NBER WORKING PAPER SERIES

IN SEARCH OF LABOR DEMAND

Paul Beaudry
David A. Green
Benjamin M. Sand

Working Paper 20568
http://www.nber.org/papers/w20568

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2014

We thank Chris Bidner, David Card, Giovanni Gallipoli, Patrick Kline, Alex Lefter, Kevin Milligan and Richard Rogerson for comments. This paper incorporate some results previously circulated under the title "The Elasticity of Job Creation". Paul Beaudry would like to acknowledge support from the Social Science and Humanities Research Council of Canada and the Bank of Canada Fellowship program. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

In Search of Labor Demand
Paul Beaudry, David A. Green, and Benjamin M. Sand
NBER Working Paper No. 20568
October 2014
JEL No. J23

ABSTRACT

We propose and estimate a novel specification of the labor demand curve incorporating search frictions and the role of entrepreneurs in new firm creation. Using city-industry variation over four decades, we estimate the employment – wage elasticity to be -1 at the industry-city level and -0.3 at the city level. We show that the difference between these estimates likely reflects the congestion externalities predicted by the search literature. Also, holding wages constant, an increase in the local population is associated with a proportional increase in employment. These results provide indirect information about the elasticity of job creation to changes in profits.

Paul Beaudry
Vancouver School of Economics
University of British Columbia
997-1873 East Mall
Vancouver, B.C.
Canada, V6T 1Z1
and University of British Columbia
and also NBER
paulbe@interchange.ubc.ca

David A. Green
Vancouver School of Economics
University of British Columbia
997-1873 East Mall
Vancouver, B.C.
Canada, V6T 1Z1
green@econ.ubc.ca

Benjamin M. Sand
Department of Economics
York University
Canada
bmsand@yorku.ca

Introduction

Policy makers interested in how wage costs affect employment decisions could be excused for being confused by what the economics literature has to tell them. At one extreme, studies using variation in minimum wages and payroll taxes tend to find only small wage elasticities of employment demand (Blau and Kahn, 1999). On the other hand, studies of regional responses to labor supply shocks generally find small wage impacts and large employment changes, which is suggestive of very elastic labor demand (Blanchard and Katz, 1992; Krueger and Pischke, 1997).[1] Further variation in estimates arises in the literature because different studies use different units of observation, time frames and identification strategies, often without a clear reference to theory to support their choice. Our goal in this paper is to propose and estimate a new specification for labor demand that is based on a comprehensive view of the labor market and that is capable of reconciling different findings in the literature.

A natural starting place to look for answers regarding the wage elasticity of employment is the micro literature on firm demand for labor (see, for example, Hamermesh (1993, Chapter 4)). The goal of this literature has traditionally been to estimate how the average firm responds to a change in wages, generally holding total output constant. It is a literature that is very close in spirit to the literature estimating production functions. Knowing the properties of a firm's production functions, such as the extent of capital labor substitutability, is certainly interesting. However, it is unlikely to provide a complete assessment of how total labor demand within a market responds to a change in wages. For example, a production function perspective of labor demand will necessarily miss any adjustment on the extensive margin since entry and exist decisions of firms are excluded. Moreover, when discussing responses at the market level, it is not very interesting to keep the output produced by firms fixed.

A firm perspective on labor demand may also differ from a market perspective because of search and matching frictions. When adopting a firm perspective, a change in wage is viewed as affecting the firm's employment decision, but this employment decision is not allowed to have any external effects on the employment decisions made by other firms. However, in the presence of search and matching frictions, an increase in the employment of one firm has a direct externality effect on the employment decisions of other firms, even holding wages fixed, since it increases market tightness and thereby increases the cost of recruitment. Such a mechanism may imply a difference between the market response to a change in labor costs and the simple sum of isolated firm responses. In summary, if one is interested in how labor demand in a market responds to wages, one must move away from a perspective focused at the individual firm level and instead adopt an approach that explicitly takes into account the many channels through

[1] The two extremes are captured in the minimum wage literature on one end (where studies commonly find either small positive or small negative elasticities) and the literature on city adjustments to shocks on the other (where, for example, Card (1990a) finds virtually no wage response to the Mariel Boat-lift supply shock in Miami).

which changes in wages cost can effect employment decisions. Accordingly, our approach will be to derive an empirically tractable specification of the market demand for labor that takes into account several different margins of adjustment.[2]

The labor demand specification we propose is built from micro-foundations and incorporates four main determinants of employment. There is obviously a direct wage effect of the kind that is central to any study of labor demand. In our framework, this effect will capture adjustments on both the intensive and entry margin of firm decisions. Second, there is a labor market tightness effect aimed at capturing the congestion externalities emphasized in the search and matching literature.[3] Third, we also include population size as a determinant of employment demand. From the perspective of the traditional labor demand literature, this is unconventional because one would typically expect population size to determine labor supply not labor demand. However, once one models the process of firm creation explicitly, and recognizes that entrepreneurs may be a limiting factor in job creation, it becomes necessary to include population size as a determinant of employment demand since it reflects the size of the pool of potential entrepreneurs. Finally, there are the effects of technological change that will appear in the error term of our specification.

In the empirical section of the paper, we estimate our labor demand specification at both the industry-city level and at the aggregate city level using data from the 1970, 1980, 1990 and 2000 US Censuses and the 2007 American Community Survey. Our approach is to treat the cities as observations on a set of local economies, allowing us to identify within-city general equilibrium effects that interest us. Since we look at changes in employment outcomes over 10-year periods, our focus will clearly be on medium-run adjustment, and, for this reason, our approach will downplay certain adjustment costs that have been central to the dynamic labor demand literature that generally focuses on much higher frequency decisions (Cooper, Haltiwanger, and Willis, 2004; King and Thomas, 2006; Kramarz and Michaud, 2010, for example).

As is common in all studies of demand or supply, the key difficulty is finding convincing data variation that allows consistent estimation of the causal impacts of the variables of interest. To this end, we rely on a set of instruments that are similar in spirit to that first proposed by Bartik (1991) to identify each of our key labor demand determinants. The instruments we build use developments at the national level to predict local outcomes and rely on the identifying assumption that changes in productivity at the local level are independent of past levels of local productivity. We discuss the plausibility of this assumption, which is certainly questionable, and provide a very informative over-identification test. To identify wage effects, we build two instruments that are based on our earlier work on search models in a multi-sector context, which we

[2]Our focus on medium run wage effects on employment differentiates our work from studies of regional adjustment to aggregate labor demand changes (Blanchard and Katz, 1992; Bartik, 1993, 2009) which mainly focus on unemployment dynamics.

[3] In many environments this type of effect is unidentified. However, by exploiting data at the industry-city level, we will show that we can identify such an effect.

discuss in more detail below (Beaudry, Green, and Sand, 2012, hereafter BGS). To identify the labor market tightness effect, we exploit the commonly used Bartik instrument.[4] Finally, to identify an effect of population size on labor demand, we use a variant of the commonly used ethnic enclave instrument from the immigration literature (which, we show, is also a Bartik-style instrument) along with instruments based on climate.

Since our main focus in this paper is on consistently estimating the wage elasticity of labor demand, it is worth providing some extra detail on our identification strategy for these wage effects up front. In our earlier work (BGS) we argue that wage patterns in the US indicate that wages are at least partially the outcome of a bargaining process that takes place at the industry-city level. In that process, the outside option of workers is an important determinant of the wage. In BGS we point out that the outside option for workers in a particular industry-city cell is better if the industrial composition of employment in the city is weighted toward high-paying industries. That is a worker in, say, construction can bargain a better wage if the city he lives in includes a high paying steel mill instead of a lower paying textile mill, since one of his outside option is to move to the steel mill. BGS show how to build instruments for wage changes that are based on this insight.[5] These instruments are of a similar form to the classic Bartik instrument in the sense that they rely on an assumption that productivity growth in a city is not related to the initial employment composition in the city. Since we can build more than one instrument based on this outside option insight, this allows us to use an over-identifying test to evaluate the plausibility of the underlying identification restrictions. We show that this test is quite strong and that it is passed easily in our data.

The main empirical results of the paper are as follows. We find a statistically significant and economically meaningful negative trade-off between city-level employment rates and wages over 10 year periods. When looking at the industry level within a city, we find that a 1% increase in the wage in an industry-city cell leads to a decrease in the employment rate in that cell of approximately 1%. This result holds both when we look at all industries and when we look at only industries producing highly traded goods. When looking at the city level, we find that a 1% increase in the wages within all in-

[4]This instrument was first presented by Bartik (1991) and has been used in much subsequent work (Bartik, 1993; Blanchard and Katz, 1992; Bound and Holzer, 2000, for example). The Bartik instrument corresponds to a prediction of employment growth in a city based on industrial growth rates at the national level combined with start-of-period employment composition in the city.

[5]The idea of obtaining identification using variation in workers' outside options has precedents in the literature examining union wage and employment contracts (e.g., Brown and Ashenfelter (1986); MaCurdy and Pencavel (1986); Card (1990b)) as these papers exploit measures of alternative wages outside the specific contract in their estimation. Card (1990b) finds that the real wage in manufacturing has a positive effect on wage changes in the Canadian union contracts he studies, which echoes the mechanism underlying our basic source of identification. In a similar spirit, MaCurdy and Pencavel (1986) obtain estimates of production function parameters from data on wage and employment setting for typesetters when allowing for an alternative wage to effect the efficient outcome through an impact on union preferences.

dustries in a city leads to only a 0.3% decrease in the employment rate.[6] We argue that the smaller effect at the city level compared to the industry level reflects the impact of search externalities. In particular, we interpret this later result as reflecting that when wages increase in all industries, this leads to a less tight labor market, thereby reducing search costs to firms. This fall in search costs partially compensates for the increase in wage costs, leading to a smaller fall in employment than would have happened if wages only increased in a worker's own industry.[7] Finally, we find that an increase in population holding wages constant leads to an approximately proportional increase in labor demand. [8] We interpret this finding as indicating that the number of entrepreneurs available to create jobs in a city moves proportionally with the size of the city. Moreover, we will argue that this population size result also indicates that local labor markets are unlikely to be significantly constrained by fixed physical factors such as land or capital when looking over a 10-year period.

An important implication of our findings relates to identification of wage cost effects. In particular, our results imply that shifts in population caused by migration shocks cannot be used as instruments for the wage in labor demand specifications because population size is a direct determinant of labor demand. Put a different way, what has been viewed in the literature as a way of tracing a wage-employment trade – off using immigration shocks is not a way of identifying the wage elasticity of labor demand that is of concern to most policy makers. In our view, the relevant wage elasticity of labor demand for many policy issues needs to be estimated holding population size constant.[9]

The crux of our findings is found in the combination of a modest negative wage elasticity and the result that, keeping wages fixed, increases in labor supply increase employment one-for-one. We believe that these findings are easiest to interpret in terms of models with explicit recognition of entrepreneurs. In particular, within our framework these results imply that 1) entrepreneurs face a span of control problem or at

[6]In Hamermesh (1993), the main estimates he reports lie in a range near -0.3, which suggest a rather low elasticity of substitution between capital and labor. While this elasticity is numerically very close to the one we obtain here, it is not appropriate to compare them as they do not address the same question.

[7]Note that our finding of smaller effects at the city versus the industry level suggests that any possible positive demand linkages across industries in a city are dominated by the negative search externalities.

[8]One implication from this is that specifications with the employment rate rather than the employment level as the dependent variable are appropriate. Our reading of the existing labor demand literature is that papers use either employment levels or employment rates without providing any direct rationale for their decision.

[9]It is interesting to think of this result in the context of the employment effects estimated in, for example, Card (1990a)'s work on the effects of the Mariel Boatlift. Card shows that the sizeable inflow of Cuban refugees into the Miami labor market had little effect on wages. In the context of our extended model, if the inflow of migrants brings with it a proportional number of entrepreneurs then one should observe something like a replication of the existing economy; that is, a one-for-one increase in employment with little change in wages. However, according to our work, this should not be interpreted as implying a perfectly elastic labor demand curve. It simply reflects the fact that holding wages constant, employment tends to increase with the size of population.

least downward demand for their product and 2) that the elasticity of the supply of entrepreneurial talent to higher profits is far from perfectly elastic. Our estimates suggest that both these mechanisms have to be present to explain the data. Overall, we view our results as highly supportive of labor market models that emphasize the role of scarce entrepreneurial talent in the job creating process.

The remaining sections of the paper are structured as follows. In Section 1, we derive our empirical specifications for labor demand. We begin deriving a labor demand specification assuming that employers can readily hire workers at the going wage. We then extend our approach to allow for search frictions and emphasize how greater tightness in the labor market should negatively affect employment at the industry level. In section 2, we discuss issues related to identification of parameters. In particular, we present and justify the instrumental variable strategy we exploit for estimation. In section 3, we discuss the data and our construction of variables. In section 4, we report our main empirical results. In Section 5, we examine the robustness of our results to breakdowns by education and to incorporating slow adjustment of labor and wages. Section 6 contains a summary of the main empirical results and our interpretation of them. In section 7, we provide concluding comments.

