Supplement to

Economic consequences of vertical mismatch

Quantitative Economics

Pietro Garibaldi, Pedro Gomes, and Thepthida Sopraseuth

A Existence and uniqueness

Assuming a Cobb-Douglas production function, substituting the wage equations (6) on the expressions of over- and under-employment:

$$u = n\left[\frac{e^{\frac{\zeta(1-\alpha)}{\nu}\left(\frac{(n-u)+\chi o}{(1-n-o)+\zeta u}\right)^{\alpha}}}{e^{\frac{\zeta(1-\alpha)}{\nu}\left(\frac{(n-u)+\chi o}{(1-n-o)+\zeta u}\right)^{\alpha}} + e^{\frac{\alpha}{\nu}\left(\frac{(1-n-o)+\zeta u}{(n-u)+\chi o}\right)^{1-\alpha}}}\right] \tag{A.1}$$

$$o = (1-n)\Theta\left[\frac{e^{\frac{\chi\alpha}{\nu}\left(\frac{(1-n-o)+\zeta u}{(n-u)+\chi o}\right)^{1-\alpha}}}{e^{\frac{\chi\alpha}{\nu}\left(\frac{(1-n-o)+\zeta u}{(n-u)+\chi o}\right)^{1-\alpha}} + e^{\frac{(1-\alpha)}{\nu}\left(\frac{(n-u)+\chi o}{(1-n-o)+\zeta u}\right)^{\alpha}}}\right] \tag{A.2}$$

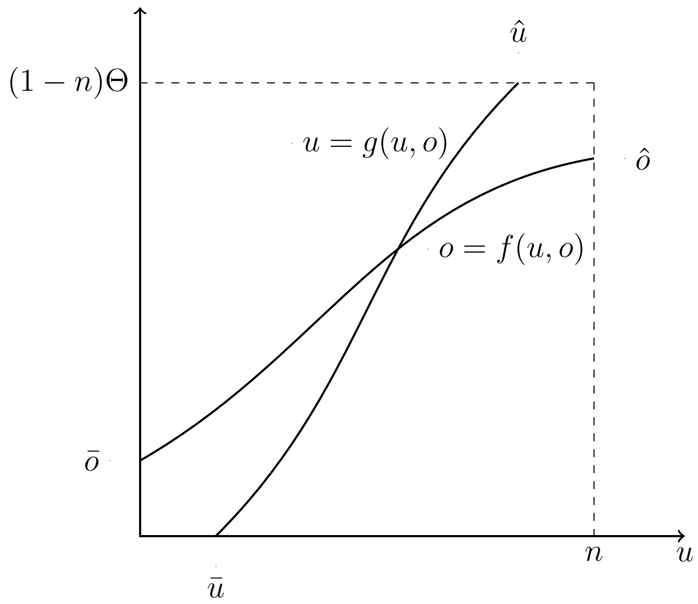
So we end up having two equations and two unknowns

$$u = g(u(-), o(+))$$
$$o = f(u(+), o(-))$$

When evaluated at the origin, $g(0,0) = n\left[\frac{e^{\frac{\zeta(1-\alpha)}{\nu}\left(\frac{n}{1-n}\right)^{\alpha}}}{e^{\frac{\zeta(1-\alpha)}{\nu}\left(\frac{n}{1-n}\right)^{\alpha}} + e^{\frac{\alpha}{\nu}\left(\frac{1-n}{n}\right)^{1-\alpha}}}\right] > 0$ and

$f(0,0) = (1-n)\Theta\left[\frac{e^{\frac{\chi\alpha}{\nu}\left(\frac{1-n}{n}\right)^{1-\alpha}}}{e^{\frac{\chi\alpha}{\nu}\left(\frac{1-n}{n}\right)^{1-\alpha}} + e^{\frac{(1-\alpha)}{\nu}\left(\frac{n}{1-n}\right)^{\alpha}}}\right] > 0$, which means the solution is above the 45 degree plane. As $g(n,0) = 0$, the solution is below the 45 degree plane and both functions are continuous, there exists an equilibrium.

The figure below plots the two functions



where $\bar{u}$, $\hat{u}$, $\bar{o}$ and $\hat{o}$ are implicitly defined by

$$\bar{u} = n\left[\frac{e^{\frac{\zeta(1-\alpha)}{\nu}\left(\frac{(n-\bar{u})}{(1-n)+\zeta\bar{u}}\right)^{\alpha}}}{e^{\frac{\zeta(1-\alpha)}{\nu}\left(\frac{(n-\bar{u})}{(1-n)+\zeta\bar{u}}\right)^{\alpha}} + e^{\frac{\alpha}{\nu}\left(\frac{(1-n)+\zeta\bar{u}}{(n-\bar{u})}\right)^{1-\alpha}}}\right] \tag{A.3}$$

