

# **THE INTERNET AND LOCAL WAGES: A PUZZLE**

**By Chris Forman, Avi Goldfarb, and Shane Greenstein**

## **Web Appendix**

**A: Data Appendix**

**B: Appendix Tables**

## A. Data Appendix

In this section we discuss the construction of our data set. We first describe the construction of our measures of internet investment, and then briefly describe our measures of county characteristics that we interact with advanced internet. Last we describe the construction of our instruments.

### *A.1 Construction of measures of advanced internet investment.*

As noted in the paper, our IT data come from the Harte Hanks Market Intelligence Computer Intelligence Technology database (hereafter CI database). Harte Hanks tracks over 300,000 establishments in the United States. For the reasons described in the paper, we exclude government, military, and nonprofit establishments. Our sample from the CI database contains commercial establishments with over 100 employees—in total 86,879 establishments.<sup>1</sup>

The CI database contains several measures of internet usage that we use to construct measures of advanced internet. Advanced internet is the type of investment that has been historically described in books on electronic commerce. Typically this involves altering sales, manufacturing, production, or distribution systems within the firm. We aggregate many applications under this umbrella. Business-to-business or business-to-consumer e-commerce fall in this category, so too does TCP/IP versions of software such as enterprise resource planning or customer relationship management. Our measure of advanced internet assumes nothing about the *intensity* of use, nor about complexity.

An establishment is counted as making an investment in advanced internet when two or more of the following conditions hold: (1) the establishment uses two or more languages common in web applications, such as Active-X, Java, CGI, Perl, VB Script, or XML; (2) the

---

<sup>1</sup> Parts of this section draw from an earlier paper on the dispersion of internet investment, Forman, Goldfarb, and Greenstein (2002).

establishment has over five internet developers; (3) the establishment has two or more “e-business” applications such as customer service, education, extranet, publications, purchasing, or technical support; (4) the establishment reports LAN software that performs one of several functions: e-commerce, enterprise resource planning, web development, or web server; (5) the establishment has an internet server that is a UNIX workstation or server, mainframe, or minicomputer, or has 5 or more PC servers, or has internet storage greater than 20 gigabytes (this was a lot of storage in 2000); (6) the establishment indicates the use of three or more applications related to internet server software, internet/web software, or intranet applications.

We tested this definition and found that it generated many false positives. These false positives arose more frequently when an establishment was experimenting with, but not actually regularly using, advanced internet applications. In other words, they were trying something small or contracting out for a test. To correct for this, we exclude establishments if: (a) They indicate they have outsourced hosting of their internet/web servers; (b) These experimenters responded affirmatively to exactly two of (1) through (6) but not any question about e-commerce. Such establishments typically had not yet done very much advanced internet as of the time of our sample (but might someday). Previous work compared the baseline measure of advanced internet with one that includes such “experimenters” (Forman, Goldfarb, and Greenstein 2002). While that latter measure shows higher levels of internet penetration (23.2% v. 12.6%), the quantitative difference between the two measures remains similar across geographic regions.

Our measure of basic internet (used in Table 4) is constructed similarly. To be counted as investing in basic internet, an establishment must engage in two or more of the following activities: (1) have an internet service provider; (2) indicate it has basic access; (3) use commerce, customer service, education, extranet, homepage, publications, purchasing or

technical support; (4) indicate that it has an intranet or email based on TCP/IP protocols; (5) indicate there are internet users or internet developers on site; or (6) outsource some internet activities. We looked for two or more activities to guard against “false positives”. As it was, this was a minor issue. Most respondents responded affirmatively to many of these criteria.

The CI database also contains information on the number of personal computers and number of employees at the establishment. We divide the number of establishment personal computers by the number of employees to obtain our PCs per employee measure.

Timing bias and sampling bias are two concerns with these measures. We first discuss timing bias. Establishments in our sample were surveyed between July 1998 and August 2000. Because advanced internet diffused between 1998 and 2000, earlier respondents are likely to have a lower adoption rate. To control for increasing adoption rates over time, we reweight our adoption data by the ratio of average adoption rates in our sample between the month of the survey and the end of 2000. Specifically, we divide our sample into six semi-annual periods between 1998 and 2000. For establishments who are surveyed in some semi-annual period  $t$  prior to the end of 2000, we reweight the adoption rate by *(average adoption rate in county at end of 2000) / (average adoption rate in county in semi-annual period  $t$ )*.

To obtain location-level measures of the extent of advanced internet investment, we compute average rates of use for a location. Because the distribution of establishments over industries may be different in our sample from that of the population, we compare the number of establishments in our database to the number of establishments in the Census to correct for sampling bias. We calculate the total number of establishments with more than 50 employees in the Census Bureau’s 1999 County Business Patterns data and the number of establishments in our database for each two-digit North American Industry Classification System (NAICS) code in

each location. We then calculate the total number in each location. Therefore, to account for over- and under-sampling in the Harte Hanks data, we weight a NAICS-location by

$$\frac{\text{Total \# of census establishments in location-NAICS}}{\text{Total \# of census establishments in location}} \times \frac{\text{Total \# of establishments in our data in location}}{\text{Total \# of establishments in our data in location-NAICS}}$$

In other words, the weights are the proportion of establishments in a location that are a given NAICS code, divided by the proportion of times it is in our database. This means that if our data undersamples a given two-digit NAICS at a location, then each observation in that NAICS-location is given more importance. We divide establishment adoption by the above weights and then sum the weighted county-NAICS-level rates of use across NAICS within a county to obtain county-level estimates of the extent of advanced internet.