1 Deriving Labor demand

Our goal in this section is to derive an empirically tractable specification for the locus describing the trade-off between wages and employment demand at the level of an industry or a whole economy. While it may seem natural to refer to that locus as a labor demand curve (and we will describe it in those terms as we proceed), there is a sense in which this terminology is misleading. In particular, the traditional labor demand literature has focused on identifying parameters of production functions that are relevant for firm-level employment decisions. While our approach will include such elements, we will also allow for effects of elements related to the entry process of firms and elements related to search frictions, as both these can affect the policy relevant trade-off that is of interest to us. As we will see, if those elements are relevant then they imply that what we will estimate is an equilibrium locus that reflects features beyond what is captured in the labor demand curve of any one firm.

To begin this endeavor, it is helpful to abstract from search frictions and consider the determination of firm employment and entry decisions in industry i in city c, taking wages as given. To this end, consider an environment where the good produced in industry i is traded on a national market at a given price p_i, and where physical capital can be rented out on the national market at rental price r. Each potential entrepreneur in this market has access to a production function $F^i(e_{ic}^j, K_{ic}^j, \theta_{ic})$, where e_{ic}^j is the number of workers employed by entrepreneur j in industry i in city c, K_{ic}^j is capital rented by the entrepreneur, and θ_{ic} is an exogenous productivity parameter capturing comparative advantage in the industry-city cell. We assume, for the moment, that there is only one type of labor. We discuss how to extend the framework to take into account worker heterogeneity in section 2.1. To ease presentation, we will assume that the production function takes the Cobb-Douglas form $F^i(e^j_{ic}, K^j_{ic}, \theta_{ic}) = (e^j_{ic})^{\alpha_1}(K^j_{ic})^{\alpha_2}\theta_{ic}$, with $0 < \alpha_1 + \alpha_2 \leq 1$. We will point out, as we proceed, where restrictions imposed by the Cobb-Douglas form affect our conclusions and describe how they are extended by relaxing that assumption. If entrepreneur j decides to enter the market, optimization implies that the employment level at his firm will be given by

$$e^j_{ic} = \left[\alpha_1 \left(\frac{\alpha_2}{r}\right)^{\frac{\alpha_2}{1-\alpha_2}}\right]^{\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}} (w_{ic})^{\frac{-(1-\alpha_2)}{1-\alpha_1-\alpha_2}} (\theta_{ic} p_i)^{\frac{1}{1-\alpha_1-\alpha_2}}.$$

The issue that interests us is how to go from this firm-level labor demand to aggregate labor demand in industry i in city c. The answer to this question depends on how we specify the firm's entry process and whether we assume the presence of a span of control problem.[10] If there is no span of control problem then going from firm demand to market demand is trivial since firm size is indeterminate and therefore the firm and the market are interchangeable. This is the traditional approach in the labor demand literature. Our approach, instead, will focus on the case where there is potentially a span of control problem. To this end, we adopt a rather flexible specification for firm entry in order to embed several of the specifications prevalent in the literature.

Before looking at our general specification, we will discuss two extreme cases. At one extreme, we could follow the firm entry literature, such as in Hopenhayn (1992), and assume that there is an infinite supply of potential entrants, with each entrant needing to pay a common fixed cost upon entry. We see this situation as extreme since it leads to a labor demand curve that is perfectly elastic. This type of specification for labor demand is not one that we want to impose on the data since it pre-supposes the answer to the question of how wages affect employment. At another extreme is the assumption that the supply of entrepreneurs is fixed exogenously, say, at the number N_{ic}. In this case, total employment demand in industry i in city c, which we will denote by E_{ic}, is given by $E_{ic} = N_{ic} \cdot e_{ic}$ and can be expressed very simply in log form as:

$$\ln E_{ic} = \alpha_{0i} - \left(\frac{1-\alpha_2}{1-\alpha_1-\alpha_2}\right) \ln w_{ic} + \epsilon_{ic}, \tag{1}$$

where $\alpha_{0i} = (1-\alpha_1-\alpha_2)^{-1} \cdot (\ln p_i - \alpha_2 \ln r + (1-\alpha_2)\ln \alpha_1 + \alpha_2 \ln \alpha_2)$ is an industry specific term which varies with p_i, and $\epsilon_{ic} = \frac{1}{1-\alpha_1-\alpha_2} \ln \theta_{ic} + \ln N_{ic}$ captures local productivity and entrepreneurial supply, where $\ln N_{ic}$ is included in the error term because it is not observed in most datasets. One of the potential restrictive features we see with such a specification for labor demand is that it is not affected by population size. While it is common to assume that labor demand is not functionally related to population size, we want to argue that such an assumption is at least questionable and should be explored empirically. For example, Equation (1) suggests that if a city is the recipient of a mass

[10] In the context of this production function, span of control problems are captured by assuming that there are decreasing returns to scale at the firm level, that is, $\alpha_1 + \alpha_2 < 1$.

migration then employment will not be affected unless the wage adjusts. This may be a correct way of describing the labor market, but it appears undesirable to us to impose such a restriction *a priori*. Instead, we believe that it is preferable to allow for the possibility that the mass of potential entrepreneurs increases with population size and, therefore, that an increase in population size may directly increase labor demand even at fixed wages. We can capture this possibility by assuming, instead of a fixed entrepreneur supply, that N_{ic} is related to the local population size, L_c, by $N_{ic} = \gamma_{0i} L_c^{\gamma_1}$, where $0 < \gamma_{0i} \leq 1$ and $0 \leq \gamma_1$. For now, we will assume that the entrepreneurs are drawn from the local population. Later we will relax this assumption to allow entrepreneurs to come from the national-level population.

While we want to allow for the possibility that the set of potential entrepreneurs increases with population size, we do not want to force all potential entrepreneurs to produce regardless of prices. Accordingly, we include a non-trival entry decision by assuming that each potential entrepreneur j faces a fixed cost, f_j, of entering the market, where f_j is drawn from the CDF, $G(f)$. The heterogeneity among entrepreneurs leads to a simple cut-off rule where only potential entrepreneurs with a fixed cost below some cut-off f^* will enter the market. To allow for simple analytic expressions, we further assume that $G(f)$ takes the form $G(f) = (\frac{f}{\Gamma})^{\phi}$, where $0 \leq \phi$ and $f \in [0, \Gamma]$.[11] Under these two extensions we get the following specification for labor demand:

$$\ln E_{ic} = \alpha_{0i} - \frac{1 - \alpha_2 + \phi\alpha_1}{1 - \alpha_1 - \alpha_2} \ln w_{ic} + \gamma_1 \ln L_c + \epsilon_{ic} \tag{2}$$

where α_{0i} captures industry effects, such as the price of the good, that are common across cities, and $\epsilon_{ic} = \frac{1+\phi}{1-\alpha_1-\alpha_2} \ln \theta_{ic}$.

The first difference to recognize between equations (1) and (2) is that local population size now appears on the right had side of (2) with the coefficient γ_1. This reflects our assumption that the set of potential entrepreneurs may increase with population size. There are several reasons why we believe it is important to introduce the potential role of population size in the determination of labor demand. First, it emphasizes that how wages adjust in response to a change in the population in a local labor market (e.g., due to an immigration shock) may reveal nothing about the wage elasticity of labor demand. In particular, note that the coefficient capturing the wage elasticity of labor demand in (2) can be very small and, nonetheless, this specification can still be consistent with an increase in population being met with a proportional increase in employment at fixed wages. In contrast, in a more standard specification for labor demand, as in (1), one would expect an increase in population to decrease wages unless the labor demand curve is perfectly elastic. Second, by looking at how population growth affects employment holding wages fixed, one can obtain substantial information about the functioning of the labor market. For example, if one finds that population enters into this equation with a coefficient of 1 then one can infer that entrepreneurship is likely propor-

[11] With this formulation of the distribution of the entry costs, the extreme case where there is only one common fixed cost can captured in the limit when ϕ goes to infinity.

tional to the population. In this latter case, it would be more appropriate to describe the wage-employment trade-off as one between wages and the employment rate as opposed to one between wages and the level of employment. Our empirical results do, in fact, support the view that the relevant labor market trade-off is between wages and employment rates, as we find that employment appears to increase one-for-one with population, holding wages fixed. Note that while we will interpret such a pattern as supportive of models where entrepreneurship is an important limiting factor, there may exist other interpretations.[12]

We next turn our focus to the coefficient on the wage in (2). This coefficient is always negative since it is given by $-\frac{1-\alpha_2+\phi\alpha_1}{1-\alpha_1-\alpha_2}$. There are two scenarios under which this coefficient equals minus infinity, i.e., where there is perfectly elastic demand. First, if there is no span of control problem, then $1-\alpha_1-\alpha_2=0$ and the wage elasticity becomes infinite. Alternatively, if potential entrepreneurs all face the same cost of entry, Γ, then ϕ must equal infinity as there is a mass point in the function $G(\cdot)$. Importantly, for the wage elasticity to be less than infinite, neither of these conditions can hold. Hence, finding evidence of a less that infinite wage elasticity in this framework is evidence of both a span of control problem and that there is not an infinitely elastic supply of entrepreneurs waiting to take advantage of any profit opportunity.

A more subtle issue in Equation (2) is the implicit restriction that the wage elasticity of labor demand should always be greater than 1 in absolute value. This feature is actually an artifact of the Cobb-Douglas structure and does not hold for more general production functions. For this reason, it should not be viewed as a relevant restriction. More importantly, we derived Equation (1) under the assumption that all goods in an industry are perfect substitutes. If, instead, we assume that goods from each entrepreneur are a differentiated product then there is further reason, beyond the span of control problem, for a fall in wages to have a limited effect on employment demand within a firm. Since the extension of the above specification to the case where the outputs of the different entrepreneurs are not perfect substitutes is rather straightforward, we omit it here. However, it should be noted that such an extension does change the interpretation of the coefficient on wages from one that is driven only by the span of control problem and firm entry decisions, to one that also takes into account the substitutability of products within the industry.

Before extending our labor demand framework to include the possibility of search frictions, we want to briefly clarify how span of control problems differ from simply assuming the presence of a fixed factor. To this end, we augment our previous production function to include a fixed physical factor (which could be land, for example) such that $F^i(e^j_{ic}, K^j_{ic}, X^j_{ic}, \theta_{ic}) = (e^j_{ic})^{\alpha_1}(K^j_{ic})^{\alpha_2}(X^j_{ic})^{\alpha_3}\theta_{ic}$, with $0 < \alpha_1+\alpha_2+\alpha_3 \leq 1$. The input X^j_{ic}

[12]The main data pattern that we find in our empirical analysis is one where the wage elasticity of labor demand is very far from infinity. At the same, time employment responds proportionally to an increase in population size at fixed wages. To explain such a pattern one needs a model with a limiting factor which is proportional to population. Our belief is that entrepreneurial talent is the most likely candidate for such a factor.

represents the use of a local fixed factor X by entrepreneur j, with X_{ic} representing the total amount of the fixed factor available in city c. Here we maintain our previous assumptions that $N_{ic} = \gamma_0 L_c^{\gamma_1}$ and that potential entrepreneurs face a fixed cost of entry equal to f drawn from $G(f) = (\frac{f}{\Gamma})^{\phi}$.[13] Under this extension, we obtain the following, slightly more general, specification for labor demand:

$$\ln E_{ic} = \alpha_{0i} - \frac{1 - \alpha_2 + \phi(\alpha_1 + \alpha_3)}{1 - \alpha_1 - \alpha_2 + \alpha_3 \phi} \ln w_{ic} + \frac{\gamma_1(1 - \alpha_1 - \alpha_2 - \alpha_3)}{1 - \alpha_1 - \alpha_2 + \alpha_3 \phi} \ln L_c + \epsilon_{ic}, \tag{3}$$

where α_{0i} again captures common industry effects and now $\epsilon_{ic} = \frac{1+\phi}{1-\alpha_1-\alpha_2+\alpha_3\phi}(\ln \theta_{ic} + \alpha_2 X_{ic})$. Thus, the error term incorporates the city-industry productivity parameter, as before, and the local supply of the fixed factor.

We see the introduction of a fixed physical factor in our set-up as having two interesting implications. First, with the presence of a fixed factor, the effect of population on labor demand is likely to be smaller than 1 even if $\gamma_1 = 1$, that is, even if entrepreneurs are proportional to the population. This is intuitive as population growth will cause the fixed factor to become more constraining even in the presence of more entrepreneurs. Second, and most importantly, if we assume away the span of control problem ($1 - \alpha_1 - \alpha_2 - \alpha_3 = 0$) then even if $\gamma_1 > 0$, population will not enter the labor demand specification. The presence of a fixed factor can justify why the wage elasticity of labor demand may be less than minus infinity. However, it cannot rationalize why an increase in population may be met with increased employment at fixed wages. To rationalize this, while maintaining the feature that the wage elasticity of labor demand is less than infinite, one needs the presence of a limiting factor that grows with population. Entrepreneurs play that role in our framework.