$$\hat{u} = n\left[\frac{e^{\frac{\zeta(1-\alpha)}{\nu}\left(\frac{(n-\hat{u})+\chi(1-n)\Theta}{(1-n)(1-\Theta)+\zeta\hat{u}}\right)^{\alpha}}}{e^{\frac{\zeta(1-\alpha)}{\nu}\left(\frac{(n-\hat{u})+\chi(1-n)\Theta}{(1-n)(1-\Theta)+\zeta\hat{u}}\right)^{\alpha}} + e^{\frac{\alpha}{\nu}\left(\frac{(1-n)(1-\Theta)+\zeta\hat{u}}{(n-\hat{u})+\chi(1-n)\Theta}\right)^{1-\alpha}}}\right] \tag{A.4}$$

$$\bar{o} = (1-n)\Theta\left[\frac{e^{\frac{\chi\alpha}{\nu}\left(\frac{(1-n-\bar{o})}{(n)+\chi\bar{o}}\right)^{1-\alpha}}}{e^{\frac{\chi\alpha}{\nu}\left(\frac{(1-n-\bar{o})}{(n)+\chi\bar{o}}\right)^{1-\alpha}} + e^{\frac{(1-\alpha)}{\nu}\left(\frac{(n)+\chi\bar{o}}{(1-n-\bar{o})}\right)^{\alpha}}}\right] \tag{A.5}$$

$$\hat{o} = (1-n)\Theta\left[\frac{e^{\frac{\chi\alpha}{\nu}\left(\frac{(1-n-\hat{o})+\zeta n}{\chi\hat{o}}\right)^{1-\alpha}}}{e^{\frac{\chi\alpha}{\nu}\left(\frac{(1-n-\hat{o})+\zeta n}{\chi\hat{o}}\right)^{1-\alpha}} + e^{\frac{(1-\alpha)}{\nu}\left(\frac{\chi\hat{o}}{(1-n-\hat{o})+\zeta n}\right)^{\alpha}}}\right] \tag{A.6}$$

To show that they only cross once, it is sufficient that the slope of the $u = g(u,o)$ is always steeper than the other. Applying the total differentiation to the condition of under-employment

$$\frac{du}{do}\mid_g = \frac{\chi j_l + j_h}{\Omega_1 + j_l + \zeta j_h} \tag{A.7}$$

and doing the same to the condition of over-employment

$$\frac{do}{du}\mid_f = \frac{j_l + \zeta j_h}{\Omega_2 + \chi j_l + j_h} \tag{A.8}$$

where $\Omega_1 = \frac{n j_h j_l \nu}{(n-u)u(\zeta\alpha w_l + (1-\alpha)w_h} > 0$ and $\Omega_2 = \frac{(1-n)\Theta j_h j_l \nu}{((1-n)\Theta - o)o(\alpha w_l + \chi(1-\alpha)w_h} > 0$.

Calculating the ratio we can show it is always larger than 1.

$$\frac{\frac{1}{\frac{du}{do}\mid_g}}{\frac{do}{du}\mid_f} = 1 + \frac{\Omega_1}{j_l + \zeta j_h} + \frac{\Omega_2}{\chi j_l + j_h} + \frac{\Omega_1\Omega_2}{(j_l + \zeta j_h)(\chi j_l + j_h)} > 1 \tag{A.9}$$

B Census IPUMS data

We use Census IPUMS data (1970, 1980, 1990, and annually between 2000 and 2017). These data sources provide substantially larger samples than either the March or May/ORG surveys. They are better suited for the analysis of educational attainment in occupations. The Census samples comprise 1% of the US population in 1970, and 5% of the population in post-1980 surveys.

College education. Education attainment is measured by the highest year of school or degree completed. We define college-educated workers as those who completed at least four years of college education.

Weekly earnings. Wage is captured by weekly earnings computed as annual wage and salary income divided by worked weeks. The number of worked weeks is given in intervals (1-13 weeks, 14-26 weeks, and so on), instead of the precise number of weeks. This is because the 1960 and 1970 samples recorded weeks worked only in intervals. We then consider that the number of worked weeks lies in the middle of the interval. For example, worked weeks equals $7 = (1 + 13)/2$ in the first interval, etc ... We use nominal weekly earnings as we are interested in wage ratios computed within the same year (well-matched premium for college and non-college workers, college premium).

Share of college in each occupation. We want to compute the share of college workers within each 4-digit occupation. This indicator is crucial in determining occupations that are majority college. However, occupations have evolved over time. How do we deal with this issue? First, the occupational categories are based on the 2010 harmonized occupation coding scheme based on the Census Bureau's 2010 ACS occupation classification scheme, with 452 categories. All through our sample, the occupation categories are the same in each survey. Secondly, an occupation is considered as majority college when more than 50% of workers within this occupation are college-educated. We consider the 2017 survey as providing a benchmark for categorizing occupations as majority college. In 2017, 35% of the 452 occupations are majority college.

We apply this 2017-benchmark for each survey (1970, 1980, 1990, 2000-2016). We then view the educational requirements for all occupations as the one prevailing in 2017. In doing so, we apply a harmonized measure of educational attainment for all surveys of our sample.