As a robustness check, we also show results with the following weights

$$\frac{\text{Total \# of establishments in our data in location}}{\text{Total \# of census establishments in location}}$$

## *A.2 Construction of variables measuring county characteristics*

The construction of our controls for county characteristics is described in Table 1b; here we describe the computation of the variables that we interact with advanced internet. As noted in the paper, we focus on the roles of *income*, *education*, *population*, and *IT-intensity*; these measures are equal to one when the corresponding continuous variable is in the highest quartile of the distribution. Income, education, and population data are from the 1990 Census: High Income is based upon median household income, High Education is based upon the fraction of the population that is university graduates, and High population is based upon 1990 county population estimates. For IT-intensity, we measure the fraction of firms in IT-using and producing industries in the county as of 1995 from the US Census County Business Patterns

data. National aggregate data shows that such industries have unusually high returns from investment in IT in the 1990s. We define these industries using the classification reported in Jorgenson, Ho, and Stiroh (2005, p. 93).<sup>2</sup>

### *A.3 Construction of the instruments*

Here we discuss the computation of our four instruments: the number of programmers in other establishments within the same firm, the “Bartik index” of expected local internet demand, the number of local ARPANET nodes, and the year when a state adopted a price cap.

To compute the number of programmers in other establishments, we use information on establishment programmers from the CI database. For each establishment that is part of a multi-establishment firm, we compute the number of programmers that reside within the same firm but in other counties. This variable is based on the “organizational capabilities” measure used in Forman, Goldfarb, and Greenstein (2008). We then compute the weighted average number of programmers for the county, using the weights described in section A.1.

Our measure of the Bartik index for local propensity of internet adoption uses our establishment-level internet data. For each county in our sample, we compute average industry adoption of advanced internet excluding the contribution of establishments in the industry-county, for industry  $j$  in county  $i$  we label this  $\widehat{Internet}_{ij}$ .<sup>3</sup> This is equal to the average propensity for an establishment in industry  $j$  to have advanced internet, excluding the contribution of establishments in county  $i$ . We then compute an index of advanced internet

---

<sup>2</sup> Specifically, they include the following industries: Communications (SIC 48), Business Services (Including Computer Services; SIC 73), Wholesale Trade (SIC 50-51), Finance (SIC 60-62, 67), Printing and Publishing (SIC 27), Legal Services (SIC 81), Instruments and Miscellaneous Manufacturing (SIC 38-39), Insurance (SIC 63-64), Machinery (Including Computers and Office Equipment; SIC 35), Gas utilities (SIC 492, parts of 493, 496), Professional and Social Services (SIC 832-839), Other Transportation Equipment (SIC 372-379), Other Electronic Machinery (including Communications Equipment and Electronic Components; SIC 36).

<sup>3</sup> Again, we use weighted adoption, where the weights are analogous to our location weights in section A.1. That is, we weight establishment adoption by  $((\text{total Census establishments in a location-industry})/(\text{total Census establishments in an industry})) \times ((\text{total establishments in our data in an industry})/(\text{total establishments in our data in a location-industry}))$ .

investment in county  $i$  by weighting these industry propensities according to the fraction of establishments in county  $i$  that are in industry  $j$ ,  $\rho_{ij}$ . That is,  $Bartik_{ij} = \sum_j \rho_{ij} \widehat{Internet}_{ij}$ . This is a generalization of similar share-weighted local demand proxies used in earlier work such as Bartik (1991) and Moretti (2009).

Our measure of number of local ARPANET nodes is simply a count of the number of local nodes, compiled from Hobbes' Internet Timeline (<http://www.zakon.org/robert/internet/timeline/> accessed Dec. 2008) and the ARPANET map (<http://som.csudh.edu/cis/lpress/history/arpamaps/> accessed Dec. 2008).

Our measure indicating the year when a state adopted a price cap or freeze, is the year that the state froze (or capped) the prices incumbent carriers could charge entrants (source: Abel and Clements 1998). These regulatory caps were attempts to facilitate entry by competitive local exchange carriers.

## References

- Abel, Jaison and Michael E. Clements. 1998. A time series and cross sectional classification of state regulatory policy adopted for local exchange carriers," NRRI report 98-25, Columbus Ohio, The National Regulatory Research Institute. December.
- Bartik, Timothy. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute.
- Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2002. Digital Dispersion: An Industrial and Geographic Census of Commercial internet Use. NBER Working Paper #9287.
- Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2008. Understanding the Inputs into Innovation: Do Cities Substitute for Internal Firm Resources? *Journal of Economics and Management Strategy* 17(2): 295–316.
- Jorgenson, Dale, Mun S. Ho, and Kevin Stiroh. 2005. *Productivity Volume 3: Information Technology and the American Growth Resurgence*. Cambridge, MA: MIT Press.
- Moretti, Enrico. 2009. Real Wage Inequality. Working Paper, University of California, Berkeley.