Up to now, we have derived the determinants of labor demand under the assumption that entrepreneurs are drawn from the local population. This allowed for a transparent and explicit discussion of individual-level entry decisions, and how those decisions affect the specification of labor demand. While this may appear as a very restrictive assumption, it turns out that Equation (3) can be derived under the alternative assumption that potential entrepreneurs are drawn from the national-level population, L, according to a rule of the form $N_{ic} = \gamma_{0i}(\frac{L_c}{L})^{\gamma_1} L^{\gamma_2}$ where $0 \leq \gamma_1$ and $0 \leq \gamma_2$; that is, we allow the local supply of potential entrepreneurs to increase with both the relative size of the local population and the size of the national population. In this alternative formulation, the national-level population is a common factor across cities and, therefore, can be incorporated into the constant term. This leaves only the size of the local population as an explicit regressor capturing entrepreneurial supply, and our main specification is unchanged. Such a formulation can be rationalized under the view that national-level entrepreneurs learn about local opportunities in proportion to the relative size of the specific locality. The case where entrepreneurial supply is not related to local population size is then captured by $\gamma_1 = 0$.

[13] We are implicitly assuming here that the factor X can be traded freely across firms in the local market.

1.1 Including search frictions

In our derivation of Equations (2) and (3), we implicitly assumed that there were no search frictions in the labor market, and that firms wanting to hire could costlessly fill vacancies at the going wage. In this subsection, we extend the above labor demand framework to allow for the possibility of search frictions out of concern that omitting that possibility may imply a biased perspective on labor demand – especially regarding the trade-off between wages and employment demand. To introduce search frictions, it is convenient to assume that our entrepreneurial firms do not hire labor directly but instead buy an intermediate good, Z_{ic}, that is specific to the industry and produced locally with labor in a one-to-one fashion. The entrepreneurial firms producing the final good now take the prices of the intermediate good, which we denote by p^z_{ic}, as given and behave as in the previous section in terms of deciding whether to produce and how much to buy of the different inputs if production takes place. The only difference is that firms buy Z_{ic} from intermediate good producers that face search frictions instead of hiring labor directly.

In order to introduce search frictions, we need to extend to our analysis to a dynamic setting. Accordingly, we will assume that time is continuous and that all the costs facing entrepreneurs discussed previously now represent instantaneous costs for flow services. We assume the existence of a large set of intermediate good producers, each of which can decide whether to post a vacancy at any point in time; where a vacancy needs to be dedicated toward producing the intermediate good for one specific industry. The flow cost of posting a vacancy for producing good Z_i is denoted h_{ic}. When an intermediate good producer finds a worker, she begins production and obtains a flow return of $p^z_{ic}-w_{ic}$. Workers are assumed to be hired from a common pool, regardless of which intermediate good they will eventually produce. Job vacancies and unemployed workers match according to a constant returns to scale matching function given by $M(L_c - E_c, V_c) = (L_c - E_c)^{\nu} V_c^{1-\nu}$, where E_c is total employment in the city and V_c is the number of vacancies. Given this matching function, the flow rate at which an intermediate good producer finds a worker is given by $(\frac{L_c-E_c}{V_c})^{\nu}$. Assuming that matches break up exogenously at rate δ, the steady state flow rate at which intermediate good firms find workers will be given by $\left[\frac{1}{\delta}(\frac{1}{\frac{E_c}{L_c}} - 1)\right]^{\frac{\nu}{1-\nu}}$. Letting ρ denote the discount rate for these firms, the equilibrium condition imposing that the value of a vacancy be zero implies the following simple expression between p^z_{ic} and w_{ic}:[14]

$$p^z_{ic} = w_{ic} + \frac{(\rho+\delta)h_{ic}}{\left[\frac{1}{\delta}\left(\frac{1}{\frac{E_c}{L_c}} - 1\right)\right]^{\frac{\nu}{1-\nu}}} \tag{4}$$

In (4), we see that the price of the input Z_{ic}, which is the cost of a flow of labor services

[14] To derive this relationship, we use the fact that the value of a filled job for an intermediate good producer, which we can denote by J, must satisfy $\rho J = p^z_{ic} - w_{ic} + \delta(W - J)$ where W is the value of a vacancy. We combine this with the fact the W must satisfy $\rho W = -h_{ic} + [\frac{1}{\delta}(\frac{1}{\frac{E_c}{L_c}} - 1)]^{\frac{\nu}{1-\nu}}(W - J)$, and $W = 0$.

to the final producer of good i in city c, is equal to the wage paid by the intermediate good producer plus a term capturing the cost of search. If h_{ic} were equal to zero, there would be no search costs and therefore p^z_{ic} would simply be equal to the wage. The importance of this search cost for the price of Z_{ic} depends on how firms discount the future, on the job destruction rate, and, most importantly, on the average time an intermediate good firm spends searching for a worker which is given by $1/\left[\frac{1}{\delta}\left(\frac{1}{\frac{E_c}{L_c}}-1\right)\right]^{\frac{\nu}{1-\nu}}$.[15] In this latter expression, it is important to note that time spent looking for a worker can be expressed as an increasing function of the employment rate in the city: the tighter is the labor market, the higher is the employment rate and the longer it takes to fill a vacancy. Hence, the cost of the labor service, Z_{ic}, will be greater in a tighter labor market, holding wages fixed. If we simplify matters further by assuming that the cost of posting a vacancy, h_{ic}, is proportional to the wage in the industry-city cell (that is, $h_{ic} = h_i \cdot w_{ic}$), then we can use (4) and (3) to get the following generalized demand for labor relationship, which now includes a term that reflects search frictions:

$$\ln E_{ic} = \beta_{0i} + \beta_1 \ln w_{ic} + \beta_1 \ln \left(1 + \frac{(\rho+\delta)h_i}{\left[\frac{1}{\delta}\left(\frac{1}{\frac{E_c}{L_c}}-1\right)\right]^{\frac{\nu}{1-\nu}}} \right) + \beta_2 \ln L_c + \epsilon_{ic}, \tag{5}$$

where $\beta_1 = -\frac{1-\alpha_2+\phi(\alpha_1+\alpha_3)}{1-\alpha_1-\alpha_2+\alpha_3\phi}$, $\beta_2 = \frac{\gamma_1(1-\alpha_1-\alpha_2-\alpha_3)}{1-\alpha_1-\alpha_2+\alpha_3\phi}$, $\epsilon_{ic} = \frac{1+\phi}{1-\alpha_1-\alpha_2+\alpha_3\phi}(\ln\theta_{ic}+\alpha_2 X_{ic})$, and β_{0i} again captures industry specific terms.

Equation (5) provides, in our view, a simple but rich framework for exploring the trade-off between wages and employment demand. In particular, this specification departs from traditional labor demand specifications by embedding elements of both the search and firm entry with span of control literatures. As a result, our specification for labor demand includes a wage effect, a search cost effect and a population effect: the latter two not being commonly included in traditional specifications of labor demand. Note that the coefficient on population, β_2, will equal 1 if $\gamma_1 = 1$ and $\alpha_3 = 0$; that is, when entrepreneurs are proportional to the population and there is no fixed physical factor. This is an important special case and one that, we will see, appears to be supported by the data. It is relevant to recall that we used steady state conditions for the search process to derive this equation. Thus, (5) is most likely appropriate for studying more of a medium-run outcome, and this is what we will do in our empirical work.[16]

In our empirical work, we will actually focus on a log-linear approximation of this equation so as to emphasize the first order effects of the aggregate employment rate, $\frac{E_c}{L_c}$,

[15] In the search literature, it is most common to use the ratio of vacancies to unemployed workers as the measure of tightness. However, at the steady state, the unemployment to vacancy ratio can be written as a simple function of the employment rate. In particular with the matching function in Cobb-Douglas form, $\frac{L-E}{V} = \frac{(1-\frac{E}{L})^{\frac{2-\nu}{1-\nu}}}{(\delta\frac{E}{L})^{\frac{1}{1-\nu}}}$.

[16] Out of steady state, the link between prices p^z_{ict} and wages given in (4) would be more complicated, as the search cost could not be summarized by a function of the current employment rate.

on the industry-specific employment rate, $\frac{E_{ic}}{L_c}$. In particular, we will generally work with the equation in the form:

$$\Delta \ln E_{ic} \approx \Delta\beta_{0i} + \beta_1 \Delta \ln w_{ic} + \beta_3 \Delta \ln \frac{E_c}{L_c} + \beta_2 \Delta \ln L_c + \Delta\epsilon_{ic}, \tag{6}$$

where β_1 and β_2 are unchanged from before. The term β_3, which can be written as $\beta_3 = \beta_1 \Phi$ (where $\Phi > 0$), reflects the log-linear effect of labor market tightness, as captured by the local employed rate, on the cost of filling a vacancy. We have written the equation in differences over time since this is the way we will estimate it in order to eliminate time invariant city-industry effects.[17] In the data work, time periods will, for the most part, be 10 years apart.

The important element in (5), relative to (3), is the presence of a negative feedback from the aggregate rate of employment to the rate of employment in one industry. This negative feedback, which reflects search externalities, may, at first pass, appear counter-intuitive since one might expect that cross-good demand linkages would imply a positive feedback. However, for goods traded on a national market, the demand effects in our formulation should be captured by the industry specific terms contained in β_{0i}, implying that the local aggregate employment rate captures the effect of search frictions.

1.2 Deriving a city level labor demand curve

Equation (6) is our baseline specification labor demand curve at the industry-city level. It will be informative to derive a city-level labor demand curve from it. To this end, let us first define η_{ict} as the fraction of employment in industry i in city c (i.e. $\eta_{ict} = \frac{E_{ict}}{\sum_j E_{jct}}$). Now consider aggregating Equation (6) using weights η_{ict}, and using the approximation $\sum_i \eta_{ict-1} \Delta \ln \frac{E_{ict}}{L_{ct}} \approx \Delta \ln \frac{E_{ct}}{L_{ct}}$, in order to get

$$\Delta \ln E_{ct} = \frac{1}{1-\beta_3} \sum_i \eta_{ict-1} \cdot \Delta\beta_{0it} + \frac{\beta_1}{1-\beta_3} \sum_i \eta_{ict-1} \cdot \Delta \ln w_{ict} + \frac{\beta_2 - \beta_3}{1-\beta_3} \Delta \ln L_{ct} + \sum_i \eta_{ict-1} \frac{\Delta\epsilon_{ict}}{1-\beta_3}. \tag{7}$$

This equation expresses the change in the employment rate within a city as being negatively affected by the average wage change in the city ($\sum_i \eta_{ict-1} \cdot \Delta \ln w_{ict}$), and positively affected by the weighted sum of the β_{0it}. Notice that β_{0it} reflects a national-level effect associated with an industry. To express β_{0it} as a function of observables, we average (7) across cities (using the weights $\frac{1}{C}$, where C is the number of cities). This gives:

$$\sum_c \frac{1}{C} \Delta \ln E_{ict} = \beta_{0it} + \beta_1 \sum_c \frac{1}{C} \Delta \ln w_{ict} + \beta_3 \cdot \sum_c \frac{1}{C} \Delta \ln E_{ct} + (\beta_2 - \beta_3) \cdot \sum_c \frac{1}{C} \Delta \ln L_{ct},$$

where we have used the assumption that $\sum_c \frac{1}{C} \Delta\epsilon_{ict} = 0$ since $\Delta\epsilon_{ict}$ reflects changes in comparative advantage.

[17] Differencing also eliminates the fixed factor component of the error term since it does not vary over time by definition.

The latter equation implies that β_{0it} can be written as

$$\beta_{0it} = \sum_c \frac{1}{C} \Delta \ln E_{ict} - \varphi_2 \sum_c \frac{1}{C} \Delta \ln w_{ict} + d_t, \tag{8}$$

where d_t is a year effect that is common across cities. The first two terms on the right side of the above equation can be approximated as the growth of employment in industry i at the national level, denoted $\Delta \ln E_{it}$, and the growth of wages in industry i at the national level, denoted $\Delta \ln w_{it}$. Thus, equation (8) indicates that the industry specific intercept in (6) is approximately equal to the national level growth in employment in the industry corrected for the average wage growth in the industry. Using (8), we can write the job creation curve at the city level as

$$\Delta \ln E_{ct} = d_t + \frac{1}{1-\beta_3} \cdot \sum_i \eta_{ict-1} \cdot \Delta \ln E_{it} + \frac{\beta_1}{1-\beta_3} \cdot \sum_i \eta_{ict-1} \Delta \ln \frac{w_{ict}}{w_{it}} + \frac{\beta_2 - \beta_3}{1-\beta_3} \Delta \ln L_{ct} + \tilde{\zeta}_{ct}, \tag{9}$$

where $\tilde{\zeta}_{ct}$ is the error term given by $\sum_i \eta_{ict-1} \frac{\Delta \epsilon_{ict}}{1-\beta_3}$.

