2013

Conceptual Basis for Identifying and Measuring Occupational Labor Shortages

Burt S. Barnow
George Washington University

John Trutko
Captial Research Corporation

Jaclyn Schede Piatak
American University

Citation

This title is brought to you by the Upjohn Institute. For more information, please contact repository@upjohn.org.
Three general issues regarding labor shortages have been widely discussed by economists and policymakers. First, because of recent declines in the U.S. birthrate, some analysts are concerned that there will simply be too few workers to maintain growth in the American economy (Levitan and Gallo 1989). Second, there has been a growing concern that there is or will be a serious mismatch between the skills of the American labor force and the needs of employers, resulting in a serious “skill gap” characterized by unfilled vacancies in many high-skill occupations along with high unemployment for less-skilled workers.1 Finally, there has long been concern that shortages sometimes develop and persist in specific occupations, leading to inefficiencies in the U.S. economy. This third topic, occupation-specific shortages, is the primary subject of this study.

Before turning to the major topics of the chapter, it is instructive to consider why it is important to study occupational labor shortages. First, occupational labor shortages are of interest to educational and training institutions and their students because this information helps students choose fields of instruction where they are most likely to obtain jobs and where wage rates are likely to rise. Information on the tightness of various occupational labor markets is useful for counselors and students alike, enabling them to improve guidance and decision making. As discussed in Chapter 6, many states and local workforce investment areas are voting with their pocketbook to purchase “real-time labor market information” that provides detailed information on local labor market conditions for specific occupations.

At the national level, documenting current occupational labor shortages and projecting future shortages can help the nation identify and mitigate pipeline issues that impede the growth of some occupations. Are there enough training institutions in the country to eliminate...
current or projected shortages? Is the supply of instructors adequate to train sufficient entrants to the occupation?

Another national policy area where understanding shortages is important is the development of immigration and temporary foreign labor visa policies. Shortages in important occupations, such as physicians and nurses, may lead to situations where U.S. residents are denied critical services such as health care. On the other hand, if too many foreign workers are admitted, the domestic labor force may suffer unemployment or wage reductions.

Occupational labor shortages are more frequently found in labor markets that depart significantly from a free market with open entry and exit. Government interventions in such markets include directly or indirectly regulating wage rates, as is the case for public school teachers and many health care workers, and restricting supply through licensing. Understanding how shortages can arise in such an environment can be useful in establishing appropriate payment schemes for services. Two of the occupations analyzed here, home health workers and special education teachers, are subject to strong government roles in setting wage rates, so understanding how inappropriate policies can lead to shortages or less than the desired number or quality of workers in the occupation can be quite valuable.

Finally, the concept of an occupational labor shortage is, at first glance, puzzling to economists and others who have learned that the “invisible hand” of the marketplace described by Adam Smith should lead to markets clearing. It is this puzzle that first led us to study occupational shortages, and we have found that there are reasonable explanations for why shortages arise in occupations we have analyzed. Indeed, as we describe in the remainder of this chapter, there are a number of reasons consistent with conventional economic theory for why occupations can experience a shortage. In addition, we find situations that do not fit the economist’s definition of a shortage but are problematic for the nation—the “social demand shortage” concept discussed later in this chapter.

In sum, labor markets do not always operate smoothly with supply and demand being equal at a given time. This book should help the reader understand why shortages arise, identify a shortage when present, and assess strategies to alleviate the shortage. As shown in the next section,
many economists, including several U.S. Nobel Prize winners, have studied occupational shortages, and this volume builds on their work.

This chapter begins with an overview of the concept of occupational labor shortages, focusing on alternative definitions that have been used to identify occupation-specific labor shortages. This is followed by a discussion of the causes and consequences of occupational labor shortages.

BACKGROUND ON LABOR SHORTAGES

The term “labor shortage” has no universally agreed upon definition. It sometimes refers to a shortfall in the total number of individuals in the labor force and sometimes denotes the possible mismatch between workers and jobs in the economy. Even when the term is used to refer to a particular occupation, several definitions have been proposed and used. In this report, the following definition of labor shortage is used: “A sustained market disequilibrium between supply and demand in which the quantity of workers demanded exceeds the supply available and willing to work at a particular wage and working conditions at a particular place and point in time.”

This definition considers a shortage as a disequilibrium condition where the amount of labor that workers are willing to supply is less than employers are willing to buy at the prevailing wage. A market is said to be in equilibrium when the amount of labor that workers (i.e., sellers) are willing to provide at the market price is equal to the amount that firms (i.e., purchasers) wish to buy at the market price. When the quantities that workers wish to provide and firms wish to buy are not identical at the prevailing price, the market is said to be in a disequilibrium situation.

If the quantity of labor offered exceeds the quantity that firms wish to purchase, there is a surplus, and if the quantity of labor desired by firms exceeds the amount workers offer at the prevailing price, there is a shortage. In general, the quantity that workers are willing to provide is an increasing function of the wages (i.e., price) they can obtain, and the relationship between wages and the amount that workers are willing
Figure 1.1 shows a typical upward-sloping supply curve for labor. As the wage rate is increased, more workers are willing to enter a particular occupation and current workers are generally willing to provide more labor.

In Figure 1.1, the amount of labor that employers will wish to hire at alternative prices is indicated by the downward-sloping demand curve. Demand curves slope down because as the price of a factor increases, the employer will generally substitute other factors of production for the factor whose price has increased. In addition, higher factor prices will generally lead to higher product prices, which in turn will lead to a reduction in the quantity of the product demanded and in the factors of production.

The point labeled E in Figure 1.1 is the market equilibrium point. If the wage is equal to \( W_E \), then the quantity of labor that workers are willing to supply at that wage (\( Q_E \)) is exactly equal to the quantity of labor that employers will wish to hire. The market is said to be in equilibrium because the quantity supplied is equal to the quantity demanded.

If, for some reason, the prevailing wage rate in the market is \( W_0 \) rather than \( W_E \), then the quantity of labor that workers are willing to supply is equal to \( Q_S \)—the point on the supply curve corresponding to \( W_0 \). Employers, however, would like to hire \( Q_D \) at that wage rate. The difference between the amount of labor that employers wish to hire and the amount that workers are willing to provide (\( Q_D - Q_S \)) is the amount of the shortage. In the next section, we discuss how such shortages might arise.