## B. Appendix Tables

**Online Appendix Table 1: Additional specifications for Table 2**

	(1)	(2)	(3)	(4)
	Alternative weighting	Multiple imputation of missing data on advanced internet	Compare advanced internet and basic internet	Compare advanced internet and PCs per employee
Advanced internet	0.0245 (0.0111)**	0.0243 (0.0143)*	0.0261 (0.0132)**	0.0248 (0.0133)*
Basic internet			0.0050 (0.0075)	
PCs per employee				0.0156 (0.0104)
Home internet use	0.0832 (0.0377)**	0.0873 (0.0393)**	0.0823 (0.0378)**	0.0822 (0.0379)**
Home internet data missing	0.0288 (0.0170)*	0.0337 (0.1778)*	0.0279 (0.0170)	0.0282 (0.0170)*
Log population in 1990	-0.0066 (0.0019)***	-0.0037 (0.0021)*	-0.0067 (0.0019)***	-0.0068 (0.0019)***
Percentage African Americans in 1990	0.0134 (0.0118)	0.0232 (0.0130)*	0.0134 (0.0119)	0.0123 (0.0118)
Percentage university graduates in 1990	0.5731 (0.0785)***	0.4219 (0.0912)***	0.5671 (0.0802)***	0.5594 (0.0802)***
Percentage high school graduates in 1990	-0.1539 (0.0521)***	-0.0590 (0.0656)	-0.1569 (0.0522)***	-0.1587 (0.0521)***
Percentage below poverty line in 1990	-0.1605 (0.0464)***	-0.0646 (0.0500)	-0.1600 (0.0464)***	-0.1600 (0.0464)***
Median income in 1990 (\$000)	-0.0006 (0.0006)	0.0002 (0.0007)	-0.0005 (0.0006)	-0.0006 (0.0006)
Percentage population attending Carnegie Type 1 schools in 1990	0.0311 (0.0467)	0.0322 (0.0463)	0.0323 (0.0476)	0.0338 (0.0480)
Percentage population enrolled in engineering program in 1990	-0.2127 (0.3588)	-0.1197 (0.3447)	-0.2162 (0.3648)	-0.2367 (0.3644)
# of patents granted to inventors in the county in the 1980s (000)	0.0165 (0.0043)***	0.0153 (0.0043)***	0.0164 (0.0043)***	0.0160 (0.0043)***
Percentage professional in 1995	-0.0124 (0.0536)	0.448 (0.0674)	-0.0078 (0.0543)	-0.0093 (0.0537)
Percentage of persons over age 65 in 1990	0.0449 (0.0512)	0.1485 (0.0650)**	0.0435 (0.0512)	0.0472 (0.0513)
Net migration into the county in 1995 (000)	0.0033 (0.0032)	0.0022 (0.0032)	0.0033 (0.0032)	0.0034 (0.0032)
Change in log total population between 1990 and 2000	0.0522 (0.0152)***	0.0820 (0.0176)***	0.0528 (0.0152)***	0.0539 (0.0153)***
Change in percentage of African American 1990 to 2000	0.0319 (0.0754)	-0.0422 (0.0814)	0.0248 (0.0759)	0.0253 (0.0758)
Change in percentage of university graduates 1990 to 2000	0.8235 (0.1605)***	0.6358 (0.1799)***	0.8169 (0.1617)***	0.8167 (0.1605)***
Change in percentage of high school graduates 1990 to 2000	-0.0214 (0.0947)	0.07760 (0.1200)	-0.0227 (0.0948)	-0.0260 (0.0947)
Change in percentage of persons over age 65 1990 to 2000	-0.5602 (0.1192)***	-0.4853 (0.1326)***	-0.5629 (0.1191)***	-0.5620 (0.1190)***
Change in net migration into the county 1990 to 2000 (000)	0.0020 (0.0037)	0.0011 (0.0037)	0.0020 (0.0037)	0.0022 (0.0037)
Constant	0.2983 (0.0458)***	0.1597 (0.0524)***	0.2978 (0.0460)***	0.3009 (0.0458)***
Observations	2743	3133	2743	2743
R-squared	0.131	N/A (F=12.4)	0.13	0.13

Dependent variable is change in logged wages from 1995 to 2000. Heteroskedasticity-robust standard errors in parentheses.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 2: Continuous measures for income, education, IT-intensive industry, and population**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Advanced internet	-0.0371 (0.0472)	-0.0424 (0.0360)	0.0251 (0.0128)*	-0.0169 (0.0299)	-0.0853 (0.0570)	0.0242 (0.0127)*	-0.0411 (0.0548)
Advanced internet x county-level income	3.02e-06 (2.13e-06)				-9.06e-07 (3.60e-06)		-1.83e-06 (3.61e-06)
Advanced internet x county-level education		0.5897 (0.3050)*			0.6510 (0.4413)		0.5850 (0.4526)
Advanced internet x county-level population			1.52e-07 (9.84e-08)		1.09e-07 (9.34e-08)		-1.79e-07 (1.16e-07)
Advanced internet x county-level IT-intensity				0.1947 (0.1006)*	0.2332 (0.1007)**		0.1738 (0.1011)*
Advanced internet x income x education x population x IT-intensity						9.88e-11 (2.66e-11)***	1.12e-10 (3.66e-11)***
Observations	2743	2743	2743	2742	2742	2742	2742
R <sup>2</sup>	0.13	0.13	0.13	0.13	0.14	0.14	0.15

Dependent variable is logged wages. Controls are the same as in table 5. Heteroskedasticity-robust standard errors in parentheses.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 3: Further robustness**

	(1)	(2)	(3)	(4)	(5)
	Includes two-way interactions	MSA only		No controls	
Advanced internet	0.0001 (0.0152)	0.0860 (0.0362)**	0.0528 (0.0392)	0.0372 (0.0132)***	0.0232 (0.0133)*
Advanced internet and High income, education, IT-intensity, and population county	0.4166 (0.1996)**		0.3949 (0.1566)**		0.6702 (0.1864)***
Advanced internet and High income county	0.0832 (0.0469)*				
Advanced internet and High education county	0.1130 (0.0653)*				
Advanced internet and High IT-intensity county	0.0269 (0.0231)				
Advanced internet and High population county	-0.0654 (0.1152)				
Advanced internet and High IT-intensity and population county	0.0164 (0.0670)				
Advanced internet and High education and IT-intensity county	-0.0814 (0.0579)				
Advanced internet and High income and IT-intensity county	-0.0708 (0.0570)				
Advanced internet and High income and population county	0.0483 (0.0579)				
Advanced internet and High education and population county	0.0910 (0.0585)				
Advanced internet and High income and education county	-0.0918 (0.0712)				
Observations	2743	843	843	2744	2743
R <sup>2</sup>	0.14	0.33	0.34	0.004	0.05

Dependent variable is logged wages. . In columns (2) and (4) controls are the same as table 2. In columns (1), (3), and (5) controls are the same as table 3. Heteroskedasticity-robust standard errors in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 4: Robustness to alternative weighting in constructing the advanced internet variable**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Weight is # observations in our data over # observations in census		No weights		No weights for county differences		No weights for time differences	
Advanced internet	0.0245 (0.0111)**	0.0217 (0.0112)*	0.0227 (0.0130)*	0.0195 (0.0131)	0.0236 (0.0119)**	0.0205 (0.0120)*	0.0264 (0.0142)*	0.0229 (0.0144)
Advanced internet and High income, education, IT-intensity, and population county		0.2220 (0.0907)**		0.5847 (0.1792)***		0.5243 (0.1627)***		0.3163 (0.1355)**
Observations	2743	2743	2743	2743	2743	2743	2743	2743
R <sup>2</sup>	0.13	0.14	0.13	0.14	0.13	0.14	0.13	0.14

