

PRODUCTIVITY AND EMPLOYMENT DENSITY:
NEW ESTIMATES AND MACROECONOMIC IMPLICATIONS*

Morris A. Davis^a, Jonas D. M. Fisher^b, and Toni Whited^c

^a University of Wisconsin-Madison, Department of Real Estate and Urban Land Economics

^b Federal Reserve Bank of Chicago

^c University of Wisconsin-Madison, Department of Finance

November, 2007

Preliminary, Please Do Not Quote

Abstract: Using MSA-level panel data on wages and employment, we estimate the impact of employment density on wages with an identification strategy that is consistent with a general equilibrium model of the spatial allocation of production. We find that a doubling of density causes the average productivity of labor to increase by between 17 and 28 percent. Our estimates are much higher than have been found previously, and we explain the difference within the context of our equilibrium model. We show that changes to density account for more than 30 percent of the growth in real wages that have occurred over the past 35 years.

*Thanks to Jeffrey Campbell, Steve Malpezzi, François Ortalo-Magné, and Tim Riddiough. Corresponding author: Morris A. Davis. Email: *mdavis@bus.wisc.edu*

1 Introduction

Cities with high wages also tend to have high employment density. Similarly, cities which experience high wage growth tend to have high employment growth also. Are these observations due to agglomeration effects on productivity, or is it just that some cities experience in-migration of high productivity workers?¹ Addressing this question has important implications for understanding the nature of economic growth and so has spawned a large literature.² Typically this literature finds that the effects of employment density on productivity are relatively small. In this paper we identify the magnitude of agglomeration on productivity and wages using the implications of a general equilibrium model of the spatial allocation of production and panel data on employment and wages across US urban areas. Contrary to previous work based on cross-sectional data, we find that agglomeration effects are very large: A doubling of employment density leads to approximately a 25% increase in productivity. We show that this implies that at least 30% of aggregate real wage growth in the US since 1969 can be attributed to the effects of agglomeration.

There are two main reasons why our estimates differ from previous findings. First, using our equilibrium framework we show that cross-section estimation strategies confound the effects of density on productivity with preferences for density. As a result, previous estimates, almost all of which employ a cross-sectional approach, correspond with some transformation of the disutility of density rather than the productivity effects of agglomeration. Consistent with this, when we combine a traditional cross-section estimation strategy with our data, we obtain much lower point estimates, in line with previous work. The second reason our findings differ from previous work is that, because we use a panel dataset, we are able to identify the relevant structural parameter

¹Agglomeration effects refer to any mechanism by which the spatial concentration of economic activity leads to increases in productivity. See Ciccone and Hall (1996) and Glaeser et. al. (1992) for discussions of the main mechanisms.

²Early studies include papers by Sveikauskas (1975), Segal (1976), and Moomaw (1981), who each attempted to measure the impact of city population on productivity. More recent work by Carlino and Voith (1991), Ciccone and Hall (1996), Ciccone (2002), and Combes et. al. (2007) measures the impact of population density on productivity. These papers tend to find that doubling city size or density increases productivity by between 3 and 8 percent. Three recent studies find effects similar in magnitude to those we find. Glaeser and Maré (2001) show that individual wages rise by 15 percent or more when workers move from a non-metropolitan to an metropolitan area. Ciccone and Peri (2006) estimate that a doubling of employment increases worker wages by about 15 percent using data from the 1970 and 1990 Decennial Census of Housing; and, Carlino et. al. (2007) show that a doubling of the density of employment leads to a 20 percent increase in patents per-capita. See Rosenthal and Strange (2004) for a thorough survey of the literature.

determining the effect of employment density on productivity. So the estimates we obtain differ from the literature because the object we estimate is different from the object estimated in the literature.

Our estimates are based on employment and wages data for 363 Census Bureau metropolitan statistical areas (MSAs) from 1969 to 2005. Using the implications of our equilibrium model and building on the work of Blundell and Bond (2000) and Arellano and Bover (1995), we develop two complimentary Generalized Method of Moments estimation strategies. Our main finding is that a doubling of employment density increases wages by between 17 and 28 percent. We are able to reconcile our findings with the previous literature by way of replicating previous findings and showing how previous estimation strategies mis-identify the relevant parameter. Furthermore, we extend our equilibrium model to show that in-migration of high productivity workers is not driving our findings.

The paper proceeds as follows. In the next two sections we outline our model and the associated identification strategy. In the fourth section we discuss the data. In the fifth section we discuss and analyze our estimation results. In the sixth section, we test to ensure our estimates are not simply proxying for unobserved changes in human capital. The last section concludes.

2 Model

Assume that there is a representative firm in MSA i that hires capital k_i and labor l_i to produce output according to the production function

$$y_i = z_i^{1-\alpha} d_i^{\delta(1-\alpha)} k_i^\alpha l_i^{1-\alpha}, \quad (1)$$

where z_i is an exogenous MSA-specific level of multi-factor productivity, call it TFP, that is common to all firms in MSA i and d_i is the density of employment in MSA i . We define density as L_i/N_i , with L_i the total number of workers in MSA i and N_i the area of the MSA. When $\delta > 0$, the representative firm's output is increasing in MSA density.

The representative firm assumes both z_i and d_i are outside of its control. Therefore, the wage rate paid to labor and the rental rate paid to capital by the representative firm in city i are

$$\begin{aligned} w_i &= (1 - \alpha) z_i^{1-\alpha} d_i^{\delta(1-\alpha)} k_i^\alpha l_i^{-\alpha} \\ r_i &= \alpha z_i^{1-\alpha} d_i^{\delta(1-\alpha)} k_i^{\alpha-1} l_i^{1-\alpha} \end{aligned} \quad (2)$$

Notice that as far as the representative firm is concerned, the effects of density on wages and rental

rates are the same as those of the exogenous level of TFP, even though in equilibrium it must be the case that the labor input of the representative firm determines overall density.

To continue, we assume that the rental rate on capital is equal to the worldwide rate times an MSA-specific deviation from the worldwide rate such that $\ln(r_i)$ can be expressed as $\ln(\bar{r}) + \left(\frac{1-\alpha}{\alpha}\right)\eta_i$. Given this assumption, the ratio of the rental rate in capital in any two MSAs can be written as:

$$\left(\frac{e^{\eta_i}}{e^{\eta_j}}\right)^{\frac{1-\alpha}{\alpha}} = \left(\frac{z_i}{z_j}\right)^{1-\alpha} \left(\frac{d_i}{d_j}\right)^{\delta(1-\alpha)} \left(\frac{k_i}{k_j}\right)^{\alpha-1} \left(\frac{l_i}{l_j}\right)^{1-\alpha}. \quad (3)$$

With terms rearranged, this is

$$\left(\frac{k_i}{k_j}\right) = \left(\frac{z_i}{z_j}\right) \left(\frac{d_i}{d_j}\right)^{\delta} \left(\frac{l_i}{l_j}\right) \left(\frac{e^{\eta_i}}{e^{\eta_j}}\right)^{\frac{-1}{\alpha}}. \quad (4)$$

From equation (2), we derive that the ratio of wages paid by the representative firm in MSAs i and j has the following expression:

$$\frac{w_i}{w_j} = \left(\frac{z_i}{z_j}\right)^{1-\alpha} \left(\frac{d_i}{d_j}\right)^{\delta(1-\alpha)} \left(\frac{k_i}{k_j}\right)^{\alpha} \left(\frac{l_i}{l_j}\right)^{-\alpha}. \quad (5)$$

Inserting equation (4) into equation (5) yields

$$\frac{w_i}{w_j} = \left(\frac{z_i}{z_j}\right)^{1-\alpha} \left(\frac{d_i}{d_j}\right)^{\delta(1-\alpha)} \left[\left(\frac{z_i}{z_j}\right)^{\alpha} \left(\frac{d_i}{d_j}\right)^{\alpha\delta} \left(\frac{l_i}{l_j}\right)^{\alpha} \left(\frac{e^{\eta_i}}{e^{\eta_j}}\right)^{-1} \right] \left(\frac{l_i}{l_j}\right)^{-\alpha}, \quad (6)$$

which reduces to

$$\frac{w_i}{w_j} = \left(\frac{z_i}{z_j}\right) \left(\frac{d_i}{d_j}\right)^{\delta} \left(\frac{e^{\eta_i}}{e^{\eta_j}}\right)^{-1}. \quad (7)$$

Taking logs, re-arranging terms, and indexing all variables at time t yields the following relationship that holds at any time t between any two MSAs i and j for wages per employee, MSA-level density, TFP, and deviations of MSA rates of return on capital from the worldwide average rate of return.

$$\ln\left(\frac{w_{t,i}}{w_{t,j}}\right) = \delta \ln\left(\frac{d_{t,i}}{d_{t,j}}\right) + \ln\left(\frac{z_{t,i}}{z_{t,j}}\right) - (\eta_{t,i} - \eta_{t,j}). \quad (8)$$

Note that the assumption of constant returns to scale production of the representative firm is not important to deriving a linear relationship between relative wages, relative density, and relative multi-factor productivity, such as in equation (8). If instead, we had assumed that production of the representative firm is equal to

$$y_i = z_i^{1-\alpha\beta} d_i^{\delta(1-\alpha\beta)} [k_i^{\alpha} l_i^{1-\alpha}]^{\beta}, \quad (9)$$

for $\beta < 1$, then our relationship between wages and density would be modified to

$$\ln \left(\frac{w_{t,i}}{w_{t,j}} \right) = \left[\frac{\delta(1 - \alpha\beta) - (1 - \beta)}{1 - \alpha\beta} \right] \ln \left(\frac{d_{t,i}}{d_{t,j}} \right) + \ln \left(\frac{z_{t,i}}{z_{t,j}} \right) - (\eta_{t,i} - \eta_{t,j}) , \quad (10)$$

assuming in this case that $\ln(r_i) = \ln(\bar{r}) + \left(\frac{1 - \alpha\beta}{\alpha\beta} \right) \eta_i$.