Economists and other analysts have proposed several alternative definitions of occupational shortages. Although these definitions are generally not used in this report, it is important to note that others use the term differently. It is particularly important to keep the definition in mind when interpreting other studies of shortages.

The Social Demand Model

Some analysts consider a shortage to be present if the number of workers in an occupation is fewer than what is considered the socially desired number. Under this definition, a shortage of engineers exists if the analyst making the determination concludes that society would be
better off if there were more engineers. This type of definition does not imply that the labor market is in disequilibrium; instead it describes a situation where the person who claims there is a shortage does not like the market’s results. Arrow and Capron (1959) explain the problem with this definition as follows:

In particular, careful reading of such statements indicates that the speakers have in effect been saying: There are not as many engineers and scientists as this nation should have in order to do all the things that need doing such as maintaining our rapid rate of technological progress, raising our standard of living, keeping us militarily strong, etc. In other words, they are saying that (in the economic sense) demand for technically skilled manpower ought to be greater than it is—it is really a shortage of demand for scientists and engineers that concerns them. (p. 307)

For example, in the late 1980s, the Secretary of Health and Human Services’ Commission on Nursing stated in an assessment of labor market conditions for registered nurses that “in the most general terms, a registered nurse [RN] shortage exists when the supply of RNs is insufficient to meet the ‘requirements’ for RNs. RN requirements can be defined based on either economic demand or clinical need” (U.S. Department of Health and Human Services [DHHS] 1988, p. 3). The
Commission rejected the use of clinical need for defining a shortage because it concluded that there is no objective method of quantifying the degree of the shortage and relating it to specific factors.

The fact that we do not use this type of definition for a shortage does not mean we believe that it is unimportant for the nation to consider whether it is satisfied with market-produced results. Quite the contrary, it is important for society to consider whether or not the market solutions are desirable and, if not, to take appropriate actions. One concern in this volume is with the operation of labor markets and the reasons why labor markets sometimes fail to achieve equilibrium, and actions that can be taken to improve their efficiency; however, in some of the occupations examined, there is evidence of a potential social demand shortage, and we identify the potential problems even if an economic type of shortage does not exist.

The Blank-Stigler Model

One of the first major studies of occupational shortages was conducted by David S. Blank and George J. Stigler (1957). Blank and Stigler define a shortage as follows: “A shortage exists when the number of workers available (the supply) increases less rapidly than the number demanded at the salaries paid in the recent past” (p. 24). Blank and Stigler then argue that to alleviate the shortage, wages in the occupation must rise, and some of the work formerly performed by the occupation with the shortage will now be performed by others.

The Blank-Stigler shortage concept is illustrated in Figure 1.2. Initially the market is in equilibrium at E with wage rate $W_E$ and $Q_E$ workers. If demand increases, the demand curve will shift to the right to line $D_1$. A shortage will result if the wage remains at $W_E$ because employers will wish to hire $Q_1$ workers—but only $Q_E$ workers will be available at that wage. Market pressures will then lead to an increase in the wage, and equilibrium will eventually be restored with a new wage of $W_2$ and with $Q_2$ workers.

There are several problems with the Blank-Stigler model. First, as discussed below, an increase in demand is only one possible cause of a shortage. Thus, the Blank-Stigler model ignores other possible causes of occupational shortages. Second, Blank and Stigler indicate that a shortage can be identified by rising wages in the affected occupation.
Wages may not rise, however, because of market imperfections such as wage controls or imperfect information.

The Arrow-Capron Dynamic Shortage Model

Arrow and Capron (1959) developed an alternative model of occupational shortages. Their definition, which they refer to as a dynamic shortage, is based on the premise that “a steady upward shift in the demand curve over a period of time will produce a shortage, that is, a situation in which there are unfilled vacancies in positions where salaries are the same as those currently being paid in others of the same type and quality” (p. 301).

The Arrow-Capron model is also illustrated in Figure 1.2. Like the Blank-Stigler model, the Arrow-Capron model is characterized by increased demand. However, Arrow and Capron note that markets are characterized by a “reaction speed,” and that institutional arrangements (such as long-term contracts) and the time it takes for information to spread will affect the time required for employers to adjust wages. Thus, Arrow and Capron conclude that shortages will be characterized by vacancies. In Figure 1.2, the number of vacancies initially result-
ing from the increase in demand will be equal to $Q_1 - Q_E$. If demand continues to grow, then the market may not achieve equilibrium. The Arrow-Capron dynamic shortage model is consistent with the general model used here, but it may be considered a specific case.

**Other Definitions of Shortages**

In addition to the definitions presented above, several other definitions for shortages have been proposed. Harrington and Sum (1984) review several other possible definitions of occupational labor shortages, and two of them are briefly discussed below.

**The rate of return model**

The “rate of return model” is based on the application of internal rate of return analysis to alternative occupations. The costs of investing in a particular occupation are defined as the sum of the direct costs for higher education, training, and supplies, plus the indirect costs of foregone wages that are incurred during periods of training. The benefits are the earnings typically derived from the occupation each year. The internal rate of return is then calculated by finding the interest rate that equates the present value of the costs and benefits. Occupations with shortages are thus defined as those occupations with higher than average rates of return.

Harrington and Sum note that the rate of return approach is “beset with numerous methodological and measurement difficulties.” One important problem is that we cannot observe future earnings streams from various occupations. Relying on cross-sectional or historical data may provide a misleading picture of what the earnings will eventually be. In addition, the returns on various occupations may differ for reasons having little to do with a shortage. For example, some occupations may pay higher wages because they have high health or safety risks—what economists refer to as compensating differentials.

**The monopsonistic labor model**

A market where there is only one buyer for a particular good or type of labor is referred to as a monopsony. The monopsonist differs from an employer in a competitive labor market because the monop-
sonist can set the wage rather than act as a price taker. The situation for a monopsonist is illustrated in Figure 1.3. Because the monopsonist is the only buyer for the occupation of interest, the monopsonist observes the labor supply curve for the occupation; this is in contrast to an employer in a competitive market who can hire all the labor desired at the market wage. Because the monopsonistic employer must pay all workers the same wage, the monopsonist faces a steep upward sloping marginal labor cost curve—if an additional worker is hired, wages must be increased for all currently employed workers as well as the marginal worker. Figure 1.3 also illustrates the marginal revenue product curve for the firm. To maximize profits, the monopsonist employer will hire labor until the marginal labor cost is equal to the marginal revenue product, corresponding to the point X in the figure. The wage paid by the monopsonist will then be $W_M$, and $Q_M$ workers will be hired. Note that the number of workers hired is fewer than in a competitive market ($Q_E$), and the wage is lower than the competitive wage ($W_E$).