Equation (9) now expresses cross-city differences in employment changes as a function of three main components. The first is a general growth effect captured by $\sum_i \eta_{ict-1} \cdot \Delta \ln E_{it}$, which reflects the notion that a city should have a better employment outcome if it is initially concentrated in industries which are growing at the national level. Second, we have a negative wage effect, which captures within-industry adjustments to a change in the cost of labor. This is given by the term $\sum_i \eta_{ict-1} \Delta \ln \frac{w_{ict}}{w_{it}}$, which is large if a city experiences wage growth across industries that is higher on average than that experienced nationally. Since β_1 is negative, a high value of $\sum_i \eta_{ict-1} \Delta \ln \frac{w_{ict}}{w_{it}}$ will result in lower employment outcomes in the city. The third term corresponds to a population growth effect. Finally, the error term reflects changes in the city's comparative advantage.

A comparison of equations (9) and (6) reveals an important difference in the wage coefficients in each. The coefficient on the city-industry specific wage change in equation (6) is the direct effect of a wage change on the employment rate in an industry-city cell holding the aggregate employment change in the city constant. This reflects the response of firms in an industry if that industry is too small to have a substantial effect on the overall equilibrium in the city. However, in general, we would expect that the immediate effect of a wage change in i, as captured in β_1, would only be a first-round response. The decrease in employment in i would imply a less tight overall labor market in the city which would raise the value of a vacancy for entrepreneurs to an extent captured by β_3. The resulting employment changes would then have further effects. The ultimate outcome of that process on total employment in the city is given by $\frac{\beta_1}{1-\beta_3}$, which is the coefficient on the aggregated wage change in (9). Given that β_3 is predicted to be negative, the total impact of the wage change at the city level will be smaller than the direct, industry specific effect, reflecting the self-correcting nature of the search externalities.

2 Identification

In general, we would not expect OLS to provide consistent estimates of the coefficients in equation 6, as the error term consists of changes in city-industry comparative advantage (the θ_{ic} terms). We expect changes in comparative advantage to be correlated with both changes in the wage in a given industry-city cell and with movements in the city level employment rate. If worker migration decisions are based only on wages and employment rates then there may be no reason to expect a correlation between the change in the city size and the error term once we condition on wage and employment rate changes; that is, there would be no correlation if a productivity change is only of interest to workers to the extent it changes wages and the chance of getting a job. However, we allow for the possibility of a more direct connection, using instrumental variables related to each of the main right hand side variables.

The main pillar of our instrumental variable strategy will be to follow and extend ideas first presented in Bartik (1993) and used in many subsequent studies.[18] The idea in Bartik is to work within a regional setting to construct instruments of the form: $\sum_i \omega_{ict}\Delta Q_{it}$, where ω_{ict} are a set of weights specific to city c, and ΔQ_{it} is a change in the variable Q_i at the national level. In the specific case considered by Bartik, the weights are the beginning-of-period employment shares across industries within a city and ΔQ_i is the growth rate in employment at the national level in industry i between $t-1$ and t. The result is a prediction of the end-of-period city employment rate based on the idea that if a particular industry grows or declines at the national level, the main effects from that change will be felt most in the cities that have the highest initial concentration in that industry. Note that this particular Bartik instrument is actually the first variable on the right side of our Equation (9). Moreover, we can see from (9) that this instrument is potentially a good candidate for instrumenting the employment rate in Equation 6 as, if β_2 is close to 1, then $\sum_i \eta_{ict-1} \cdot \Delta \ln E_{it}$ should be correlated with the change in the city level log employment rate. We will call that instrument, Z_{1ct}.

Given our reliance on Bartik-type instruments, it is important to clarify the conditions under which they are valid. We will specify those conditions for Z_{1ct}, first, then set them out in more general terms. Recall that the error term in (6) is given by $\Delta\epsilon_{ict}$ and corresponds to changes in local (industry-city level) productivity. It seems reasonable to be concerned that this error term is correlated with changes in the employment rate in the city. Now consider the potential correlation of this error term with Z_{1ct}. Since Z_{1ct} varies across cities, we are concerned with the cross-city correlation between it and the error term, which we can write as,

$$\sum_c \frac{1}{C} \sum_i \eta_{ict-1}\Delta \ln E_{it}\Delta\epsilon_{jct} = \sum_i \Delta \ln E_{it} \sum_c \frac{1}{C}\eta_{ict-1}\Delta\epsilon_{jct}.$$

Taking the limit of the correlation as C goes to infinity implies that the instrument is

[18] See in for example Blanchard and Katz (1992).

asymptotically uncorrelated with the error term if

$$\text{plim}_{C\to\infty}\sum_{c}\frac{1}{C}\eta_{ict-1}\Delta\epsilon_{jct}=0 \tag{10}$$

It is intuitive (and straightforward to show) that η_{ict-1} is a function of the values of the ϵ_{jct-1}'s. Thus, this latter condition can be written in terms of the ϵ's, in which form it is equivalent to the following condition holding for all c and i:

$$\text{plim}_{C\to\infty}\sum_{c}\frac{1}{C}\epsilon_{ict-1}\Delta\epsilon_{jct}=0 \tag{11}$$

where $\Delta\epsilon_{jct}=(\epsilon_{jct}-\epsilon_{jct-1})$. Thus, the validity of the instrument depends on a random walk-type assumption. This is clearly a stringent assumption, and we would like to be able to test it. This is possible if there is more than one instrument, allowing for over-identifying tests of the underlying assumptions. This is precisely how we will proceed. We will take as a maintained assumption that the driving forces in the model, given by the set of ϵs, satisfy the conditions for Bartik-type instruments to be potentially valid. We will then propose a set of such instruments and test the over-identifying restrictions to see if such an assumption is reasonable.

We view several of the features of this example as reflecting general characteristics of Bartik-type instruments; (1) the estimation is done in over-time differences, (2) the error term often is a function of differences in productivities, and (3) the weights (which we called ω's earlier) are plausibly functions of the lagged productivity levels.[19] From this, two lessons carry over to other implementations of Bartik-type instruments. First, validity of the instruments requires a random walk-type assumption, typically in terms of productivity processes. Second, the national-level change component of the Bartik instrument (the ΔQ) does not enter the asymptotic consistency condition. This is true because the validity of the instrument depends on cross-city correlations and the cross-city variation in the Bartik instruments comes from differences in the ω_{ic} vectors and not from ΔQ_i, which takes a common value across cities. This means that, asymptotically, there is no reason to worry about how city-level changes aggregate to a national value for Q. It is important, though, that this is an asymptotic statement that is based on an assumption that as the number of cities goes to infinity, industries are spread across many of them (i.e., there is no industry that operates only in one, or a handful of cities, as the number of cities gets large).

We now turn to discussing instruments of the Bartik form that are likely correlated with the change in wages. We have argued previously that labor supply shifters provide dubious instruments for the wage since they may be correlated with shifts in the

[19]For example, in what is commonly called the Ethnic Enclave instrument used in examining the impacts of immigration on a local economy, the concern is that immigrants move to the economy because of changes in productivity (captured, at least partially, in the error term). The ω's in that example correspond to the proportion of immigrants from some source country that were located in a given city in an earlier period. That distribution of immigrants is plausibly correlated with productivity in the city in the earlier period, and the identifying assumption is that those earlier productivity levels are uncorrelated with the changes in productivity in the sample period.

supply of entrepreneurs. Hence, we need to turn to other forces that may drive wage changes. To this end, we draw on search and bargaining theory and exploit insights presented in BGS regarding the role of industrial composition in affecting workers' outside options and, through bargaining, wages. The idea in BGS is straightforward. Consider two identical workers who meet with potential employers in the same industry but in different cities. Upon meeting, the worker and employer can form a match and begin production or they can continue to search. With search frictions, a match will produce a bilateral monopoly, and workers and firms can bargain over the available match surplus to determine the wage paid. For the worker, the value of continuing to search serves as an outside option in the bargaining process. If there are frictions hindering perfect and costless mobility across cities, the value of continued search will depend on local labor market conditions. Within a local labor market, this value will depend, in part, on the expected quality of other potential matches and the expected duration of search. In particular, BGS show that when workers can potentially meet with firms in any industry, the value of workers' outside options will depend on the industrial composition of cities. Differences in local industrial composition will translate into differences in wages via bargaining, even if the tightness of the labor markets are the same, since higher outside options allow workers to capture more of the surplus. For example, workers in, say, the chemical industry should be able to bargain a higher wage if they live in a city with high-paying steel mills than if they live in a city where the steel mills are replaced with low-paying textile mills. We exploit this idea to justify two instrumental variables that will help to consistently estimate (6) and (9). The two instruments will be valid under the same assumption as we stipulated for Z_{1ct}.

Within the context of a multi-sector search and bargaining model, BGS formalize the idea that, within a given industry, outside options (and, hence, wages) will be higher in cities with an industrial composition that is tilted toward higher-paying industries because it increases the value of search for workers; that is, outside options are greater in cities where the probability of meeting a high-wage industry is higher. Therefore, industry-city wages, w_{ict}, will tend to be higher in cities where $\sum_j \eta_{jct} w_{jt}$ are higher (where w_{it} represents wages in sector i at the national level and η_{ict} is the relative size of industry i at the city level, and, therefore, $\sum_j \eta_{jct} w_{jt}$ proxies the outside options of workers). Notice that this is not a mechanical result since the ability of workers to switch industries implies that it would arise even if we just focused on other industries by dropping i when calculating the city average wage.

It is useful to decompose the movements in $\sum_j \eta_{jc} w_j$ as follows:

$$\Delta \sum_j \eta_{jct} w_{jt} = \left(\sum_j \eta_{jct-1}(w_{jt} - w_{jt-1}) \right) + \left(\sum_j w_{jt-1}(\eta_{ict} - \eta_{ict-1}) \right). \tag{12}$$

Equation (12) indicates that for a worker in a particular city, outside options will increase over time if employment in that city is concentrated in industries where wages are increasing at the national level or if the worker is in a city where there is a shift in industrial composition toward relatively high-paying sectors. Importantly, BGS show that workers value each source of change in the value of outside options equally; a worker bargaining a wage in given sector doesn't care whether her outside options change because of shifts in industrial structure or shifts in industry wages since all that matters is the expected wage in the city outside the current firm. In our empirical work, we use each component of shifts in outside options to form the basis of an instrument for wages in (6) and (9), exploiting the fact that each component relies on very different sources of variation.

We construct our first wage instrument, which we will call Z_{2ct}, based on the first term in (12):

$$Z_{2ct} = \sum_j \eta_{jct-1}(\ln w_{jt} - \ln w_{jt-1}).$$

BGS show that this instrument is a good predictor of wage growth at the industry-city level and give a formal justification for its relevance based on the wage bargaining story discussed previously. Importantly, Z_{2ct} varies across cities and obtains its variation entirely from the η_{ict-1}'s (the initial period local industrial composition). As in our discussion of Z_{1ct}, the national-level wage changes are not relevant for our consistency considerations since they are common across cities. As such, Z_{2ct} will be uncorrelated with the error terms in (6) and (9) (and, hence, will be a valid instrument) under the assumption given in (11), that the comparative advantage terms, ϵ_{ict}, behave as random walks with changes independent of past levels.[20]

The second instrument we propose for wages builds on the second term of (12), $\sum_j w_{jt-1}(\eta_{ict} - \eta_{ict-1})$. This term would not be an appropriate instrument since its dependence on the current industrial structure as captured by the η_{ict}'s implies that it will not be orthogonal to the error terms in (6) or (9). Instead, consider the closely related variable given by:

$$Z_{3ct} = \sum_i \ln w_{it-1} \cdot (\hat{\eta}_{ict} - \eta_{ict-1}) = \sum_i \eta_{ict-1} \cdot (g^*_{it} - 1) \cdot \ln w_{it-1}, \tag{13}$$

where $g^*_{it} = \frac{1+\Delta \ln E_{it}}{\sum_j \eta_{jct-1}(1+\Delta \ln E_{it})}$. For the variable Z_{3ct}, we have replaced the current industrial composition term η_{ict} with its predicted value base on η_{ict-1} and the national-level trend in employment patterns.[21] As with Z_{1ct} and Z_{2ct}, the resulting variable's cross-city variation stems from the η_{ict-1}'s and the same random walk assumption is needed for consistency. Furthermore, it should have predictive power for industry-city wage

[20]BGS presents a formal derivation of the form of the error term in the wage equation and prove that the conditions listed here imply that these instruments are valid.

[21]To create the predicted share term, we first predict the level of employment for industry i in city c in period t as:

$$\hat{E}_{ict} = E_{ict-1}\left(\frac{E_{it}}{E_{it-1}}\right).$$

Thus, we predict period t employment in industry i in city c using the employment in that industry-city cell in period $t-1$ multiplied by the national-level growth rate for the industry. We then use these predicted values to construct predicted industry-specific employment shares, $\hat{\eta}_{ict} = \frac{\hat{E}_{ict}}{\sum_i \hat{E}_{ict}}$, for the city in period t.

changes as it should capture the higher value of outside options for workers in a city where we predict that the industrial composition is tilting toward higher-paying jobs.

