a modern economy has both unemployed workers and posted vacancies. Thinking about this fact in a static setting might lead one to think there is a serious mismatch between the workers and the jobs, in skills or in location, for example. Or it might lead one to think that the primary problem is that wages are not at the right levels to clear the markets, referred to as “sticky wages” since wages are not changing adequately as circumstances change. A search perspective puts these two ideas into a richer, more informative context. That is, the quality of matches between

7 In Arrow-Debreu theory, subjective probabilities are part of preferences, which are respected in normative evaluations. With consumers searching across stores with different marginal costs without knowing the distribution of prices, I examined the socially optimal prices in the stores, evaluated with the correct distribution of prices. I found three reasons for the optimum to differ from marginal cost pricing—nonoptimal stopping rules, mistaken consumer prior beliefs on the distribution of prices, and correct prior beliefs that are then revised based on observed prices.

8 Rogerson, Shimer, and Wright (2005) discuss the differences in models with wage posting and wage bargaining. As they state, the differences across the models also depend on the nature of information flows (typically modeled as random search in bargaining models and as directed search in price posting models). Paralleling the retail market literature with the dependence of prices on the nature of information flows, I suspect that with multiple sources of information flow (and reputations for hard and soft bargaining) the differences between the two approaches would not be large. In any event, Robert E. Hall and Alan B. Krueger (2010) note that posting and bargaining (and bargaining within posting) are all widespread in the United States today. Firms with nonnegotiable wage schedules may bargain on the title given to a worker when wages vary with title, which supposedly reflects duties or experience, but may be somewhat flexible.
workers and jobs matters, and limited rates of wage change matter, but understanding the extent and effect of these two issues is best done in a search setting.

I start by considering the monthly CPS Household Survey. At the time of their interviews, workers are classified into three categories—employed, unemployed, and out-of-the-labor force. A month later many of the workers who were unemployed will have found jobs. And many people who were classified as out-of-the-labor force also will have found jobs. Of those employed in the earlier month, most will still be in the same job, but many will have left their previous jobs, some voluntarily and some not, some to new jobs, and some to unemployment or to a departure from the measured labor force. Indeed, the flow of workers directly from one job to another is large and is an important part of the efficiency of the labor market. On average, 2.6 percent of the employed have a different employer a month later—there are more hires from the employed than the unemployed.9

For someone to be defined as being unemployed, he or she must be actively trying to find a job—not just being without a job and wishing for one. Many people labeled as being out-of-the-labor force are in fact available for employment, and many of them do find jobs. Figure 1 shows the average monthly flow rates in the United States over the last 20 years. On average, 37 percent of the unemployed were employed as of the following month (labeled UE in Figure 1). Almost as large a fraction, 33 percent, left unemployment by leaving the measured labor force, labeled UI, with I for inactive. Similarly, those who were inactive flow into both employment and unemployment (IE and IU). And the employed flow into unemployment and also outside the labor force (EU and EI). On average over this 20-year period, roughly 6 million workers moved into employment each month and roughly the same number moved

9 The estimate is for 1994 and 1996–2003 (Bruce Fallick and Charles A. Fleischman 2004).
out of employment. The large differences in the rates of flow in the chart reflect the much smaller number of unemployed than of employed or of inactives.

Each of the rates in the figure is an average for all the individuals in the category, with the rates varying greatly across groups, whether measured by characteristics, such as education, race, and gender, or by the duration of the unemployment spell. Thus there is not a stationary stock of unemployed, but a steadily shifting set of workers, with large flows both in and out but with some people staying unemployed for a long time.

The US economy is an outlier, having larger labor market flow rates than other advanced economies, although flow rates are substantial elsewhere as well. Figure 2 shows monthly flow rates into and out of unemployment that have been estimated by Elsby, Hobijn, and Şahin (2008) using OECD data. The calculation of these figures examined the flows to and from being unemployed, and did not distinguish between the employed and those outside the labor force. The rates are much higher in the United States than elsewhere, and only a portion of the differences comes from the larger fractions involved in movements to and from out-of-the-labor force. The difference in flows is, in part, a reflection of government policies that affect hiring and firing, a subject that has received considerable attention. And the resulting level of job availability affects the willingness to seek new employment.

10 Dates for the averaging vary by country, depending on data availability.
11 According to calculations for a subsample of these countries, drawn from multiple sources and provided to me by Elsby, Hobijn, and Şahin, the fraction of unemployment outflows that become employed is lowest in the United States (at 53 percent), but this figure does not appear to be a large outlier: analogous numbers for the other countries for which they could find estimates hover between 55 and 60 percent.
Turning to a perspective from data on employers, Figure 3 draws on the extensive pioneering work of Davis and Haltiwanger and coauthors. In a typical quarter, many firms are increasing employment while others are decreasing it. On average over the last 20 years, additional employment at growing and new firms, labeled job creation, was 7.5 percent of employment, with job destruction almost as large. This is another view of the way that the change in employment is a result of the netting of large gross flows. Hires and separations are roughly twice as large as job creation and destruction, as workers leave and firms replace them. Note that, of workers leaving, there are roughly equal numbers of quits and layoffs on average. However, the relative sizes vary greatly over the course of the business cycle—at a time when layoffs rise, quits drop as the ability to land a new job also drops.

Search theory is designed to help make sense of these flows and to frame analysis of related government policies such as unemployment benefits. Considering “ordinary” times, captured by a rational expectations model set in a steady state, a primary purpose of unemployment benefits is to provide insurance to workers against involuntary job loss. The provision of insurance affects the willingness of workers to accept particular jobs, making it more attractive to pass up some opportunities in expectation of doing better later. And it affects the diligence of job search. This is always the case with insurance with asymmetric information—the provision of insurance affects behavior, commonly referred to as a moral hazard problem. Of course, there are behavioral changes that reflect income effects and can have a positive efficiency implication as well as those that reflect substitution effects and can have a negative efficiency implication. The negative effects do not mean one should not have insurance, but that the strength and design of the insurance should take into account the behavioral responses it induces.

**Figure 3. Quarterly Job and Worker Flow Rates for the US Private Sector 1990:II–2009:IV**
(As a percent of employment)

Notes: Series drawn from methodology used in Davis, Faberman, and Haltiwanger (2010b) Series measured from Business Employment Dynamics (BED) and Job Openings and Labor Turnover Survey (JOLTS). Pre-2001:III hires, separations, layoffs, and quits are model-based estimates.