The monopsonist might consider the resulting situation to be a shortage because the monopsonist would like to hire more workers at the monopsony wage. However, because the monopsonist faces an upward

**Figure 1.3 Illustration of Labor Demand by a Monopsonist**
sloping labor supply curve, the wage must be increased to attract additional labor into the occupation. Ehrenberg and Smith (2009) conclude that the labor shortage faced by a monopsonist is “more apparent than real” because the monopsonist is hiring the quantity of labor desired at the wage offered. In addition, Ehrenberg and Smith point out that monopsony situations are probably very rare.

**Concepts from Butz et al.**

Butz et al. (2003) describe five possible meanings of an occupational shortage:

1) if production is lower than in the recent past;
2) if competitors’ share of total production is growing;
3) if production is lower than what the people doing the producing would like;
4) if less is produced than the nation is deemed to need; and
5) if production is not meeting market demand, as indicated by a rising price.

Although all of these concepts may signify a problem from some perspective, only the final definition corresponds to the economic concept of an occupational shortage. Although the first three concepts listed by Butz et al. may well be cause for concern for a firm, they are more indicative of a shortfall in desired production than any type of occupational shortage. The fourth concept, producing less than the nation is deemed to need, corresponds to the social demand concept of a shortage introduced by Arrow and Capron (1959).

**Summary of Shortage Concepts**

Some of the labor shortage concepts that have been proposed, such as looking at the total amount of labor supplied and the potential economy-wide mismatch between employer needs and worker qualifications, are important, but they are not relevant to the study of occupational shortages. As the Secretary of Health and Human Services’ Commission on Nursing concluded, the social demand concept is difficult
to apply because there is no objective way of determining the optimal number of workers in various professions.

Definitions proposed in the 1950s by Blank and Stigler (1957) and by Arrow and Capron (1959) are closer to the concept of a labor shortage that is used here. The principal advantage of these definitions is that they provide relatively straightforward tests for the existence of a shortage—rising relative wages in the case of the Blank-Stigler definition and increasing vacancies in the case of the Arrow-Capron definition. However, these definitions are too narrow to capture all the types of shortages of interest. Both the Blank-Stigler and Arrow-Capron definitions do not include labor market situations classified as shortages by the other definition, and both omit situations where excess demand results from market imperfections. The more general definition employed here covers such cases.

The use of a broad definition does have some disadvantages. As Franke and Sobel (1970) note in using a similar definition, “The definition is neither altogether concrete and precise nor is it susceptible to precise measurements.” However, we also concur with Franke and Sobel’s conclusion that “viewed in the context of a study whose purpose is to examine the degree to which labor market institutions respond to and facilitate adjustment to varying degrees of labor market tightness, the definition is, however, meaningful and operational” (p. 7).

CAUSES OF LABOR SHORTAGES

There are several reasons why it is important to address the causes and consequences of labor shortages. First, because we have adopted a fairly broad definition of a shortage, we will have no single indicator that a shortage exists. By reviewing the economic theory of the causes and consequences of shortages, we will be aware of the appropriate market signals to look for in assessing whether or not a shortage exists. This is especially important because under certain conditions, various interest groups have incentives to argue that a shortage is present or absent. For example, employers and trade associations sometimes have an incentive to claim that there is a shortage to increase immigration
quotas for particular occupations, giving them access to a broader pool of applicants. At other times, employers might find it in their interest to claim there are no shortages, in order to gain better leverage in contract negotiations with their workforce.

Another important reason for analyzing the causes and consequences of shortages is to help identify and assess potential public and private policies for dealing with shortages. Being able to identify causes will help interested parties focus on the relevant developments in labor and product markets. Understanding the consequences will help us to assess what interventions, if any, are appropriate by government, employers, and workers.

Before turning to the causes and consequences of labor shortages, it is useful to note some of the dimensions of shortages:

- **Geographic scope of the shortage.** Depending on the occupation and the nature of the market, labor markets can be national or regional in scope. Similarly, a particular occupation may have a nationwide shortage, or the shortage may be confined to a few labor markets or a single region of the country (e.g., the shipyard industry on the Gulf Coast).

- **Longevity of the shortage.** As will be discussed below, various forces act to bring markets into and out of equilibrium. Thus, shortages can be relatively brief, lasting for a few weeks or months, or prolonged, lasting for one or more years.

- **Severity of the shortage.** Unlike the two dimensions discussed above, it is not easy to develop good measures of the severity of a shortage. Conceptually, we can measure the severity of a shortage in terms of the magnitude of the changes in wages required to restore equilibrium or in terms of the number of workers needed to alleviate the shortage. There are several difficulties with these concepts. First, we do not generally observe the supply and demand curves for specific occupations. Thus, we cannot directly estimate the size of the employment or wage gap of a shortage. Second, even if we could measure supply and demand, it would not be easy to classify a particular gap as large or small, especially when comparing across occupations—occupations vary significantly in their normal vacancy rates and wage dispersion. Thus, a high vacancy rate for one
occupation with a shortage may be characteristic of another occupation in equilibrium.

- **Subspecialty shortages.** Up to this point we have considered occupations as if they are uniform. For some occupations this may be correct, but for others there may be differentiation by subspecialty (e.g., emergency room nurses), years of work experience, or specialized training. In such cases, a shortage may exist for the entire occupation or only for workers with selected characteristics. For example, training for engineers has changed considerably over the past 20 years, and older electrical engineers may not be good substitutes for new engineers who have more training in designing integrated circuits. Likewise, new tool and die makers may not be good substitutes for experienced tool and die makers who have gained additional skills through their work. The key determinant of whether there can be shortages for some parts of an occupation is whether all workers within the occupation are reasonable substitutes for each other. If not, a shortage can exist within an occupation while other subcategories are in equilibrium or even in surplus. As will be shown below, subspecialty shortages can be difficult to document because labor market trends are only captured at a broader level. For example, the Standard Occupational Classification (SOC) system in the United States considers all computer programmers to be in a single occupation. However, when employers have an opening, it will be for a programmer who can write code in a particular language, such as HTML, and a COBOL programmer would not meet those needs. Thus, looking at data for all types of programmers combined could mask the presence of a shortage of a particular type of programmer.