The availability of two instruments for wages raises the possibility of implementing an over-identification test. Z_{2ct} and Z_{3ct} are both predicted to have an impact on city-industry wages through channels related to workers' outside options. But the channels that each exploits are quite different – one related to shifts in industrial structure and one to within industry wage movements. As discussed above, theory predicts that each source of variation in outside options should have the same impact on wages since what matters for workers' bargaining positions is the change in the average wage in other industries, regardless of whether that change stems from changes in industrial composition or the industrial wage premia. Likewise, since what matters for employers is the bargained wage, variation in wages induced by either Z_{2ct} and Z_{3ct} should produce the same employment response. Since Z_{2ct} and Z_{3ct} rely on different forms of variation but are predicted to have the same employment impacts, this set-up lends itself naturally to an over-identification test of the validity of our identification assumptions. Recall that both Z_{2ct} and Z_{3ct} are valid under the same random-walk assumption, that the η_{ict-1}'s are uncorrelated with the $\Delta\epsilon_{ic}$'s in equations (6) and (9). If this assumption were violated, the offending correlations will be weighted differently by the two instruments (with changes in national-level industrial wages in Z_{2ct} and national-level employment changes in Z_{3ct}). This would, in turn, imply that the two instruments should result in quite different estimated coefficients if the key correlations do not equal zero. Thus, we can test our identification assumption by testing that estimation of (6) and (9) using either Z_{2ct} or Z_{3ct} produces similar results. We view this test as quite strong because Z_{2ct} and Z_{3ct} work from quite different sources of variation; in fact, in our data their correlation is only 0.18 after removing year effects.

Recall that in Equation 6 we have three explanatory variables for employment (besides the industries dummies). As we suspect all three of these variables to be potentially correlated with the error term, we need at least three instruments to estimate this equation. We have now proposed three instruments, so in principle we could move to estimation. However, we choose to propose two more sets of instruments for two main reasons. First, we want to have more instruments than variables in order to perform over identification tests, as in the absence of credible over-identification tests, we could not provide any evidence in support of the needed identification assumptions. Second, with the current set of instruments, we are worried that we will not meet the rank condition necessary for identification. In particular, since both instruments Z_{2ct} and Z_{3ct} are aimed at isolating admissible variability in wages, while Z_{1ct} is aimed mainly at isolating variability in the city-level employment rate, it is plausible that this set of instruments does not span the space necessary to isolate independent variation in all three regressors. For this reason, we now propose two sets of instruments aimed at helping isolate admissible variation in population growth.

The first of these two instruments is again of the Bartik form, and will be referred to it as Z_{4ct}. The idea behind this instrument is to use historical patterns of interstate migration to predict inflows and outflows of people to a city.[22] For example, suppose a city has a large proportion of its population at the beginning of a period which is born out of state, young, female and black. We infer that such a city is likely attractive to young, black females. Our proposed instrument is based on the prediction that such a city will grow if the out-of-state population of young female black people grows. To be more precise, Z_{4ct} is constructed as follows:

$$Z_{4ct} = \sum_j \omega_{jct-1}(1 + g_{jst}),$$

where ω_{jct-1} is the fraction of the population in city c at time $t-1$ that is both born out of state and is in demographic group j; and g_{jst} is the growth between $t-1$ and t of the out-of-state population in demographic group j. We segmented the population into 40 demographic groups based on indicators for female and black and age grouped into 5-year bins, using only those born in the U.S..[23] Note that one of the sources of variation for this instrument is the ageing of the baby boom, with this instrument predicting high population growth in cities where people of a given age group have tended to locate in the past as the baby boom moves through that age range.

Since we would like to have more than one instrument aimed at isolating variation on population growth, we also propose a second set of such instruments. However, for this latter set, instead of building on the Bartik logic, we follow the Urban Economics literature and build instruments aimed at capturing effects of local amenities. It seems natural to assume that people move in part to gain access to local amenities that may be independent of productivity. However, most amenities not related to employment and wages are relatively constant over time, making them unhelpful as instruments in our difference specification. Nonetheless, measures of amenities can still be used as instruments in this case if the value of the amenity has changed over time. For example, if the value of living in a nice climate has increased over time then the level of an indicator variable corresponding to a city having a nice climate can be used as an instrumental variable for labor force growth.[24] Building on this insight, we collected data from a number of sources to construct an instrument set consisting of average temperatures and precipitation for each city in our sample. Consistent with the idea that workers are increasingly drawn to cities by amenity factors, we find that indicators of mild climates are significant predictors of city labor force growth.[25] The city level climate variables we

[22] Reference here the immigrant enclave lit.

[23] The weights ω_{jct-1} in this case do not sum to one.

[24] This idea comes from Dahl (2002) who empirically tests a Roy (1951) model of self-selection of workers across states. He finds that while migration patterns of workers are partially motivated by comparative advantage, amenity differences across states also play a role in worker movements.

[25] The validity of the climate instruments rests on the assumption that the relationship between city climate and city-industry job creation and cost advantages (the θ_{ict}s) is constant over time. In this case, the relationship is entirely captured in time-invariant city-specific effects that are differenced out of the estimating equation. This assumption may not be valid if the evolution of these advantages are related to long-term climate conditions.

extracted are from "Sperling's Best Places to Live."[26] The variables we use are the average daily high temperatures for July and January in degrees Fahrenheit, their squares and the number of sunny and rainy days.[27]

2.1 Worker Heterogeneity

As we have emphasized, our aim in this paper is to provide an estimate of how employment decisions, on average, are affected by an across-the-board increase in the cost of labor. By its very nature, this question is about an aggregate labor market outcome. In the model developed so far workers are identical and so all parameters are "aggregate" by definition. However, in our data, workers are heterogeneous in many dimension including, among others, education and experience. We therefore need to address this heterogeneity in order to proceed appropriately. Depending on the assumptions that one makes, there are several ways to approach this issue.

The first approach, which we use for our main set of results, is to treat individuals as representing different bundles of efficiency units of work, where these bundles are treated as perfect substitutes in production. Therefore, in our baseline results we control for skill differences in wages via a rich regression adjustment and we correct for selection of workers across cities. This approach implicitly introduces an additional term in (6) which represents changes in average efficient units per worker. In our baseline specification we treat this extra term as a part of the error structure, while in the robustness section we will show that our results are not sensitive to explicitly controlling for measures of efficiency units per capita at the local level. An alternative assumption is that labor markets are segregated along observable skill dimensions and that our model applies to homogeneous workers within these markets. Thus, we also perform our analysis separately by education group as a specification check.

3 Data Description and Implementation Issues

The data we use in this paper come from the U.S. decennial Censuses for the years 1970 to 2000 and from the American Community Survey (ACS) for 2007. For the 1970 Census data, we use both metro sample Forms 1 and 2 and adjust the weights for the

[26] See http://www.bestplaces.net/docs/DataSource.aspx. Their data is compiled from the National Oceanic and Atmospheric Administration.

[27] An alternative variable available from the same source is a 'comfort' index. The comfort index is a variable created by "Sperling's Best Places to Live" that uses afternoon temperature in the summer and local humidity to create an index in which higher values reflect greater "comfort". Using this as an alternative instrument gives similar results. We have also compiled climate data from an alternative source to use as a robustness check. These data come from CityRating.com's historical weather data, and include variables on average annual temperature, number of extreme temperature days per year, humidity, and annual precipitation.

fact that we combine two samples.[28] We focus on individuals residing in one of our 152 metropolitan areas at the time of the Census. Census definitions of metropolitan areas are not comparable over time. The definition of cities that we use in this paper attempts to maximize geographic consistency across Census years. Since most of our analysis takes place at the city-industry level, we also require a consistent definition of industry affiliation. Details on how we construct the industry and city definitions are left to Appendix A.

As discussed earlier, our approach to dealing with worker heterogeneity is to control for observed characteristics in a regression context. Since most of our analysis takes place at the city-industry level, we use a common two-step procedure. Specifically, using a national sample of individuals, we run regressions separately by year of log weekly wages on a vector of individual characteristics and a full set of city-by-industry dummy variables.[29] We then take the estimated coefficients on the city-by-industry dummies as our measure of city-industry average wages, eliminating all cells with fewer than 20 observations.

Our interpretation of the regression adjusted wage measure is that it represents the wage paid to workers for a fixed set of skills. However, since we only observe the wage of a worker in city k if that worker chooses to live and work in k, self-selection of workers across cities may imply that average city wages are correlated with unobserved worker characteristics such as ability. In this case, our wage measure will not only represent the wage paid per efficiency unit but will also reflect (unobservable) skill differences of workers across cities. To address this potential concern, when we estimate our wage equations we control for worker self-selection across cities with a procedure developed and implemented by Dahl (2002) in a closely related context.

Dahl proposes a two-step procedure in which one first estimates various location choice probabilities for individuals, given their characteristics such as birth state. In the second step, flexible functions of the estimated probabilities are included in the wage equation to control for the non-random location choice of workers.[30] The actual procedure that we use is an extension of Dahl's approach to account for the fact we are concerned with cities rather than states, as in his paper, and that we also include individuals who are foreign born. When we estimate the wage equations, the selection correction terms enter significantly, which suggests that there are selection effects. Our

[28]Our data was extracted from IPUMS, see Ruggles, Alexander, Genadek, Goeken, and Schroeder, Matthew B. Sobek (2010)

[29]We take a flexible approach to specifying the first-stage regression. We include indicators for education (4 categories), a quadratic in experience, interactions of the experience and education variables, a gender dummy, black, hispanic and immigrant dummy variables, and the complete set of interactions of the gender, race and immigrant dummies with all the education and experience variables.

[30]Since the number of cities is large, adding the selection probability for each choice is not practical. Therefore, Dahl (2002) suggests an index sufficiency assumption that allows for the inclusion of a smaller number of selection terms, such as the first-best or observed choice and the retention probability. This is the approach that we follow.

results with or without the Dahl procedure are very similar. Nevertheless, all estimates presented below include the selection corrected wages.[31]

Our Z_{2ct} and Z_{3ct} instruments are constructed as functions of the national-industrial wage premia and the proportion of workers in each industry in a city. We estimate the wage premia in a regression at the national level in which we control for the same set of individual characteristics described for our first-stage wage regression and also include a full set of industry dummy variables. This regression is estimated separately for each Census year. The coefficients on the industry dummy variables are what we use as the industry premia in constructing our instruments.

The dependent variable in our analysis is the log change in industry-city-level employment. We construct this variable by summing the number of individuals working in a particular industry. Our measure of L_c for a city is the city working-age population.[32] For most of our estimates, we use decadal differences within industry-city cells for each pair of decades in our data (1980-1970, 1990-1980, 2000-1990) plus the 2007-2000 difference, pooling these together into one large dataset and including period specific industry dummies. In all the estimation results, we calculate standard errors allowing for clustering by city and year.[33]

4 Estimates of Labor Demand: Basic Results

In Table 1, we present estimates of our main equation of interest, (6). All the reported regressions include a full set of year-by-industry dummies. Column 1 reports OLS results. For the OLS results, both the coefficients on the wage and the city-level employment rate are positive and highly significant. This is the opposite of what our theory predicts for the coefficients in (6). However, the employment equation derived from the model implies that OLS estimation of this equation should not provide consistent estimates. The fact that productivity shocks, $\Delta\epsilon_{ict}$, enter the employment equation's error terms, and that wages are likely positively related to productivity, explains why the OLS regression coefficient on wages is positive.

Columns 2-4 contain results associated with estimating (6) using different sets of instruments, where we treat all three variables as endogenous. In Column (2) we use all the instruments discussed in section 2, that is, we use as instruments Z_{1ct} to Z_{4ct}, plus our local climate variables. Since we have more than one endogenous variable, we use a test suggested by Angrist and Pichke (2009) to assess the strength of our instruments in a setting with multiple endogenous variables. These tests, reported in the bottom rows of Table 1, under the heading 'AP p-val,' indicate that a weak instrument problem is

[31]Details on our implementation of the Dahl's procedure are contained in Appendix E. Results without the selection corrections are available upon request.

[32] We have verified the robustness of our results to restricting the population to include only those individuals that report themselves as being in the labor force.

[33]We cluster at the city-year level because this is the level of variation in our data. Clustering only by city has little effect on the estimates of standard errors that we report.

unlikely to be present. For completeness, we also report conventional F-statistics in the table. The F-statistics show that our instruments are particularly good at predicting wage changes and population changes.