Source: John Haltiwanger, personal communication.

---

12 Much of the literature simplifies matters greatly by assuming risk-neutral workers and employers, all having the same discount rate. Of course, one cannot do justice to insurance in such a context, and, happily, the literature has developed to include both risk aversion and savings (e.g., Daron Acemoglu and Shimer 1999).

13 Experience rating the taxes that finance unemployment benefits also influences the behavior of firms.
Less diligent search and greater willingness to wait for a future job make it harder for employers to find workers but easier for other workers to find jobs. The relative importance of these positive and negative feedback effects will vary with the large swings in the vacancy-unemployment ratio that happen over the business cycle, so the impact will be different at different times—at times of high unemployment, a little less search by some workers will not have much impact on the difficulty of filling vacancies and so will have little effect on total unemployment. These effects on workers and firms are externalities, affecting people with whom there are no direct transactions that might counter the effect. Thus, the “natural rate of unemployment,” coming from adding labor market frictions to a standard competitive model, is not generally an efficient level of unemployment. That is, having a worker search less and be more choosy about job acceptance as a result of unemployment benefits may raise or lower efficiency in the labor market. One cannot assume that the no-benefit levels of search and job acceptance would be efficient, in contrast with the usual assumption when considering moral hazard issues in insurance analyses for ordinary goods without externalities.

By contrasting the experiences of workers having access to more or less generous unemployment insurance coverages, such as extensions of coverage during recessions, one can estimate how the average duration of unemployment of covered workers varies cross-sectionally with details of insurance provision. However, by itself, this does not provide an estimate of the impact of unemployment insurance (or an extension of the period of benefits) on total unemployment. Because of the externalities, looking only at the change in effort to become employed does not provide an estimate of the equilibrium impact of unemployment insurance on total unemployment. Such an approach would be tantamount to assuming that employment equals labor supply. That is, it ignores the impact of less diligent search on the ability of others to find jobs and on the creation of vacancies. This is clearly incomplete. Equating employment to labor supply is an amusing contrast to traditional Keynesian analyses that assume that employment equals labor demand. These can’t both be right. And from a search theory perspective, neither is—the matching function, discussed below, gives the change in hiring from changes in the number of workers seeking jobs and the number of vacancies seeking workers; it depends on both demand and supply.

In addition, the presence of unemployment benefits and the induced changes in the flows of hires, layoffs, and quits affect the level of compensation. Whether modeled as a bargained wage or a wage set on a take-it-or-leave-it basis, the availability of workers and jobs will affect the determination of compensation. And that will affect the decisions of employers and workers as to the willingness to create jobs, to stay in the labor force, and to seek a better match. All of these issues are naturally approached through an equilibrium search model.

---

14 Friedman (1968, p. 8) defined the natural rate of unemployment as follows: “The ‘natural rate of unemployment,’ in other words, is the level that would be ground out by the Walrasian system of general equilibrium equations, provided there is imbedded in them the actual structural characteristics of the labor and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on. Footnote: It is perhaps worth noting that this ‘natural’ rate need not correspond to equality between the number unemployed and the number of job vacancies. For any given structure of the labor market, there will be some equilibrium relation between these two magnitudes, but there is no reason why it should be one of equality.”
While insights have come from models with ex ante homogeneity of both jobs and workers, a central element in the labor market is that workers and jobs vary greatly. And the extent to which a given worker is a good fit for a particular job not only varies but may take time and expense to determine, especially as some worker training, sometimes a great deal of training, is generally needed for a worker to do well at a job. Thus, a hiring decision is a form of investment, just as is the acquisition of plant and equipment. Both workers and firms need to have views as to the value of future alternatives. The worker needs to be concerned about how long the job might last and what alternative jobs might be found. The firm needs to be concerned about the future value of the services that will be provided by the worker and what alternative workers might become available. All of these depend on what other firms and workers will do and how well the economy will do.

With heterogeneity of both workers and jobs, there is a “matching problem” that affects the efficiency of the economy. Search theory helps us to make sense of this complex environment. It has been natural to focus on rational expectations to isolate the impact of search per se on resource allocation. But modeling only with rational expectations is incomplete in not recognizing the heterogeneity in expectations that are always present (e.g., Johannes Spinnewijn 2009, 2010). Also, it does not apply well to circumstances, as at present, of a recession deeper than any experienced in many decades, which makes extrapolation from past recessions an incomplete foundation for expectations.

A key shortcut that makes search analyses tractable is the “matching function.” For theoretical work the matching function gives the rate of meetings of workers and firms in the process of seeking employment as a function of the numbers of searching workers and job vacancies. The theory combines this with analysis of which meetings result in new hires. Empirically, we do not have information on meetings, but on hiring. The same term, the matching function, is used for the empirical relationship relating hires to the numbers of unemployed and vacancies. From the context, I don’t think there is much risk of confusion from the different meanings for “matching” of meeting and hiring. The matching function plays a similar role in search analyses as does the aggregate production function in growth theory. Empirically, for the economy as a whole, the matching function appears to have constant returns to scale.

The matching function is not solely technologically driven, but the outcome of a process reflecting the ways in which hiring occurs, which vary greatly across the economy. The simplest model has workers and jobs meeting one to one, although urn-ball models are also used to examine some implications of multiple applicants for particular jobs. And the presence of firms with multiple vacancies matters for the matching process. Some of these issues will be mentioned again in the discussion of the Beveridge curve.

---

15 For an extensive discussion of the matching function, see Barbara Petrongolo and Christopher A. Pissarides (2001).
16 And possibly quits and layoffs triggered by new hires, as in Diamond and Maskin (1979, 1981).
17 An urn-ball model sees job applications as balls being tossed in a randomized way into urns, representing vacancies awaiting applications. See, for example, Olivier Blanchard and Diamond (1994).
III. Aggregate Demand

Basic search analyses of the labor market for “normal times” proceed by taking the value of a worker to a firm as given primarily by technology; and the value to a worker of being without a job as given by preferences (and unemployment insurance). This is partial equilibrium modeling, leaving out the role of the output market. As a method of capturing labor market outcomes around turning points in the business cycle, Pissarides (1985) and Mortensen and Pissarides (1994) have advanced the analysis by examining the dynamics of an economy when these exogenous values shift. This is an important step forward, a way of examining the labor market around turning points. However, that still leaves the critical task of endogenizing the assumed values of production and the assumed costs of lack of work. That task is particularly important for times of high unemployment.