For an occupation to have a shortage, two conditions are necessary. First, the occupation must be in disequilibrium, where the number of workers employers wish to hire exceeds the number willing to work at the prevailing wage. Second, the market must adjust slowly, if at all, with the achievement of equilibrium requiring a substantial period of time. We first discuss the reasons why markets are sometimes in disequilibrium. We then examine the adjustments that employers make to alleviate the disequilibrium, followed by a discussion of the reasons
why disequilibria may persist. We then discuss the consequences of prolonged shortages.

**REASONS WHY OCCUPATIONAL LABOR MARKETS ARE IN DISEQUILIBRIUM**

Labor shortages can result from a number of different causes. In this section, we discuss the reasons why the labor market for a particular occupation might leave an equilibrium situation where the market wage equates supply and demand.\(^7\)

**Increase in the Demand for Labor**

Figure 1.4 illustrates how a labor shortage can result from an increase in the demand for labor; several variants of this scenario were discussed previously. Suppose that the labor market is initially in equilibrium at point E. If the demand for labor increases, the demand curve will shift to the right. If the supply curve remains the same and the prevailing price (wage) remains at \(W_E\), employers would like to hire \(Q_D\) workers, but only \(Q_E\) will be available. Thus, there will be a shortage of \(Q_D - Q_E\) workers.

The demand for labor by employers can increase for several reasons. Perhaps the most likely reason for an increase in the demand for labor is an increase in the demand for the goods or services produced by employers (e.g., a substantial expansion in contracts to shipyards for construction and/or repair of ships). An increase in the demand for the product can result from an increase in the number of consumers, an increase in the income or wealth of consumers, a change in the composition of the population of buyers, or changes in the tastes of consumers.

Another potential reason for an increase in the demand for labor is an increase in the prices of substitute factors of production. For example, in a hospital the demand for nurses might increase if the wage rates of doctors or nurse’s aides increases. The demand for a given type of labor will also increase if the price of a nonlabor factor (e.g., raw materials or machinery) increases and the labor can be used as a substitute
Basis for Identifying and Measuring Occupational Labor Shortages

An increase in demand for labor in a particular occupation does not necessarily lead to a shortage. If the supply of labor to an occupation can respond to the increased demand, the result will be a new equilibrium with more workers employed and a higher wage rate than at the previous equilibrium, as is illustrated in Figure 1.4. An increase in demand will almost certainly require some time for the market to reach a new equilibrium, but if vacancies persist for a sustained period, the occupation can be characterized as experiencing a shortage. Reasons why occupational labor markets may adjust slowly are discussed below.

**Decrease in the Supply of Labor**

A decrease in the supply of labor to a particular market can also create a labor shortage. This situation is illustrated in Figure 1.5. Once again suppose that the market is originally in equilibrium at point E. If
the labor supply curve is shifted to the left, indicating fewer workers available at each wage rate, there will be a labor shortage if the prevailing wage remains at $W_E$. Employers will still be trying to hire $Q_E$ workers, but only $Q_S$ workers will be available after the supply decreases. Thus, there will be a shortage of $Q_E - Q_S$ workers.

The labor supply curve for the labor market in question might shift for several reasons. One potential cause is a decrease in the size of the population that works in the relevant jobs. For example, as the baby-boom generation has aged, employers who generally hire young people as they complete high school have suddenly faced a much smaller supply of entry-level workers from the so-called baby-bust generation, which has a much smaller population.

The supply curve might also shift to the left because wages in other occupations have risen, making employment in the market of interest less attractive, or because nonwork opportunities, such as welfare, crime, and retirement, have become more attractive. Finally, the labor supply curve for an occupation might shift to the left because of restrictions on entry into the relevant labor market. Such restrictions may be
implemented by the government (through licensing requirements and restrictions on the number of licenses granted), professional organizations that set standards for practice, labor unions, or training institutions (e.g., universities or community colleges).

**Restrictions on Prices**

Although most prices are determined competitively by markets in the United States, the price of labor or the price of the final product is regulated in some industries. For example, cities generally regulate the price that taxi drivers can charge. In such instances, the supply curve is truncated at the regulated price. This situation is illustrated in Figure 1.6. The wage rate is restricted to be no higher than $W_M$, so the supply curve at higher wages is indicated by a dashed line. The labor that will be supplied at that wage is $Q_S$. At that wage, however, the demand is for $Q_D$ workers, so there is a shortage of $Q_D - Q_S$ workers. An example of this type of shortage during some periods is the U.S. government’s market for entry-level PhD economists. The federal government traditionally hires entry-level economists at the GS-12 pay level, and agencies

---

**Figure 1.6 Illustration of Labor Shortage Arising from Restrictions on Wages**

Diagram showing the labor shortage caused by restrictions on wages, with the supply curve truncated at the regulated wage $W_M$, resulting in a shortage of workers $Q_D - Q_S$. The diagram includes a supply curve $S$, a demand curve $D$, and a wage rate $W_M$. The quantity of workers is indicated on the x-axis, and wages on the y-axis.
are generally not permitted to pay a higher wage rate. Sometimes the market wage for entry-level economists is higher, so there is sometimes a shortage of entry-level PhD economists in government agencies.

More commonly, the government regulates the prices of products and services rather than labor. In industries where labor comprises a relatively small share of the product’s price, such as in the generation of electric power, the product price regulation is not likely to cause a labor shortage. In very labor-intensive industries, however, output price regulation can be tantamount to regulating the price of labor. Examples include the health care industry in general and the home care industry in particular. A large share of the U.S. health industry is financed by the Medicare and Medicaid programs. In the case of Medicare, the federal government limits the reimbursements that providers can obtain for treating covered elderly patients. State governments provide similar regulation under Medicaid programs for the poor. Because the government restricts the charges that providers can make, the providers face limits on what they can pay workers and still cover their costs.