Dependent variable is logged wages. . In columns (1), (3), (5), and (7) controls are the same as table 2. In columns (2), (4), (6), and (8) controls are the same as table 3. Heteroskedasticity-robust standard errors in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 5: Robustness to alternative treatments of missing data**

	(1)	(2)	(3)	(4)
	<i>Missing data coded as zero</i>		<i>Multiple imputation of missing data</i>	
Advanced internet	0.0296 (0.0127)**	0.0265 (0.0127)**	0.0238 (0.0143)*	0.0233 (0.0150)
Advanced internet and High income, education, IT-intensity, and population county		0.4902 (0.1603)***		0.3779 (0.2049)*
Observations	3133	3133	3133	3133
R <sup>2</sup>	0.10	0.11	N/A	N/A
F statistic	N/A	N/A	12.43	13.57

Dependent variable is logged wages. In columns (1) and (3) controls are the same as table 2. In columns (2) and (4) controls are the same as table 4. Heteroskedasticity-robust standard errors in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 6: Robustness to alternative definitions of “advanced internet”**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Wider advanced internet before adjustment for the likely “false positives” described on p. 2 of this appendix		Has at least one advanced internet application		Has at least one internet or LAN-enabled e-commerce application		Has at least one internet application or internet storage	
Advanced internet	0.0270 (0.0098)***	0.0232 (0.0099)**	0.0247 (0.0097)**	0.0210 (0.0098)**	0.0261 (0.0095)***	0.0225 (0.0096)**	0.0233 (0.0093)**	0.0195 (0.0094)**
Advanced internet and High income, education, IT-intensity, and population county		0.4665 (0.1266)***		0.4435 (0.1133)***		0.4221 (0.1103)***		0.4385 (0.1128)***
Observations	2743	2743	2743	2743	2743	2743	2743	2743
R <sup>2</sup>	0.13	0.14	0.13	0.14	0.13	0.14	0.13	0.14

Dependent variable is logged wages. . In columns (1), (3), (5), and (7) controls are the same as table 2. In columns (2), (4), (6), and (8) controls are the same as table 3. Heteroskedasticity-robust standard errors in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 7: Mean and correlations between different definitions of “advanced internet”**

	Mean value	Correlation				
		(1)	(2)	(3)	(4)	(5)
Advanced internet (main definition)	0.089	1.0000				
Wider advanced internet before adjustment for the likely “false positives” described on p. 2 of this appendix	0.140	0.7996	1.0000			
Has at least one advanced internet application	0.143	0.7937	0.9043	1.0000		
Has at least one internet or LAN-enabled e-commerce application	0.147	0.7836	0.9117	0.9902	1.0000	
Has at least one internet application or internet storage	0.149	0.7751	0.9335	0.9753	0.9670	1.0000

**Online Appendix Table 8: Full set of coefficients for Table 3**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Advanced internet	0.0168 (0.0137)	0.0120 (0.0125)	0.0214 (0.0159)	0.0246 (0.0127)*	0.0049 (0.0149)	0.0239 (0.0128)*	0.0067 (0.0150)
Advanced internet and High income county	0.0960 (0.0389)**				0.0442 (0.0492)		0.0377 (0.0496)
Advanced internet and High education county		0.1101 (0.0455)**			0.0770 (0.0548)		0.0757 (0.0547)
Advanced internet and High IT-intensity county			0.0206 (0.0228)		0.0134 (0.0235)		0.0102 (0.0237)
Advanced internet and High population county				0.3631 (0.0934)***	0.2378 (0.1018)**		0.0182 (0.1027)
Advanced internet and High income, education, IT-intensity, & population county						0.4588 (0.1585)***	0.3393 (0.1904)*
High income county	-0.0136 (0.0062)**				-0.0093 (0.0066)	-0.0067 (0.0050)	-0.0102 (0.0067)
High education county		-0.0158 (0.0061)***			-0.0134 (0.0066)**	-0.0067 (0.0048)	-0.0132 (0.0066)**
High IT-intensity county			0.0101 (0.0039)***		0.0086 (0.0040)**	0.0094 (0.0034)***	0.0083 (0.0041)**
High population county				-0.0231 (0.0117)**	-0.0119 (0.0125)	0.0078 (0.0061)	0.0058 (0.0121)
High income, education, IT-intensity, and population county						-0.0370 (0.0214)*	-0.0250 (0.0239)
Log population in 1990	-0.0062 (0.0017)***	-0.0060 (0.0018)***	-0.0068 (0.0017)***	-0.0084 (0.0022)***	-0.0083 (0.0022)***	-0.0080 (0.0022)***	-0.0080 (0.0022)***
Percentage African Americans in 1990	0.0128 (0.0118)	0.0122 (0.0118)	0.0125 (0.0118)	0.0100 (0.0118)	0.0098 (0.0118)	0.0098 (0.0118)	0.0100 (0.0118)
Percentage university graduates in 1990	0.5712 (0.0784)***	0.6022 (0.0845)***	0.5693 (0.0787)***	0.5576 (0.0785)***	0.5855 (0.0848)***	0.5821 (0.0855)***	0.5774 (0.0848)***
Percentage high school graduates in 1990	-0.1491 (0.0519)***	-0.1537 (0.0517)***	-0.1466 (0.0514)***	-0.1407 (0.0513)***	-0.1351 (0.0509)***	-0.1301 (0.0513)**	-0.1309 (0.0510)**
Percentage below poverty line in 1990	-0.1449 (0.0473)***	-0.1483 (0.0462)***	-0.1398 (0.0460)***	-0.1490 (0.0465)***	-0.1326 (0.0471)***	-0.1327 (0.0470)***	-0.1309 (0.0470)***
Median income in 1990 (\$000)	-0.0000 (0.0007)	-0.0003 (0.0006)	-0.0003 (0.0006)	-0.0004 (0.0006)	-0.0000 (0.0007)	-0.0000 (0.0007)	0.0000 (0.0007)
Percentage population attending Carnegie Type 1 schools in 1990	0.0311 (0.0478)	0.0237 (0.0427)	0.0297 (0.0469)	0.0310 (0.0480)	0.0251 (0.0448)	0.0289 (0.0475)	0.0257 (0.0450)
Percentage population enrolled in engineering program in 1990	-0.1939 (0.3652)	-0.1418 (0.3390)	-0.1582 (0.3629)	-0.1318 (0.3612)	-0.0542 (0.3440)	-0.0884 (0.3595)	-0.0549 (0.3436)
# of patents granted to inventors in the county in the 1980s (000)	0.0161 (0.0043)***	0.0161 (0.0043)***	0.0160 (0.0044)***	0.0130 (0.0041)***	0.0121 (0.0040)***	0.0104 (0.0040)***	0.0102 (0.0041)**
Percentage professional in 1995	-0.0118	-0.0112	-0.0322	-0.0244	-0.0427	-0.0443	-0.0431