In the econometrics that follows, we will work with the constant returns to scale case of equation (8), although our results conceptually also apply to equation (10).³ Note that since we can not identify capital's share of income, α , we will not be able to identify $\delta(1 - \alpha)$, which is the coefficient on density in equation (1). Assuming we ultimately want to report $\delta(1 - \alpha)$, one way to proceed would be to estimate δ and then fix $\alpha = 0.3$ based on estimates of capital's share of income from the National Income and Product Accounts (NIPA) data.⁴ However, δ is of direct interest because it determines how overall labor productivity changes with changes to MSA-level density if capital is mobile across MSAs. To see this, set $\eta_i = 0$ and suppose that the worldwide rate of return is exogenously given as \bar{r} such that

$$\bar{r} = \alpha z_i^{1-\alpha} d_i^{\delta(1-\alpha)} \left(\frac{k_i}{l_i} \right)^{\alpha-1} . \quad (11)$$

and thus

$$\left(\frac{k_i}{l_i} \right) = \alpha^{\frac{1}{1-\alpha}} \bar{r}^{\frac{-1}{1-\alpha}} z_i d_i^{\delta} . \quad (12)$$

From equation (1), output per worker of the representative firm in MSA i is

$$\frac{y_i}{l_i} = z_i^{1-\alpha} d_i^{\delta(1-\alpha)} \left(\frac{k_i}{l_i} \right)^{\alpha} . \quad (13)$$

Substituting equation (12) into (13) yields

$$\frac{y_i}{l_i} = \alpha^{\frac{\alpha}{1-\alpha}} \bar{r}^{\frac{-\alpha}{1-\alpha}} z_i d_i^{\delta} . \quad (14)$$

Thus, a doubling of density in MSA i , assuming z_i and \bar{r} do not change, yields an increase in average labor productivity of $100 * \delta$ percent in that MSA. The reason that labor productivity increases by $\delta > \delta(1 - \alpha)$ percent is that the marginal product of capital increases in response to the change in density, which draws in new capital and further boosts labor productivity.

³For example, if we estimate the coefficient in brackets of equation (10) to be 0.25, and set $\alpha = 0.3$ and $\beta = 0.95$, then our estimate of δ would be 0.32.

⁴See Kydland and Prescott (1982) or Gomme and Rupert (2007).

3 Econometrics

We start by rewriting the density terms in equation (8) in terms of MSA-level employment and geographic area,

$$\ln \left(\frac{w_{t,i}}{w_{t,j}} \right) = \delta \ln \left(\frac{L_{t,i}}{L_{t,j}} \right) - \delta \ln \left(\frac{N_i}{N_j} \right) + \ln \left(\frac{z_{t,i}}{z_{t,j}} \right) - (\eta_{t,i} - \eta_{t,j}) . \quad (15)$$

We assume that the geographic areas N_i and N_j are fixed over time. This assumption is consistent with our data, discussed later.

Next, we assume that the logarithms of MSA-level wages and employment are observed with additive measurement error denoted $e_{t,i}^w$ and $e_{t,i}^L$, respectively, for MSA i in period t ,

$$\begin{aligned} \ln(w_{t,i}^o) &= \ln(w_{t,i}) + e_{t,i}^w \\ \ln(L_{t,i}^o) &= \ln(L_{t,i}) - e_{t,i}^L \end{aligned} \quad (16)$$

We assume that $e_{t,i}^w$ and $e_{t,i}^L$ are uncorrelated with the true but unobserved values $\ln(w_{t,i})$ and $\ln(L_{t,i})$. Substituting observed wages and observed employment into equation (8) yields

$$\ln \left(\frac{w_{t,i}^o}{w_{t,j}^o} \right) = \delta \ln \left(\frac{L_{t,i}^o}{L_{t,j}^o} \right) - \delta \ln \left(\frac{N_i}{N_j} \right) + \ln \left(\frac{z_{t,i}}{z_{t,j}} \right) - (\eta_{t,i} - \eta_{t,j}) + (e_{t,i} - e_{t,j}) . \quad (17)$$

For convenience, we have defined $e_{t,i} = e_{t,i}^w + e_{t,i}^L$, with $e_{t,j}$ defined analogously. We assume that $e_{t,i}$ and $\eta_{t,i}$ are i.i.d. random variables.

If $z_{t,i}$ is correlated with its lag, then it is not clear a valid instrument for direct estimation of equation (17) can be identified. The reason is that if $z_{t,i}$ is correlated with its first lag, it is also correlated with all of its lags. Thus, $z_{t,i}$ will be potentially correlated with all lagged variables, no matter how far back in history we consider; we discuss this issue in more detail in section 5.2. To produce unbiased estimates of δ , we therefore work with a transformed version of equation (17). We start by defining the level of log productivity for city i at time t as

$$\ln(z_{t,i}) = z_{0,i} + gt + \tilde{z}_{t,i}, \quad (18)$$

in which $z_{0,i}$ is the initial level of productivity for MSA i , g is the trend growth rate of productivity, and $\tilde{z}_{t,i}$ is the stationary but cyclical component of $\ln(z_{t,i})$. We assume the stationary component is AR(1),

$$\tilde{z}_{t,i} = \rho \tilde{z}_{t-1,i} + u_{t,i}, \quad (19)$$

in which $u_{t,i}$ has zero mean and variance of σ_u^2 and is independently drawn over time. Given equations (18) and (19), equation (17) can be rewritten, after subtracting ρ times once-lagged values, to

$$\begin{aligned} \ln \left(\frac{w_{t,i}^o}{w_{t,j}^o} \right) &= \kappa_{i,j} + \rho \ln \left(\frac{w_{t-1,i}^o}{w_{t-1,j}^o} \right) + \delta \left[\ln \left(\frac{L_{t,i}^o}{L_{t,j}^o} \right) - \rho \ln \left(\frac{L_{t-1,i}^o}{L_{t-1,j}^o} \right) \right] + (u_{t,i} - u_{t,j}) \\ &\quad - [(\eta_{t,i} - \rho\eta_{t-1,i}) - (\eta_{t,j} - \rho\eta_{t-1,j})] + (e_{t,i} - \rho e_{t-1,i}) - (e_{t,j} - \rho e_{t-1,j}), \end{aligned} \quad (20)$$

where $\kappa_{i,j}$ is a constant that is equal to $(1 - \rho) [(z_{0,i} - z_{0,j}) - \delta \ln(N_i/N_j)]$.

To make further headway, we now define variables with “hats” as deviations from sample averages, assuming N MSAs in the sample. For example, in the case of log observed wages and employment, the hatted variables are defined as

$$\ln \left(\widehat{w_{t,i}^o} \right) = \ln(w_{t,i}^o) - \frac{1}{N} \sum_{j=1}^N [\ln(w_{t,j}^o)] \quad (21)$$

$$\ln \left(\widehat{L_{t,i}^o} \right) = \ln(L_{t,i}^o) - \frac{1}{N} \sum_{j=1}^N [\ln(L_{t,j}^o)] \quad (22)$$

Equation (20) can then be rewritten as

$$\ln \left(\widehat{w_{t,i}^o} \right) = \rho \ln \left(\widehat{w_{t-1,i}^o} \right) + \delta \left[\ln \left(\widehat{L_{t,i}^o} \right) - \rho \ln \left(\widehat{L_{t-1,i}^o} \right) \right] + \nu_{t,i}, \quad (23)$$

where we have defined the error term $\nu_{t,i}$ as

$$\nu_{t,i} = \widehat{\kappa}_i + \widehat{u}_{t,i} - [\widehat{\eta}_{t,i} - \rho\widehat{\eta}_{t-1,i}] + [\widehat{e}_{t,i} - \rho\widehat{e}_{t-1,i}]. \quad (24)$$

Unlike the case of equation (17), the composite error term in equation (23) has a finite moving-average representation that allows us to use lagged variables as instruments. Along this line, we follow Blundell and Bond (2000), who estimate production functions at the firm level using the dynamic panel GMM-based approach in Arellano and Bond (1991) and Arellano and Bover (1995).