While commonly referred to both as changes in the “value of output” and as shocks to “productivity,” the broad range of changes around turning points is supportive of an aggregate demand interpretation (Blanchard and Diamond 1989). While the term “productivity” may be useful metaphoric shorthand, we should not think that output per worker provides a reasonable measure of the value of additional output for business cycle analysis. With a Walrasian output market, productivity does measure the value of output; but with a search model of the output market, and so a limited ability to make sales, productivity is not necessarily a good measure of the value of output.

In other words, a labor market model is a partial equilibrium model, not a general equilibrium model. As such it can shed light on partial equilibrium questions and partial equilibrium aspects of general equilibrium issues, but cannot, by itself, fully evaluate general equilibrium questions, such as the role of aggregate demand stimulation. There are multiple ways of combining a search model of the labor market with a model of the output market. Both positive and normative evaluations of interventions in the labor market, such as the considerable extension of unemployment benefits during the last few years, can vary critically with how the output market is modeled as well as how the labor market is modeled. Combining a frictional labor market with a Walrasian output market seems likely to miss some important links that matter for policy design for extended periods of high unemployment. Looking only at output per worker is thus an inadequate measure of the value of an additional hire. While the bulk of search analyses have focused on the labor market per se, search also has been used to see how the presence of frictions affects the aggregate economy.

Complete-market Arrow-Debreu theory has coordination of production decisions and purchase decisions by having all prices and quantities settled ahead of time. While there is considerable production after contracts are signed for delivery of the output, overwhelmingly, firms make investment and production decisions in anticipation of future sales without contact with potential future customers, suggesting a possibly significant role for search theory.18 The Arrow-Debreu model conditions household demands on lifetime budget constraints. But current income plays a much

18 Insofar as contracted production is for other firms that base their demand on anticipated sales, the same considerations hold.
larger role in consumption decisions at business cycle frequencies than is consistent with the market structure in the Arrow-Debreu model (Orazio P. Attanasio and Gugliermo Weber 2010).19

Keynesian theory gives important roles to both sticky wages and contemporaneous income. Some New Keynesian analyses consider sticky prices and wages without a significant role for contemporaneous income by using infinite-horizon budget constraints for consumption demand (e.g., Blanchard and Jordi Galí 2010). My research hunch when starting on this model was that sticky wages were only one source of macro difficulties. I thought that search frictions in the output market opened a way to capture the macro role of contemporaneous purchasing power as a complementary basis for analyzing unemployment, which would then also clarify the role of sticky wages. That is, my choices for modeling the entire economy (and not just the labor market) came from a belief in the incompleteness of some macro, Keynesian, sticky-wage analyses, which was to be demonstrated by similar properties in a model with no wage or price stickiness—with the completion of all mutually advantageous trades that the parties were aware of. The approach was not meant to remove sticky wages from the consideration of unemployment, but by considering a model without sticky wages to make the point that sticky wages were not the sole basis for macro problems. While I started working on search theory out of dissatisfaction with general equilibrium theory, I gravitated to seeing search also as a way to address my dissatisfaction with macro theory. My dissatisfaction did not relate to basic Keynesian concepts, but to the nature of modeling. I wanted to see a microfoundation that would enhance the ability to do normative analysis and to develop policy insights.

A. Multiple Equilibria without Sticky Wages

The complexity of addressing frictions in both output and labor markets makes it difficult to address both in a single model. So, my 1982b paper on the role of aggregate demand suppressed a role for the labor market by assuming only self-employment. This approach was based on a presumption that a model with a distinct labor market would have similar properties in the output market—that production contributes to demand, whether its proceeds are owned by a self-employed worker or divided between a worker and an employer. Frictions arise from lags between a production decision and a potential sale, reflecting both the time of production and the time to sell. Since time to sell was the key endogenous variable, the variable reflecting the state of aggregate demand, I left out time to produce, making that instantaneous. To capture the need to trade produced goods I assumed that individuals could not consume what they produced, but must barter output with another producer. Thus production provides purchasing power. As a barter model without credit, this overplays the importance of contemporaneous purchasing power in order to clarify its role.20 Since all producers with something to sell were assumed to be in

19 Furthermore, the Arrow-Debreu model has no budget constraints on investment (beyond the need to break even). But investment is more sensitive to cash flows than is consistent with that model (Steven M. Fazzari, R. Glenn Hubbard, and Bruce C. Petersen 1988).

20 The model did not allow for trade credit, something I analyzed in a 1990 paper. The ability to extend credit alters the workings of the model but, not surprisingly, does not change the basic qualitative findings.
the same setting, all trade would be one-for-one and no potentially mutually advantageous trade would be passed up.\textsuperscript{[21]}

Assuming that trade is quicker with more potential trading partners, higher total production (implying more people with purchasing power) shortens expected sale time and there is an external economy with positive feedback. That is, when others are producing little, the ability to sell is low, and so the incentive to produce is low. When others are producing much, the incentive to produce is high. The goal was to capture an effect of the availability of more purchasing power on equilibrium, not particularly the frictions in an actual shopping process. That is, search modeling would capture the impact of the difficulty of selling when aggregate incomes were low.

This externality implies that the equilibrium level of production is inefficiently low. The positive feedback opens up the possibility of multiple equilibria. And with multiple equilibria, the economy can trace out what would resemble a business cycle (Diamond and Drew Fudenberg 1989) and there is a role for government policy to affect aggregate demand. Key to the multiple equilibria result is an assumption of increasing returns in the combination of production and sales, not particularly in the labor market. Thus, it is not appropriate to reject the importance of multiple equilibria based solely on the typical finding of constant returns of the aggregate matching function in the labor market, since that ignores the output market. It is hard to think of what would be a comparable directly informative empirical analysis of returns to scale of the full production and sales process. One place to look would be the gap version of Okun’s law, which has historically seen measured productivity rising as the gap closes. That is, when aggregate demand increases at a time of high unemployment, the percentage increase in output exceeds that in employment, a form of increasing returns.

Government policies, like unemployment insurance, affect the workings of the labor market in normal economic times and in times of high unemployment.\textsuperscript{[22]} The distinction matters because the ratio of vacancies to unemployment, $v/u$, is different at different times and because of the role of insurance in supporting aggregate demand at times of high unemployment. While policies can be modified for times of high unemployment, automatically or by new legislation, it is still useful to consider diverse circumstances when setting general policies—their effects will come into play more quickly than triggered changes or legislated ones. That is, modeling policies that apply to a range of diverse states of nature should be able to improve policy analyses.