	(0.0535)	(0.0533)	(0.0525)	(0.0538)	(0.0529)	(0.0532)	(0.0530)
Percentage of persons over age 65 in 1990	0.0564	0.0605	0.0577	0.0566	0.0549	0.0507	0.0536
	(0.0508)	(0.0508)	(0.0513)	(0.0515)	(0.0510)	(0.0507)	(0.0506)
Net migration into the county in 1995	0.0038	0.0037	0.0037	0.0040	0.0040	0.0037	0.0038
(000)	(0.0031)	(0.0032)	(0.0031)	(0.0030)	(0.0030)	(0.0029)	(0.0029)
Change in log total population between	0.0530	0.0531	0.0589	0.0570	0.0636	0.0647	0.0647
1990 and 2000	(0.0151)***	(0.0152)***	(0.0158)***	(0.0152)***	(0.0156)***	(0.0154)***	(0.0154)***
Change in percentage of African American	0.0208	0.0275	0.0068	0.0048	-0.0051	0.0001	-0.0014
1990 to 2000	(0.0736)	(0.0738)	(0.0746)	(0.0734)	(0.0736)	(0.0739)	(0.0738)
Change in percentage of university graduates	0.8301	0.8466	0.8044	0.8190	0.8400	0.8228	0.8394
1990 to 2000	(0.1618)***	(0.1598)***	(0.1596)***	(0.1590)***	(0.1594)***	(0.1601)***	(0.1592)***
Change in percentage of high school	-0.0106	-0.0198	-0.0154	-0.0045	-0.0025	0.0006	-0.0017
graduates 1990 to 2000	(0.0941)	(0.0937)	(0.0939)	(0.0936)	(0.0928)	(0.0934)	(0.0928)
Change in percentage of persons over age	-0.5471	-0.5380	-0.5198	-0.5210	-0.4755	-0.4881	-0.4745
65 1990 to 2000	(0.1191)***	(0.1196)***	(0.1202)***	(0.1198)***	(0.1203)***	(0.1205)***	(0.1202)***
Change in net migration into the county	0.0025	0.0024	0.0023	0.0030	0.0031	0.0030	0.0031
1990 to 2000 (000)	(0.0037)	(0.0037)	(0.0037)	(0.0035)	(0.0035)	(0.0034)	(0.0034)
Constant	0.3075	0.3107	0.3190	0.3372	0.3270	0.3205	0.3216
	(0.0407)***	(0.0394)***	(0.0390)***	(0.0428)***	(0.0445)***	(0.0447)***	(0.0447)***
Observations	2743	2743	2743	2743	2743	2743	2743
R-squared	0.13	0.13	0.13	0.13	0.14	0.14	0.14

Dependent variable is logged wages. Heteroskedasticity-robust standard errors in parentheses.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 9A: Full set of coefficients for Table 4 (First stage)**