We estimate ρ and δ consistently using two complementary approaches. In the first approach, we efficiently combine two sets of moment conditions using GMM, an approach highly analogous to that of Blundell and Bond. We start by first-differencing equation (23), which removes the MSA-specific intercept from $\nu_{t,i}$, obtaining

$$\Delta \ln \left(\widehat{w_{t,i}^o} \right) = \rho \Delta \ln \left(\widehat{w_{t-1,i}^o} \right) + \delta \Delta \left[\ln \left(\widehat{L_{t,i}^o} \right) - \rho \ln \left(\widehat{L_{t-1,i}^o} \right) \right] + \Delta \nu_{t,i}. \quad (25)$$

We then consider as instruments the levels of $\ln \left(\widehat{w_{t-s,i}^o} \right)$ and $\ln \left(\widehat{L_{t-s,i}^o} \right)$ for $s \geq 3$. If the parameter ρ is near unity, these level variables may be weak instruments because in this case the variables

$\Delta \ln(\widehat{w_{t,i}^o})$ and $\Delta \ln(\widehat{L_{t,i}^o})$ are likely to be primarily unpredictable innovations to non-stationary processes. They cannot, therefore, be correlated with lagged instruments. Not only will low instrument quality decrease estimator precision, but, as shown in Blundell and Bond, it can also bias downward estimates of the parameters of interest in finite samples. For our second set of moment conditions, we follow Blundell and Bond, who suggest working directly with equation (23) and using as instruments $\Delta \ln(\widehat{w_{t-s,i}})$ and $\Delta \ln(\widehat{L_{t-s,i}})$ for $s \geq 2$. For these instruments to be valid, the city-specific fixed effect, $\widehat{\kappa}_i$, must be uncorrelated with future wage growth and employment growth. This assumption is likely consistent with recent evidence that Gibrat's Law can not be rejected by available data (Eeckhout, 2004).

In the second approach, we eliminate the MSA-specific intercept using the orthogonal deviations approach in Arellano and Bover (1995). First-differencing (23) to eliminate the MSA-specific intercept does not use all available information to remove the fixed effect. As an alternative, Arellano and Bover express each observation as a deviation from the average of all future observations for MSA i in the sample, reweighting each deviation to standardize the variance. In this case, $\ln(\widehat{w_{t-s,i}})$ and $\ln(\widehat{L_{t-s,i}})$ for $s \geq 2$ can be used as instruments.

4 Data

Our data on MSA-level wages and employment are from the Regional Economic Accounts as published by the Bureau of Economic Analysis (BEA).⁵ We are the first to use this data set to estimate the impact of density on productivity. The BEA data are annual and are available from 1969 to 2005. We use the entire sample in our data work.

For each year from 1969 to 2005, the BEA holds fixed the counties comprising an MSA, and thus the square miles of land area for each MSA are fixed over time. The counties comprising each MSA are consistent with the definitions issued in December 2006 by the U.S. Office of Management and Budget (OMB).⁶ The data on square miles of land for each county are taken from the web site of the U.S. Census Bureau.⁷ Thus, we compute employment density in an MSA as the number of jobs in that MSA, as reported by the BEA and discussed next, divided by the square miles of land area for the counties comprising the MSA.

⁵These data are publicly available at <http://www.bea.gov/regional/reis/>.

⁶See http://www.census.gov/population/www/estimates/metro_general/2006/List4.txt for the counties comprising each MSA in the December 2006 definitions.

⁷These data are available at <http://www.census.gov/geo/www/gazetteer/places2k.html>.

For MSA-level employment, we use the BEA variable “Wage and salary employment” and for wages we use “Average wage per job,” both of which are available as part of Table CA34 in the Regional Accounts. The BEA notes in a footnote that the employment estimates count jobs and not people, and thus any person holding more than one job is counted multiple times.⁸ The BEA computes average wage per job in a given MSA as total wage and salary disbursements in that MSA divided by wage and salary employment. Total wage and salary disbursements are, in turn, defined as earnings by place of work less supplements to wages and salaries and less proprietors’ income.

In the top two sections of Table 1 we report summary statistics for the raw levels of BEA-reported employment $L_{t,i}^o$ and average wage per job $w_{t,i}^o$ for 1970, 1980, 1990, and 2000. Note that the summary statistics in any year are for the entire sample of 363 MSAs. For example, the reported mean of $w_{t,i}^o$ in any year (row 5) is the average, across 363 MSAs, of the average wage per job in each MSA as reported by the BEA. We report statistics for the four years ending in “0” to easily illustrate trends in the data and to facilitate comparisons with statistics on employment and wages that we derive from micro data from the 1970-2000 Decennial Census of Housing in a later section. The bottom two sections of Table 1 report the summary statistics of the key variables of equation (23), deviations of log employment and log average wage per job from their sample averages.

We take away a few main results about employment and wages from this sample of 363 MSAs from the data in Table 1. There are few pronounced time-trends in the deviations of log wages and employment from average, but two changes are noticeable. First, the distribution of employment among this set of MSAs has narrowed. This can be seen from comparisons of the standard deviation of log deviations of employment from average in 1970 and 2000, shown in row 10. Also, from 1970-2000 the smallest MSAs in terms of employment have experienced more rapid employment growth relative to the average, shown in row 11. Second, the variance of the MSA-average wage per job across these 363 MSAs may have increased. The standard deviation of deviations of MSA-level log wages from the average across MSAs, row 14, has increased from 0.14 in 1970 to 0.16 in 2000.

⁸We show later that the BEA estimates of jobs in an MSA are nearly perfectly correlated with independent estimates of the number of full time workers in that MSA that we derive from micro data from Decennial Census of Housing files.

5 Results

5.1 Estimates of ρ and δ

Our GMM estimates for the full sample of MSAs over the entire 1969-2005 time period are reported in Table 2. Before analyzing the estimates, we note that our estimation procedure is designed to be essentially identical to the procedure outlined by Holtz-Eakin et al. (1988). Summary statistics that describe the correlation of the instruments and the regressors (the right-hand side of equation 23 or 25) are shown in Table 3. In general, the R^2 and F-test values shown in this table indicate the instruments are significantly correlated with the regressors.

Focusing on the p-values shown in the bottom sets of rows of Table 2, we employ three different specification tests of the model. The first is the standard J-test of the overidentifying restrictions of the model from Hansen (1982) and Sargan (1958). The second two are from Arellano and Bond (1991), who report that the power of the J-test to detect misspecification can be quite low. As a more powerful alternative they suggest a test of the serial correlation of the residuals. Under correct model specification, the residuals of (23) should only exhibit autocorrelation up to order one. The p-values for the tests of serial correlation under this specification are listed in the table under the heading “Residuals in levels.” The residuals of (25) should only exhibit autocorrelation up to order two, with p-values reported in the table under the heading “Residuals in differences.” For each model we report the $m2$ test of second order serial correlation from Arellano and Bond (1991), as well as an analogous test for third-order serial correlation, which we dub $m3$.

Consider first the results for column 1, “BB” (standing for Blundell and Bond), joint estimation of equations (23) and (25) using twice and three-times lagged differences of the regressors as instruments for equation (23) and three- and four-times lagged levels of the regressors as instruments for equation (25). We estimate that the AR(1) coefficient on $\widetilde{z}_{t,i}$, ρ , is 0.988 (annual), which is a bit higher than the standard assumption on the persistence of aggregate multi-factor productivity from the macro literature of 0.96 (Cooley and Prescott, 1995). For δ , arguably the coefficient of most interest, we estimate a value of 0.280. The standard error around this estimate is 0.013, indicating it is precisely estimated. The interpretation of this coefficient is that if employment were to double, all else equal, output per worker would increase by 28 percent. As mentioned, conventional wisdom among urban economists is that this estimate should be in the range of 3 to 8 percent. We separately estimate $-\rho\delta$ to be -0.276 , which is almost exactly equal to the product of the estimates of $-\rho$ and δ .

Before discussing the results of the J-test, m2 test and m3 tests, we consider the estimation results of the orthogonal deviations technique of Arellano and Bover (1995), shown in column 2, “OD”, of Table 2. The estimates of ρ and δ , 0.907 and 0.243, are precisely estimated but are lower than reported in column 1. Thus, the orthogonal deviations approach suggests that a doubling of density within an MSA leads to a 24 percent increase (rather than a 28 percent increase) in the average product of labor in that MSA. The estimate of $-\rho\delta$ is -0.224 , which, as with the estimates from the BB procedure, is almost exactly equal to the product of the individual coefficient estimates.

Focusing on the regression diagnostics at the bottom of Table 2, the J-test p-values suggest that the the overidentifying restrictions of the model can be rejected. We show later in this section that a likely reason the overidentifying restrictions can be rejected is that the effects of changes of density on productivity may be non-linear in density. When residuals are in levels, the m2 and m3 p-values suggest that residuals are not second- and third- order serially correlated. When residuals are expressed in differences, the p-value for the m2 test suggests there is second-order autocorrelation in the residuals – as expected – and the p-value for the m3 test suggests that residuals do not have third-order serial correlation.

In columns 3 and 4 of Table 2, the “Nonlinear” columns, we re-estimate the models of columns 1-2, except we impose the the restriction that the product of the first two coefficients, ρ and δ , is the opposite of the third, $-\rho\delta$, and estimate the model nonlinearly. Another key difference is that we cannot use the *m2* statistic from Arellano and Bond (1991), which is formulated for a linear model. Modifying this test for a linear model is straightforward because this statistic is a t-statistic, and all that is required is the variance of the second-order autocovariance of the residuals. This autocovariance is a function of moments of current and lagged values of $\ln(\widehat{w_{t-s,i}})$ and $\ln(\widehat{L_{t-s,i}})$, as well as estimates of ρ and δ .⁹ The coefficient estimates and p-values of the J-test, m2 test, and m3 test that are shown in columns 2-3 are almost identical to those for the appropriate unrestricted estimates reported in columns 1-2. This is not surprising given that in the unrestricted results, the third coefficient, is almost exactly equal to the opposite of the product of the first two coefficients.