ADJUSTMENTS TO OCCUPATIONAL SHORTAGES BY EMPLOYERS

This section describes some of the actions that firms are likely to take to deal with labor imbalances. Employers will first note problems when they are unable to fill vacancies at current wage rates, and they will therefore likely take actions to deal with the unfilled positions. Some actions are less costly and more reversible than others, so employers are likely to undertake these actions first. The potential actions that employers can take are listed roughly in order of desirability from the employer’s point of view. In particular cases, of course, some of the potential actions may be inapplicable or employers may undertake the actions in a different order. Note that all the potential actions employers can take are the opposite of the actions that lead to vacancies: increasing supply, decreasing demand, and increasing wages. Finally, many of the actions described below may be undertaken by employers for reasons other than trying to fill vacant positions.
Increase Recruiting Efforts

A logical first step to fill vacancies is to increase recruiting efforts. Although employers will incur short-term costs in expanding recruiting, there are no long-term or permanent costs involved. Recruiting can be increased through several approaches:

- **Increase advertising in the usual outlets.** For example, employers who advertise in newspapers can increase the frequency of the advertisements or the size of the advertisements to attract more attention to vacancies.

- **Advertise in other media.** To reach a wider audience of potential employees, firms can expand their advertising campaigns. Firms that traditionally recruit through newspaper advertisements can add other newspapers in the community or increase the use of radio and television advertising. Use of job fairs is another such technique. Over the past decade, employers have increasingly advertised available positions over the Internet, using the company’s Web site, trade association Web sites, and other more generalized job search Web sites (e.g., Monster.com, CareerBuilder.com, and CareerLink.com).

- **Expand the recruiting area.** Employers who believe that the problem is local rather than regional or national can increase the geographical scope of their recruiting efforts. For example, a firm having difficulty recruiting shipfitters in Louisiana might expand its recruiting efforts to other nearby areas along the Gulf Coast (e.g., Texas, Mississippi, and Florida) and, if necessary, in other parts of the country or world. Some occupations, generally those with highly skilled jobs, already have national labor markets. For these firms, and for firms recruiting for occupations with a national shortage, the only way to increase the recruiting market is to recruit abroad. The extreme case of expanding the recruiting area is to recruit immigrants either for temporary or permanent jobs. Because the number and characteristics of immigrants that can be admitted is determined by law, individual employers may be limited in the extent to which they can make use of this option.
• **Use public and private employment agencies.** Firms that do not already do so can make use of public and private employment agencies. Public agencies, referred to as the employment service or job service, are free to both workers and employers. In some states, the employment service may tend to specialize in serving particular types of workers, but all employers can list their openings with the employment service. Private employment agencies charge a fee to either the worker or the firm, with the fee based either on the time spent by the agency or as a percentage of the hired worker’s salary.

• **Pay recruiting bonuses to employees who bring in new workers.** For many employers, current workers are often the best source of potential new hires. Employees are likely to be hesitant to recommend individuals who are unqualified, and the candidates they recommend are likely to know more about the work and working conditions than other job candidates. Thus, for many firms, current employees are a major source of job applicants. To encourage workers to assist in the recruiting process, employers sometimes offer a bonus for referring qualified applicants or applicants who are hired.

**Increase Use of Overtime**

A relatively simple solution to the problem of filling vacancies is to have current employees work more hours. Employers who anticipate that the problem will not last for a substantial period of time are likely to use this approach. If the workers are exempt from the overtime provisions of the Fair Labor Standards Act (FLSA) and do not receive a premium for hours in excess of 40 hours per week, overtime may actually save money relative to hiring additional workers. Even for non-exempt workers, overtime may be less expensive because many fringe benefits (such as health insurance, unemployment insurance, and workers compensation) are fixed per worker, and the firm will not experience any increase in costs for these benefits when current employees work additional hours, and other costs (such as recruitment and training) can be avoided.
As a long-term measure, however, increased use of overtime may not be a viable option. For workers not exempt from the FLSA, the employer must pay a premium of at least 50 percent for overtime work, which gives employers a strong financial incentive to try other means to deal with vacancy problems. In addition, many workers prefer not to work overtime, so increased use of overtime may lead to employee dissatisfaction and increased turnover, thereby exacerbating the vacancy problem instead of reducing it.

**Reduce Minimum Qualifications for the Job**

Another method of filling vacancies is to reduce the minimum hiring standards for the occupation. At first this may appear damaging, but this is not necessarily the case. The firm may have set the minimum hiring qualifications higher than necessary when labor was abundant. For example, a firm may have required a college degree for sales workers when a high school diploma would have been adequate. For professional jobs, the firm may have selected graduates from the most prestigious schools, or have had a minimum grade point average or test score cutoff, and these requirements may not be necessary.

If the productivity of less qualified workers is lower, the firm may be able to train the workers to reach the productivity levels of the more qualified workers after a reasonable period of time. When a firm reduces the minimum hiring qualifications, the labor supply is effectively increased, and the firm may be able to reduce the wages offered or at least avoid increasing wages.

**Restructure Work to Use Current or New Employees in Other Occupations**

If employers have difficulty filling vacancies with workers in one occupation, it is sometimes possible to restructure the work to make use of workers in other occupations. For example, hospital services are performed by workers in a number of occupations, such as physicians, nurses, nurse’s aides, and orderlies. Although some duties cannot be readily reassigned (only physicians can perform major surgery), nurses can perform some of the testing, and care-taking functions can be assumed by virtually any of the staff. Likewise, some engineering
tasks can be performed by drafters, and some tasks performed by teachers can be handled by aides.

For several reasons, firms will not always make use of this option. Although it is often possible to reassign some tasks from higher-skilled to lower-skilled workers, it is seldom possible to go in the other direction. For example, hospitals are unlikely to use physicians to perform care-taking tasks because physicians are so costly that other measures will generally be less expensive. In addition, assigning what is perceived to be low-level work to employees may hurt morale and productivity. Finally, reassigning tasks may involve considerable expense and disruption because of training and rescheduling that must be conducted.

In some cases, complex jobs can be decomposed into simpler tasks that can be handled by less-skilled workers. For example, a tool and die maker’s work could be split among metal workers who possess some, but not all, of the skills of a tool and die maker. In general, shortages are more likely to occur for high-skill occupations than low-skill occupations.