<i>Instrument</i> →	(1)	(2)	(3)	(4)
	Programmers in other establishments within the same firm	Bartik index	ARPANET nodes	All three instruments
Programmers in other establishments within the same firm	0.0002 (0.00005)***			0.0002 (0.00005)***
Bartik index		0.2990 (0.1774)*		0.2612 (0.1790)
ARPANET nodes			0.0058 (0.0048)	0.0052 (0.0046)
Average number of establishments in Harte Hanks firms		-0.0116 (0.0141)	-0.0101 (0.0140)	-0.0251 (0.0142)*
Home internet use	0.0350 (0.0324)	0.0345 (0.0327)	0.0380 (0.0323)	0.0317 (0.0326)
Home internet data missing	0.0191 (0.0150)	0.0191 (0.0151)	0.0207 (0.0149)	0.0178 (0.0151)
Log population in 1990	0.0099 (0.0032)***	0.0107 (0.0033)***	0.0116 (0.0031)***	0.0092 (0.0033)**
Percentage African Americans in 1990	-0.0341 (0.0267)	-0.0325 (0.0267)	-0.0304 (0.0267)	-0.0358 (0.0266)
Percentage university graduates in 1990	0.0923 (0.1230)	0.1010 (0.1223)	0.1178 (0.1229)	0.0731 (0.1228)
Percentage high school graduates in 1990	0.0568 (0.1236)	0.0624 (0.1240)	0.0586 (0.1240)	0.0615 (0.1228)
Percentage below poverty line in 1990	0.0363 (0.0983)	0.0559 (0.0996)	0.0400 (0.0990)	0.0511 (0.0989)
Median income in 1990 (\$000)	0.0003 (0.0010)	-0.0004 (0.0010)	-0.0003 (0.0009)	-0.0003 (0.0010)
Percentage population attending Carnegie Type 1 schools in 1990	0.0323 (0.0593)	0.0315 (0.0578)	0.0308 (0.0579)	0.0325 (0.0593)
Percentage population enrolled in engineering program in 1990	-0.3824 (0.5492)	-0.4261 (0.5531)	-0.4383 (0.5553)	-0.3650 (0.5488)
# of patents granted to inventors in the county in the 1980s (000)	0.0034 (0.0026)	0.0035 (0.0027)	0.0029 (0.0027)	0.0025 (0.0027)
Percentage professional in 1995	-0.0606 (0.0870)	-0.0438 (0.0873)	-0.0619 (0.0869)	-0.0445 (0.0875)
Percentage of persons over age 65 in 1990	0.0395 (0.0928)	0.0527 (0.0933)	0.0452 (0.0934)	0.0472 (0.0927)
Net migration into the county in 1995 (000)	-0.0018 (0.0015)	-0.0019 (0.0015)	-0.0020 (0.0015)	-0.0019 (0.0015)
Change in log total population between 1990 and 2000	-0.0055 (0.0266)	-0.0058 (0.0268)	-0.0105 (0.0268)	-0.0011 (0.0266)
Change in percentage of African American 1990 to 2000	0.2824 (0.1465)*	0.2829 (0.1465)*	0.2819 (0.1481)*	0.2909 (0.1470)**
Change in percentage of university graduates 1990 to 2000	-0.2006 (0.2781)	-0.2016 (0.2781)	-0.1869 (0.2786)	-0.2159 (0.2778)
Change in percentage of high school graduates 1990 to 2000	0.0846 (0.2193)	0.0719 (0.2197)	0.0820 (0.2200)	0.0761 (0.2190)
Change in percentage of persons over age 65 1990 to 2000	-0.0861 (0.2263)	-0.0407 (0.2292)	-0.0750 (0.2278)	-0.0559 (0.2279)
Change in net migration into the county 1990 to 2000 (000)	-0.0023 (0.0017)	-0.0024 (0.0017)	-0.0027 (0.0016)	-0.0025 (0.0016)
Constant	-0.0523 (0.0804)	-0.1081 (0.0826)	-0.0748 (0.0804)	-0.0842 (0.0829)
Partial R-squared	0.0067	0.0022	0.0001	0.0084
F-Statistic	12.41	2.84	1.4690	5.77
Observations	2743	2743	2743	2743
R-squared	0.027	0.022	0.020	0.029

Heteroskedasticity-robust standard errors in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 9B: Full set of coefficients for Table 4 (Second stage)**

<i>Instrument</i> →	(1)	(2)	(3)	(4)
	Programmers in other establishments within the same firm	Bartik index	ARPANET nodes	All three instruments
Advanced internet	0.2781 (0.1490)*	0.0156 (0.2374)	2.4859 (2.1301)	0.2752 (0.1587)*
Home internet use	0.0765 (0.0377)**	0.0865 (0.0387)**	0.0055 (0.0403)	0.0766 (0.0377)**
Home internet data missing	0.0250 (0.0172)	0.0305 (0.0175)*	-0.0078 (0.1065)	0.0251 (0.0172)
Average number of establishments in Harte Hanks firms	-0.0168 (0.0074)**	-0.0195 (0.0067)***	-0.0206 (0.0553)	-0.0168 (0.0073)**
Log population in 1990	-0.0085 (0.0028)***	-0.0055 (0.0031)*	-0.0342 (0.0258)	-0.0085 (0.0028)***
Percentage African Americans in 1990	0.0238 (0.0142)*	0.0158 (0.0135)	0.0909 (0.0920)	0.0237 (0.0141)*
Percentage university graduates in 1990	0.5450 (0.0833)***	0.5767 (0.0835)***	0.2777 (0.4091)	0.5453 (0.0831)***
Percentage high school graduates in 1990	-0.1813 (0.0608)***	-0.1661 (0.0537)***	-0.3090 (0.3308)	-0.1811 (0.0607)***
Percentage below poverty line in 1990	-0.1837 (0.0518)***	-0.1734 (0.0468)***	-0.2710 (0.2583)	-0.1836 (0.0516)***
Median income in 1990 (\$000)	-0.0006 (0.0006)	-0.0007 (0.0006)	0.0003 (0.0026)	-0.0006 (0.0006)
Percentage population attending Carnegie Type 1 schools in 1990	0.0211 (0.0369)	0.0293 (0.0490)	-0.0474 (0.1290)	0.0212 (0.0371)
Percentage population enrolled in engineering program in 1990	-0.0850 (0.3257)	-0.2015 (0.3832)	0.8941 (1.5685)	-0.0863 (0.3271)
# of patents granted to inventors in the county in the 1980s (000)	0.0153 (0.0042)***	0.0162 (0.0044)***	0.0069 (0.0113)	0.0153 (0.0042)***
Percentage professional in 1995	-0.0005 (0.0589)	-0.0168 (0.0551)	0.1367 (0.2568)	-0.0007 (0.0586)
Percentage of persons over age 65 in 1990	0.0196 (0.0550)	0.0313 (0.0515)	-0.0791 (0.2458)	0.0197 (0.0548)
Net migration into the county in 1995 (000)	0.0038 (0.0032)	0.0032 (0.0032)	0.0080 (0.0060)	0.0038 (0.0032)
Change in log total population between 1990 and 2000	0.0590 (0.0166)***	0.0562 (0.0153)***	0.0828 (0.0725)	0.0590 (0.0165)***
Change in percentage of African American 1990 to 2000	-0.0395 (0.0950)	0.0333 (0.0993)	-0.6522 (0.6487)	-0.0387 (0.0967)
Change in percentage of university graduates 1990 to 2000	0.8643 (0.1779)***	0.8157 (0.1665)***	1.2731 (0.7923)	0.8637 (0.1785)***
Change in percentage of high school graduates 1990 to 2000	-0.0569 (0.1088)	-0.0354 (0.0960)	-0.2375 (0.5711)	-0.0567 (0.1086)
Change in percentage of persons over age 65 1990 to 2000	-0.5538 (0.1312)***	-0.5735 (0.1198)***	-0.3884 (0.5775)	-0.5540 (0.1311)***
Change in net migration into the county 1990 to 2000 (000)	0.0026 (0.0037)	0.0019 (0.0038)	0.0080 (0.0072)	0.0026 (0.0037)
Constant	0.3263 (0.0520)***	0.3069 (0.0478)***	0.4890 (0.2503)*	0.3261 (0.0517)***
Overidentification test (p-value)	N/A	N/A	N/A	0.158
Hausman test (p-value)	1.000	1.0000	1.0000	1.0000
Observations	2743	2743	2743	2743
R-squared	0.13	0.13	0.13	0.13