All of the J-tests shown in Table 2 suggest that the overidentifying restrictions of the model can be rejected. We believe this occurs because of possibly nonlinear effects of density on wages. Specifically, it appears that changes to density in smaller MSAs have less of an impact on wages than

⁹To construct the required variance, we use the influence function approach in Erickson and Whited (2002) to calculate the covariance matrix of these moments and of the estimates of ρ and δ . We then use the delta method to calculate the variance of the second-order autocovariance. The procedure for the *m3* test is identical.

changes to density in larger MSAs. In Table 4, we report coefficient estimates for the “Nonlinear” case – when $-\rho\delta$ is forced to equal the negative product of ρ and δ – when we split the sample of MSAs in half by MSA-level employment in 1969.¹⁰ For the split sample, the coefficient estimates arising from the Blundell and Bond procedure are shown in columns 1 (small MSA) and 3 (large MSA) of Table 4, and the estimates arising from the orthogonal deviations approach are shown in columns 2 and 4. For both small and large MSAs, the estimate of δ is larger with the BB approach than with the OD approach. However, for both the BB and OD estimates, the estimate of δ is significantly greater for large MSAs than for the small MSAs. Further, based on the p-values of the J-tests, we would not reject the overidentifying restrictions of the model for the small MSAs. For the larger MSAs, based on the p-values of the J-tests, the case is less clear as to whether we reject the model or not.

In summary, the GMM estimates from Table 2 and Table 4 suggest the following: (1) The annual autocorrelation coefficient on city-specific multifactor productivity ρ , is somewhere between 0.90 and 0.98; (2) A doubling of employment density increases wages by between 18 and 28 percent; (3) The impact of density on wages appears to be more pronounced for larger MSAs.

5.2 Analysis

In this section, we explain why our range of GMM estimates of δ is much higher than has been found in previous research. There are two reasons we can identify. First, the use of panel data is key: As we will show, the cross-section does not easily allow for the estimation of production-function parameters separately from preference parameters. Switching from cross-sectional to panel data (with fixed effects, or variables expressed in differences) doubles simple OLS estimates of δ from 0.07 to 0.15. Second, measurement error on employment may cause an additional downward bias in the estimate of δ , such that that OLS estimates may be lower than GMM estimates. We now explain each of these ideas in turn.

Cross Section vs. Panel:

Previous studies of the impact of density on wages have typically estimated a coefficient analogous to δ using only one wave of data. The use of one wave of data implies that all identification comes from comparisons within a cross-section. In Figure 1, we graph the sequence of estimates of δ that result from OLS regressions of either log wages on log employment (the solid red line) or log wages

¹⁰For reference, Table 5 shows the outcomes of the first-stage regressions from the split sample.

on log density (the dashed blue line) when only one year of data is used in estimation at a time. The average value across years of the point estimates for each of the estimates of δ is 0.07.

Once we allow for multiple waves of data, we uncover what appear to be very different estimates of δ . Figure 2 shows the point estimates of δ that result from year-by-year OLS regressions of the first-difference of $\ln \left(\widehat{w_{t,i}^o} \right)$ on the first-difference of $\ln \left(\widehat{L_{t,i}^o} \right)$. Although the estimates fluctuate, and the standard error around each estimate is about 0.05, the average of the point estimates, shown by the dashed black line, is double the average of the cross-section OLS estimates, 0.15.

A summary of a few full-sample OLS estimators is shown in Table 6. Column 1 shows the regression coefficient when the level of log wages is regressed on the level of log employment and no fixed effects or lags are included, 0.074. When the panel aspect of the data is exploited, either through the use of fixed-effects (column 3) or through appropriate differencing as in equations (23) or (25), columns 2 or 4, the coefficient estimate doubles to about 0.15.

We believe that a simple economic model illustrates why cross-sectional data provides a different estimate of the impact of density on wages than the panel. We have in mind a straightforward extension of recent models of location choice and city size by Eeckhout (2004) and Ortalo-Magné and Davis (2007). The details of the model follow.

Production: We assume that an MSA i is populated with identical firms that all produce output according to

$$y_{t,i} = z_{t,i} \left(\frac{L_{t,i}}{N_i} \right)^\delta l_{t,i}, \quad (26)$$

where $z_{t,i}$ is the level of multi-factor productivity specific to city- i in period t , $l_{t,i}$ is the number of workers the firm chooses to employ, and $L_{t,i}/N_i$ is employment density, where N_i is land area of city i and $L_{t,i}$ is aggregate employment in city i in period t . For convenience, we will normalize $N_i = 1$ for all MSAs. The firm takes employment as outside its control, and it maximizes profits by setting the marginal product of labor equal to the wage rate, denoted $w_{t,i}$. In logs, this is

$$\ln(w_{t,i}) = \ln(z_{t,i}) + \delta \ln(L_{t,i}) . \quad (27)$$

Households: There are a measure μ of homogeneous households. Each household inelastically supplies 1 unit of labor to the marketplace. Households receive utility from consumption and disutility from employment density. Utility in city i at time t is assumed to be,

$$\ln(c_{t,i}) - f(L_{t,i}) , \quad (28)$$

Assuming households cannot save, the budget constraint for households is

$$c_{t,i} = w_{t,i} . \quad (29)$$

Households living in city i optimally choose to consume all labor income. Households are assumed to choose freely among a set of N cities as to where to live, but households choose a city before wages are realized. Thus, they choose a city among a set of cities $i = 1, \dots, N$ to maximize:

$$E [\ln (c_{t,i})] - f (L_{t,i}) . \quad (30)$$

Equilibrium: An allocation in this economy is characterized by the consumption chosen by agents in each MSA $\{c_{t,i}\}_{i=1}^N$ and the employment in each MSA $\{L_{t,i}\}_{i=1}^N$. An equilibrium is a set of wages $\{w_{t,i}\}_{i=1}^N$ and an allocation such that (1) firms maximize profits taking wages as given, (2) agents maximize expected utility, (3) total MSA employment is equal to the population, $\sum_{i=1}^N L_{t,i} = 1$, and (4) no household wants to move, i.e. all cities provide the same expected utility.

Implications for Cross Section Regressions: In equilibrium, assuming more than one city is occupied, households receive the same expected utility in every city. Given that households optimally consume their entire labor income, the following condition holds for any two cities i and j :

$$E [\ln (w_{t,i}) - \ln (w_{t,j})] = f (L_{t,i}) - f (L_{t,j}) . \quad (31)$$

From equation (27), we know:

$$E [\ln (w_{t,i})] = E [\ln (z_{t,i})] + \delta \ln (d_{t,i}) . \quad (32)$$

Since realized TFP is equal to its expectation plus a shock

$$\ln (z_{t,i}) = E [\ln (z_{t,i})] + u_{t,i} . \quad (33)$$

Given equations (32) and (33), equation (31) implies

$$\widehat{\ln (w_{t,i})} = f (L_{t,i}) - \frac{1}{N} \sum_{j=1}^N f (L_{t,j}) + \widehat{u_{t,i}} , \quad (34)$$

where the hatted variables denote deviations from average, as in the previous section: For example, $\widehat{\ln (w_{t,i})} = \ln (w_{t,i}) - \frac{1}{N} \sum_{j=1}^N \ln (w_{t,j})$.

To continue, implicitly define \bar{L}_t and \widetilde{L}_t such that

$$f (\bar{L}_t) = \frac{1}{N} \sum_{j=1}^N f (L_{t,j}) \quad (35)$$

$$\ln (\widetilde{L}_t) = \frac{1}{N} \sum_{j=1}^N \ln (L_{t,j}) .$$

Then it follows that

$$\begin{aligned}
\widehat{\ln(w_{t,i})} &= f(L_{t,i}) - f(\bar{L}_t) + \widehat{u_{t,i}} \\
&\approx f'(\bar{L}_t) [L_{t,i} - \bar{L}_t] + \widehat{u_{t,i}} \\
&= \left[f'(\bar{L}_t) \widetilde{L}_t \right] \left[\frac{L_{t,i} - \bar{L}_t}{\widetilde{L}_t} \right] + \widehat{u_{t,i}} \\
&\approx \xi \left[\ln(L_{t,i}) - \frac{1}{N} \sum_{j=1}^N \ln(L_{t,j}) \right] + \tilde{\kappa} + \widehat{u_{t,i}} \\
&= \xi \left[\widehat{\ln(\bar{L}_{t,i})} \right] + \tilde{\kappa} + \widehat{u_{t,i}} .
\end{aligned} \tag{36}$$

where ξ is defined as $\left[f'(\bar{L}_t) \widetilde{L}_t \right]$ and $\tilde{\kappa}$ is equal to $\xi \left(\bar{L}_t / \widetilde{L}_t - 1 \right)$. The last line of equation (36) shows that given our assumptions on preferences and production, a researcher running a cross-sectional regression of log wages on log density will not uncover δ , the impact of density on wages, but rather will uncover ξ , some function of preferences related to the disutility of density.