Substitute Machinery and Equipment for Labor

Employers can sometimes alter the production process to replace workers with equipment. As technology has advanced in recent years, the types of tasks performed by machines have also changed. Formerly, machines typically replaced humans in unskilled tasks such as lifting and moving. More recently, computer-based technology permits machines to perform more sophisticated tasks, including voice recognition, drawing, designing, and (to some extent) teaching. Artificial intelligence “expert system” models even permit computers to substitute for professional judgment under certain circumstances.

There are obviously limits to how much technology can substitute for labor, and in many situations technology will be used to substitute for labor for reasons other than difficulty in filling job openings. However, substituting technology for labor is sometimes a viable method of dealing with difficulty in filling vacancies.
Train Workers for the Jobs

For some occupations, training is traditionally performed by employers, either formally through apprenticeship or other training programs or informally through on-the-job training. For many other occupations, however, training for entry-level jobs is performed by other means—typically colleges and universities for professional occupations, and vocational schools and trade schools for skilled craft and service occupations. Employers who traditionally do not train their own workers may resort to offering or sponsoring training if they are experiencing difficulty in filling vacancies.

Offering training for an occupation is often a major commitment for employers, and it is typically not provided unless most other approaches fail. There are several related reasons why firms are reluctant to offer occupational training. First, training is generally time-consuming. Training new employees for a skilled occupation can sometimes take years, and by the time the workers are trained, the problem of filling vacancies may have disappeared. Second, establishing and operating a training program to bring new employees into an occupation is costly. Employers must feel confident that they can recoup their investment before they are willing to underwrite these costs. Finally, training new hires for occupations with vacancies carries several risks for employers. The individuals selected may not be able to successfully complete the training, or if the skills are transferable to other employers, they may quit shortly after they are trained.9

For occupations that do not require a college degree, establishing an apprenticeship program is one potential method of training workers for occupations through a combination of classroom and on-the-job training. Other possibilities include training current or new workers in-house or in cooperation with local colleges, vocational schools, and proprietary schools. In some cases the employer may not pay for the training—the courses can be partially or fully funded under federal programs such as the Workforce Investment Act (WIA), state training programs, or educational institutions.

Training for entry into an occupation can be illustrated by an extreme but interesting case. The uniformed services need physicians, but they are prohibited from paying market wages to physicians (a shortage induced by market restrictions). To get around this problem, the
uniformed services established their own medical school, Uniformed Services University, to train physicians at no cost to the students. To prevent the students from leaving soon after being trained, the students are required to sign contracts agreeing to stay in the military for a specified number of years.

**Improve Working Conditions**

Improving working conditions sometimes is an effective way to attract new workers and/or reduce turnover. Working conditions include factors such as hours worked, upgrades in equipment and facilities used by workers, level and type of supervision, involvement in operation of the firm, training to deal with stress related to the job, and recognition of the importance of workers in the occupation. Improvements in working conditions can be especially useful in situations where vacancies are created by high turnover. High turnover is often associated with occupations that have high stress, low wages, or low prestige. A concomitant benefit of improving working conditions is that productivity may increase as well.

Improving the number or timing of work hours can also help in recruiting new workers and in reducing vacancies. Some occupations may require split shifts (e.g., driving buses), night and weekend work (e.g., health occupations), or down time between productive periods (e.g., home care). Employers sometimes deal with these unpleasant working conditions by offering premiums for work at undesirable times, but they often believe they cannot afford a sufficiently high shift differential to eliminate the problem. Although shift differentials are still often necessary for undesirable shifts, employers can sometimes improve recruiting and reduce turnover by working with employees to structure shifts to be as desirable as feasible. For example, hospitals have experimented with a number of shift structures to fill the most undesirable shifts. In the home care industry, where workers sometimes have a great deal of travel time and down time between cases, some employers have been successful in restructuring caseloads to minimize these problems.
Offer Bonuses to New Employees

Although this approach is not commonly used, firms sometimes offer new employees bonuses for joining the firm. Signing bonuses are similar to paying current employees bonuses (or “bounties”) for recruiting new employees for occupations that are difficult to fill, except that bonuses go to the new employees rather than the current employees.

For workers, this option provides an extra incentive to join the firm offering the bonuses. This approach is more advantageous for employers than raising wages because it is a one-time cost and only affects the employees added in the occupation of interest. The disadvantage for employers is that the employees lured by such bonuses may not be as interested in long-term careers with the firm, and they may be “pirated” away by other firms offering similar bonuses. Signing bonuses are most frequently used when employers feel that they are under intense pressure to fill vacancies in the short run. When employers recognize this to be the case, they sometimes resort to using hiring bonuses to lure employees from other firms. Hiring bonuses have been used by hospitals to recruit nurses and by data processing firms to recruit programmers.

Improve Wages and Fringe Benefits

Based on the simple supply and demand curve analysis, increasing wages is an obvious way to increase the number of workers willing to work in a particular occupation. Employers are generally reluctant to increase wages for several reasons. First, an increase in wages will affect the entire workforce in the occupation with vacancies, not just the new workers the firm wishes to attract. Thus, the employer incurs costs for more than just the added workers.

Second, the employer might have to increase wages for workers in other occupations as well. Employers generally attempt to maintain equity among workers in various occupations. Thus, if an employer increases wages for one occupation because of difficulties in filling vacancies, wages may have to be increased for other occupations as well to maintain what are viewed as appropriate differentials. Another problem with raising wages is that wages tend to be “sticky” in terms of moving down. That is, once market conditions change, employers will generally have less flexibility to reduce wages later. Finally, rais-
ing wages might not be an effective means of recruiting in the short run if supply is not responsive to changes in wages (i.e., the supply is inelastic). In the extreme case, if the supply is totally fixed in the short run, higher wages cannot induce any change in the number of workers qualified to work in the occupation.

Improving fringe benefits is similar to increasing wages, but in some instances employers will reduce their vacancy rates more by improving benefits rather than by increasing wages by a similar amount. For example, health insurance is often an important fringe benefit to provide. Because group health insurance rates are usually substantially less expensive than individual policies, the value of health insurance to the employee will often be greater than the cost to the employer. Health insurance is especially a concern for employers trying to fill vacancies for relatively low-paying jobs if Temporary Assistance for Needy Families (TANF) recipients are potential workers. If employers who do not provide health insurance are small and pay low wages. Thus, adding benefits such as health insurance may be most burdensome in those cases where it would be most important. Other potential fringe benefits include subsidized housing and child care.