The first aspect to note about the IV results, relative to OLS, is that now the coefficients on wages and the city-employment rate enter with the predicted negative sign. In particular, the coefficient on wages is estimated to be -0.78, while the coefficient on the employment rate is estimated to be -1.33. For population changes, we find a strongly positive relationship, with a coefficient not significantly different from 1. To explore the robustness of this last finding, we report results where we use, alternatively, either Z_{4ct} or the climate variables to help isolate admissible variation in population in Columns 3 and 4. Both instruments provide the same message; holding wages constant, an increase in the labor force is associated with a close to proportional change in employment. Recall from Section 2 that a coefficient of population growth of 1 likely indicates that there are no important fixed factors at the industry-city level beyond that associated with a span of control problem. In columns 5 and 6 of Table 1, we follow up on this result by imposing a coefficient of 1 on population growth. We implement this by using as our dependent variable the employment rate in an industry-city cell instead of the level of employment. Once we impose this restriction, we again see that the OLS estimates remain inconsistent with the theory since both the wage effect and the employment rate effect are estimated to be positive. However, once we instrument this equation using Z_{1ct}, Z_{2ct} and Z_{3ct}, in column 6, we again find a significantly negative wage elasticity (β_1) which is now very close to -1.0. Moreover, we find evidence, as suggested by search theory, of a negative congestion effect, with the effect of the local employment rate on industry-level employment (β_3) having a coefficient near -2.

Since we are especially interested in the wage elasticity of labor demand, in Table 2 we report the first-stage results for wages in order to provide support for our IV approach. From Table 2, we see that our instruments Z_{1ct} and Z_{2ct} predict wages independently, with each exhibiting a strong positive relationship. Recall that that these two instruments exploit very different data variation: in the data, we find that they are only weakly correlated (with a correlation of 0.18 once time dummies removed). Hence, they offer a good set-up for exploring over-identification restrictions. In particular, if our identification assumptions are right then we should get very similar results for the wage elasticity if we use either one of these instruments. This conjecture is confirmed in Columns 6 and 7 of Table 1, where we see that the wage elasticity is close to -1 using either set of instruments. The last row of Table 1 provides the p-value for the Hanson's J over-identification test, which can be interpreted as testing whether the regression results using the two different sets of instruments give similar results. In column 8, for example, the p-value for this test is 0.68, easily failing to reject. This indicates that results estimated using either variation from Z_2 or Z_3 are not statistically different, and is very supportive of the search theory discussed above. We view the fact that our IV estimates are both changing the coefficients quite drastically compared to OLS results and are stable across instrument sets, as strong support for our IV approach. In BGS, we show the same sort of over-identifying result for wage equations and provide a more detailed interpretation. The other key prediction from the model is that an increase in labor market tightness in a city (as represented by the city-level employment rate) should negatively affect within industry-employment rates. Once we instrument, we do, in fact, find evidence of this negative effect. This is a striking result since one may have expected a positive relationship between these variables. In our opinion, it is rather difficult to explain this later result without relying on search costs.

We are now in a position to interpret the results in terms of their implications for the wage elasticity of labor demand at the industry versus city level. First, consider a wage increase in a particular industry, holding overall employment rates constant. If the industry in question is not large enough to have a significant impact on overall employment rates, the IV estimates in Table 1 imply a labor demand elasticity at the industry level of about -1. What about wage increases for a city as a whole? Since all industries will adjust employment downward in response to a general wage increase, there will be feedback effects on overall employment rates. Allowing these equilibrium effects to play out using our estimates of equation (6) implies a city-level labor demand elasticity of $\frac{\beta_1}{1-\beta_3}$ or of about -0.30. In other words, since β_3 is predicted to be less than zero in the presence of search frictions, overall wage increases in a locality have a built in dampening effect on employment responses because they simultaneously increase the availability of workers. In our model, this leads to reduced search costs for firms. Thus, our framework suggests that the city-level labor demand curve should be less elastic than the industry level demand curve by a factor of $1-\beta_3$.

Recall that we can also obtain an estimate of the city-level demand elasticity through direct estimation of the city-level specification (9). Estimates of (9) are presented in Table 3, columns 1-4, with estimates where we use the employment rate as the dependent variable in columns 5-8. All estimations in the table contain a full set of year dummies, whose coefficients we suppress for brevity. Our IV estimates of the coefficient on log changes in average city wages, which represents an estimate of $\frac{\beta_1}{1-\beta_3}$, range from about -0.26 to -0.31. The coefficients on labor force growth in the first four columns are extremely close to 1, regardless of the source of variation we use to isolate movements in population. In the last four columns of the table, the wage elasticity obtained using Z_{2ct} and Z_{3ct} are again nearly identical to each other, and the over-identification test again fails to reject the null hypothesis associated with these being valid instruments. Thus, in this city-level specification, the results continue to support our proposed framework for studying labor demand. Furthermore, it is important to emphasize that the estimates of the city-level demand elasticities using the aggregated data are almost identical to what we just calculated using the estimated coefficients from the industry level specification (6). Since estimation of (6) and (9) use very different levels of aggregation, and since there is no mechanical reason the two specifications should provide the same results for $\frac{\beta_1}{1-\beta_3}$, we view the similarity of the estimates of the city-level elasticity obtained from the two different approaches as evidence supporting our framework.[34] Finally, note that in Table 3, Z_{1ct} has a positive and strongly statistically significant direct effect on the city-level employment rate, supporting the idea that it is a good instrument for that employment rate in the dis-aggregated equation estimation presented in Table 1.

4.1 Breakdown between traded and non-traded goods

In our model and interpretation of the data, we have assumed that all goods are traded across cities. This assumption allows us to treat the price of goods as being common across cities and, therefore, to fully capture their effects through time-varying industry effects. If there are goods produced that are not tradeable across cities, it will create a city-specific component in prices that will appear in the error term of our labor demand regressions. A simple way to get around this problem is to focus only on labor demand in tradeable goods sectors. To this end, we define tradeable and non-tradeable sectors using an approach from Jensen and Kletzer (2006). They argue that the share of employment in tradeable goods should vary widely across regional entities (cities in our case) since different cities will concentrate in producing different goods which they can then trade. For non-tradeable goods, on the other hand, assuming that preferences are the same across cities, one should observe similar proportions of workers in their production across cities. We therefore rank industries by the variance of their employment shares across cities in the 1970 Census and label the industries in the top, middle and bottom third as high-, medium- and low-trade industries.

In Table 4 and 5, we present estimates of Equation 6 carried out separately for the low-, medium- and high-trade industries. The difference between Table 4 and 5 is that the coefficient on population growth is constrained to be 1 in Table 5. The odd numbered columns of these two tables report OLS estimates, while the even number columns report IV estimates. The striking aspect one immediately notices from these two tables is the strong stability of our estimates across the different industry groupings. For example, the wage elasticity of labor demand varies only between -0.73 and -1.05 across the two tables for our IV estimations. If we focus on highly tradeable industries, we find an elasticity of -0.79 if we do not constrain the effect of population to be 1 (column 6 of Table 4), while we obtain a coefficient of -0.88 when we do constrain this coefficient (column 6 of Table 5). To push potential industry differences further, in Table 6 we report estimates of the wage elasticity of labor demand for seven common industry groupings. In the first column of this table we report OLS estimates of this elasticity and in the second column we report IV estimates. For these estimates we have constrained the coefficient of population growth to equal 1.[35] The only industry in which we do not find a significant negative wage elasticity is Agriculture, Mining and Construction. For the other

[34] Note that the OLS estimates of $\frac{\beta_1}{1-\beta_3}$ obtained from the city-level estimation is not close to that obtained from the OLS estimates at the industry city equation. This supports the claim that there is no obvious mechanical link forcing a similar result from the two estimates.

[35] We have omitted the estimates on the city-level employment rate to save space.

six industry groupings wage elasticities range between -0.73 and -1.28. The average of the IV estimates using the industry shares as weights, reported in the last row of the table, is -0.85. Hence, it appear reasonable to conclude that the wage elasticity of labor demand at the industry-city level is close to -1.

5 Robustness

In this section we explore the robustness of our results along two dimensions. In order to save space and provide more precise estimates of the wage elasticity, we report results for specifications where we impose the effect of population growth on wage to be 1. Results where we do not impose this restriction are similar but less well defined.

5.1 Breakdown by Education groups

The model we developed in section 1 conceptually applies to workers of a single skill group. In section 2.1 we discussed how we address worker heterogeneity in our baseline results by adjusting wages in accord with treating individuals as bundles of efficiency units. In this section, we report results from estimating our labor demand curve separately by education group. The education groups we consider are those with high school education or less and those with some post secondary or more.[36] When we perform this exercise, we are assuming that there are two completely segregated markets defined by education.[37] The dependent variable in Table 7 is the change in log city-industry employment for a particular education group. Similarly, wages and their instruments are constructed separately by education group.[38] In these equations, we have constrained the coefficient on population growth to equal one in order to favor more precise estimates. Columns 1-4 pertain to the low-education group and columns 5-8 to the high-education group. Inspection of the table reveals that the results for the high school educated group are very similar to those for the full sample. The results for the (smaller) college or more group are more erratic but also tend to imply a similar sized wage elasticity.

5.2 Allowing for lagged wage effects

In the derivation of our labor demand specification, we downplayed potential dynamic effects arising from adjustment costs as our goal was to derive a labor demand spec-

[36]We have assessed the sensitivity of our results to finer breakdowns in education which typically resulted in very imprecise estimates. Finer skill definitions dramatically reduce the number of city-industry cells to work with, and results in sample size problems.

[37]Empirical evidence suggests that workers within our education classes are perfect substitutes, but that there is imperfect substitution of workers between the high- and low-education groups (Card, 2009). This latter type of substitution is ruled out in this framework.

[38]For example, Z_{2ct} and Z_{3ct} are constructed using city-industry shares and national wage premia that are estimated with education specific samples.

ification appropriate for long-differences aimed at capturing the main, low frequency determinants of employment. In this section we want to briefly explore whether this perspective may be biased due to the presence of dynamic effects that could extend over periods of more than 10 years. In particular, in our theoretical framework we did not allow for potential entrepreneurs/firms to move across localities in search of low-wage areas. Firms, for example, may have gradually adjusted from the higher wage North-eastern labor market to the lower wage south and west. If this type of adjustment is present and it operates at low frequencies then this could bias our results. To explore this possibility, we re-estimate our labor demand equation allowing for the initial level of wages to affect the change in employment. The rational for this extension is that the initial wage should capture incentives for entrepreneurs to move in low-wage cities. Since our measure of initial wages is likely affected by measurement error, we will also treat the initial wage level as an endogenous variable and add to our instrument set the level of wages ten years prior. It turns out that this instrument is an extremely strong predictor of initial wage levels as suggested by the F-statistics reported in Table 8.[39] In addition, in Table 8, we have constrained the coefficient on population growth again to be 1. The first column of the table reports OLS estimates. Columns 2, 3 and 4 provide three different combinations for the instrument set.

There are two observations that emerge from these regressions. First, the estimate of the wage elasticity of employment at the city level remains close to -1. Second, there is very little evidence suggesting that initial wage levels play an important part in determining subsequent changes in employment. Although this does not imply that other types of dynamic effects are not present, it does provide some support that our rather static specification of labor demand may be appropriate for studying change in employment over decades.

6 Summary and Interpretation of Empirical results

From our estimation of Equation 6 using data over four decades, we have found strong support for the following three patterns. First, we have found a significant and robust negative wage elasticity of labor demand. This wage elasticity is estimated to be close to -1 at the industry-city level and -0.3 at the city level. Second, we have found that, holding wages constant, an increase in the size of labor force is associated with an increase in employment in a proportion close to one-to-one. Finally, we have observed that tighter labor markets at the city level reduce industry-level employment.

The issue we now want to discuss is how best to interpret these results. The finding we believe to be most interesting is the joint observation of a wage elasticity of labor demand far from infinity combined with an estimated elasticity of labor demand to pop-

[39] One drawback of using this additional instrument is that it forces us to drop the data for the 1970s.

ulation close to 1. The model presented in section 2 suggests that one should infer from the latter observation on the effects of population that fixed inputs such as, for example, land, are unlikely to be placing important constraints on employment at the local level. This estimated population effect therefore also implies that the non-infinite wage elasticity of labor demand we estimate should not be interpreted as reflecting decreasing returns to scale due to some fixed physical factor. Recall from Equation 3 that, in the absence of a fixed physical factor, the wage elasticity of labor demand should be equal to infinity if there is either an infinite supply of potential entrepreneurs or if there is no span of control problem.[40] Hence, observing a far from infinite wage elasticity of labor demand combined with a proportional effect of population on employment, implies that there is a limited supply of entrepreneurs willing to open shop in response to profit opportunities and that those entrepreneurs face span of control problems within their firms. We emphasize this finding because it is rather common in the macro-labor literature to assume that the supply of entrepreneurs is infinitely elastic with respect to any profit opportunities, while our estimates suggest that this is likely an un-warranted assumption even when looking over rather long time spans.