Of course a model with multiple equilibria is an incomplete model. Whatever determines which equilibrium occurs is simply outside what is being modeled. In the model, coordinated expectations select the equilibrium, and poor outcomes have been dubbed a “coordination failure.” Actual expectations don’t get fully coordinated, but the distribution of expectations, proxied perhaps by the degree of confidence about the future of the economy, is critical for the economy and fits with the

\textsuperscript{[21]}Moreover, with only self-employment, the division of output between employer and employee also plays no role.

\textsuperscript{[22]}I raised the issue of policies that affect both allocation and stabilization in 1994a. Another example is that marginal tax rates are part of the equity-efficiency trade-off in normal times and part of built-in stabilizers for times of high unemployment.
insights of this modeling. Confidence about the future is primarily confidence about the behavior of others and so reflects coordination issues.

A model with multiple equilibria can be used as a starting place in different ways, depending on how the missing elements are filled in. One approach is through sunspot equilibria, with the economy coordinating on different equilibrium structures for no real reason. While useful for understanding the potentials in a model, this does not seem to me a sound way to think about addressing the incompleteness inherent in multiple equilibria. Rather, I see two potentially useful approaches. One is to incorporate an empirically based evolution of the distribution of expectations. Another is to recognize potentially large responses to shocks by considering alternative rational expectations equilibria, rather than modeling the development of expectations as the economic environment (slowly) changes. While history is a prime ingredient in considering how the economy will go forward from a given position, it is not the case that we can draw on sufficient data to view parameters (or outcomes) of the economy as following a stochastic process that can be reasonably well estimated—hence the basic insufficiency, by themselves, of rational expectations analyses. Each circumstance is somewhat different in detail from similar episodes in the past, and some, like the current circumstances, considerably so.23

B. Sticky Wages

As noted above, my 1982b paper expressly omitted any role for sticky prices or wages to isolate a role for search frictions in the output market. Yet wages on most jobs are sticky—they are adjusted infrequently and so do not respond quickly to changing demand and supply conditions.24 Little use is made of wages that are conditional on aggregate data, with the exception of inflation indexing in economies with a history of high inflation. And I find Truman F. Bewley (1999) convincing that layoffs are often used rather than wage renegotiations or reductions in hours—such layoffs are consistent with a spike in layoffs early in a recession. With my continuing belief in the importance of sticky wages, I turn to that subject.

Starting from the standard demand-supply, price-clears-the-market framework, a natural step for considering unemployment is to have a wage that does not clear the market, one that does not adjust adequately and so can be too high, a sticky wage. This results in workers who are willing to start employment at the going wage, but are unable to do so. Much literature has been devoted to analyzing market models where prices and/or wages do not adjust at a rate sufficient to continuously clear markets. In addition, there has been considerable attention to issues that arise in making efficient use of current employees and how that affects both layoffs and compensation for new hires. Efficiency-wage and insider-outsider models recognize the important link between intensive and extensive margins for employment—between making efficient use of the current labor force and changing the number of employees.

---

23 And rational expectations ignores the systematic errors in stochastic thinking that are the focus of behavioral economics as well as differences in prior beliefs, which contribute to diversity in expectations at a point of time.

24 Prices and wages do change frequently in economies with hyperinflation.
Consideration of wages in a search setting is more complicated than the simplest demand-supply approach, reflecting two issues—first, that hiring is the start of a multiperiod relationship and, second, that an explicitly modeled search-matching-bargaining process influences wages. The standard basic model assumes that a decision to begin an employment relationship is based on an (implicit?) agreement for the indefinite future. It assumes that subjective probabilities are the same for worker and firm and that there is no asymmetric information during the ongoing relationship—both sides anticipate the same (stochastic) future and both sides recognize the same reality as it occurs. It assumes that all later decisions are consistent with this shared expectation—there are no individual actions that violate the common expectations and there are no surprises, such as a state of nature that was not anticipated or not covered in the initial agreement because of limits in the complexity of changes that are incorporated and the extent of renegotiation.

Since both sides are risk neutral and have the same discount rate, the actual trajectory of the flow of compensation is not determined, as long the expected present discounted value is at the right level. The level of the expected present discounted value of compensation satisfies the Nash bargaining solution with threat points of not completing the match (returning to the states of unemployed and vacant), and with some assumed parameter for relative shares in the Nash bargaining division of the surplus from starting production. This condition is how the search-matching-bargaining process influences wages. Wage determination affects the value of creating a new vacancy, which equals zero with free entry and identical potential new vacancies (a horizontal supply). Of the many conditional wage paths that are consistent with the Nash bargaining rule, it is common to analyze the one that is equivalent to continuously renegotiated wages (satisfying the Nash bargaining condition with the same structure) and a specified stochastic process for the value of output, and so the timing of the related ending of the employment relation. There are assumed to be no feedback effects on the economy from the decision to start employment or from the timing of the payment of wages.

The standard model has the simplification that each firm has at most a single vacancy, so there is no tension between the intensive and extensive margins and no distinction between average and marginal products of a worker. The efficiency assumed in the expectation about future employment includes only actions that are jointly optimal—no quits or layoffs that are not an improvement for the two parties together (a concern examined in Mortensen 1978). This latter condition is where sticky wages, in the traditional sense, are ruled out as part of the basic model.

Given this structure, if one picks a dataset for the value of output (and other key variables) one can compare the ratio of vacancies to unemployment \(v/u\) in a calibrated model with historic data. This is what Robert Shimer (2005) does. His conclusion that the model shows too little variation in \(v/u\) set off a flurry of activity. Responses include reexamining the calibrations (Marcus Hagedorn and Iourii Manovskii 2008; Mortensen and Éva Nagypál 2007 and the papers cited therein), exploring alternative bargaining solutions with an eye on having a less variable wage (Hall 2005; Hall and Paul R. Milgrom 2008; Shimer 2004), and staggered renegotiation with the current wage applying to new hires as well as the current labor force (Mortensen and Nagypál 2007; Mark Gertler and Antonella Trigari 2009). Without
getting into the details of this lively literature, I want to make a few methodological points about thinking through the lens of these models.

Shimer’s calibration uses output per worker for the marginal value of output. With a Walrasian output market and one-worker firms, this approach makes sense. Recognizing larger firms calls for distinguishing between average and marginal products. A change to a monopolistic competitive output market would give similar dynamics if there were no systematic pattern to the elasticity of demand over the business cycle, but not otherwise. However, with a different approach to modeling both the ability to sell and the price in the event of a sale, the marginal value of output could be quite different. As an extreme alternative, consider my 1971 model of a retail market described above. If the shock is to the number of customers shopping, but not the demand curve of those who do shop, then an integrated producer-retailer will not change the retail price, and with no change in the technology being used, will want to produce less, and will still have the same measured productivity. While the value to the employer of the marginal unit produced would not change, there would be layoffs of extramarginal workers for whom the value of output is zero if there are no storage possibilities. While this model is extreme, I think it is not adequate to use productivity as a measure of the marginal value of output. Shimer’s analysis rejects the match to the data of the calibration using the combination of assumptions employed; the follow-up literature has explored which one or ones are suspect.

Search frictions in the standard model imply that there is a surplus to be divided. Thus, Shimer (2004, 2005) and Hall (2005) note that a deviation from Nash bargaining could select a different level of compensation that is within the range where both prospective employer and prospective employee are willing to make a match. Such a different bargaining solution would imply that no pairwise efficient match was passed up (or existing efficient match terminated), a condition Shimer and Hall want to meet. Such a changed wage rule could result in a wage that is less sensitive to productivity than the Nash solution, and so would be one way to make the calibration fit the data better. The lower sensitivity to productivity changes than the Nash solution has been taken as a definition of a sticky wage. By affecting the compensation of new hires as well as current employees, an altered wage bargaining rule can affect the value of creating a vacancy, even though it does not directly affect the willingness of a given match to result in a new hire.

In the discussion of alternative bargaining rules, there appears to be wide acceptance of the Robert J. Barro (1977) stricture that a model should not have “an inefficiency that intelligent actors could easily avoid,” and should not be “invoking unexplained inefficiencies in economic arrangements” (Hall 2005, pp. 51, 56.) The Barro view is similar to the view expressed in Frank H. Hahn (1973), “that it is a mistake to import unexplained second-best constraints into a model which leaves no room for their justification.”

I disagree with these views because they ignore the incompleteness of models and the role of simplification for tractability. For simplicity, many search

25 Of course, one could pretend to derive constraints by invoking an underlying consideration that leads directly to a restriction one might have simply imposed. An example in some tax analyses is to invoke an observability constraint to restrict the available choices of tax base. Even though all candidate bases have some cost and some inaccuracy, they are labeled either costlessly and perfectly available or simply technically unavailable.
models have one-employee firms to simplify the analysis. Yet employment is overwhelmingly in firms with two or more employees. Are we going to learn more from one-employee modeling by invoking considerations that seem plausible in a literal one-employee environment or from involving considerations that seem plausible in many-employee firms and applying them to the one-employee environment? It seems to me that the latter is more likely to yield useful insights. And the alternative of requiring analysis with many-employee firms will yield some new insights, but may not yield additional insights for some questions in return for the extra complexity (e.g., tracking the distribution of firm sizes); indeed, it may make it harder to find some types of insights. Using plausible constraints seems to me to be in the spirit of the Marshall quote in the introduction. Model simplification is done as part of “breaking up a complex question, studying one bit at a time, and at last combining … partial solutions.” It seems to me this works best when the “exact and firm handling of a narrow issue” is done in a way that fits with the image of the “complex question” being addressed.

In many-employee firms, relations among employees are generally critical for productivity. And uniform compensation for similar workers in similar jobs is standard practice (and underlies the approach of Gertler and Trigari). Introducing a two-tier wage structure happens, but is not common and is a big deal for a firm. Thus there may not be a wage that is within the bargaining range for the full set of existing employees without some layoffs. And if there is one, it may be too high for the firm when considering new hires. If paying new workers less than current ones is not overall efficient, then a wage restriction that blocks a pairwise efficient hiring may not be “one that intelligent actors can easily avoid.” So, if you think that consistency of treatment of all workers is a real constraint in many-worker firms, then it seems right to import some implications of such a constraint while using the simplification of one-employee firms. I think that model tractability sometimes makes it appropriate to assume rather than derive plausible conditions when one thinks the two approaches would lead to the same central conclusion, even though, of course, some other conclusions would not carry over.

As in the retail market discussion above, consistency of treatment is a link between the intensive and extensive margins. Obtaining the right hours and effort from existing workers and attracting additional workers are separate concerns of an employer, both typically addressed by the same wage policy. It seems likely that as with optimal tax theory, recognition of the separate roles of intensive and extensive margins would change implications compared to a model that recognizes only one of them. I am not sure that these considerations, simple contracts and infrequent renegotiation, matter a great deal for search modeling of “normal” times, but I suspect they matter greatly for analysis of both output and labor markets in times of high unemployment.

---

26 A similar issue arises in optimal taxation—whether to formally model the complications that come from high complexity of a tax structure when selecting a tax base, or more simply to merely rule out some complex tax structures before optimizing. For more discussion of methodology see James Banks and Diamond.

27 Also relevant, as identified in the literature on asymmetric information and the assignment of decision making, is the inability to settle the arrangements for the rest of the employment period at the start.


29 While written with an eye on unions, some of the considerations in Hall and David M. Lilien (1979) seem to me relevant in nonunion settings as well.
The literature responding to Shimer explores the implications for cyclic sensitivity in alternatives to the Nash bargaining solution with threat points calculated from not having a match. Hall and Milgrom (2008) draw on Ken Binmore, Ariel Ruben, and Asher Wolinsky (1986), which considers bargaining with threat points based on delay in the start of employment rather than giving up the match. Hall and Milgrom argue that this is more appropriate for continuing an existing match and show that this improves the Shimer calibration. For this to work for vacancy creation as well as layoffs, the rule needs to apply to new hires as well. Note that this preserves the basic assumption that firms and workers negotiate with one person at a time. However, at some times, some workers have multiple offers, setting off a competition, perhaps captured by the Bertrand solution (James Albrecht, Pieter A. Gautier, Susan Vroman 2006). The large role of currently employed workers in filling vacancies makes this issue of prime importance (Nagypal 2008; Fabien Postel-Vinay and Jean-Marc Robin 2002). And sometimes, some firms have multiple qualified applicants as well. The mix of firms with multiple applicants (and so a lower wage) and workers with multiple offers (and so a higher wage) varies over the cycle. And sometimes wages are posted on a take-it-or-leave-it basis with no bargaining available, or no bargaining attempted, as shown in the survey in Hall and Krueger (2010). Bargaining circumstances vary across firms and workers so that there is not a single bargaining rule that can be viewed as dominating the landscape of circumstances—treat all bargaining by a single rule does not do justice to the diversity in the economy. This suggests the need to rely on multiple models and not give excessive weight to one particular calibration when thinking about the economy.