Dependent variable is logged wages. Heteroskedasticity-robust standard errors in parentheses.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 10: Other instrumental variables specifications applied to table 2 column 2**

<i>Instrument</i> →	(1)	(2)	(3)	(4)
	First stage	All four instruments		Year when state adopted a price cap
		Second stage	First stage	Second stage
Advanced internet		0.3318 (0.2118)		0.7180 (0.7306)
Programmers in other establishments within the same firm	0.0002 (0.00005)***			
Bartik index	0.2586 (0.1384)*			
ARPANET nodes	0.0056 (0.0049)			
Year when state adopted a price cap	0.0013 (0.0008)		0.0012 (0.0009)	
Home internet use	0.0355 (0.0396)	0.0745 (0.0384)*	0.0418 (0.0380)	0.0597 (0.0544)
Home internet data missing	0.0180 (0.0179)	0.0239 (0.0184)	0.0208 (0.0171)	0.0159 (0.0261)
Average number of establishments in Harte Hanks firms	-0.0256 (0.0134)*	-0.0163 (0.0073)**	-0.0104 (0.0147)	-0.0124 (0.0122)
Log population in 1990	0.0094 (0.0038)**	-0.0092 (0.0041)**	0.0119 (0.0036)***	-0.0137 (0.0094)
Percentage African Americans in 1990	-0.349 (0.0244)	0.0254 (0.0147)*	-0.0295 (0.0240)	0.0372 (0.0325)
Percentage university graduates in 1990	0.0628 (0.1083)	0.5385 (0.0789)***	0.1115 (0.1044)	0.4917 (0.1291)***
Percentage high school graduates in 1990	0.0583 (0.1462)	-0.1844 (0.0754)**	0.0548 (0.1464)	-0.2067 (0.1163)*
Percentage below poverty line in 1990	0.0520 (0.0854)	-0.1859 (0.0687)***	0.0405 (0.0895)	-0.2011 (0.0833)**
Median income in 1990 (\$000)	-0.0002 (0.0010)	-0.0005 (0.0009)	-0.0003 (0.0010)	-0.0004 (0.0011)
Percentage population attending Carnegie Type 1 schools in 1990	0.0322 (0.0523)	0.0194 (0.0382)	0.0308 (0.0502)	0.0075 (0.0369)
Percentage population enrolled in engineering program in 1990	-0.3278 (0.5804)	-0.0612 (0.2399)	-0.4089 (0.5837)	0.1101 (0.4364)
# of patents granted to inventors in the county in the 1980s (000)	0.0020 (0.0027)	0.0150 (0.0036)***	0.0034 (0.0027)	0.0136 (0.0046)***
Percentage professional in 1995	-0.0402 (0.1045)	0.0029 (0.0752)	-0.0579 (0.1021)	0.0269 (0.1129)
Percentage of persons over age 65 in 1990	0.0664 (0.0877)	0.0172 (0.0691)	0.0630 (0.0853)	-0.0001 (0.0892)
Net migration into the county in 1995 (000)	-0.0016 (0.0016)	0.0039 (0.0034)	-0.0016 (0.0017)	0.0046 (0.0038)
Change in log total population between 1990 and 2000	-0.0039 (0.0248)	0.0596 (0.0226)***	-0.0134 (0.0267)	0.0638 (0.0308)**
Change in percentage of African American 1990 to 2000	0.2881 (0.1637)*	-0.0544 (0.0964)	0.2747 (0.1584)	-0.1616 (0.2509)
Change in percentage of university graduates 1990 to 2000	-0.2160 (0.2393)	0.8742 (0.1630)***	-0.1852 (0.2410)	0.9457 (0.2544)***
Change in percentage of high school graduates 1990 to 2000	0.0698 (0.2717)	-0.0613 (0.1327)	0.0757 (0.2733)	-0.0929 (0.2223)
Change in percentage of persons over age 65 1990 to 2000	-0.0491 (0.2385)	-0.5498 (0.1211)***	-0.0680 (0.2454)	-0.5208 (0.1887)***
Change in net migration into the county 1990 to 2000 (000)	-0.0020 (0.0018)	0.0027 (0.0040)	-0.0020 (0.0019)	0.0037 (0.0045)
Constant	-0.2107 (0.1217)*	0.3302 (0.0741)***	-0.1940 (0.1248)	0.3587 (0.0953)***
Partial R-squared		0.0092		0.0007
F-Statistic		5.67		1.88
Overidentification test (p-value)		0.2034		N/A
Hausman Test		1.0000		1.0000
R-squared	0.029	0.13	0.021	0.13

Dependent variable is logged wages. Heteroskedasticity-robust se clustered at the state level in parentheses. \* sig at 10%; \*\* sig at 5%; \*\*\* sig at 1%