**Contract Out the Work**

If a firm is unable to hire all the employees it needs in particular occupations, the firm may be able to contract out the work to another employer who is not experiencing shortages. In some instances, the labor problem may be regional in nature, and the firm can contract out the work to a firm in another part of the country. If the problem is nationwide, the firm can sometimes have the work performed overseas.

**Turn Down Work**

If a firm has exhausted all means that it considers reasonable and can find no reasonable way around its occupational vacancies, the firm always has the option of turning down work. Employers generally use this “solution” only as a last resort because they do not like to give up
customers to competitors, and, more basically, the only way to make a profit is to sell goods and services.

If the firm has limited capacity to conduct its business because of occupational shortages, there are more subtle measures than simply refusing work. For example, the firm might reduce its marketing activity, thereby reducing the demand for its products as well as its advertising costs.

**REASONS LABOR MARKETS MAY ADJUST SLOWLY**

As discussed above, labor markets, and other markets as well, constantly experience changes in supply and demand that cause them to deviate from an equilibrium situation. In most cases, firms and workers will take actions that will move the labor market toward equilibrium. In some instances, however, the market adjusts slowly, and equilibrium is not restored, resulting in a shortage for the occupation. The literature suggests several factors that may result in the market failing to clear reasonably quickly. These factors are discussed below.

**Slow Reaction Time by Employers**

In most industries, each individual firm employs a small share of the workers in a particular occupation. Thus, individual employers may be unaware of an increase in demand, and they are almost certainly unaware of the magnitude of the increase. As the firm recognizes that workers cannot be attracted at what they believe to be the market wage, they may then take the actions described above to deal with the vacancies.

A number of factors can influence the reaction time of employers. If the firm does not recruit frequently for the occupation, either because of low turnover or because it employs few workers in the occupation, the firm may not know the typical period for filling vacancies for that occupation. The firm also may not have a good idea of what the market wage is and may tend to set its offer wage too low.

Several institutional factors are likely to affect reaction time by employers. Occupations characterized by long vacancy periods are
more likely to have slow reaction times by employers because employers expect to take a significant amount of time before they fill vacancies. Lengthy recruiting periods are more characteristic of occupations with high salaries, typically professional and managerial occupations and highly skilled craft jobs. Occupations where employment is concentrated in small firms are likely to be characterized by slow reaction times because the employers are likely to recruit for fewer positions and less frequently than larger employers.

Other institutional factors that can influence employer reaction time include the extent to which employers and workers in the occupation are organized and exchange information. For example, if employers have a trade association that monitors and publishes data on wages, vacancies, and other employment-related factors, employers will be aware of the occupational situation early in the search process. Of course, receiving national-level data is not as useful as local data for an employer who recruits locally. For some occupations, hiring is done in conjunction with the trade union representing the workers. Even if most firms are small, the centralization of the hiring process will help employers gain a quicker grasp of the supply available.

**Slow Response Time by Employers**

After firms recognize that there is excess demand for an occupation, they may delay taking actions to fill their vacancies. Most strategies that a firm might try could be risky, expensive, or both. Relatively minor responses, such as intensifying the recruiting effort, will waste the firm’s money if the positions would be filled without them. More significant responses, such as changing the occupational structure of the firm and training workers, require major commitments of time and resources to plan and implement. Such actions are unlikely to be taken unless the employer believes that the firm is facing a prolonged period of difficulty in hiring.

Increasing wages can also be a major step for employers. As noted above, the wage increases must also be passed on to current workers as well as the newly hired workers, and sometimes workers in other occupations must receive increases as well. If the firm is in a competitive product market, it must carefully balance two competing interests. If it sets the wage too high, the firm’s costs will be higher than the costs
of its competitors, and the firm is likely to either lose market share (if it passes the costs on to consumers) or profits (if it absorbs the increased costs) or both. Thus, firms are likely to be conservative in increasing wages as a method of filling vacancies.

**Slow Reaction Time by Workers**

Workers in other occupations and individuals who are unemployed or out of the labor force may not immediately recognize that wages or working conditions have improved in the occupation with the developing shortage. If workers who might be attracted to jobs in the occupation with the excess demand are unaware of the opportunities, they will not be able to consider entering that occupation. The time required for workers to become aware of the new opportunities depends on how effective firms’ recruiting strategies are and how sensitive workers are to the recruiting effort. Also, workers may have a certain amount of loyalty to their current employer, occupation, or industry, and the greater such loyalty is, the slower the reaction time by workers.

**Slow Response Time by Workers**

Once workers are aware of the opportunities, their response time will depend on the time required to qualify for the positions and the costs and benefits of obtaining any needed qualifications, applying for the positions, and changing jobs. Typically, the greater the incentives provided by employers to induce workers into the occupation of interest, the quicker and greater the response by potential entrants.

For many occupations, training time is the most important factor slowing worker response time. Occupations requiring a specialized college degree, such as engineering, will be very slow in adjusting because the “pipeline” for producing new engineers is four years. The lag might be more extensive if some potential engineers must adjust their mathematics course load in high school. Some specialized occupations, such as architecture and medicine, require even longer preparation.

Many occupations requiring less than a college education still demand several years of training and will have a substantial lag before interested individuals qualify for the occupation. For example, many technician and skilled craft positions require two or more years of train-
ing. At the other extreme, some low-skill jobs, including paraprofessional home health care workers, may require as little as one week of formal training. Thus, the worker response lag generally will be shorter.

Response time can also be slowed if training institutions lack the capacity to train additional workers. For instance, the supply of nurses cannot be readily expanded if there are too few nursing instructors.

**Restrictions on Occupational Entry**

In some cases, institutional barriers to occupational entry will slow down the adjustment process. These restrictions are generally instituted to achieve certain purposes, so removing or modifying the barriers is not always appropriate. However, in times of occupational shortages, consideration is often given to modifying these restrictions.

One example of a barrier to occupational entry is limits in the enrollment capacity of training institutions that supply workers for the occupation. Suppose, for instance, that hospitals needed to hire more physicians and there were enough individuals interested in attending medical school to meet the hospitals’ demands. If the nation’s medical schools could not admit the extra students because of limited capacity, the supply of physicians could not increase. Note that hospitals do not regulate the capacity of medical schools, so it would be difficult for this market to adjust.