Researchers that have worked in this area know that $\ln(z_{t,i})$ and $\ln(L_{t,i})$ are likely correlated and have tried to use an instrumental variables approach in estimation. Our equilibrium analysis suggests that finding a valid instrument, call it x , is unlikely at best. The reasoning is straightforward once we combine equations (32), (33), and the last line of (36):

$$\widehat{\ln(z_{t,i})} \approx \underbrace{(\xi - \delta) \widehat{\ln(L_{t,i})}}_A + \underbrace{\tilde{\kappa} + \widehat{u_{t,i}}}_B . \tag{37}$$

For x to be a valid instrument for equation (27), two conditions must hold: (a) the instrument is correlated with the regressor in equation (27), deviation of log employment from average, implying $E \left[\widehat{\ln(L_{t,i})} x \right]$ is not equal to 0, and (b) the instrument is uncorrelated with deviation of log TFP from average, implying $E[(A + B)x] = 0$. Given the first condition, as long as ξ is not equal to δ then $E[Ax] \neq 0$. Since $\widehat{u_{t,i}}$ is a function of unforecastable shocks to expectations, and given $\tilde{\kappa}$ is a constant, $E[Bx] = \tilde{\kappa}E[x]$. Therefore, for $E[(A + B)x] = 0$ to be the case $E[Ax]$ has to be exactly equal $-\tilde{\kappa}E[x]$. It seems unlikely at best we can find an instrument such that this equality *exactly* holds – and, regardless, how would we know?. We thus conclude an instrumental variables approach to a cross-sectional regression of log wages on log density is unlikely at best to produce a consistent estimate of δ .

Implications for Regressions of First Differences. Now consider a scenario where we have two waves of data and attempt to estimate δ by first-differencing equation (27):

$$\Delta \widehat{\ln(w_{t,i})} = \Delta \widehat{\ln(z_{t,i})} + \delta \widehat{\ln(L_{t,i})} . \tag{38}$$

Suppose for simplicity that log $z_{t,i}$ is a random walk, such that

$$\ln(z_{t,i}) = \ln(z_{t-1,i}) + u_{t,i} . \quad (39)$$

The above implies that $E[\ln(z_{t,i})] = \ln(z_{t-1,i})$, and thus $u_{t,i}$ is independent of optimal location decisions for all i . Equation (38) transforms to

$$\Delta \widehat{\ln(w_{t,i})} = \delta \Delta \widehat{\ln(L_{t,i})} + \widehat{u}_{t,i} . \quad (40)$$

Since the set of $u_{t,i}$ are shocks that are realized after the household decisions are made as to where to live, estimation of equation (40) provides unbiased estimates of δ . Note that, if we had assumed $z_{t,i}$ was an AR(1) process rather than a random walk, we would have specified an estimation equation similar to equation (23).

In summary, with one wave of data we think it unlikely that researchers can produce an unbiased estimate of δ . Rather, we suspect researchers have uncovered some transformation of households' disutility for density. The reason that δ may be impossible to identify using only wave of data is that shocks are unobservable, and because of this, finding valid instruments is likely impossible. With multiple waves of data, and assumptions about the timing of decisions, it is possible to identify shocks and instruments, and therefore possible to identify δ .

Measurement Error in Employment:

In the remainder of this section, we explain why our preferred GMM estimates for δ , in the range of 0.17 to 0.28, are higher than the OLS estimates that exploit the panel nature of our data, about 0.15. At least some of the reason is due to the fact that employment is potentially measured with error; a well known result for univariate regressions is that measurement error biases down coefficient estimates.

Assuming that any measurement error in log employment, $e_{t,i}^L$, is independently distributed over time, and assuming (counterfactually) that we know ρ with certainty, we can calculate the bias for the OLS estimate of δ from a regression based on equation (23) as

$$\text{bias} = 1.0 - \frac{(1 + \rho^2) \text{Var}(e_{t,i}^L)}{\text{Var}\left[\widehat{\ln(L_{t,i}^o)} - \rho \widehat{\ln(L_{t-1,i}^o)}\right]} . \quad (41)$$

Table 7 shows the size of the bias, calculated according to equation (41), for various values of ρ and for the standard deviation of $e_{t,i}^L$. If the reported bias in the table is 0.708, then a value of $\delta = 0.25$ would be hypothetically estimated to be $0.177 = 0.25 * 0.708$ using OLS. As Table 7 shows, the OLS

estimates of δ become increasingly biased downward as the standard deviation of the measurement error increases or as ρ approaches 1.0. To add perspective to the numbers on this table, if the value of the standard deviation of $e_{t,i}^L = 0.001$, then the standard deviation on the level of employment is about 1 percent. Thus, what might appear to be relatively small amounts of measurement error in the level of employment can cause a meaningful downward bias in OLS estimates of δ .

6 Human Capital

In this final section, we check that our estimate of δ is truly capturing production externalities of density, and is not proxying for something like differences in the average level of human capital across MSAs. Our current econometric procedure allows for the fact that MSAs may permanently differ in the educational attainment (and thus productivity) of its workforce, and such differences are explicitly captured by the MSA fixed effects (the $\hat{\kappa}_i$ terms). What cannot occur, however, is that the MSAs that experience the most rapid employment and thus density (relative to the average) cannot simultaneously have experienced the most rapid change in the human capital of its workforce, also relative to the average. Such a correlation would make it appear as if the relative change in employment caused the relative increase in wages, whereas in actuality wages increased relative to the sample average because of the relative increase in the human capital of the workforce.

To see why this is true, consider an environment in which there are two types of workers in each MSA, low skill and high skill, denoted $l_{t,i}^u$ and $l_{t,i}^e$ respectively. Further, suppose that one unit of work of high-skill workers is equivalent to $1 + e$ units of work of low skill workers. Finally, suppose low and high skill workers are perfectly substitutable in production, and consider production of the representative firm in MSA i to equal

$$y_{t,i} = z_{t,i} \left(\frac{L_{t,i}}{N_i} \right)^\delta [l_{t,i}^u + (1 + e) l_{t,i}^e]. \quad (42)$$

When $e = 0$, this set-up exactly produces the equilibrium relationship of relative wages between MSAs i and j that is shown in equation (7), assuming $\eta_i = 0$ for all i .

Given the production function of equation (42), it is immediate that total wages paid by the representative firm in MSA i is equal to total output in i . Given that, we can define average wage

per employee, $\widetilde{w}_{t,i}$, as output $y_{t,i}$ divided by the total number of employees, $l_{t,i}^u + l_{t,i}^e$,

$$\widetilde{w}_{t,i} = z_{t,i} \left(\frac{L_{t,i}}{N_i} \right)^\delta \left[\frac{l_{t,i}^u + (1+e)l_{t,i}^e}{l_{t,i}^u + l_{t,i}^e} \right] \quad (43)$$

$$= z_{t,i} \left(\frac{L_{t,i}}{N_i} \right)^\delta [1 + e f_{t,i}]. \quad (44)$$

In the above equation we have denoted the fraction of high skill in the workforce as $f_{t,i}$, that is $f_{t,i} = l_{t,i}^e / (l_{t,i}^u + l_{t,i}^e)$.

After taking logs, subtracting ρ times once-lagged values, including total MSA land area as part of fixed effects, and expressing all variables as deviations from sample averages, equation (44) approximately reduces to

$$\begin{aligned} \ln(\widehat{w}_{t,i}) & \quad (45) \\ &= \rho \ln(\widehat{w}_{t-1,i}) + \delta \left[\ln(\widehat{L}_{t,i}) - \rho \ln(\widehat{L}_{t-1,i}) \right] + \widehat{u}_{t,i} + \widehat{\kappa}_i + e \left(\widehat{f}_{t,i} - \rho \widehat{f}_{t-1,i} \right). \end{aligned}$$

For our econometric procedures to yield unbiased results, $\widehat{f}_{t,i}$ should not be correlated with our instruments – either lagged levels or lagged differences, depending on whether we are working with equation (23) or (25).¹¹

We cannot proceed to test this correlation using data from the BEA Regional Accounts because these data do not have any information on the skill level or human capital of workers. So we turn to micro data from the Decennial Census of Housing (DCH).¹² The DCH are collected every ten years and contain detailed self-reported data, by person, on wage and salary income, weeks worked the previous year, and educational attainment. In our analysis we will use the four waves of DCH data (1970-2000) that overlap with our BEA sample period. The DCH also identifies the MSA of residence for each person, but for a limited sample of MSAs. We include the 81 MSAs that can be consistently identified in all four waves of the DCH data. Finally, we restrict our sample to include only persons working at least 40 weeks per year. We do this so that our DCH-based estimate of average wage per person in an MSA is likely highly correlated to the BEA estimate of average wage per job in the same MSA.

Table 8 reports regression results that compare the DCH and BEA employment and wage data for the 81 MSAs in each of the four years. Shown in the top half of the table, the DCH and BEA data on employment by MSA are highly correlated. The coefficient in a regression of log DCH

¹¹Due to data limitations, checking the validity of the instruments for equation (25) is identical to checking the validity of the instruments for the orthogonal deviations approach.

¹²The DCH data are available at the IPUMS web site, <http://www.ipums.org/usa/>.

persons working at least 40 weeks per year (called “Log DCH Employment” in the table) on the log of the number of jobs as reported by the BEA is equal or very close to 1.0 in each year. The R^2 of these regressions are 0.97 in each year except 2000 and 0.90 in 2000. The bottom half of Table 8 shows results from regressions of average wage per person working at least 40 weeks in the DCH data (“Log DCH Avg. Wages”) against the log of average wage per job from the BEA. The regression coefficients indicate that average wage per job from the BEA is less dispersed than our DCH estimate – the coefficients are 0.83 in 1970-1990 and 0.65 in 2000. However, the R^2 values from these regressions are quite high in every year, 0.84 and above. We conclude that the DCH and BEA data sets are not identical, but exhibit enough similarity in cross-sectional distributions of jobs and wages that analysis of the DCH data may be informative about the BEA Regional data.

Table 9 show some basic correlations of log employment and the fraction of workers in each MSA with educational attainment of “some college” or more.¹³ The top panel shows the coefficient estimates and standard errors of regressions of deviations of log employment from average (“relative log employment”) on the deviations of the fraction of high-skill workers from average (“relative fraction of high-skill.”) In each year of the data, the regression coefficients show that the log employment and the fraction of high-skill workers are positively correlated, and the correlation is statistically significant except in 1970. The bottom panel shows the estimates and standard errors of regressions of the decade-on-decade change in relative log employment on the change in the relative fraction of high-skill workers. We should note that almost every MSA experienced positive gains in every decade to the fraction of its employees with some college or more.¹⁴ But, as the bottom panel of the table shows, the change in the relative fraction of high-skill workers is *negatively* correlated with the change in relative log employment. Thus, before doing any other testing, we could reason that if any of our instruments are highly correlated with both the change in the relative fraction of high-skill workers and the change in employment, then estimates of δ from equation (25) might be biased down.

Finally, we test to see if our instruments are uncorrelated with $\widehat{f}_{t,i}$. In the top panel of Table 10, we report the coefficient estimates and standard errors of regressions of $\widehat{f}_{t,i}$ and $\widehat{f}_{t-1,i}$ on $\Delta \ln(\widehat{L}_{t-2,i})$ and $\Delta \ln(\widehat{w}_{t-2,i})$. This is a test that the instruments of equation (23), the “Levels” specification, are orthogonal to the part of the error term that we can identify that is related to human capital

¹³We use “some college” as a proxy for high-skill, but note that when we define high skill as workers with four years of college or more, we produce qualitatively similar results.

¹⁴The average across MSAs of the fraction of workers within an MSA (working 40 weeks or more) with some college or more increased from 0.29 in 1970 to 0.64 in 2000.

differences across MSAs. In the bottom panel of the table, we report the coefficient estimates and standard errors of regressions of $\Delta \widehat{f}_{t,i}$ and $\Delta \widehat{f}_{t-1,i}$ on $\ln(\widehat{L}_{t-3,i})$ and $\ln(\widehat{w}_{t-3,i})$. This is a test of the instruments of equation (25), the “Differences” specification.

The two tests yield mixed results. The test of the levels specification indicates that the lagged change in relative log employment is significantly correlated with the relative fraction of workers with some college, suggesting our estimates from the “Levels” specification of the previous section may have an upward bias. The test of the differences specification indicates that the lagged level in relative log employment or relative log wage is not correlated at all with the changes in the relative fraction of workers with some college: The R^2 is just about zero for both $\Delta \widehat{f}_{t,i}$ and $\Delta \widehat{f}_{t-1,i}$. Given that Table 2 shows that the “differences” and “levels” estimates both produce approximately the same estimate of δ , and given the instruments appear to be uncorrelated with the human capital variables in the differences specification, our estimates of δ may not be obviously biased due to omission of MSA-level estimates of average human capital.

7 Conclusions

A common reason given for the existence of cities is that productivity is increasing in the total number of workers in a given urban area. Previous estimates of the quantitative impact of these “urbanization externalities” have been low, such that the conventional wisdom in the field is that a doubling of population, or population density, in a given area causes somewhere between a 3 and 8 percent increase in worker productivity. Using a long panel of data and modern GMM techniques for the estimation of the coefficients of a production function, we find that a doubling of employment causes a 25 percent increase in wages. We show that the other estimates in this literature are likely biased down for two reasons. First, the other estimates typically use a single cross section in estimation. A simple model of location choice suggests that the use of multiple waves of data to identify production externalities may be critical. Second, measurement error in employment likely further biases down coefficient estimates.

The fact that gains in MSA-level employment may cause significant gains in MSA-level wages may have important implications for researchers studying other topics. Two recent examples come to mind. Ortalo-Magné and Prat (2007) and Van Nieuwerburgh and Weill (2007) examine the implications of models where MSA-level supply constraints impact house prices. In both papers it is assumed that residents in any MSA do not experience any gains in wages if there is an increase

in the total number of residents of that MSA. Fisher (2007) shows that if residential capital is complementary to business capital and labor in the production of market output, then residential investment can be shown to lead business investment in an RBC-style model, as it does in the data. Assuming residential investment encourages migration from low-employment to high-employment places, then Fisher's result may be compatible with our finding of big productivity gains arising from changes to area employment and density.

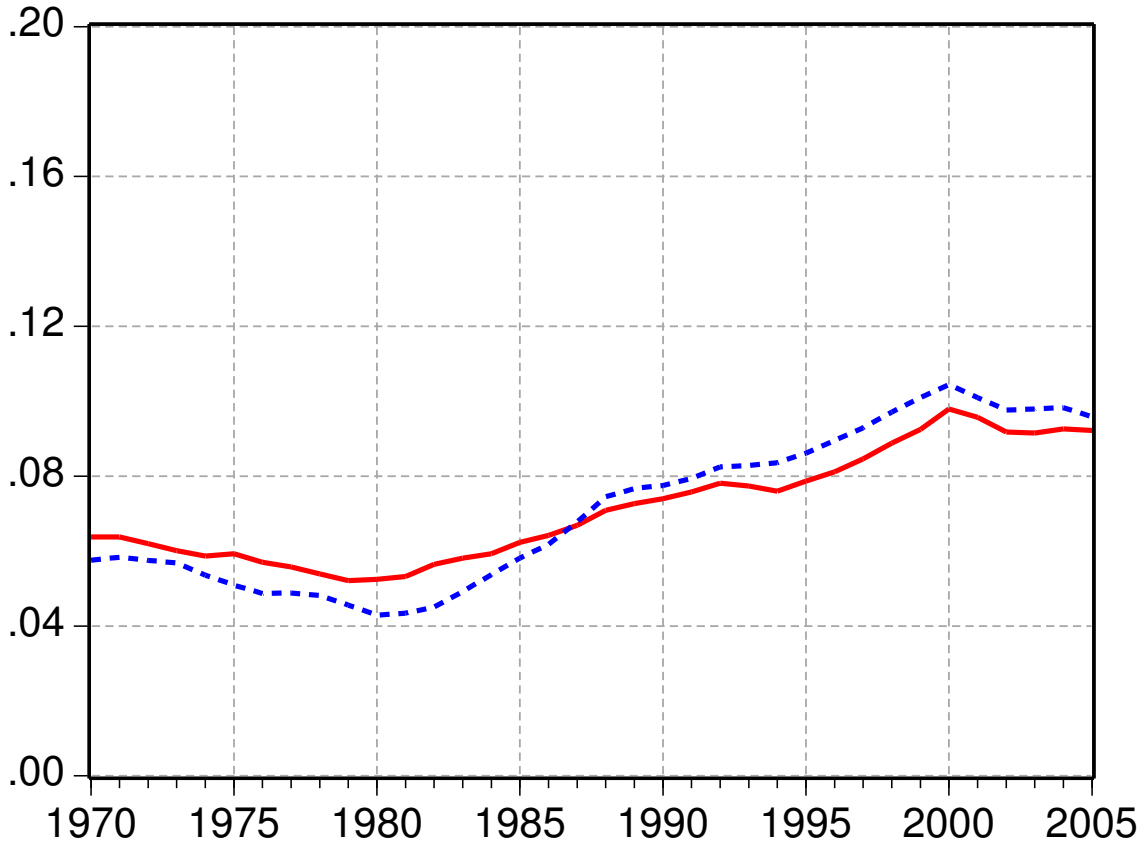
References

- [1] Arellano, Manuel and Bond, Stephen, 1991, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies* 58(2), 277-97.
- [2] Arellano, Manuel and Bover, Olympia, 1995, "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics* 68(1), 29-51.
- [3] Blundell, Richard and Stephen Bond, 2000, "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews* 19(3), 321-340.
- [4] Carlino, Gerald A., Chatterjee, Satyajit, and Hunt, Robert M., 2007, "Urban Density and the Rate of Invention," *Journal of Urban Economics* 61, 389-419.
- [5] Carlino, Gerald A. and Voith, Richard, 1992, "Accounting for Differences in Aggregate State Productivity," *Regional Science and Urban Economics* 22, 597-617.
- [6] Ciccone, Antonio, 2002, "Agglomeration Effects in Europe," *European Economic Review* 46, 213-227.
- [7] Ciccone, Antonio, and Hall, Robert E., 1996, "Productivity and the Density of Economic Activity," *American Economic Review* 86(1), 54-70.
- [8] Ciccone, Antonio, and Peri, Giovanni, 2006, "Identifying Human-Capital Externalities: Theory with Applications," *Review of Economic Studies* 73, 381-412.
- [9] Combes, Pierre-Philippe, Duranton, Gilles, and Gobillon, Laurent, 2007, "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics*, forthcoming.

- [10] Cooley, Thomas F. and Prescott, Edward C., 1995, "Economic Growth and Business Cycles," In *Frontiers of Business Cycle Research*, edited by Thomas F. Cooley. Princeton University Press.
- [11] Davis, Morris A. and Ortalo-Magné, François, 2007, "Wages and Rents: The Factor of Four," Mimeo, University of Wisconsin-Madison.
- [12] Eeckhout, Jan, 2004, "Gilbrat's Law for All Cities," *American Economic Review* 94(5), 1429-1451.
- [13] Erickson, Timothy and Whited, Toni, 2002, "Measurement Error and the Relationship between Investment and q ," *Journal of Political Economy* 108(5), 1027-1057.
- [14] Fisher, Jonas D. M., 2007, "Why Does Household Investment Lead Business Investment over the Business Cycle?," *Journal of Political Economy* 115(1), 141-168.
- [15] Glaeser, Edward L., Kallal, Hedi D., Scheinkman, Jose A., and Shleifer, Andrei, 1992, "Growth in Cities," *Journal of Political Economy* 100(6), 1126-1152.
- [16] Glaeser, Edward L., and Maré, David C., 2001, "Cities and Skills," *Journal of Labor Economics* 19(2), 316-342.
- [17] Gomme, Paul and Rupert, Peter, 2007, "Theory, Measurement and Calibration of Macroeconomic Models," *Journal of Monetary Economics* 54(2), 460-497.
- [18] Hansen, Lars P., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50(4), 1029-1054.
- [19] Henderson, J. Vernon, 1986, "Efficiency of Resource Usage and City Size," *Journal of Urban Economics* 19(1), 47-70.
- [20] Holtz-Eakin, Douglas, Newey, Whitney, and Rosen, Harvey S, 1988, "Estimating Vector Autoregressions with Panel Data," *Econometrica* 56(6), 1371-95.
- [21] Kydland, Finn E. and Prescott, Edward C., 1982, "Time to Build and Aggregate Fluctuations," *Econometrica* 50(6), 1345-1370.
- [22] Marshall, Alfred, 1920. *Principles of Economics*. Macmillan.

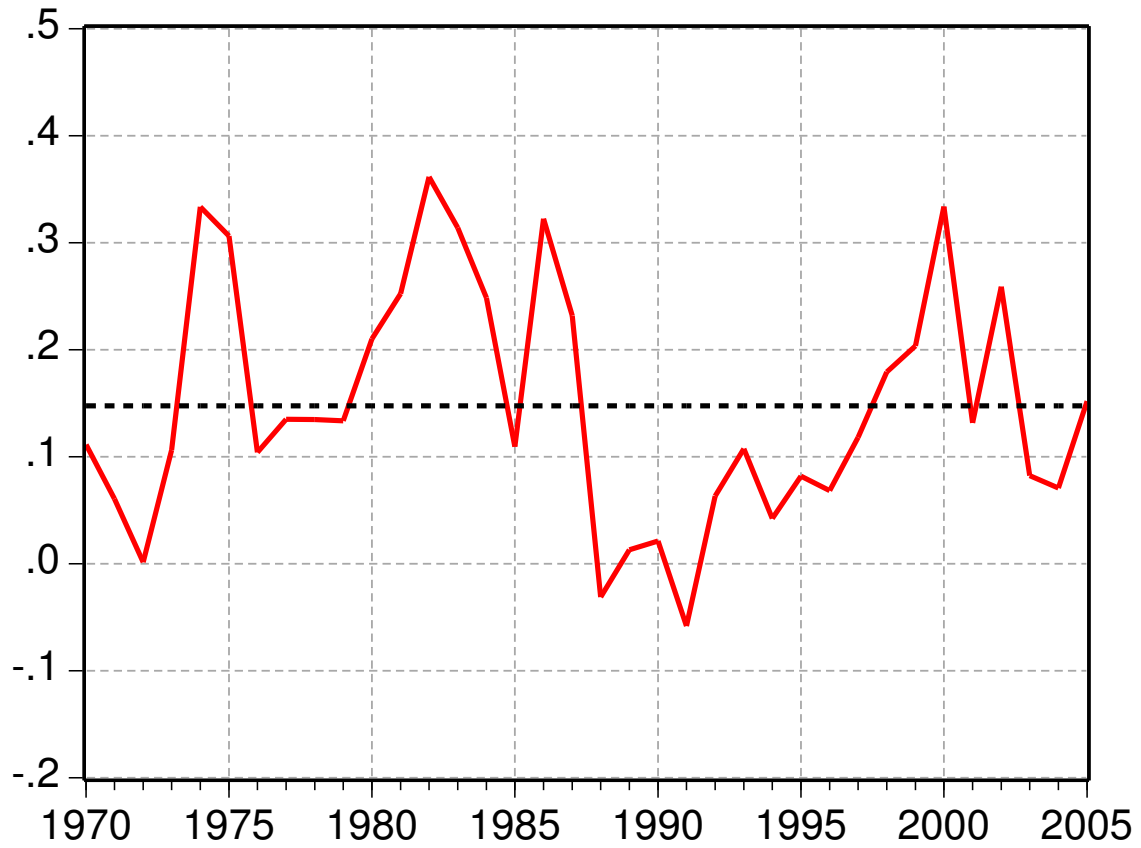
- [23] Moomaw, Ronald L., 1981, "Productivity and City Size: A Critique of the Evidence," *Quarterly Journal of Economics* 96(4), 675-688.
- [24] Ortalo-Magné, François, and Prat, Andrea, 2007, "The Political Economy of Housing Supply," Mimeo, University of Wisconsin-Madison.
- [25] Prescott, Edward C., 1986. "Theory Ahead of Business Cycle Measurement," *Quarterly Review*, Federal Reserve Bank of Minneapolis, Fall, 9-22.
- [26] Rosenthal, Stuart S. and Strange, William C., 2004, "Evidence on the Nature and Sources of Agglomeration Economies," In *Handbook of Regional and Urban Economics*, edited by J. Vernon Henderson and Jacques-Francois Thisse. Elsevier North-Holland.
- [27] Sargan, John D., 1958, "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica* 26(3), 393-415.
- [28] Segal, David, 1976, "Are there Returns to Scale in City Size?" *Review of Economics and Statistics* 58(3), 339-350.
- [29] Sveikauskas, Leo A., 1975, "The Productivity of Cities," *Quarterly Journal of Economics* 89(3), 393-413.
- [30] Van Nieuwerburgh, Stijn and Weill, Pierre-Olivier, 2007, "Why Has House Price Dispersion Gone Up?," Mimeo, New York University.

Figure 1: Year-by-Year OLS Estimates of δ , Levels Regressions, 1969-2005, 363 MSAs



Notes: The solid red line traces through the year-by-year point estimates of δ from the OLS regressions of $\ln(\widehat{w_{t,i}^o})$ on $\ln(\widehat{L_{t,i}^o})$. The dashed blue line shows the point estimates of δ when the regressor is log density, $\ln(\widehat{d_{t,i}^o})$, rather than log employment. The average across years of each of the estimates is 0.07.

Figure 2: Year-by-Year OLS Estimates of δ , Change Regressions, 1970-2005, 363 MSAs



Notes: The solid red line traces through the year-by-year point estimates of δ from the OLS regressions of the first difference of $\ln(\widehat{w_{t,i}^o})$ on the first difference of $\ln(\widehat{L_{t,i}^o})$. The solid black line shows the average across years of the estimates of δ , 0.15.

Table 1: Summary of BEA MSA-Level Data for 363 MSAs

Row	Variable	Statistic	Year:			
			1970	1980	1990	2000
1	$L_{t,i}^o$	Mean	183,655	228,176	277,748	329,154
2		S.D.	515,575	574,951	676,339	748,075
3		Min	1,305	3,327	8,072	13,375
4		Max	7,273,381	7,372,909	8,278,604	8,795,116
5	$w_{t,i}^o$	Mean	\$6,301	\$13,049	\$20,767	\$29,618
6		S.D.	\$896	\$1,782	\$2,780	\$5,387
7		Min	\$4,311	\$9,188	\$14,979	\$21,278
8		Max	\$9,636	\$22,166	\$33,796	\$73,715
9	$\widehat{\ln(L_{t,i}^o)}$	Mean	0.00	0.00	0.00	0.00
10		S.D.	1.161	1.110	1.108	1.098
11		Min	-4.02	-3.37	-2.68	-2.37
12		Max	4.60	4.33	4.25	4.12
13	$\widehat{\ln(w_{t,i}^o)}$	Mean	0.00	0.00	0.00	0.00
14		S.D.	0.142	0.132	0.128	0.158
15		Min	-0.37	-0.34	-0.32	-0.32
16		Max	0.43	0.54	0.50	0.93

Notes: $L_{t,i}^o$ and $w_{t,i}^o$ denote the number of jobs in an MSA and average wage per job in an MSA, respectively as reported by the BEA in the Regional Economic Accounts. $w_{t,i}^o$ is reported in current dollars. $\widehat{\ln(L_{t,i}^o)}$ and $\widehat{\ln(w_{t,i}^o)}$ denote, respectively, the deviation from sample average of the natural log of the number of jobs and the natural log of average wage per job.

Table 2: Production Function Estimates: GMM, Full Sample

Coefficient	Linear		Nonlinear	
	BB (1)	OD (2)	BB (3)	OD (4)
ρ	0.988 (0.005)	0.907 (0.004)	0.984 (0.003)	0.910 (0.004)
δ	0.280 (0.013)	0.243 (0.018)	0.273 (0.013)	0.248 (0.018)
$-\rho\delta$	-0.276 (0.014)	-0.224 (0.017)		
J-test p-value	0.000	0.000	0.000	0.000
Residuals in levels:				
m2 test p-value	0.490	0.483	0.490	
m3 test p-value	0.491	0.486	0.491	
Residuals in differences:				
m2 test p-value	0.006		0.006	0.489
m3 test p-value	0.113		0.112	0.492

Notes: In column 1, “BB” (Blundell and Bond), we estimate equation (23) using twice and three-times lagged differences of the regressors as instruments and equation (25) using three- and four-times lagged levels of the regressors as instruments. In the second column, we estimate equation (23) using the orthogonal deviations (“OD”) technique of Arellano and Bover (1995). Columns 3 and 4 are the same as 1 and 2, respectively, except in estimation $-\rho\delta/(1 - \alpha)$ is forced to be the product of $-\rho$ and δ . Standard Errors are in parentheses. The J-test is a chi-square test of the overidentifying restrictions of the model. The m2 and m3 tests are tests of second- and third-order serial correlation of the residuals.

Table 3: Instrument Quality, Full Sample

Instruments in Levels, Regressors in Differences		
Regressor	R^2	F
$\Delta \log(\widehat{w_{t-1}^o})$	0.112	11.333
$\Delta \log(\widehat{L_t^o})$	0.143	15.020
$\Delta \log(\widehat{L_{t-1}^o})$	0.142	14.880
Instruments in Differences, Regressors in Levels		
Regressor	R^2	F
$\log(\widehat{w_{t-1}^o})$	0.138	14.390
$\log(\widehat{L_t^o})$	0.084	8.277
$\log(\widehat{L_{t-1}^o})$	0.085	8.331
Instruments in Levels, Regressors in Orthogonal Deviations		
Regressor	R^2	F
$\log(\widehat{w_{t-1}^o})$	0.375	53.772
$\log(\widehat{L_t^o})$	0.278	34.478
$\log(\widehat{L_{t-1}^o})$	0.281	34.994

Notes: The top panel refers to the first-stage regression results of the instruments of (25) using three- and four-times lagged levels of the regressors as instruments. The middle panel refers to first-stage regression results of the instruments of equation (23) using twice and three-times lagged differences of the regressors as instruments. The bottom panel refers to first-stage regression results of the instruments of equation (23) using instruments appropriate for the Arellano and Bover (1995) orthogonal deviations technique.

Table 4: Production Function Estimates: GMM, Nonlinear, Split Sample

Coefficient	Small MSAs		Large MSAs	
	BB (1)	OD (2)	BB (3)	OD (4)
ρ	0.983 (0.006)	0.893 (0.007)	0.975 (0.004)	0.928 (0.005)
δ	0.216 (0.018)	0.173 (0.021)	0.267 (0.015)	0.240 (0.020)
J-test p-value	0.682	0.397	0.051	0.018
Residuals in levels:				
m2 test p-value	0.495	0.489	0.485	
m3 test p-value	0.500	0.484	0.482	
Residuals in differences:				
m2 test p-value	0.094		0.012	0.497
m3 test p-value	0.087		0.441	0.484

Notes: Columns 1 and 2 are the same as columns 3 and 4 of Table 2, except coefficients are estimated only for small MSAs, with sample divided in half by employment levels in 1969 (see paper for definition). Columns 3 and 4 are the same as 1 and 2, except coefficients are estimated only for large MSAs. Standard Errors are in parentheses. The J-test, m2, and m3 tests are as described in Table 2.

Table 5: Instrument Quality, Split Sample

Instruments in Levels, Regressors in Differences				
Regressor	Small MSAs		Large MSAs	
	R^2	F	R^2	F
$\Delta \log(\widehat{w_{t-1}^o})$	0.107	5.294	0.212	11.914
$\Delta \log(\widehat{L_t^o})$	0.079	3.815	0.130	6.616
$\Delta \log(\widehat{L_{t-1}^o})$	0.081	3.886	0.129	6.533
Instruments in Differences, Regressors in Levels				
Regressor	Small MSAs		Large MSAs	
	R^2	F	R^2	F
$\log(\widehat{w_{t-1}^o})$	0.103	5.093	0.175	9.389
$\log(\widehat{L_t^o})$	0.177	9.541	0.161	8.462
$\log(\widehat{L_{t-1}^o})$	0.178	9.557	0.155	8.141
Instruments in Levels, Regressors in Orthogonal Deviations				
Regressor	Small MSAs		Large MSAs	
	R^2	F	R^2	F
$\log(\widehat{w_{t-1}^o})$	0.405	30.140	0.354	24.281
$\log(\widehat{L_t^o})$	0.423	32.393	0.257	15.281
$\log(\widehat{L_{t-1}^o})$	0.427	33.034	0.260	15.522

Notes: See Table 3 for details.

Table 6: City-level Production Function Estimation: OLS

Coefficient	No Fixed Effects		Fixed Effects	
	(levels) (1)	(differenced) (2)	(levels) (3)	(differenced) (4)
ρ		0.987 (0.003)		0.938 (0.006)
δ	0.074 (0.001)	0.161 (0.010)	0.147 (0.007)	0.171 (0.011)
$-\rho\delta$		-0.159 (0.010)		-0.149 (0.011)
R^2	0.365	0.987	0.884	0.988

Notes: Columns 1 and 2 are the coefficient estimates, standard errors (in parentheses), and R^2 values resulting from simple OLS for equations (17) and (23), respectively. Columns 3 and 4 are the same as 1 and 2, but include MSA-level fixed effects.

Table 7: Bias for OLS Estimate of δ Due to Measurement Error in Log Employment

ρ	Standard Deviation of e_t^L		
	0.006	0.008	0.010
1.00	0.895	0.813	0.708
0.98	0.934	0.882	0.816
0.95	0.980	0.965	0.945
0.91	0.994	0.988	0.982

We report the bias of δ from an OLS regression of equation (23), under different assumptions of ρ and the standard deviation of measurement error to log employment, e_t^L . For example, if we report the bias to be 0.708, a value of $\delta = 0.25$ would be hypothetically estimated to be $0.177 = 0.25 \cdot 0.708$ using OLS.

Table 8: Comparisons of DCH and BEA data for 81 MSAs, 1970-2000

Regression	Year	Coefficient	
		on Log BEA data	R^2
Log DCH Employment	1970	1.02	0.97
	1980	1.00	0.97
	1990	1.00	0.97
	2000	0.96	0.90
Log DCH Avg. Wages	1970	0.84	0.89
	1980	0.83	0.90
	1990	0.83	0.86
	2000	0.65	0.84

Notes: In the top half of this table, we report the results of regressions of log employment by MSA, as we measure using DCH data for each of the years listed, against log wage and salary employment by MSA as reported in the BEA for the same years. In the bottom half, we report the results of regressions of log average wage per person by MSA, as we measure using DCH data, against log average wage per job by MSA as reported in the DCH.

Table 9: Correlations of $\widehat{\ln(L_{t,i})}$ and $\widehat{f_{t,i}}$, DCH data for 81 MSAs, 1970-2000

Regressions of $\widehat{\ln(L_{t,i})}$ on $\widehat{f_{t,i}}$				
	1970	1980	1990	2000
Coefficient	1.11	3.08*	3.66*	3.99*
Standard Error	(1.54)	(1.42)	(1.13)	(1.11)

Regressions of $\Delta\widehat{\ln(L_{t,i})}$ on $\Delta\widehat{f_{t,i}}$				
	1970	1980	1990	2000
Coefficient	x	-2.70*	-0.76	-2.63*
Standard Error	x	(0.79)	(0.48)	(1.22)

Notes: The top panel shows the regression coefficients and standard errors from a regression of $\widehat{\ln(L_{t,i})}$ on $\widehat{f_{t,i}}$. The bottom panel shows coefficients and standard errors from a regression of $\Delta\widehat{\ln(L_{t,i})}$ on $\Delta\widehat{f_{t,i}}$. An asterisk denotes the coefficient estimate is statistically significant at the 5% level.

Table 10: Regressions of Human Capital on Instruments, DCH data for 81 MSAs, 1970-2000

Regressions of $\widehat{f}_{t,i}$ and $\widehat{f}_{t-1,i}$ on Instruments ("Levels" specification, equation 23)				
Variable/Instrument	$\Delta \ln(\widehat{L}_{t-2,i})$	$\Delta \ln(\widehat{w}_{t-2,i})$	R^2	F
$\widehat{f}_{t,i}$	0.067* (0.034)	-0.110 (0.121)	0.05	2.03
$\widehat{f}_{t-1,i}$	0.108* (0.039)	-0.124 (0.123)	0.09	3.88

Regressions of $\Delta \widehat{f}_{t,i}$ and $\Delta \widehat{f}_{t-1,i}$ on Instruments ("Differences" specification, equation 25)				
Variable/Instrument	$\ln(\widehat{L}_{t-3,i})$	$\ln(\widehat{w}_{t-3,i})$	R^2	F
$\Delta \widehat{f}_{t,i}$	0.004 (0.004)	-0.064 (0.044)	0.02	1.09
$\Delta \widehat{f}_{t-1,i}$	0.001 (0.004)	0.006 (0.039)	0.00	0.04

Notes: The top panel ("Levels") shows coefficients and standard errors from regressions of $\widehat{f}_{t,i}$ and $\widehat{f}_{t-1,i}$ on $\Delta \ln(\widehat{L}_{t-2,i})$ and $\Delta \ln(\widehat{w}_{t-2,i})$. The bottom panel ("Differences") shows coefficients and standard errors from regressions of $\Delta \widehat{f}_{t,i}$ and $\Delta \widehat{f}_{t-1,i}$ on $\ln(\widehat{L}_{t-3,i})$ and $\ln(\widehat{w}_{t-3,i})$. An asterisk denotes the coefficient estimate is statistically significant at the 5% level.