It is interesting to reconsider the wide range of available estimates of the elasticity of labor demand in light of these results. At one extreme, studies examining local labor market effects of migration related supply shocks tend to find large increases in the number of workers employed in the receiving labor market but little change in wages. This is what David Card found in his famous study of the 7% increase in the population of Miami generated by the Mariel boatlift. In a standard neoclassical framework, this could be interpreted as implying a nearly perfectly elastic labor demand curve. However, we argue that the population inflow would likely bring with it more entrepreneurs and that this, alone, would imply an increase in employment. Importantly, in our very general specification the resulting wage and employment changes cannot be used to identify the effects of a wage change on employment at the level of the local labor market. Instead, one would need to focus directly on mechanisms for generating reliable variation in wage costs. This is the goal of the minimum wage literature, but one might be worried that the resulting estimates are specific to the low wage labor market. We, instead, make use of insights from the search and bargaining literature to obtain identifying variation based on wage spillovers from changes in the industrial composition of a city. The resulting estimates indicate that the city-level labor demand curve is much less than perfectly elastic.

The second insight we believe should be taken away from our estimates of labor demand is the relevance of search frictions. Allowing for such frictions in the estimation of labor demand curves has certainly not been the norm. However, our results suggest they are important. In particular, we saw from our estimates of industry-city level labor demand curves that, holding wages constant, employment at the industry level decreased

[40]As noted previously, our approach does not allow us to differentiate evidence of decreasing return at the firm level between a span of control problem or a limited demand for differentiated goods.

when employment at the city increased. Viewed through the lens of our framework, this pattern also implies that the wage elasticity of labor demand at the city level should be smaller than that at the industry level, which is precisely what we found when estimating the city-level demand curve. While there may exist other explanations for such a pattern, search frictions offer a simple rationalization of the observed effects. In summary, our framework explains the rather small wage elasticity of labor demand that we estimate at the city level as reflecting a combination of three factors: deceasing returns to scale at the firm level, limited supply of new firms, and search externalities.

7 Conclusion

In this paper, we present an empirically tractable labor demand framework which incorporates several insights from the macro-labor literature. The data we use to evaluate the framework involves city-industry level observations that span over a period of four decades. Although our proposed labor demand framework is extremely parsimonious, we find considerable empirical support for it in the sense that (i) estimates of the main forces implied by the model are of the theoretically predicated sign and are statistically significant, (ii) over-identifying restrictions implied by the theory are not rejected by our data, and (iii) the results are robust and consistent across different levels of aggregation.

Our main motivation for re-exploring the issue of labor demand was to shed light on the question: how does a reduction in the labor costs borne by firms affect the employment prospects of individuals. As noted in the introduction, there remains considerable debate over this quesiton. Some researchers infer that labor demand is very elastic based on how economies react to migration flows while others infer that it is quite inelastic based on, for example, the observed effects of minimum wage changes. Our framework offers a reconciliation of these two views by separating out wage effects and population growth effects. Looking at the data through the lens of our model, we found there to be a significant negative effect of wages on employment, with an elasticity of close to -1 at the industry level and an elasticity of -0.3 at the city level. We argue that the lower elasticity at the city level is consistent with congestion externalities driven by search frictions. We also find that, holding wages constant, an increase in population is associated with a proportional increase in employment. We argue this latter pattern is consistent with the view that potential job creators are a special scarce factor because it is a scarce factor that is likely proportional to the population. An important insight we draw from our analysis is the importance of allowing a role for scarce entrepreneurial talent in the determination of labor demand.

References

ANGRIST, J. D., AND J.-S. PICHKE (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

BARTIK, T. (2009): "How Do the Effects of Local Growth on Employment Rates Vary With Initial Labor Market Conditions?," Discussion Paper No. 09-148, Upjohn Institute Working Paper.

BARTIK, T. J. (1991): *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

——— (1993): "Who benefits from local job growth: migrants or the original residents?," *Regional studies*, 27(4), 297–311.

BEAUDRY, P., D. A. GREEN, AND B. M. SAND (2012): "Does Industrial Composition Matter for Wages? A test of search and bargaining theory," *Econometrica*, 80(3), 1063–1104.

BLANCHARD, O., AND L. KATZ (1992): "Regional Evolutions," *Brookings Papers on Economic Activity. Microeconomics*.

BLAU, F. D., AND L. M. KAHN (1999): "Institutions and laws in the labor market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3, chap. 25, pp. 1399–1461.

BOUND, J., AND H. HOLZER (2000): "Demand shifts, population adjustments, and labor market outcomes during the 1980s," *Journal of labor Economics*, 8(1), 20–54.

BROWN, J., AND O. ASHENFELTER (1986): "Testing the Efficiency of Employment Contracts," *The Journal of Political Economy*, 94(3), S41–S87.

CARD, D. (1990a): "The impact of the Mariel boatlift on the Miami labor market," *Industrial and Labor Relations Review*, 43(2), 245–257.

——— (1990b): "Unexpected Inflation, Real Wages, and Employment Determination in Union Contracts," *The American Economic Review*, 80(4), 669–688.

——— (2009): "Immigration and Inequality," *American Economic Review: Papers & Proceedings*, pp. 1–21.

COOPER, R., J. HALTIWANGER, AND J. WILLIS (2004): "Dynamics of Labor Demand: Evidence from Plant-level Observations and Aggregate Implications," Discussion Paper 102974, National Bureau of Economic Research, Cambridge, MA.

DAHL, G. (2002): "Mobility and the return to education: Testing a Roy model with multiple markets," *Econometrica*, 70(6), 2367–2420.

DEATON, A., AND D. LUBOTSKY (2003): "Mortality, inequality and race in American cities and states.," *Social science & medicine*, 56(6), 1139–53.

HAMERMESH, D. (1993): *Labor Demand*. New Jersey:Princeton University Press.

HOPENHAYN, H. (1992): "Entry, Exit, and firm Dynamics in Long Run Equilibrium," *Econometrica*, 60(5), 1127–1150.

JENSEN, J., AND L. KLETZER (2006): "Tradable Services: Understanding the Scope and Impact of Services Outsourcing," in *Brookings trade forum, offshoring white-collar work*, ed. by M. Collins, and L. Brainard, vol. 6, pp. 75–134. Brookings Institution, Washington, D.C.

KING, R., AND J. THOMAS (2006): "Partial adjustment without apology," *International Economic Review*, 47(3), 779–809.

KRAMARZ, F., AND M.-L. MICHAUD (2010): "The shape of hiring and separation costs in France," *Labour Economics*, 17(1), 27–37.

KRUEGER, A., AND J. PISCHKE (1997): "Observations and Conjectures on the US Employment Miracle," Discussion Paper 390, Industrial Relations Section, Princton University.

MACURDY, T. E., AND J. H. PENCAVEL (1986): "Testing between Competing Models of Wage and Employment Determination in Unionized Markets," *Journal of Political Economy*, 94(S3), S3.

PARK, J. H. (1994): "Estimation of Sheepskin Effects and Returns to Schooling Using he Old and the New CPS Measures of Educational Attainment," Discussion Paper 717, Princeton University, Industrial Relations Section.

ROY, A. (1951): "Some thoughts on the distribution of earnings," *Oxford economic papers*, 3(2), 135–146.

RUGGLES, S., T. J. ALEXANDER, K. GENADEK, R. GOEKEN, AND M. SCHROEDER, MATTHEW B. SOBEK (2010): "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]," Discussion paper, University of Minnesota.

Table 1: Estimates of Labour Demand Equation (6)

	OLS	IV			OLS	IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \log w_{ict}$	0.14*	-0.78*	-0.79*	-0.78*	0.12*	-1.02*	-0.93*	-0.95*
	(0.016)	(0.23)	(0.26)	(0.22)	(0.016)	(0.28)	(0.23)	(0.22)
$\Delta \log \frac{E_{ct}}{L_{ct}}$	0.82*	-1.33*	-1.49*	-1.28*	0.81*	-1.83*	-2.11*	-1.98*
	(0.048)	(0.61)	(0.73)	(0.61)	(0.051)	(0.86)	(0.79)	(0.75)
$\Delta \log L_{ct}$	0.90*	0.90*	0.91*	0.89*				
	(0.011)	(0.065)	(0.086)	(0.070)				
Observations	33984	33548	33548	33984	33984	33984	33984	33984
R^2	0.59				0.51			
Instruments		Z_1, Z_2, Z_3, Z_4, CL	Z_1, Z_2, Z_3, Z_4	Z_1, Z_2, Z_3, CL		Z_1, Z_2	Z_1, Z_3	Z_1, Z_2, Z_3
F-Stats:								
$\Delta \log w_{ict}$		14.11	29.19	13.55		21.22	38.03	29.42
$\Delta \log \frac{E_{ct}}{L_{ct}}$		3.93	7.24	4.54		8.16	13.17	9.63
$\Delta \log L_{ct}$		26.28	41.29	25.37				
AP p-val:								
$\Delta \log w_{ict}$		0.00	0.00	0.00		0.00	0.00	0.00
$\Delta \log \frac{E_{ct}}{L_{ct}}$		0.00	0.00	0.00		0.00	0.00	0.00
$\Delta \log L_{ct}$		0.00	0.00	0.00				
Over-id. p-val		1.00	0.81	1.00		.	.	0.68

NOTES: Standard errors, in parentheses, are clustered at the city-year level. (*) denotes significance at the 5% level. All models estimated on a sample of 152 U.S cities using Census and ACS data for 1970-2007. The dependent variable is the decadal change in log industry-city employment (columns 1-4) log industry-city employment rates (column 5).

Table 2: First Stage Results

	OLS		
	(1)	(2)	(3)
Z_{1ct}	0.023	0.26*	0.051
	(0.073)	(0.065)	(0.076)
Z_{2ct}	3.37*		2.38*
	(0.65)		(0.64)
Z_{3ct}		3.28*	2.90*
		(0.45)	(0.41)
Observations	33984	33984	33984
R^2	0.49	0.50	0.51

NOTES: Standard errors, in parentheses, are clustered at the city-year level. (*) denotes significance at the 5% level. All models estimated on a sample of 152 U.S cities using Census and ACS data for 1970-2007. The dependent variable is the decadal change in log industry-city wages.

Table 3: Estimates of the Aggregate Labour Demand Equation (9)

	OLS	IV			OLS	IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \log w_{ct}$	0.13*	-0.27*	-0.28*	-0.27*	0.13*	-0.31*	-0.26*	-0.28*
	(0.032)	(0.090)	(0.087)	(0.088)	(0.031)	(0.13)	(0.090)	(0.082)
$\Delta \log L_{ct}$	0.99*	0.96*	0.95*	0.97*				
	(0.0090)	(0.031)	(0.041)	(0.035)				
Z_{1ct}	0.10*	0.27*	0.28*	0.25*	0.092*	0.22*	0.21*	0.21*
	(0.038)	(0.062)	(0.074)	(0.062)	(0.035)	(0.056)	(0.044)	(0.044)
Observations	608	593	593	608	608	608	608	608
R^2	0.97				0.50			
Instrument Set		Z_2, Z_3, Z_4, CL	Z_2, Z_3, Z_4	Z_2, Z_3, CL		Z_2	Z_3	Z_2, Z_3
F-Stats:								
$\Delta \log w_{ict}$		11.47	29.30	13.06		33.82	59.43	39.94
$\Delta \log L_{ct}$		14.78	20.57	13.39				
AP p-val:								
$\Delta \log w_{ict}$		0.00	0.00	0.00		0.00	0.00	0.00
$\Delta \log L_{ct}$		0.00	0.00	0.00				
Over-id. p-val		1.00	0.99	1.00		.	.	0.73

Notes: Standard errors in parentheses. (*) denotes significance at the 5% level. All models estimated on a sample of 152 U.S cities using Census and ACS data for 1970-2007. The dependent variable is the decadal change log city employment (columns 1-4) or employment rates (columns 5-8).

Table 4: Estimates of Labor Demand Equation (6) by Trade Groups

	Low Trade		Medium Trade		High Trade	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{ict}$	0.13*	-0.73	0.11*	-0.79*	0.18*	-0.79*
	(0.028)	(0.47)	(0.020)	(0.30)	(0.027)	(0.23)
$\Delta \log \frac{E_{ct}}{L_{ct}}$	0.51*	-1.96	0.78*	-1.65*	0.89*	-1.33*
	(0.13)	(1.36)	(0.068)	(0.84)	(0.065)	(0.66)
$\Delta \log L_{ct}$	0.82*	0.87*	0.84*	0.86*	0.94*	0.95*
	(0.036)	(0.15)	(0.017)	(0.098)	(0.014)	(0.082)
Observations	5230	5220	14078	13929	14676	14399
R^2	0.62		0.60		0.56	
Instrument Set		Z_1, Z_2, Z_3, Z_4		Z_1, Z_2, Z_3, Z_4		Z_1, Z_2, Z_3, Z_4
F-Stats:						
$\Delta \log w_{ict}$		16.59		21.79		34.83
$\Delta \log \frac{E_{ct}}{L_{ct}}$		3.74		6.33		8.63
$\Delta \log P_{ct}$		35.60		43.09		36.76
AP p-val:						
$\Delta \log w_{ict}$		0.00		0.00		0.00
$\Delta \log \frac{E_{ct}}{L_{ct}}$		0.01		0.00		0.00
$\Delta \log P_{ct}$		0.00		0.00		0.00
Over-id. p-val		0.33		0.95		0.84

NOTES: Standard errors, in parentheses, are clustered at the city-year level. (*) denotes significance at the 5% level. All models estimated on a sample of 152 U.S cities using Census and ACS data for 1970-2007. The dependent variable is the decadal change in log industry-city employment.