As noted above, a lower sensitivity to productivity changes than the Nash solution has been taken as a definition of a sticky wage. This is a change in definition from wages that do interfere with the efficiency of hiring and layoffs to wages that are less sensitive to productivity than the usual Nash bargain. Using the same term for two different concepts seems inappropriate and should not result in losing sight of the importance of layoffs and limited hiring because of intensive-extensive margin concerns. The impact of alternative wage bargaining rules on the value of vacancies is important, whether called a sticky wage or not, and well worth exploring. One thrust of the discussion above is that reaching a conclusion on the cyclicalilty of wages for new hires is not readily done by competing calibrated models of the entire economy, since there are too many plausible candidates for different pieces of the model. That is, actual bargaining circumstances are diverse. Models containing just one of the set of different bargaining outcomes that exist can be jointly informative, but settling on just one for answering empirical questions seems unsatisfactory. Thus it is natural to try to get data shedding light directly on the reservation wages of marginal vacancies, which is the key variable to compare with either reservation wages (for new hires) or actual wages (for layoffs).

Possible sources of transaction information that might shed light on firm reservation wages are the wages of people actually hired, drawn from employer data.

---

30 For models with multiple simultaneous offers, see Matthew Elliott (2011) and the papers cited therein.
31 For models with bargaining with asymmetric information, see Björn Brügemann and Moscarini (2010) and the papers cited therein.
or drawn from worker data. The problem is that there is a serious selection issue. Comparing the wages of movers and stayers from worker data suffers from this particularly, since the mix of quits and layoffs varies so much over the business cycle. Since we expect quitters often move to higher wages and we expect layoffs often to be followed by a drop in wages, merely comparing average wage change numbers of movers and stayers does not seem informative, plausibly being overwhelmed by the selection issue, even if made conditional on some aspects of the workers.\textsuperscript{32} That is, not all hires are to marginal vacancies, once we recognize heterogeneity in productivity, as is in more complex models. Quits directly to new employment are likely to be to nonmarginal vacancies, and those will sometimes be created by other quits.

Looking at the wages of new hires from firm data seems potentially more promising, and might allow a more forward looking calculation than just the current wage. One would want to compare the compensation of new hires at different times for the central issue—contrasting new hires and current employees could shed light on other issues. I note that there is a literature looking at data from the firm side (Marianna Kudlyak 2010). Particularly interesting in the literature discussed there is the link of later pay to unemployment when hired.

My retail market analyses, in 1971 and later, assumed prices on a take-it-or-leave-it basis, knowledge of the distribution of prices, but no prior knowledge of which price was at which supplier. A central finding was that prices exceeded what would be efficient. My labor market analyses preserved random search but assumed bargaining. They focused on the efficiency of the search process per se, and not the efficiency of the wages that resulted. There is a literature on directed search in the labor market, which is focused on wage determination (Espen R. Moen 1997). Some of the directed search literature, like the earlier Robert E. Lucas Jr. and Edward C. Prescott (1974) modeling with separate labor markets, assumes perfect knowledge about the different places to search, although the outcomes after having made such a choice remain stochastic. Simplification is critical for insightful modeling. But I think it is also important not to lose sight of the multiple elements that people need to use time and resources to learn and the multiple ways people do get some information, sometimes costlessly.

Analyses of sticky wages naturally consider the intensive-extensive margin in employment and the nature of an ongoing employment relationship. That such an approach is incomplete is suggested by the market for houses. The business cycle shows up in the market for houses in variation in the length of time from putting houses on the market to completing sales. Thought of as sticky prices in the house market, this does not involve an ongoing relationship nor a similar intensive-extensive margin.

### IV. Beveridge Curve

The Beveridge curve shows the pattern of vacancies and unemployment over time. In economically good times we expect lots of vacancies and low unemployment, with bad times showing fewer vacancies and more unemployment. In the

\textsuperscript{32} Also relevant is the varying mix of job opportunities over the business cycle (Wayne Vroman 1977, 1978).
course of a business cycle, a movement from good times to bad and back again, we expect to see a loop around a curve, as shows up in a differential equation setting of a basic search model.\(^3\)

Figure 4 shows the empirical Beveridge curve for the United States for the decade up to August 2010, with the open circles, connected by lines, for 2008–2010. Until the last 12 months in the figure, you can see the expected pattern of a recession, as vacancies shrink and unemployment rises, moving southeast roughly along a curve. Since then, we have had a rise in vacancies without a fall in unemployment.\(^4\) With rising vacancies and stable high unemployment, we are hearing claims that the United States has just had a leap in structural unemployment—that the economy may have a long-term higher level of unemployment as the “new normal.” This inference is taken to imply that we should not be so concerned with stimulating aggregate demand through monetary and fiscal policies. For example, here is an August 17, 2010, statement by Narayana Kocherlakota, President of the Minneapolis Federal Reserve Bank:

*What does this change in the relationship between job openings and unemployment connote? In a word, mismatch. Firms have jobs, but can’t find*

---

\(^3\) See, for example, Figure 3 in Blanchard and Diamond (1989), which shows the dynamic path moving from a steady state with poor opportunities to a steady state with good ones.

\(^4\) Taking a longer time horizon, the curve around which a business cycle moves has shown shifts over time.
appropriate workers. The workers want to work, but can’t find appropriate jobs. There are many possible sources of mismatch—geography, skills, demography—and they are probably all at work. Whatever the source, though, it is hard to see how the Fed can do much to cure this problem. Monetary stimulus has provided conditions so that manufacturing plants want to hire new workers. But the Fed does not have a means to transform construction workers into manufacturing workers.

This statement has set off a flurry of reactions, to which I will add.

There is no surprise that we are hearing claims of higher structural unemployment—such statements appear when unemployment is high. A similar debate unfolded as I was a new student of economics. And in 1964, Bob Solow devoted his Wicksell Lectures to rebutting claims that the high unemployment in the late 1950s and early 1960s was structural rather than a result of inadequate aggregate demand. Indeed, there is a long history of claims that the latest technological or structural developments make for a new long-term high level of unemployment, but these have repeatedly been proven wrong (Gregory R. Woirol 1996).35

It is likely to be more informative to think about the state of the labor market by focusing on the matching function, relating hires to unemployment and vacancies, rather than the Beveridge curve, which considers only the latter two. The natural interpretation is that the Beveridge curve movements would appear as a decrease in the efficiency of matching workers and jobs. Figure 5 shows the ratio of hires

35 For example, consider this 1931 statement: “The real issue is not whether technological displacement causes workers to lose their jobs. It undoubtedly does. The real issue is whether over a period of years the continual introduction of new and improved machines and processes is causing a total net increase or decrease in mass employment. … On this issue there are two opposing points of view, each held by large numbers of earnest people,” US Senate, Select Committee on Unemployment insurance, Unemployment Insurance, Part 2, “Report of the Committee on Technological Unemployment to the Secretary of Labor,” November 1931, 72nd Congress, 1st Session, 1931, 560, cited in Woirol (1996, p. 36).
to vacancies over the last decade. Consistent with the picture from the Beveridge curve, over the last year we have had a drop in the rate of hiring relative to vacancies, even though unemployment has stayed steady—a drop in the level of the matching function.

Note Kocherlakota’s statement that “Firms have jobs, but can’t find appropriate workers. The workers want to work, but can’t find appropriate jobs.” This is a static view of the labor market that does not make sense when thinking of the millions of hires that happen each month. While many workers, far too many workers, remain unemployed for a long time, many workers are finding jobs and many vacancies are being filled. Figure 1 shows that over the last two decades, on average, 37 percent of unemployed found employment each month. That percentage has dropped, but is still roughly 20 percent. Moreover, the large increase in the number of unemployed roughly offsets the fall in the exit rate, leaving monthly hires at a similar level to before—for the 12 months from November 2009 to October 2010, 5.7 million workers found a job per month, not hugely different from the 6 million average over the last 20 years. So we still have to think about large flows into and out of employment.

The matching function is not a technologically given structural relationship. Rather it is a reflection at the aggregate level of a complex and varied pattern of hiring at the level of individual employers and workers. Thus it is useful to examine some of the details at a less aggregative level to see how the current slump might be affecting the aggregate relationship—empirically we are outside the range of values of the ratio of vacancies to unemployment that were used in most estimates of the matching function. That is, a key question for interpreting the data in this recession and recovery compared to earlier ones is how the pattern of hires, unemployment, and vacancies is different in recessions of different sizes and also different because of specific events, such as the large and continuing issues in both banking and housing markets.

The severity of the current recession in both depth and length has resulted in a great deal of long-term unemployment. Figure 6 shows the distributions of unemployment durations as of October 2010 and a year earlier, October 2009, before the Beveridge curve started moving vertically. The low vacancies we experienced raised
long-term unemployment. In addition, we have had extended unemployment benefits. Such benefits somewhat reduce job search efforts and also discourage movement out of the labor force. Any lowering of the job search effort of the long-term unemployed is not likely to have much effect on aggregate unemployment, as there are many other workers who are seeking jobs and relatively few vacancies. And a reduction in the flow of unemployed out of the labor force increases measured unemployment, while having little effect on hiring, both because of the large numbers of remaining unemployed and because those outside the labor force take jobs as well. Reducing the flow of unemployed out of the labor force shifts the Beveridge curve up and the measured matching function down as hires divided by the number of unemployed is lower because the denominator is higher.

Long-term unemployment is very hard on the workers experiencing it and on their families. Moreover, over time, extended durations of unemployment affect behavior—the long-term unemployed are less good at maintaining their connection to employment and so we may have a slower-responding labor force after the economy grows significantly, which may be relevant for inflation concerns once we are nearing full employment, but not now. The deleterious effects of long-term unemployment are a reason to be particularly concerned about how long the economy does badly. Historically, recovery is slow after financial crises. The impact of a slow recovery on the long-term unemployed emphasizes the importance of stimulating aggregate demand enough to speed up recovery. And it emphasizes the importance of experimenting with programs to help the long-term unemployed find and hold jobs.

Just as measured unemployment does not fully reflect the availability of workers to be hired, so too the measured level of vacancies does not fully reflect the availability of jobs. Some hiring is done by firms that do not have measured vacancies, with some of these happening at firms that hire without posted vacancies, and some at firms that fill posted vacancies too quickly to be picked up in the data. John Haltiwanger provided me an estimate that about 40 percent of hires in the raw data are associated with establishments that begin the month with zero vacancies, with an estimated two-thirds due to the timing issue and the rest due to hiring without posting. Thus, measurement of the aggregate matching function may well vary with shifts in the makeup of hiring.

On a cross-section basis, the speed of hiring varies widely in systematic ways (Davis, Faberman, and Haltiwanger 2010a). There are large differences across industries, with construction having a very high ratio of hires to vacancies, compared to industries like education and health. While generally cyclically sensitive, construction has been particularly hard hit this recession, which would lower the measured efficiency of the matching function compared with a time with a smaller relative impact on construction. Establishments that are growing fast fill vacancies much more quickly than those growing more slowly. I do not know of data, but there may be a larger change in the mix of vacancies at fast- and slow-growing firms in this slump compared with smaller and less prolonged periods of high unemployment. Small firms are much more likely to hire without measured vacancies than

---

36 If someone constructed a measure of those outside the labor force who have a significant probability of becoming employed, one could analyze a three-argument matching function.
large firms. Giuseppe Moscarini and Postel-Vinay (2010) report on the relative roles of large and small firms over the business cycle. Moscarini reports that gross job creation by large firms minus that by small firms has been unusually high, more so than in the other recoveries since 1980. This is consistent with a differential impact of credit market changes on firms of different size that seems to be happening. The drop in house values has also affected the ability of small firms to finance hiring by borrowing against the houses of the firm owners. The resulting smaller share of hiring by small firms lowers the measured matching function. Some types of positions are filled much more readily and rapidly than others. Hall (2010) has suggested that positions that a firm wants to fill after a quit are filled more quickly than newly created positions because quits are most likely to occur in high-turnover jobs with low and generic skills, such as fast-food restaurants. And of course quits are way down, possibly reducing the average speed of filling jobs.

A key question for interpreting the pattern of aggregate unemployment and vacancy rates in this recession and recovery compared to earlier ones is whether the prime difference is in a changed difficulty of hiring at the disaggregated level or from a changed mix of diverse, but basically unchanged, hiring patterns across different firms and sectors, given that this is such a large and prolonged slump and with large and continuing issues in both the capital and housing markets. Complementing this analysis of hiring on a disaggregated basis is consideration of what Kocherlakota’s assertion would suggest might be found. Is there really a widespread difficulty in hiring in some industries or locations? I have not seen such reports. Thus we may be having shifts in the Beveridge curve and the matching function that do not signal change for the underlying functioning of the economy once a recovery is well established. That is, the pattern would return to normal after a sufficient rise in aggregate demand, apart from the lingering effects of long-term unemployment.

Having looked at the data, let me now look at possible policy inferences from whatever shifts may still be there. First, whatever one’s view on the magnitude of recent slippage in matching efficiency, more education, better education, good retraining all make for a more productive labor force and, done well at a reasonable cost, are policies to pursue. And carefully evaluated experiments in helping

---

37 Bernanke (2010, p. 4) notes: “The availability of credit to finance investment and expand business operations remains quite uneven: Generally speaking, large firms in good financial condition can obtain credit in capital markets easily and on favorable terms. Larger firms also hold considerable amounts of cash on their balance sheets. By contrast, surveys and anecdotes indicate that bank-dependent smaller firms continue to face significantly greater problems in obtaining credit, reflecting in part weaker balance sheets and income prospects that limit their ability to qualify for loans as well as tight lending standards and terms on the part of banks.”

38 This is consistent with the finding in Davis, Faberman, and Haltiwanger (2010a) that the job-filling rate rises with the worker turnover rate.

39 William Dickens (2010, p. 10) notes: “Figure 3 presents the ratio of vacancies to unemployment in eight different industries. While it is possible to discern the increase in vacancies over recent months in some industries, the ratio remains substantially depressed in all industries. What we do not see is any industries with high vacancy unemployment ratios. This suggests that it would be hard to make a case for structural mismatch being a major problem today.” In personal correspondence, he reports on ongoing work with Bob Trieste that looked at geographic mismatch indices based on both the JOLTS and the Conference Board’s new help wanted online data. They explored occupational mismatch, geographic mismatch at a much more detailed level than in the original paper, and industry mismatch. None shows any evidence of increasing mismatch coincident with the apparent outward shift in the Beveridge curve. Regis Barnichon et al. (2010, p. 1) “decompose the recent deviation from the Beveridge curve … [and] find that most of the current deviation from the Beveridge curve can be attributed to a shortfall in … hires per vacancy. This shortfall is broad-based across all industries and is particularly pronounced in construction, transportation, trade, and utilities, and leisure and hospitality. Construction alone accounts for more than a third of the Beveridge curve gap.”
the long-term unemployed get and hold jobs seem likely to be worthwhile. Indeed, a time of high unemployment is likely to be a time when further education is less socially costly by using time that would otherwise not be so well spent. The policy debate is not about whether to do more on the structural side; it is about what to do on the aggregate demand side, which is particularly an issue now with concern about projected long-run debt levels.

Second, for the current moment, the argument about the aggregate demand side is academic, in the negative sense of the word. Current estimates I have seen of how much of the increase in unemployment from a few years ago is “structural,” rather than due to inadequate aggregate demand, still leaves enough need for aggregate demand stimulation that it is clear what direction is needed for further policies.

Third, I am skeptical of the value of attempting to separate cyclical from structural unemployment over a business cycle. When firms evaluate candidates for positions, they consider the quality of the match of available candidates, projections of the availability of new candidates, and the value to the firm of filling the slot. That is, the willingness to hire for a given quality of match depends on expectations about the profitability of investing in a new worker and about the likely pool of future applicants.

The tighter the labor market and the more valuable the filling of a vacancy, the more a firm is willing to hire a worker who is a less good match and who may need more training. In other words, a worker who might be viewed as structurally unemployed, as facing serious mismatch in the current state of the economy, may be readily employable in a tight labor market. The common practice of thinking about the extent of unemployment as a sum of frictional, structural, and cyclical parts misses the point that the tightness of the labor market affects worker quitting decisions and affects employers’ willingness to hire an applicant who needs more training. In so far as direct measures of frictional or structural unemployment are dependent on the tightness of the labor market, they have limited relevance for the design of demand stimulation policies. The idea that the US economy is not adaptable and capable of dealing with the need for skills and jobs to adapt to each other is peculiar, given the long history of unemployment going up and down.40

When the labor market is tight and firms have trouble finding workers, they reach out to places they have not looked before and extend training in order to find workers who can fill their needs. Supporting current stimulus policies as very good for the economy is entirely compatible with taking care to avoid future inflation.

V. Concluding Remarks

Having been away from this topic for a long time, I was surprised during my crash course in search analysis this fall at what a long way search has come since its early days. Without the high quality work using a search-based approach of many researchers, it is safe to say there would not have been a prize recognizing

---

40 A basically static perception of the economy that fits with a perception of structural unemployment that is not helped by aggregate demand policies is akin to the badly mistaken idea that policies that encourage a high rate of early retirement will be helpful for youth unemployment on a sustained basis (Diamond 2006; Jonathan Gruber and David A. Wise 2010).
the analysis of markets with search frictions. And the work is far from done. For addressing unemployment, there are clear needs to incorporate credit markets and (non-Walrasian) output markets and to include nominal thinking and nominal contracting as well as a larger role for current income. Filling such needs would better place partial equilibrium search analyses of the labor market in a full general equilibrium setting. Indeed, this essay has stressed the importance of not treating a partial equilibrium model as if it were a satisfactory general equilibrium model.

More inclusive modeling aside, I want to reiterate the perspective of Marshall quoted at the start. Understanding of the economy, and policy recommendations and decisions, should reflect analysis through multiple models. And they should incorporate insights that seem right even though they have not yet been modeled.

REFERENCES


Dickens, William. 2010. “Has the Recession Increased the NAIRU?” Paper presented to the Federal Reserve Board of Governors, academic advisory meeting, October 12.


Shimer, Robert. 2007. “Reassessing the Ins and Outs of Unemployment.” http://sites.google.com/site/shimerresearch/flows