**Online Appendix Table 11: Other instrumental variables specifications applied to Table 3 column 6**

	(1)	(2)
	All four instruments	Year when state adopted a price cap
FIRST STAGE—Advanced internet ARPANET nodes	0.0091 (0.0176)	
Programmers in other establishments within the same firm	0.00016 (0.00005)***	
Bartik index	0.2326 (0.1404)*	
Year when state adopted a price cap	0.0012 (0.0008)	0.0012 (0.0009)
ARPANET nodes and high all factors	-0.0052 (0.0172)	
Programmers in other establishments within the same firm and high all factors	0.00003 (0.0001)	
Bartik index and high all factors	0.2181 (0.6335)	
Year when state adopted a price cap and high all factors	0.0007 (0.0011)	0.0005 (0.0012)
Partial R-squared	0.009	0.0007
F-Statistic	3.25	2.27
FIRST STAGE—Advanced internet and high all factors		
ARPANET nodes	-0.0002 (0.0008)	
Programmers in other establishments within the same firm	-3.13E-07 (3.72E-07)	
Bartik index	0.0012 (0.0011)	
Year when state adopted a price cap	-0.000035 (0.000024)	-0.000047 (0.000026)*
ARPANET nodes and high all factors	0.0053 (0.0029)*	
Programmers in other establishments within the same firm and high all factors	0.0002 (0.0001)**	
Bartik index and high all factors	1.0043 (0.5475)*	
Year when state adopted a price cap and high all factors	0.0021 (0.0007)***	0.0016 (0.0006)**
Partial R-squared	0.123	0.021
F-Statistic	5.44	6.42
SECOND STAGE		
Advanced internet	0.2764 (0.2421)	0.7416 (0.7328)
Advanced internet and High income, education, IT-intensity, and population county	1.0476 (0.5874)*	0.3859 (1.5895)
Overidentification test (p-value)	0.3539	N/A
Hausman test (p-value)	1.0000	1.0000
Observations	2743	2743
R <sup>2</sup>	0.13	0.13

Dependent variable is logged wages. Controls same as table 4. Heteroskedasticity-robust standard errors clustered at the state level in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

**Online Appendix Table 12: Benefits of early internet use do not spill over to adjacent locations**

	(1)	(2)
Advanced internet and in same MSA as High income, Education, IT-intensity, and population county	-0.0118 (0.0300)	-0.0008 (0.0309)
Advanced internet	0.0354 (0.0293)	0.0075 (0.0314)
Advanced internet and High income county		0.0377 (0.0497)
Advanced internet and High education county		0.0757 (0.0548)
Advanced internet and High population county		0.0178 (0.1045)
Advanced internet and High IT-intensity county		0.0101 (0.0236)
Advanced internet and High income, education, IT-intensity, and population county	0.4478 (0.1606)***	0.3390 (0.1903)*
Observations	2743	2743
R <sup>2</sup>	0.14	0.14

Dependent variable is logged wages. Controls are the same as in Table 3.

Heteroskedasticity-robust standard errors in parentheses.

+ significant at 10%; \* significant at 5%; \*\* significant at 1%.

**Online Appendix Table 13: Coefficients used to generate Figure 2**

	Advanced internet	Advanced internet and HighAllFactors
Interacted with 1991	0.0014 (0.0068)	-0.0226 (0.0263)
Interacted with 1992	0.0134 (0.0089)	0.0157 (0.0257)
Interacted with 1993	0.0108 (0.0099)	0.0050 (0.0246)
Interacted with 1994	0.0108 (0.0098)	-0.0007 (0.0255)
Interacted with 1995	0.0162 (0.0115)	0.0147 (0.0278)
Interacted with 1996	0.0178 (0.0114)	0.0677 (0.0312)**
Interacted with 1997	0.0289 (0.0131)**	0.1270 (0.0386)***
Interacted with 1998	0.0218 (0.0128)*	0.1521 (0.0373)***
Interacted with 1999	0.0313 (0.0163)*	0.1777 (0.0441)***
Interacted with 2000	0.0418 (0.0144)***	0.2656 (0.0587)***
Observations		30173
R <sup>2</sup>		0.87

Dependent variable is logged wages. This table presents the results of one panel regression of 11 years (1990-2000) and 2743 counties. Base year in 1990. Controls are the same as in Table 4, but separately interacted with each year (e.g. 1991×MedianIncome1990, 1992×MedianIncome1990, 1993×MedianIncome1990, etc.). Year dummies also included. County fixed effects are differenced out. Heteroskedasticity-robust standard errors in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

**Online Appendix Table 14: Advanced internet and number of establishments**

	<i>NUMBER OF ESTABLISHMENTS</i>	
	<i>(1)</i>	<i>(2)</i>
	Full set of controls	High all factors interaction and full set of controls
Advanced internet	-0.0023 (0.0123)	-0.0033 (0.0124)
Advanced internet and High income, education, IT-intensity, population county		0.0637 (0.1248)
Observations	2743	2743
R <sup>2</sup>	0.46	0.46

In columns (1) and (3) controls are the same as table 2. In columns (2) and (4) controls are the same as table 3.

Heteroskedasticity-robust standard errors in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%.

**Online Appendix Table 15: Labor market tightness and advanced internet investment**

	<i>Unemployment rate</i>		<i>Unemployment rate among college graduates</i>		<i>The Rust Belt</i>	
	<i>(1)</i>	<i>(2)</i>	<i>(3)</i>	<i>(4)</i>	<i>(5)</i>	<i>(6)</i>
	Below median	At or above median	Below median	At or above median	In rust belt	Outside rust belt
Advanced internet	0.1027 (0.0599)*	0.0787 (0.0682)	0.1665 (0.0507)***	-0.0203 (0.0730)	0.0030 (0.0327)	0.0311 (0.0138)**
Observations	266	266	263	269	642	2101
R <sup>2</sup>	0.31	0.29	0.33	0.30	0.10	0.16

Dependent variable is change in logged wage growth from 1995 to 2000. Heteroskedasticity-robust standard errors in parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Columns (1) through (4) show estimates of the association between advanced internet and local wages for subsamples of counties where the overall unemployment rate and unemployment rate among skilled workers was above and below the median. Dependent variable is logged wages. Controls are the same as in Table 2. We found reliable county-level unemployment numbers for 532 counties.

Columns (5) and (6) show estimates of the association between advanced internet and local wages for subsamples of counties inside and outside the rust belt. Dependent variable is logged wages. Controls are the same as table 2. The rust belt is defined as Illinois, Indiana, Ohio, Wisconsin, Michigan, New York outside of the New York City CMSA, Pennsylvania outside of the Philadelphia CMSA, Maryland, and West Virginia.