Table 5: Estimates of Labor Demand Equation (6) by Trade Groups

	Low Trade		Medium Trade		High Trade	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{ict}$	0.11*	-0.96*	0.086*	-1.05*	0.18*	-0.88*
	(0.028)	(0.44)	(0.021)	(0.29)	(0.027)	(0.18)
$\Delta \log \frac{E_{ct}}{L_{ct}}$	0.55*	-2.64	0.78*	-2.48*	0.89*	-1.59*
	(0.13)	(1.82)	(0.069)	(1.03)	(0.066)	(0.57)
Observations	5230	5230	14078	14078	14676	14676
R^2	0.59		0.54		0.44	
Instrument Set		Z_1, Z_2, Z_3		Z_1, Z_2, Z_3		Z_1, Z_2, Z_3
F-Stats:						
$\Delta \log w_{ict}$		18.00		21.87		37.43
$\Delta \log \frac{E_{ct}}{L_{ct}}$		4.34		8.09		11.71
AP p-val:						
$\Delta \log w_{ict}$		0.00		0.00		0.00
$\Delta \log \frac{E_{ct}}{L_{ct}}$		0.01		0.00		0.00
Over-id. p-val		0.34		0.87		0.68

NOTES: Standard errors, in parentheses, are clustered at the city-year level. (*) denotes significance at the 5% level. All models estimated on a sample of 152 U.S cities using Census and ACS data for 1970-2007. The dependent variable is the decadal change in log industry-city employment rates.

Table 6: Basic Results by Industry Aggregates

	OLS	IV
	(1)	(2)
Agriculture, Mining, Cons.	0.37*	0.15
	(0.057)	(0.24)
Manufacturing	0.27*	-1.28*
	(0.047)	(0.36)
Transport, Com., Util.	0.091*	-0.91*
	(0.044)	(0.41)
Retail, Wholesale	0.074*	-0.77*
	(0.021)	(0.21)
F.I.R.E	0.080*	-0.84*
	(0.037)	(0.21)
Personal, Entertainment.	0.060	-0.90*
	(0.035)	(0.32)
Professional	0.067	-0.73*
	(0.034)	(0.25)
Observations	32350	32350
R^2	0.52	
Instruments		z_1, z_2, z_3
Average	0.15	-0.85

NOTES: Standard errors, in parentheses, are clustered at the city-year level. (*) denotes significance at the 5% level. All models estimated on a sample of 152 U.S cities using Census and ACS data for 1970-2007. The dependent variable is the decadal change in regression adjusted city-industry wages.

Table 7: Estimates of Labour Demand Equation (6) By Education Group

	High School or Less				College or More			
	OLS	IV			OLS	IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \log w_{ict}$	0.056*	-0.80	-0.91*	-0.87*	0.12*	-0.68	-1.37*	-1.39*
	(0.015)	(0.43)	(0.33)	(0.30)	(0.020)	(1.45)	(0.34)	(0.36)
$\Delta \log \frac{E_{ct}}{L_{ct}}$	0.84*	-1.94	-1.80*	-1.86*	0.69*	-7.73	-3.27	-3.60
	(0.045)	(1.06)	(0.84)	(0.88)	(0.11)	(9.80)	(1.95)	(2.14)
Observations	24717	24717	24717	24717	11768	11768	11768	11768
R^2	0.48				0.50			
Instruments		z_1, z_2	z_1, z_3	z_1, z_2, z_3		z_1, z_2	z_1, z_3	z_1, z_2, z_3
F-Stats:								
$\Delta \log w_{ict}$		19.26	22.60	23.25		6.91	23.94	16.18
$\Delta \log \frac{E_{ct}}{L_{ct}}$		4.75	8.65	5.78		1.68	5.25	3.69
$\Delta \log L_{ct}$								
AP p-val:								
$\Delta \log w_{ict}$		0.00	0.00	0.00		0.02	0.00	0.00
$\Delta \log \frac{E_{ct}}{L_{ct}}$		0.00	0.00	0.00		0.31	0.00	0.01
$\Delta \log L_{ct}$								
Over-id. p-val		.	.	0.82		.	.	0.51

NOTES: Standard errors, in parentheses, are clustered at the city-year level. (*) denotes significance at the 5% level. All models estimated on a sample of 152 U.S cities using Census and ACS data for 1970-2007. The dependent variable is the decadal change in log industry-city employment (columns 1-4) log industry-city employment rates (column 5).

Table 8: Estimates of Labor Demand, Allowing for Dynamics

	OLS	IV		
	(1)	(2)	(3)	(4)
$\Delta \log w_{ict}$	0.098*	-1.23*	-1.12	-1.03*
	(0.016)	(0.54)	(0.60)	(0.46)
$\Delta \log \frac{E_{ct}}{L_{ct}}$	0.79*	-4.00	-5.82	-4.26
	(0.048)	(3.38)	(5.27)	(3.37)
w_{ict-1}	-0.070*	0.0014	0.020	-0.0059
	(0.014)	(0.13)	(0.17)	(0.13)
Observations	33984	27673	27673	27673
R^2	0.51			
Instrument Set		Z_1, Z_2, w_{ict-2}	Z_1, Z_3, w_{ict-2}	Z_1, Z_2, Z_3, w_{ict-2}
F-Stats:				
$\Delta \log w_{ict}$		11.14	29.17	22.54
$\Delta \log \frac{E_{ct}}{L_{ct}}$		1.58	1.50	1.64
w_{ict-1}		178.28	158.87	138.77
AP p-val:				
$\Delta \log w_{ict}$		0.00	0.00	0.00
$\Delta \log \frac{E_{ct}}{L_{ct}}$		0.09	0.15	0.22
w_{ict-1}				
Over-id. p-val		.	.	0.54

NOTES: Standard errors, in parentheses, are clustered at the city-year level. (*) denotes significance at the 5% level. All models estimated on a sample of 152 U.S cities using Census and ACS data for 1970-2007. The dependent variable is the decadal change in log industry-city employment rates.

A Data

The Census data was obtained with extractions done using the IPUMS system (see Ruggles, Alexander, Genadek, Goeken, and Schroeder, Matthew B. Sobek (2010)). The files were the 1980 5% State (A Sample), 1990 State, 2000 5% Census PUMS, and the 2007 American Community Survey. For 1970, Forms 1 and 2 were used for the Metro sample. The initial extraction includes all individuals aged 20 - 65 not living in group quarters. All calculations are made using the sample weights provided. For the 1970 data, we adjust the weights for the fact that we combine two samples. We focus on the log of weekly wages, calculated by dividing wage and salary income by annual weeks worked. We impute incomes for top coded values by multiplying the top code value in each year by 1.5. Since top codes vary by State in 1990 and 2000, we impose common top-code values of 140,000 in 1990 and 175,000 in 2000.

A consistent measure of education is not available for these Census years. We use indicators based on the IPUMS recoded variable `EDUCREC` that computes comparable categories from the 1980 Census data on years of school completed and later Census years that report categorical schooling only. To calculate potential experience (age minus years of education minus six), we assign group mean years of education from Table 5 in Park (1994) to the categorical education values reported in the 1990 and 2000 Censuses.

Census definitions of metropolitan areas are not comparable over time since, in general, the geographic areas covered by them increase over time and their definitions are updated to reflect this expansion. The definition of cities we use attempts to maximize geographic comparability over time and roughly correspond to 1990 definitions of MSAs provided by the U.S. Office of Management and Budget.[41] To create geographically consistent MSAs, we follow a procedure based largely on Deaton and Lubotsky (2003) which uses the geographical equivalency files for each year to assign individuals to MSAs or PMSAs based on FIPs state and PUMA codes (in the case of 1990 and 2000) and county group codes (for 1970 and 1980). Each MSA label we use is essentially defined by the PUMAs it spans in 1990. Once we have this information, the equivalency files dictate what counties to include in each city for the other years. Since the 1970 county group definitions are much courser than those in later years, the number of consistent cities we can create is dictated by the 1970 data. This process results in our having 152 MSAs that are consistent across all our sample years. Code for this exercise was generously provided by Ethan G. Lewis. Our definitions differ slightly from those in Deaton and Lubotsky (2003) in order to improve the 1970-1980-1990-2000 match.

We use an industry coding that is consistent across Censuses and is based on the IPUMS recoded variable `IND1950`, which recodes census industry codes to the 1950 definitions. This generates 144 consistent industries.[42] We have also replicated our results using data only for the period 1980 to 2000, where we can use 1980 industry

[41]See http://www.census.gov/population/estimates/pastmetro.html for details.

[42]See http://usa.ipums.org/usa-action/variableDescription.do?mnemonic=IND1950 for details.

definitions to generate a larger number of consistent industry categories.[43] We are also able to define more (231) consistent cities for that period.

B Selection Correction

The approach we use to address the issue of selection on unobservables of workers across cities follows Dahl (2002). Dahl argues that, under a sufficiency assumption, the selection-related error mean term in the wage equation for individual i can be expressed as a flexible function of the probability that a person born in i's state of birth actually chooses to live in city c in each Census year.[44] Dahl's approach is a two-step procedure that first requires estimates of the probability that i made the observed choice and then adds functions of these estimates into the wage equation to proxy for the error mean term. Dahl also presents a flexible method of estimating the migration probabilities that groups individuals based on observable characteristics and uses mean migration flows as the probability estimates. We closely follow Dahl's procedure aside from several small changes to account for the fact that we use cities rather than states and to account for the location of foreign born workers.

Dahl's approach first groups observations based on whether they are "stayers" or "movers". Dahl defines stayers as individuals that reside in their state of birth in the Census year. Since we use cities instead of states, we define stayers as those individuals that reside in a city that is at least partially located in individual's state of birth in a given Census year. Movers are defined as individuals that reside in a city that is not located in that individual's state of birth in a given Census year. We also retain foreign born workers, whereas Dahl drops them. For these workers, we essentially treat them as "movers" and use their country of origin as their "state of birth".[45] Within the groups defined as stayers, movers, and immigrants, we additionally divide observations based on gender, education (4 groups), age (5 groups), black, and hispanic indicators. Movers are further divided by state of birth. For stayers, we further divide the cells based on family characteristics.[46] Immigrants are further divided into cells based on country of origin as described above.

[43] The program used to convert 1990 codes to 1980 comparable codes is available at http://www.trinity.edu/bhirsch/unionstats . That site is maintained by Barry Hirsch, Trinity University and David Macpherson, Florida State University. Code to convert 2000 industry codes into 1990 codes was provided by Chris Wheeler and can be found at http://research.stlouisfed.org/publications/review/past/2006. See also a complete table of 2000-1990 industry crosswalks at http://www.census.gov/hhes/www/ioindex/indcswk2k.pdf

[44] This sufficiency assumption essentially says that knowing the probability of an individual's observed or "first-best" choice is all that is relevant for determining the selection effect, and that the probabilities of choices that were not made do not matter in the determination of ones wage in the city where they actually locate.

[45] We use the same country of origin groups as for the enclave instrument.

[46] Specifically, we use single, married without children, and married with at least one child under the age of 5.

As in Dahl (2002), we estimate the relevant migration probabilities using the proportion of people within cells, defined above, who made the same move or stayed in their birth state. For each group, we calculate the probability that an individual made the observed choice and for movers, we follow Dahl in also calculating the retention probability (i.e. the probability that individual i was born in a given state, and remained in a city situated at least partly in that state in general). For movers, the estimated probabilities that individuals are observed in city c in year t differ based on individuals' state of birth (and other observable characteristics). Thus, identification of the error mean term comes from the assumption that the state of birth does not affect the determination of individual wages, apart from through the selection term. For stayers, identification comes from differences in the probability of remaining in a city in ones birth state for individuals with different family circumstances. For immigrants, we assign the probability that an individual was observed in city c in a given Census year using the probabilities from immigrants with the same observable characteristics in the preceding Census year.[47] This follows the type of ethnic enclave assumption used in several recent papers on immigration, essentially using variation based on the observation that immigrants from a particular region tend to migrate to cities where there are already communities of people with their background.

Having estimated the observed choice or "first-best" choice of stayers, movers, and immigrants and the retention probability for movers, we can then proceed to the second step in adjusting for selection bias. To do this, we add functions of these estimated probabilities into the first stage individual-level regressions used to calculate regression adjusted average city-industry wages. For movers, we add a quadratic of the probability that an observationally similar individual was born in a given state and was observed in a given city and a quadratic of the probability that an observationally similar individual stayed in their birth state. For stayers, we add a quadratic of the probability that an individual remained in their state of birth. For immigrants, we add a quadratic of the probability that an similar individual was observed in a given city in the preceding Census year. Dahl allows the coefficients on these functions to differ by state, whereas we assume that they are the same across all cities.

[47] For cities in the 1980 Census not observed in the 1970 Census, we use the 1980 probabilities.