Other institutional barriers include licensing and certification requirements. Employers might be willing to lower the standards for a particular occupation, but if entry to the occupation is regulated, the regulatory or licensing board would have to agree. These boards, which are often state bodies, might not wish to lower the standards, and current members of the occupation might object to relaxation because it would cheapen their credentials and possibly result in lower wages. Restrictions on immigration may operate as a similar institutional barrier to achieving equilibrium in occupational labor markets. Trade unions or associations, at the time of contract negotiations or through other activities, may restrict the supply of workers or hiring requirements for workers. An example of this type of barrier is restrictions on the ratio of apprentices to journeyman workers in an occupation. In some instances, such restrictions could constrain employers and poten-
tial entrants from increasing the number of entrants in an occupation making use of apprenticeships.

All the barriers mentioned above were established for particular reasons, ostensibly to assure quality for workers in the occupation. Although consideration should be given to changing or eliminating the barriers, their original intent should not be forgotten.

**Continuous Increases in Labor Demand**

If the labor demand schedule continuously increases faster than the amount supplied can increase, then the market will not achieve equilibrium. This scenario is the basis of the Arrow-Capron dynamic model of labor shortages, and it can occur in periods of rapid sustained growth in one or more industries that employ workers in the shortage occupation. Such a period of sustained rapid growth for a particular sector of the economy can prevent the market from clearing for a substantial period of time. According to Arrow and Capron (1959), this situation occurred for engineers following World War II; more recently, similar rapid growth in demand occurred for information technology (IT) workers in the 1990s in the period leading up to the IT bubble. Note that in this situation the problem is not necessarily that workers or employers cannot adjust; rather, the problem is due to continued shocks to the equilibrium levels of employment and wages.

**CONSEQUENCES OF LABOR SHORTAGES**

Labor shortages can lead to a number of consequences for the firms experiencing them and the rest of the economy as well. In economic terms, the major consequence of a sustained shortage is that the economy will be operating less efficiently than it could. Until the market achieves equilibrium, resources are not put to their most productive use. Thus, aggregate production for the nation is below capacity. Workers may have to work more hours than they desire, or they may be assigned to jobs they do not want. Employers may have to use their workers and equipment less efficiently than they desire, and this may result in lower output and reduced profits. In some cases, the impact on consumers will
be relatively modest, but if consumers cannot obtain needed health care because of a labor shortage, the consequences can be severe. Finally, the impact of a shortage can extend beyond the firms directly experiencing the problem. A shortage of home health workers or nursing home workers, for example, may result in hospitals having to keep patients longer than is desirable or releasing them without adequate care after discharge. Thus, it is difficult to trace all of the effects of an occupational shortage.

By systematically studying occupational labor shortages, and tight labor markets in general, it may be possible to take actions to improve the performance of supporting institutions. As noted at the beginning of the chapter, knowledge about labor shortages can help training institutions, schools, and students make better choices about programs to expand and training programs to enroll in. Identifying occupations with shortages can also be useful for setting government policies regulating entry into occupations and in admitting temporary workers and immigrants.

**OUTLINE OF THE REPORT**

The remainder of this book has five chapters. The next four chapters provide case studies of special education teachers, pharmacists, physical therapists, and home care workers. For each occupation, we provide background on the occupation, describe the reasons why the occupation was selected for study, and summarize evidence from databases, literature, and interviews on whether the occupation is currently experiencing or recently experienced a labor shortage, the reasons for the shortage, and possible ways to alleviate the shortage. The final chapter provides our conclusions and discusses potential uses for occupational shortage data. More specifically, we discuss the limitations of current occupational data collected by the Bureau of Labor Statistics (BLS) and the extent to which the federal and state governments and other organizations could collect additional data that would be useful for policy purposes. The chapter also discusses how occupational data can be used in immigration policy.
Notes


2. The remainder of this chapter is largely based on work originally presented in Barnow (1996) and Trutko et al. (1993). Many of the references are fairly dated, but some of the most important research on shortages was conducted in the 1950s.

3. This definition, which was provided by the U.S. Department of Labor in a Request for Proposals for a study of labor shortages, is essentially identical to the definition used by Franke and Sobel (1970) in their study of labor shortages: “a situation existing over an extended period of time in which employers were unable to hire at going wages or salaries sufficient numbers of qualified persons to fill positions for which there were budgeted funds and for which personnel were required to meet existing demands for services.”

4. Technically, the supply curve for labor may be “backward-bending,” which means that at very high wages workers actually reduce the amount of labor they are willing to supply. We do not consider this concept further in our discussion because it is unlikely to be relevant in a study of labor shortages. Labor supply could also be perfectly inelastic, where an increase in the wage rate does not result in an increase in labor supplied. For highly skilled workers, this is often the case in the short run.

5. There are different definitions of labor surpluses as well. For example, in the 1990s, Daniel Hecker of the Bureau of Labor Statistics (1992) concluded that the nation had a surplus of college graduates, whereas Bishop (1996) concluded that there was a shortage. For a review of this issue see Barnow and Bawden (1991).

6. Formally, the internal rate of return is found by solving the equation 0 = \((W_0 - C_0) + (W_1 - C_1)/(1 + I) + (W_2 - C_2)/(1 + I)^2 + \ldots + (W_n - C_n)/(1 + I)^n\), where \(W_t\) represents earnings in year \(t\), \(C_t\) represents costs incurred in year \(t\), and \(I\) is the internal rate of return that is solved for.

7. For a similar discussion of the issues raised here, see Burke (2005).

8. This is the situation that occurred in the U.S. labor market for shipfitters in the 1990s. Because shipfitters and oil pipeline workers use many of the same skills, an increase in demand for pipeline workers by the oil companies led to a decrease in the supply of shipfitters (see Trutko and Barnow 1998).

9. The arguments presented above apply primarily to training new hires for entry into an occupation that they are not qualified for. The arguments do not apply, or do not apply to the same extent, to training workers already on the payroll to improve their skills. Moreover, even if training does not pay off for an individual employer, it might result in a return for society as a whole (see Barnow, Chasanov, and Pande 1990).

10. In 1996, TANF replaced Aid to Families with Dependent Children (AFDC).
Occupational Labor Shortages
Concepts, Causes, Consequences, and Cures

Burt S. Barnow
John Trutko
Jaclyn Schede Piatak

2013

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan