Journal of Urban Economics **51,** 542–562 (2002)
doi:10.1006/juec.2001.2257, available online at http://www.idealibrary.com on IDEAL®

Urban Wages and Labor Market Agglomeration

William C. Wheaton[1] and Mark J. Lewis

Department of Economics, MIT, Cambridge, Massachusetts
E-mail: wheaton@mit.edu

Received January 22, 2001; revised September 10, 2001; published online December 11, 2001

Using the 5% public use micro sample of the 1990 U.S. census, we find that observationally equivalent workers in the manufacturing sector earn higher wages when they are in urban labor markets that have a larger share of national or metropolitan employment in *their same* occupation and industry groups. Quantitatively, the effect is large, with an elasticity (measured at the means) of between 1.2 and 3.6 for these effects. We interpret the willingness of firms to pay more for equivalent workers in dense markets as evidence of an agglomeration economy in urban labor.

I. INTRODUCTION

The last decade has seen renewed interest in the idea that the traded goods sector of urban areas exhibits ''increasing returns'' because of some form of agglomeration economy. The characterization of these advantages is typically broken down into three categories: localization economies resulting from the concentration of similar economic activity, urbanization economies from the concentration of diverse economic activity, and establishment economies from plant level increasing returns. Such agglomeration may produce one time or ''static'' increases in productivity as well as greater rates of innovation, technological change, and hence productivity growth. When these latter dynamic economies result from localization they are referred to as MAR (Marshall–Arrow–Romer) externalities. When they are based on urbanization they are called Jacobs externalities (after Jacobs [11]). Empirical evidence for agglomeration economies has been found by many researchers. Early studies found evidence that both localization and urbanization generated static productivity improvements (e.g., Carlino [1], Moomaw [15], Henderson [7]). More recently, Ciccone and Hall [2] find evidence that it may be the spatial density of activity that boosts productivity. Studies testing for dynamic impacts on rates of economic growth include Glaeser *et al*. [4] which finds evidence for Jacobs

[1] Communications should be directed to William Wheaton, E52-252b, MIT, Cambridge, MA 02139. Phone: (617) 253-1723.

0094-1190/01 $35.00

externalities, Henderson *et al.* [9] for both MAR and Jacobs externalities, Henderson [8] for localization economies, and Nakamura [17] for localization and urbanization economies.

The theoretical underpinnings for urbanization economies have focused on the argument that large cities can provide more direct industrial linkages and support services (Jacobs [11]). The theory underlying localization economies has emphasized two quite different ideas. The first is that knowledge is transferred between firms in the same industry through direct contact or spatial proximity (Jaffee *et al.* [10]). The second is that labor market search and matching are improved with scale, that this encourages workers to specialize more, which then increases both productivity and its rate of growth. This latter view has been best formalized by Kim [13]. To our knowledge there has been little successful effort to disentangle these two arguments, except perhaps for Jaffee *et al.*'s direct evidence of spatial linkages in patents.

In contrast with this previous work, we test more directly the idea that it is local labor markets which form the basis of increasing returns. If workers are more productive, then it is wages which should reflect this gain. Further, the labor market gains hypothesized by Kim [12] are directly of a localization form: when there are more workers in a particular industry/occupation (within a labor market area) then matching is improved, workers opt to increase the ''depth'' rather than ''breadth'' of their human capital, and this is what generates productivity and innovation. The novel feature of our research is the creation and use of two new localization variables: the fraction of SMA employment that is in the *same* industry and occupation as the observed worker (*specialization*), and the fraction of national employment in the observed worker's industry/occupation that is in the worker's SMA (*concentration*). Theories based on economies from labor market operation suggest that improved matching and human capital accumulation should be based on the size of one's *own* occupation/industry labor pool. Thus, for example, computer programmers should specialize and be more productive when a labor market is deep in *their* services. Having lots of workers in other, unrelated industry/occupations might generate some more general kind of urbanization economy, but it is difficult to imagine that it improves the operation of the labor market for computer programmers.

In our study, we use the 1990 U.S. census and test for these agglomeration economies by creating measures of both industry and occupation localization for each metropolitan worker. The large number of observations (over 400,000 metropolitan manufacturing workers) allows us to measure agglomeration at a finer level than the two-digit SIC industry codes used in many other studies. Also we are able to test for agglomeration economies by occupation as well as industry. One drawback to this approach is that we must use hourly wage data imputed from annual income. However, we feel that any error in computed wage estimates is less likely to be correlated with the agglomeration variables than the types of measurement error which plague capital and output calculations. As Ciccone and Hall [2] comment, most agglomeration studies, ‘‘are seriously flawed by their reliance on unsatisfactory measures of output from the Census of Manufactures.’’

It might be argued that productivity differences should not be inferred from wage differences. For example, there is ample evidence of inter-industry wage differentials that are related to institutional features (such as unionization) and not to productivity. Likewise, area specific wages have long been associated with cost of living differences. We control for these issues by including both industry and area specific effects in our wage equations and believe that within industries, the geographical mobility of production is strong enough so that some productive advantage must underlie the decision to employ more costly labor.

This is not the first study to use wage data to investigate agglomeration economies. Using 1980 census data, Rauch [18] tests for human capital spillovers in cities by linking individual wage effects to the average years of education and experience in one’s city. Maré [14] extends this paper and Glaeser and Maré [3] use wage data to investigate the wage premium paid workers in larger cities. There also are related studies in other fields. For example, Hanson [6] uses average wage data from Mexican states to test the predictions of increasing returns-based trade theories. Again, our innovation lies in testing directly for the impact of labor market scale on worker wages by using same industry/occupation employment, not just city size or urbanization.

Our basic results are:

- In the 1990 Census, our measure of occupational specialization across 220 SMAs and 424 occupations varies from zero to 7%. Our estimated wage equation has this sample range generating 23% higher wages.
- In the same sample, our measure of occupational concentration varies from zero to 46% which yields wages that are 12% greater.
- Our measure of industrial specialization across 220 SMAs and 77 (three digit SIC) categories varies from zero to 68%, with the wage equations generating 30% higher wages across this range.
- Industrial concentration varies from zero to 41%, and this range leads to 16% higher wages.

The paper is organized as follows. In the next two sections, we review a number of theoretical issues that arise when estimating cross-section wages equations, focusing on whether rents in addition to wages capture productivity differences, and whether omitted ability bias might be an issue. The fourth section goes over our data, construction of variables, and model specification. The fifth section gives our basic results under a number of alternative specifications.

II. WAGES VERSUS RENTS

Since the model of Roback [19], it has been realized that the inter-metropolitan markets for labor and land are closely linked. If firms and workers are to be indifferent across SMAs then two ‘‘prices’’ are needed to achieve equilibrium: the wages paid workers and the rent paid to land. In Roback’s model, worker utility depends on the wage level, the price of land, and a vector of potential lifestyle or amenity characteristics. This generates an upward sloping wage−rent indifference curve that shifts upward for areas with ‘‘less desirable’’ amenities. For firms, equal profit requires a downward sloping wage−rent curve that in turn shifts upward if the region has lower unit costs or is otherwise more attractive to firms. Thus in full equilibrium, SMAs that are more attractive only to firms have both higher wages and rents, while areas with only more desirable lifestyle have lower wages and higher rents. Various combinations of firm attractiveness and lifestyle will generate different configurations of wages and rents.

The outcome of this interregional equilibrium is a set of reduced form equations that relate wages and rents (on the left hand side) to some set of firm cost variables and worker lifestyle amenities on the right hand side. Most estimate wage equations represent one of these reduced forms. It is important to note that any well specified wage equation should always include SMA level variables that represent potential cost and amenity variables. Also, if a wage equation were to include any direct measure of housing costs or other cost-of-living variable (which could proxy for land rent), then it would turn into a structural equation for either firms or workers. This of course would raise an identification issue.

If industry or occupational specialization generates lower unit costs through greater firm productivity, *and if workers are perfectly mobile*, then the reduced form impacts from the Roback model allow us to say something about the *structural* impact of this productivity enhancement on firm costs. Given the two schedules in Roback’s model, the equilibrium impact of a productivity variable (such as specialization) on labor wages will always be *less than* the magnitude that this same variable actually shifts the firm’s iso-cost (i.e., wage−rent) schedule. The upward sloping worker iso-utility schedule dampens the structural impact in determining the equilibrium outcome. *The reduced form impact of a productivity variable on wages should be a lower bound on the true structural impact of that productivity variable on actual firm costs*. This is shown in the comparison of equilibrium A with B in Fig. 1.

In comparing equilibrium A with B, it is clear that specialized or concentrated areas should have higher real estate rents as well as wages. While larger cities have higher rents, we know of no evidence that this is true, in general for MSA economies that are more specialized or concentrated. Put differently, if specialization generates higher wages, but not higher rents, then why don’t all

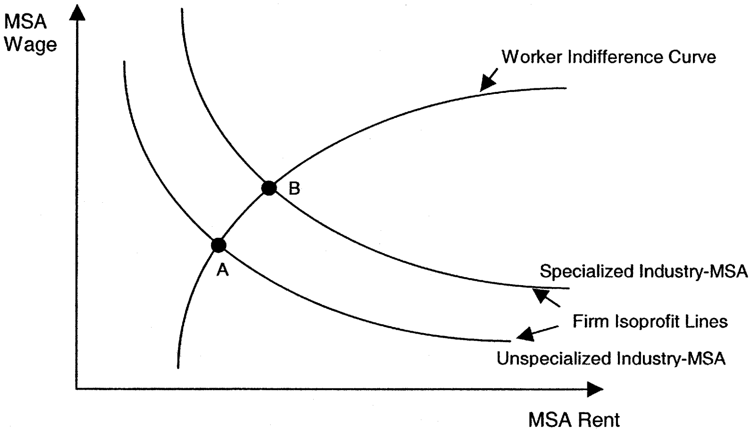

FIGURE 1

workers in a particular industry or occupation specialize? One answer is that the worker wage–rent indifference curve is close to being vertical because wages matter much more than housing costs. In this case, rents do differ, but by small amounts that might escape empirical scrutiny. Another answer is that there could be imperfect or idiosyncratic worker mobility. If not all workers in a given industry or occupation wish to live in the specializing region, then as specialization proceeds, the marginal worker finds the region less lifestyle desirable. The end result is that the marginal worker is indifferent between the higher wages (in the specialized region) and the lower wages in his or her more ‘‘lifestyle preferred’’ region.

But what about other exogenous amenity variables that might shift worker utility or other factors outside of the labor market that might affect firm productivity? In our analysis, these are easily accounted for with SMA level structural effects. We should expect that the inclusion of these fixed effects will somewhat reduce the impact of our measures of labor specialization and concentration, but certainly not eliminate them. This is because our measures are *worker specific* and reflect differences across industries and occupations within the same SMA as well as across SMAs.

III. OMITTED ABILITY BIAS

While we are able to control for all relevant observable personal characteristics in our wage equation, any such equation potentially omits unobserved ‘‘ability’’ or ‘‘motivational’’ factors that logically do impact labor market outcomes. If such omitted variables are also correlated with any right hand side variables that measure productivity, then bias can result. This criticism has been especially raised about the so-called ‘‘city size wage premium.’’ It is plausible, for example, to imagine that workers with above average ability and motivation have a ‘‘lifestyle’’ preference for living in larger SMAs. Glaeser and Maré [3], however, reject the argument that such omitted ability bias explains the large city wage premium. They do so because including direct measures of ability (the AFQT test), instrumenting for place of birth, and then studying wages separately for recent urban migrants all fail to eliminate the wage premium.

Similarly, we do not believe that omitted ability is likely to be correlated with our measures of agglomeration. It must be remembered that our measures reflect differences between industry/occupations within SMAs as well as differences between SMAs within industry/occupations. Thus, any argument for such a correlation must be based on some tendency for workers of higher ability—in a given occupation/industry—to want to reside in an area where there are many workers in this same industry/occupation. Alternatively, workers of higher ability in a given SMA must want to select an industry/occupation which is highly prevalent or concentrated in that SMA. Furthermore, this tendency has to be fairly common or universal across all occupation and industry groups.

Suppose, however, that such an argument did exist, for example from labor market search theory. The argument would be that higher ability workers are attracted to markets that are dense in their own industry/occupation, presumably because such markets offered better opportunities for them to achieve or stand out. Alternatively, higher ability workers choose the dense occupations/industries within their SMA for the same reason. But in these cases, surely the increased return is due as much to the density of the market as to the worker’s higher ability. They represent true joint products that would not refute the existence of labor market agglomeration. Omitted ability bias is a problem only if the correlation between ability and labor market scale results from non-agglomeration related factors. In this sense, we know of no such argument but are open to ideas.

IV. DATA AND SPECIFICATION

The data used to estimate the wage equations are from the 1990 United States Census 5% Public Use Micro Sample.[2] The sample includes both males and females, aged 16 to 65, who worked in the private sector for wages or salary. Only individuals who reported working 50 or more weeks in 1989 and who usually work 35 or more hours per week are included. The sample excludes workers who do not speak English, who have a disability which limits or prevents work, or who reported usually working more than 98 hours per week (99 hours is the maximum coded value of this variable). To allow the addition of specialization and concentration variables, the sample is further limited to workers in a manufacturing industry who are identified as living in

[2] The data used were obtained from the IPUMS website (http://www.ipums.umn.edu); see Ruggles and Sobek [20].

an MSA.[3] Residents of Alaska and Hawaii are also excluded. After applying these filters, the remaining sample contains more than 400,000 individuals living in 220 metropolitan areas. There are 424 occupational categories and 77 industrial SIC groups identified in the sample.

An initial consideration is the definition of each respondent's labor market, or city. The census defines single Metropolitan Statistical Areas (MSAs) as well as Consolidated Metropolitan Statistical Areas (CMSAs) when there are two or more economically and socially linked MSAs. We have chosen to use the broadest definition and consider each CMSA as a single city. In general we will simply use the term SMA or city to apply to either MSA or CMSA.

The occupation specialization and concentration variables were constructed from the full 1990 Census data (excluding Alaska and Hawaii). All individuals who reported being in the labor force were counted. The 5% PUMS is stratified: not all individuals had the same probability of being surveyed. Each record in the sample contains an integer weight which is the inverse of the probability of being sampled. Totals for the number of individuals in a category were creating by adding the weights for all the people in that category. Counts were made of the number of people in each occupation in the continental United States and each MSA, and of the labor force size in each MSA. Occupation concentration is the number of workers in an MSA/occupation divided by the total national workers in that occupation. Occupation specialization is the MSA/occupation number of workers divided by total MSA labor force. Note that the concentration variables will not sum to one across each occupation since the national total includes workers not in an MSA. Each individual in the Census sample is assigned a specialization and concentration value based on his/her cell (one of 220 $\times$ 424 values).

Since census data do not contain information on establishments, the industry data are drawn from the County Business Patterns (CBP) survey. This annual survey of employers contains information on employment and establishments by county for manufacturing industries. Income data in the 1990 Census refer to the 1989 calendar year so the 1989 CBP data were used. MSA and national (continental U.S.) totals of both employment and establishments were constructed for each industry.[4] For confidentiality reasons, employment was usu-

[3] The 5% Census sample does not identify geographical units with population less than 100,000, but it does list each individual's state of residence. Thus if either the MSA or the portion of the MSA in a worker's state has fewer than 100,000 people, the MSA of residence is not included. Additionally, some New England MSAs had to be excluded due to data problems; see footnote 4.

[4] The CBP data are presented by county, without any metropolitan identifiers. Data were aggregated from the county level to create MSA totals. Outside of New England, MSAs are defined by county so this was easily done. In New England, MSAs are defined by smaller units than counties. Excluding New England was not acceptable (note that the New York MSA extends into New England). Each New England MSA was individually examined and a judgement made about which counties to include and exclude. A few small MSAs which were not closely approximated by counties were removed from the sample. Further details are available on request.

ally given in a range (e.g., 25–49 employees). The midpoint of each range was used when estimating employment for each industry. Actual values were usually available for total manufacturing employment.[5] Industry employment specialization and concentration are defined in the same way as for occupation, as were measures of industry establishment specialization and concentration. Each individual in the Census sample was assigned specialization and concentration values based on his/her cell (one of 220 × 77 values).

Hourly wages were estimated by dividing total wage and salary income by the product of weeks worked last year and usual hours worked per week. The imputed wage for some individuals fell well below the federal minimum wage for 1989 of $3.35/hour. Approximately 1% of the sample has an imputed wage below this level. These observations remain in the sample for all results reported here; removing them was not found to alter the results in any significant way. Income data was also top-coded. In 1990 any value greater than $140,000 per year was replaced with the state median of all the reported values greater than this amount. We did not control for this problem.

Summary statistics for the variables can be found in Table 1. The matrix of correlations between the various constructed occupation and industry variables is presented as Table 2.

The equation to be estimated is

$$\ln(w_i) = \alpha + X_i B + Z_{jl}\Lambda + Y_{kl}\Gamma + \Theta_j + \Phi_k + \Psi_l + \varepsilon_i,$$

where i indexes individuals, j refers to an individual's occupation, k refers to an individual's industry, l refers to an individual's MSA, X_i is individual specific characteristics, Y_{kl} is industry specialization and concentration (employment and establishments), Z_{jl} is occupation specialization and concentration, and B, Λ, Γ, Θ_j, Φ_k, Ψ_l are coefficients.

The vector of individual characteristics includes dummies for being female, black, single, a student, a veteran, as well as female interacted with black and single. Continuous years of education is not available in the 1990 Census, only four categories of education. There are also no data on current job experience. Instead, to control for both education and experience, a different quadratic age–earnings profile was allowed for eight groups: two sexes times four education groups. The education groups are less than grade 12, grade 12, one to three years of college, and four or more years of college. The log-linear specification implies the following interpretation for the individual Λ and Γ's: They are the approximate percentage change in wages from having one

[5] There is an additional quirk in the CBP data: not all establishments are classified at the finest level of detail, so group totals often exceed the sum of all the constituent subgroups. Thus, in general, industry specialization will not sum to one in each MSA, for example. We do not believe this has any significant effect on the results.

TABLE 1
Summary Statistics

Variable	Min	Max	Weighted Mean	Weighted St. dev.	Unweighted Mean	Unweighted St. dev.
Weight	2	176	25.364	10.085	21.511	9.104
Imputed hourly wage	0.00	111.71	14.456	9.170	14.502	9.295
Weeks worked in 1989	50	52	51.858	0.495	51.859	0.494
Usual hours worked per week	35	98	43.260	6.170	43.274	6.179
Age	16	65	39.348	11.056	39.468	11.063
Female dummy	0	1	0.296	0.457	0.296	0.457
Black dummy	0	1	0.089	0.284	0.078	0.268
Single dummy	0	1	0.184	0.388	0.176	0.381
Student dummy	0	1	0.071	0.257	0.069	0.254
Veteran dummy	0	1	0.234	0.423	0.236	0.424
Single female dummy	0	1	0.062	0.241	0.059	0.236
Black female dummy	0	1	0.033	0.178	0.029	0.169
Occupation code	6	889				
Industry code	100	391				
Occ. employment specialization (%)	0.000	7.304	1.090	1.379	1.092	1.385
Occ. employment concentration (%)	0.002	46.741	2.358	3.103	2.360	3.123
Ind. employment specialization (%)	0.001	100[a]	6.677	9.223	6.632	9.115
Ind. employment concentration (%)	0.000	40.989	4.270	6.404	4.276	6.517
Ind. establishment specialization (%)	0.006	38.475	3.883	5.037	3.847	4.998
Ind. establishment concentration (%)	0.006	29.371	3.614	5.216	3.631	5.261
MSA code	40	9360				
Total MSA employment	44,701	8,832,000	2,387,242	2,761,102	2,398,826	2,798,031
Occupation employment in MSA	3	461,634	27,471	65,120	27,938	66,708
Industry employment in MSA	2	153,684	20,137	31,760	19,787	31,302
Industry establishments in MSA	1	5695	326	812	322	805
Total MSA manufact. employment	1,338	1,297,052	401,729	444,662	402,006	451,370
Total MSA manufact. establishments	50	32,902	8428	10,470	8476	10,170

[a] This is a result of the ''binning'' of employment figures: the midpoint of the Ship Building and Repair employment range in Pascagoula, MN is larger than the value given for total manufacturing employment. The second highest value is approximately 68.

percentage point more specialization or concentration in an individual's occupation or industry.

The structural effects Θ_j, Φ_k, and Δ_l do not appear in all specifications. Δ_l is an MSA fixed effect which has the full dimensionality of 220. The vectors Θ_j and Φ_k are coefficients for occupation and industry dummies. As mentioned above, these are included to account for systematic differences in industry or occupation wage levels not accounted for by the other variables. In regressions that include these, there is the full dimensionality of 424 detailed occupations and 77 detailed industries. Later in the analysis, separate regressions were run for different industry and occupation groups, to test for parameter stability. So as not to strain the data, these groups represented aggregations of the original codes. The 424 occupations were grouped into eight aggregate groups, while

TABLE 2
Correlation Matrix

	Occ. L.F.	Ind. emp.	Ind. est.	Occ. spec.	Occ. conc.	Ind. spec.	Ind. conc.	Est. spec.	Est. conc.
Occ. L.F.	1.00								
Ind. emp.	0.27	1.00							
Ind. est.	0.26	0.65	1.00						
Occ. spec.	0.63	0.03	0.02	1.00					
Occ. conc.	0.44	0.57	0.44	0.05	1.00				
Ind. spec.	−0.08	0.33	0.05	0.01	−0.08	1.00			
Ind. conc.	0.25	0.67	0.22	0.03	0.52	0.28	1.00		
Est. spec.	0.00	0.26	0.53	0.00	0.01	0.26	−0.05	1.00	
Est. conc.	0.40	0.70	0.43	0.05	0.73	0.01	0.75	0.02	1.00

Note. Occ L.F. is MSA occupation labor force size; ind. emp. is MSA employment by industry; ind. est. is number of establishment in MSA by industry; the remaining variables are the specialization and concentration measures for occupation labor force, industry employment, and industry establishments, respectively.

the 77 industries were combined into nine aggregate categories. These are defined later in Tables 6 and 7.

Regressions were estimated with ordinary least squares. No attempt was made to correct for attenuation due to the measurement error present in the data. However, measurement error in the industry employment data are likely to cause heteroscedasticity because the ranges in which the data were reported are different sizes. The coefficient estimates should still be consistent,[6] but White's heteroscedasticity robust estimation of the covariance matrix was used (White [21]).[7] The standard errors may also be biased downward due to clustering (see Moulton [16]); all workers in the same SMA and industry have the same value for the industry specialization and concentration variables. Normally this would be easily remedied by clustering the standard errors by industry-SMA. But the data is also clustered by occupation-SMA and the two sets of clusters overlap.

We deal with this issue in the following manner. We cluster separately by industry-SMA and occupation-SMA as well as by industry-occupation-SMA. Since it might be argued that none of these strategies is sufficient, we also try a more conservative approach. The model is estimated with a full set of occupation-SMA dummies and the standard errors are clustered by industry-SMA.

[6] I.e. $P\lim(X\Omega X/n)$ should still be positive definite; see Greene [5, p. 499].

[7] Weighted least squares is often used with stratified data to control for the fact that the sample is not entirely random. Since a large part of the error in our study is from mis-reporting of wages and the imputation of hourly wages from annual data, this is not of great concern. However, the Census sampling probability is a function of metropolitan status (large cities are under-sampled). Since this is directly related to the variables of interest, the weighted versions of the OLS specifications were estimated. As expected, there were no significant differences.

Note that this eliminates all occupation variation, so the occupation specialization and concentration variables are not included in these specifications. A similar regression is performed saturating the model by industry-SMA and clustering on occupation-SMA.[8]

V. EMPIRICAL RESULTS

The full results from a regression including the six specialization and concentration variables and both occupation and industry dummies are reported in Table 3. With all of the fixed effects included, the fit of the equation is quite good for samples of this size and is consistent with other wage/earnings equations. Most of the interactive parameters for the combined age−sex−education variables are significant, as are all of the industry and occupation dummy variables.

Of most interest, both industrial and occupational specialization as well as industrial and occupational concentration are highly significant. The coefficients in the fifth column of Table 5 show the percentage increase in wages from a 1% increase in specialization or concentration—for the equation reported in Table 3. They are quite large. Occupational specialization varies across 220 MSAs and 424 occupations from zero to 7%. This range generates a 23% increase in wages. Similarly, industrial specialization ranges across 220 MSAs and 77 industries from zero to 68%, and this generates a huge 30% increase in wages. The sample ranges for occupational and industrial concentration are zero to 46% and zero to 41%, respectively. Interestingly, these generate somewhat smaller, 12% and 16% increases in wages.

Interpreting the combined effects of (say) industry specialization and concentration is a bit tricky owing to the fact that MSA size (total workforce) is controlled for (in this equation) with structural effects. Thus a worker who is in an industry with both high concentration and specialization is effectively in a ‘‘one industry city’’ as well as a ‘‘one city industry.’’ The same interpretation applies for occupations, and it is these theoretical situations that give rise to the highest wages.

The measures of establishment specialization and concentration have negative signs, but it must be remembered that this is ceteris paribus with respect to employment specialization and concentration. When there is high establishment specialization/concentration, there are more firms employing the same number of workers in a portion of the labor market, and firm size is hence lower. The result is that wages are lower, and this suggests that monopsony is *not* a problem in local labor markets. Instead it supports some form of increasing returns at the establishment level—independent of increasing returns in the labor market.

[8] The clustered regressions are performed using the ‘‘cluster’’ option on STATA’s ‘‘areg’’ command.

TABLE 3

Full Regression Results for Preferred Specification

Fixed effects regression with clustered standard errors
Dependent variable: ln(imputed hourly wage)

Total observations: 403,680
$F(330,402926) = 326.95$, $R - \text{Sq.} = 0.5126$
S.E. clusters: metro. area * occupation * industry
Age profiles for sex-education groups

Variable	Coefficient	Std. error
Constant	0.9978	0.0576
Age	0.0547	0.0008
Age^2	−0.0005	0.0000
Less than high school	{0.0317}	0.0309
1−3 Years of college	{−0.0058}	0.0247
4+ Years of college	{−0.0495}	0.0323
Age * less than H.S.	−0.0105	0.0016
Age * 1−3 years of college	{0.0014}	0.0013
Age * 4+ years of college	0.0079	0.0016
Age^2 * less than H.S.	0.0001	0.0000
Age^2 * 1−3 years of college	{$1.9*10^{-6}$}	$1.5*10^{-5}$
Age^2 * 4+ years of college	{$-1.9*10^{-5}$}	$1.9*10^{-5}$
Female * less than H.S.	0.3812	0.0702
Female * 1−3 years of college	0.3367	0.0561
Female * 4+ years of college	0.2365	0.0592
Female * age * less than H.S.	−0.0160	0.0024
Female * age * high school	−0.0148	0.0013
Female * age * 1−3 a. college	−0.0063	0.0017
Female * age * 4+ a. college	0.0126	0.0026
Female * age^2 * less than H.S.	0.0001	0.0000
Female * age^2 * high school	0.0001	0.0000
Female * age^2 * 1−3 a. college	{$-8.7*10^{-5}$}	$2.1*10^{-5}$
Female * age^2 * 4+ a. college	−0.0003	0.0000

Metropolitan area fixed effects: 219
Occupation fixed effects: 423
Industry fixed effects: 76
Demographic dummies

Variable	Coefficient	Std. error
Female dummy	−0.2902	0.05806
Black dummy	−0.0908	0.0034
Single dummy	−0.1346	0.0024
Student dummy	−0.0443	0.0025
Veteran dummy	{−0.0034}	0.0019
Single female dummy	0.1243	0.0040
Black female dummy	0.0826	0.0053
Specialization and concentration variables		
Occ. L.F. specialization	0.0336	0.0034
Occ. L.F. concentration	0.0025	0.0009
Ind. employ. specialization	0.0041	0.0002
Ind. employ. concentration	0.0035	0.0006
Ind. estab. specialization	−0.0038	0.0004
Ind. estab. concentration	{−0.0019}	0.0012

Note. { } indicates insignificance at the 5% level.

TABLE 4
Regression Coefficients (×100)

Independent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Occupation	3.87	3.80	3.63	3.36	3.36	3.36	3.36		3.04
specialization	0.05	0.05	0.17	0.18	0.34	0.34	0.37		0.32
Occupation	1.55	0.98	1.00	0.25	0.25	0.25	0.25		(0.13)
concentration	0.03	0.03	0.04	0.04	0.09	0.11	0.10		0.09
Industry specialization	0.26	0.39	0.26	0.41	0.41	0.41	0.41	0.42	
	0.01	0.01	0.01	0.01	0.02	0.03	0.02	0.04	
Industry concentration	0.89	0.84	0.53	0.35	0.35	0.35	0.35	0.29	
	0.02	0.02	0.02	0.03	0.06	0.07	0.07	0.08	
Establishment	−0.50	−0.56	−0.33	−0.38	−0.38	−0.38	−0.38	−0.37	
specialization	0.02	0.02	0.03	0.04	0.04	0.07	0.04	0.06	
Establishment	−0.54	−0.65	−0.07	−0.19	(−0.19)	(−0.19)	(−0.19)	(−0.06)	
concentration	0.03	0.03	0.03	0.04	0.12	0.11	0.12	0.15	
MSA fixed effects?	N	Y	N	Y	Y	Y	Y	N	N
Occupation fixed effects?	N	N	Y	Y	Y	Y	Y	N	N
Industry fixed effects?	N	N	Y	Y	Y	Y	Y	N	N
MSA ∗ occupation F.E.?	N	N	N	N	N	N	N	Y	N
MSA ∗ industry F.E.?	N	N	N	N	N	N	N	N	Y
Cluster	None	None	None	None	M∗O∗I	M∗I	M∗O	M∗I	M∗O

Note. Standard errors are in small type below the coefficient estimates and have been multiplied by 100. ( ) indicates insignificance at the 5% level. Each regression contains seven demographic dummy variables and variables which allow a different quadratic age−earnings profile for each of eight sex-education groups. There are 220 MSAs, 424 occupations, 77 industries, and $N = 403{,}669$.

Alternative Fixed Effects Specifications

It must be remembered that the specialization/concentration measures have a large dimensionality: 220 × 424 in the case of occupations and 220 × 77 for industries. Thus much of the power behind the main results in Table 3 comes from the *interaction* between MSA geography and industrial or occupational structure. In Table 3, the importance of these variables comes from the fact that, *with a given occupation*, when that occupation is concentrated in particular MSAs then wages are higher. Alternatively, *within an MSA*, those industries that MSA specializes in again pay higher wages (to observationally equivalent workers). It certainly would seem prudent to explore how sensitive these results are to the inclusion of the industry/occupation/MSA structural effects. This is done in Table 4, and note that the coefficients have been multiplied by 100 to give a finer level of detail.

In Table 4, column (4) gives the full specification results of Table 3 for the six specialization and concentration variables. Columns (1) through (3) show results with different combinations of structural effects. With the exception of establishment concentration, which is generally insignificant, the coefficients are remarkably stable. The coefficients on occupational specialization are largely unchanged with the inclusion of (literally) hundreds of structural effects, while those for industry specialization and establishment specialization *increase* in value. In only two cases, industry concentration and occupational concentration, do the values decrease—although they are still highly significant. All of this reinforces the conclusion that the results obtained are *not* due to basic differences across sectors or occupations. Nor are the results explainable by simple SMA size or geography—which are fully controlled for with their structural effects. Rather, it is the *interaction* of these variables that proves important. When any portion of an economy is concentrated, or any area specialized, then higher wages for those in these sectors are the result. Since there are compelling reasons to include both the dummies and fixed effects, Specification (4) clearly is our preferred equation.

The remaining columns show the results for the regressions with clustered standard errors. Note that clustering only affects standard errors and not point estimates, so the coefficients in columns (5) to (7) are the same as column (4). Clustering by industry-occupation-SMA, industry-SMA, or occupation-SMA increases the standard errors by up to a factor of three. All but the establishment concentration variable remain highly significant. As shown in column (8), even when the model is saturated with occupation-SMA fixed effects, the estimated coefficients and standard errors are of similar magnitude. Finally, in the regression with a full set of industry-SMA fixed effects the occupation concentration variable becomes insignificant, but the standard errors are similar to the other clustered regressions. We conclude from these regressions that the standard errors are biased downward, but, with the exception of establishment concentration, the bias is not significant enough to question the statistical significance of the results.

Elasticities

To better show the size of the estimated effects, Table 5 presents the coefficients from Table 4 multiplied by the mean value for the explanatory variable in question. With log-linear specification, this is equal to the elasticity at the mean ($\xi = \partial \ln(w_i)/\partial \ln(x) = x \times \partial \ln(w_i)/\partial x = x \times \beta$). Again, the results are multiplied by 100 for convenience. This reveals that the economic significance is similar for each of the specialization and concentration variables. So, for example, in the preferred specification (column (4)) a doubling of industry employment concentration at the mean is associated with a wage increase of about 2%. Note also that the establishment concentration variable is not only not as statistically significant as the other variables, but not as economically significant either.

TABLE 5
Elasticities at the Mean of the Dependent Variable (×100)

Independent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Occupation specialization	4.22	4.14	3.96	3.66	3.66	3.66	3.66		3.31
Occupation concentration	3.65	2.31	2.35	0.59	0.59	0.59	0.59		(0.31)
Industry specialization	1.76	2.64	1.76	2.78	2.78	2.78	2.78	2.84	
Industry concentration	3.82	3.61	2.27	1.50	1.50	1.50	1.50	1.24	
Establishment specialization	−1.94	−2.17	−1.28	−1.47	−1.47	−1.47	−1.47	−1.44	
Establishment concentration	−1.96	−2.36	−0.25	−0.69	(−0.69)	(−0.69)	(−0.69)	(−0.22)	
MSA fixed effects?	N	Y	N	Y	Y	Y	Y	N	N
Occupation fixed effects?	N	N	Y	Y	Y	Y	Y	N	N
Industry fixed effects?	N	N	Y	Y	Y	Y	Y	N	N
MSA ∗ occupation F.E.?	N	N	N	N	N	N	N	Y	N
MSA ∗ industry F.E.?	N	N	N	N	N	N	N	N	Y
Cluster	None	None	None	None	M ∗ O ∗ I	M ∗ I	M ∗ O	M ∗ I	M ∗ O

Note. Reported elasticities come from regressions that each contain seven demographic dummy variables and variables which allow a different quadratic age-earnings profile for each of eight sex-education groups. There are 220 MSAs, 424 occupations, and 77 industries categories.

Differences across Sectors and Occupations

One of the virtues of this study is that we have measured agglomeration by constructing variables of labor market scale which can be applied to all industries and occupations at the same time. The downside of this approach is that we restrict the agglomeration economies to be constant across all of the industries and occupations. We can examine the impact of this restriction by allowing the specialization and concentration effects to vary (somewhat) by occupation and industry groups. This can be accomplished in several ways, and the one chosen here is simply to stratify the regressions. One must be careful in such stratification for the specialization and concentration measures themselves are merely interactions between SMA and industry (or occupation). Since full SMA structural effects are included, stratifying by all 77 industries, for example, would remove all of the independent variation in the industry specialization or concentration variables, and this in turn would make the original results unstable. Industry stratification, however, would not have this effect on the occupation specialization/concentration variables. The opposite situation would arise when the stratification is by occupation. To try and ameliorate this problem we stratify by only more aggregate categories of eight

TABLE 6

Definition of Aggregate Occupation Groups Used for Stratification

Group	% of sample	Census code	Occupations
1	21.6%	000–042	Executive, administrative, and managerial occupations
		084–112	Health diagnosing, assessment, and treating occupations
		113–162	Teachers
		163–165	Counselors, librarians, archivists, and curators
		166–173	Social scientists and urban planners
		174–177	Social, recreation, and religious workers
		178–182	Lawyers and judges
		183–202	Writers, artists, entertainers, and athletes
		203–212	Health technologists and technicians
		403–472	Service occupations
		473–476	Farm operators and managers
		477–493	Other agricultural and related occupations
		494–496	Forestry and logging occupations
		497–502	Fishers, hunters, and trappers
		553–612	Construction trades
		613–627	Extractive occupations
2	12.6%	043–063	Engineers, architects, and surveyors
		064–068	Mathematical and computer scientists
		069–083	Natural scientists
		213–242	Technologists and technicians, except health
3	4.3%	243–302	Sales occupations
4	12.7%	303–402	Administrative support occupations, including clerical
5	4.6%	503–552	Mechanics and repairers
6	13.4%	628–702	Precision production occupations
7	24.4%	703–802	Machine operators, assemblers, and inspectors
8	6.5%	803–863	Transportation and material moving occupations
		864–902	Handlers, equipment cleaners, helpers, and laborers

Note. There are 424 occupations in the sample.

occupations and nine industries, described in Tables 6 and 7. Table 8 reports the coefficients for the six specialization and concentration variables when the stratification is done both by industry and occupation. Table 9 gives the associated elasticities—when measured at the mean value for each category.

The impact of stratifying and regressions is quite pronounced on the concentration–specialization measures that are based on the same stratifying variable. Thus in the upper frame of Table 8, where we stratify the equations by occupation, the coefficients for occupational concentration/specialization become quite unstable—varying greatly across occupations, and in one case turning negative. On the other hand, the industrial concentration/specialization variables have coefficients that are reasonable and of similar magnitude across the occupation-specific equations. Similarly in the bottom frame, where the

TABLE 7

Definition of Aggregate Industry Groups Used for Stratification

Group	% of sample	Census code	Industry
1	2.3%	130–131	Tobacco manufactures
		220–229	Leather and leather products
		391	Miscellaneous manufacturing industries
		392–399	Not specified manufacturing industries
2	5.9%	100–129	Food and kindred products
3	5.8%	132–150	Textile mill products
		151–159	Apparel and other finished textile products
4	7.0%	160–170	Paper and allied products
		230–241	Lumber and wood products, except furniture
		242–249	Furniture and fixtures
5	10.0%	171–179	Printing, publishing, and allied industries
6	12.0%	180–199	Chemicals and allied products
		200–209	Petroleum and coal products
		210–219	Rubber and miscellaneous plastics products
7	2.7%	250–269	Stone, clay, glass, and concrete products
8	10.5%	270–309	Metal industries
9	43.8%	310–339	Machinery and computing equipment
		340–350	Electrical machinery, equipment, and supplies
		351–370	Transportation equipment
		371–389	Professional and photographic equipment, and watches
		390	Toys, amusement, and sporting goods

Note. There are 77 industries identified in the sample.

stratifying is based on industry, the coefficients for industry concentration/specialization variables begin to vary widely while the occupation measures remain more stable.

The conclusion we draw is that any attempt to estimate agglomeration effects that are specific to select industries or occupations is intrinsically difficult, due to identification problems, if SMA structural effects are also included in the equation.

VI. INTERPRETATIONS AND CONCLUSIONS

Using wage data has allowed us to overcome some of the measurement problems inherent in estimating agglomeration economies with output data. We find that observationally equivalent workers in cities with a larger share of national or metropolitan employment in their same occupation or industry earn higher wages. The effects are statistically and economically significant and provide an interesting characterization of agglomeration economies. Similar to the findings of most other studies, employment shows strong localization economies, and there also are strong gains to specialization. Our results also are

TABLE 8

Stratification by Occupation and Industry Groups Coefficient ×100

	Occupations					
	Occ. lab. force		Ind. empl.		Ind. establ.	
Occupation group	Specialization	Concentration	Specialization	Concentration	Specialization	Concentration
All	3.36	0.25	0.41	0.35	−0.38	(−0.19)
1 Miscellaneous	2.16	0.87	0.45	(0.05)	−0.29	(0.15)
2 Engineers, scientists, technicians, etc.	3.73	(−0.05)	0.16	0.26	(−0.06)	(−0.08)
3 Sales occupations	(−3.97)	(−0.17)	0.44	(−0.16)	(−0.05)	(0.52)
4 Administrative support, incl. clerical	(1.67)	(−0.41)	0.40	0.32	−0.31	(−0.13)
5 Mechanics and repairers	(1.27)	(0.14)	0.45	0.39	(−0.36)	−0.43
6 Precision production occupations	(−1.81)	(−0.03)	0.41	0.40	−0.45	(−0.21)
7 Machine operators, assemblers, inspectors	2.15	(−0.26)	0.46	0.78	−0.47	−0.54
8 Transportation, handlers, helpers, laborers	(0.16)	(0.41)	0.54	1.16	−0.61	−1.19

	Industries					
	Occ. lab. force		Ind. empl.		Ind. establ.	
Industry group	Specialization	Concentration	Specialization	Concentration	Specialization	Concentration
All	3.36	0.25	0.41	0.35	−0.38	(−0.19)
1 Miscellaneous (incl. tobacco and leather)	4.25	(0.63)	(0.98)	0.99	−1.64	−1.45
2 Food and kindred products	(0.72)	(−0.13)	0.73	1.48	(−0.15)	−1.70
3 Textile mill products, apparel, etc.	4.17	(−0.26)	0.46	(−0.17)	−0.47	(0.06)
4 Paper, lumber, wood products, furniture	4.87	(0.04)	0.54	1.37	−0.52	−1.54
5 Printing, publishing, and allied industries	(0.26)	0.74	0.56	(−0.26)	−0.48	(−1.10)
6 Chemicals, petroleum, rubber, and plastics	3.23	1.12	0.39	0.93	(−0.38)	−1.65
7 Stone, clay, glass, and concrete products	(1.83)	(0.10)	1.00	1.85	−2.07	−2.62
8 Metal industries	3.78	(0.46)	0.90	0.86	−1.89	−0.92
9 Machinery, equipment, electronics, etc.	3.07	(−0.01)	0.29	0.36	−0.34	(−0.15)

Note. ( ) indicates insignificance at the 5% level. Each regression contains 220 MSA fixed effects, as many occupation and industry dummies as possible, seven demographic dummy variables, and variables which allow a different quadratic age−earnings profile for each of eight sex-education groups. Definitions of occupation and industry groups used for stratification are given in Tables 6 and 7.

TABLE 9

Stratification by Occupation and Industry Groups Elasticity at Group Mean ×100

	Occupations					
	Occ. lab. force		Ind. empl.		Ind. establ.	
Occupation group	Specialization	Concentration	Specialization	Concentration	Specialization	Concentration
All	3.66	0.59	2.78	1.50	−1.47	(−0.69)
1 Miscellaneous	4.98	2.38	2.79	(0.24)	−1.15	(0.61)
2 Engineers, scientists, technicians, etc.	1.51	(−0.16)	1.32	1.63	(−0.15)	(−0.39)
3 Sales occupations	(−5.17)	(−0.45)	1.83	(−0.51)	(−0.24)	(1.79)
4 Administrative support, incl. clerical	(1.99)	(−1.00)	2.49	1.44	−1.30	(−0.52)
5 Mechanics and repairers	(0.41)	(0.24)	3.71	1.65	(−0.99)	−1.32
6 Precision production occupations	(−1.17)	(−0.06)	2.78	1.53	−1.90	(−0.71)
7 Machine operators, assemblers, inspectors	1.52	(−0.48)	3.22	2.72	−2.10	−1.66
8 Transportation, handlers, helpers, laborers	(0.15)	(0.60)	3.12	3.57	−1.81	−3.00

	Industries					
	Occ. lab. force		Ind. empl.		Ind. establ.	
Industry group	Specialization	Concentration	Specialization	Concentration	Specialization	Concentration
All	3.66	0.59	2.78	1.50	−1.47	(−0.69)
1 Miscellaneous (incl. tobacco and leather)	4.81	(1.54)	(3.85)	6.21	−5.50	−6.42
2 Food and kindred products	(0.78)	(−0.22)	1.98	2.73	(−0.16)	−2.98
3 Textile mill products, apparel, etc.	4.90	(−0.76)	3.57	(−0.6)	−2.17	(0.41)
4 Paper, lumber, wood products, furniture	5.24	(0.07)	2.46	2.54	−1.63	−3.09
5 Printing, publishing, and allied industries	(0.26)	2.13	4.01	(−0.87)	−6.17	(−3.29)
6 Chemicals, petroleum, rubber, and plastics	3.71	2.73	1.50	3.84	(−0.63)	−5.55
7 Stone, clay, glass, and concrete products	(2.04)	(0.16)	1.82	3.23	−2.55	−4.50
8 Metal industries	3.70	(0.83)	3.27	2.692	−4.43	−2.49
9 Machinery, equipment, electronics, etc.	3.42	(−0.03)	2.68	2.00	−1.13	(−0.63)

Note. ( ) indicates insignificance at the 5% level. Each regression contains MSA fixed effects, as many occupation and industry dummies as possible, seven demographic dummy variables, and variables which allow a different quadratic age−earnings profile for each of eight sex-education groups. Definitions of occupation and industry groups used in stratification are given in Tables 6 and 7.

consistent with most other studies that find little evidence of economic diversity or urbanization economies. Establishment specialization and concentration measures are negative, but since we control for employment when measuring these effects, we interpret them as showing possible evidence of increasing returns at the firm or establishment level—independent of agglomeration in the local labor market.

If our results truly measure agglomeration, a natural question arises as to why there is not complete concentration and specialization. Based on our earlier discussion, our suggested answer is that there are other production considerations for firms that are fixed at certain locations, such as access to markets, raw materials, or higher educational resources. Hence not all firms within the same industry necessarily want to be in the same location, and operating along an iso-cost curve, some choose to pay more for workers in locations with higher productivity, while others choose the combination of low wages and low productivity locations. It is the other factor cost considerations that determine which firms make which locational selections.

If production costs can be equal across markets with different agglomeration levels, the question arises as to why workers within a given industry and occupation will not seek the higher wages of the location that is most specialized and concentrated in their field. Our answer here is that there must exist some form of supply rigidity. Specialization occurs at least in part because the majority of workers in an MSA are convinced to be employed in one industry or occupation. Concentration occurs because that MSA is able to attract the national workforce in that industry-occupation. If workers have idiosyncratic preferences as to living sites, it is easy to imagine that such spatial labor market ‘‘focus’’ requires rising wages, and that these supply elasticities are not uniform across economic sectors. In some parts of the nation and some occupations, it is easier to attract the requisite labor supplies than others. Hence specialization and concentration are far from complete and agglomeration varies across locations. Firms are willing to pay for the agglomeration they get at each location, and other factor considerations, along with high factor substitutability, allow firms to compete across the different wage-agglomeration locations.

This study provides a starting point for much additional research. The finding that workers earn more when there are few establishments (ceteris paribus) needs further investigation as an approach to investigating firm level production returns. The study also could be extended to non-manufacturing industries. There is particular interest in the contrast between manufacturing and service industries. Finally, the approach could be used to measure dynamic externalities, through the analysis of wage growth.

REFERENCES

1. G. Carlino, Increasing returns to scale in metropolitan manufacturing, *Journal of Regional Science*, **19**, 363–373 (1979).
2. A. Ciccone and R. Hall, Productivity and the density of economic activity, *American Economic Review*, **86**, 54–70 (1996).
3. E. Glaeser and D. Maré, ''Cities and Skills,'' NBER Working Paper 4728 (1994).
4. E. Glaeser, H. Kallal, J. Scheinkman, and A. Shleifer, Growth in cities, *Journal of Political Economy*, **100**, 1126–1152 (1992).
5. W. Greene, ''Econometric Analysis,'' Prentice Hall, Englewood Cliffs, NJ (1997).
6. G. Hanson, Increasing returns, trade and the regional structure of wages, *Economic Journal*, **107**, 113–133 (1997).
7. V. Henderson, Efficiency of resource usage and city size, *Journal of Urban Economics*, **19**, 47–70 (1986).
8. V. Henderson, ''Marshall's Scale Economies,'' NBER Working Paper 7358 (1999).
9. V. Henderson, A. Kuncoro, and M. Turner, Industrial development in cities, *Journal of Political Economy*, **103**, 1067–1090 (1995).
10. A. Jaffee, M. Trajtenberg, and R. Henderson, Geographic localization of knowledge spillovers, *QJE*, CVIII, **3**, 259–293 (1993).
11. J. Jacobs, ''The Economy of Cities,'' Vintage, New York (1968).
12. S. Kim, Labor heterogeneity, wage bargaining, and agglomeration economies, *Journal of Urban Economics*, **28**, 160–177 (1990).
13. S. Kim, Labor specialization and the extent of the market, *Journal of Political Economy*, **98**, 692–705 (1989).
14. D. Maré, ''Living with Graduates,'' Mimeo, Harvard University (1994).
15. R. Moomaw, Productivity and city size: A critique of the evidence, *Quarterly Journal of Economics*, 675–688 (1981).
16. B. Moulton, Random group effects and the precision of regression estimates, *Journal of Econometrics*, 385–397 (1986).
17. R. Nakamura, Agglomeration economies in urban manufacturing industries: A case of Japanese cities, *Journal of Urban Economics*, **17**, 108–124 (1985).
18. J. Rauch, Productivity gains from geographic concentration of human capital: Evidence from the cities, *Journal of Urban Economics*, **34**, 380–400 (1993).
19. J. Roback, Wages, rents, and the quality of life, *Journal of Political Economy*, **90**, 1257–1278 (1982).
20. S. Ruggles and M. Sobek, ''Integrated Public Use Microdata Series: Version 2.0,'' Minneapolis: Historical Census Projects, University of Minnesota (1997).
21. H. White, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, **48**, 817–838 (1980).