This strategy does not introduce a strong bias in our empirical exercise. To illustrate this point, we compare occupations that are majority college in 1970 and in 2017. Educational requirement are similar in both surveys for 92% of occupations. 8% of occupations are characterized by an increase in educational requirement. Occupations in arts (such as photographer, musicians, designer, actors, producers and directors) were majority non-college in 1970 and are majority college in 2017. The same observation is made for occupations in sales involving complex products (such as financial or business services, or real estate) or occupations that have become more complex over time (such as accountants and auditors, aircraft pilots and flight engineers, computer programmers) or health-related occupations (medical and health services managers, health diagnosing and treating practitioners, podiatrists).

C Entry-level education requirements from BLS Occupational Outlook Handbook

The BLS Occupational Outlook Handbook provides information on entry-level education requirements for 792 occupational profiles, which could be viewed as an alternative for education requirements as measured in our paper.

In this section, we argue that our measure of education requirements is consistent with entry-level education requirements as documented by BLS Occupational Outlook Handbook.

To establish this point, we first collect entry-level education requirements for occupational profiles available in the 2020 BLS Occupational Outlook Handbook (OOH) provides [1]. For instance, for "accountants and auditors", the BLS reports entry level education as "Bachelor's degree", and for "bakers" "No formal educational credential".

Using our methodology described in section 4.2, "accountants and auditors" are considered a majority-college education and "bakers" as majority non-college. For these occupations, our categorization and BLS education requirements are consistent. We repeat the procedure for a large set of occupations. [2]

Among the occupations categorized as "majority-college" using our measure, 94% of them indeed require "Bachelor's degree", "Master's degree" or "Doctoral or professional degree" according to the BLS. [3]

Symmetrically, among the occupations categorized as "majority-non college" using our measure, 96% of them indeed require "High school diploma or equivalent", "No formal educational credential" or "Postsecondary nondegree award" according to the BLS. [4]

[1] We use the currently available OOH from the BLS website, as there is no archive on past OOH. We compare the 2020 OOH to our most recent Census data.

[2] BLS OOH does not report any occupational classification code. We then use text matching algorithms to match occupational titles from Census 2017 with BLS occupational names. We were then able to successfully match 97% of 2017 Census occupations. The occupations that were not matched relate to the Gaming jobs, that are not yet documented in the BLS 2020 OOH.

[3] The discrepancy between BLS OOH and our measure comes from occupations such as "Private Detectives and Investigators" that require "High school diploma or equivalent" according to the BLS, but is actually "majority college" as 56% of Private Detectives and Investigators held a Bachelor's degree or more.

[4] The discrepancy between BLS OOH and our measure comes from occupations such as "Statistical assistants" or "Construction managers" that require "Bachelor's degree" according to the BLS, but is actually "majority non-college" as only 33% of Statistical assistants, and 35% of Construction managers held a Bachelor's degree or more, in 2017.

D CPS Data on Switchers

The Current Population Survey (CPS) Basic Monthly Data provides information on labor market status. The survey is conducted on a monthly basis. A housing unit in the CPS is interviewed for four consecutive months and then dropped out of the sample for the next eight months and is brought back in the following four months. Hours and earning questions are asked only at households in their 4th and 8th interview (1 year later). These outgoing interviews are the only ones who report information on weekly earnings. The Merged Outgoing Rotation Groups (MORG) allow to match workers between the 4th and 8th interview, with information, in each interview, on labor market status, occupation and wages.

We use the IPUMS monthly CPS data. Since we look at weekly earnings, the data only starts in 1982m1. The time period stops in 2017m12 for the sake of consistency with our empirical exercise on Census data. As we need to compare wages of different years, we compute real earnings using the Consumer Price Index for All Urban Consumers: All Items in U.S. City Average, Index 1982-1984=100, Seasonally Adjusted. We categorize workers as college / non-college workers and well-matched / under-employed / over-employed using the same definitions as the ones reported in section 4.2. [5] For the purpose of consistency with our empirical study on Census data, the categorization of occupations as majority non-college / majority college is based on the same list as in section 4.2.

We focus on individuals who are employed in the 4th and 8th interviews. We compute yearly transitions by linking the 4th to the 8th interview of CPS respondents. College CPS respondents are identified as switchers when

i. they are well-matched in the 4th interview and under-employed in the 8th interview

ii. they are under-employed in the 8th interview and well-matched in the 8th interview

For each type of transition, earnings of each well-matched worker can be compared to earnings when the very same worker is under-employed. The well-matched wage premium can be computed for each college CPS switcher. The data suggests that, on average, workers who start as mismatched and become well-matched 1 year later are as numerous as workers doing the reverse transition. The well-matched wage premium of college workers is then the mean of well-matched wage premium of workers with transition i. and their counterparts with transition ii.

We repeat the procedure on non-college CPS respondents who are identified as switchers when

i. they are well-matched in the 4th interview and over-employed in the 8th interview

ii. they are over-employed in the 8th interview and well-matched in the 8th interview

The well-matched wage premium can be computed for each non-college CPS switcher.

[5] Namely, individuals are categorized as college-educated when their educational attainment in Bachelor degree or more (4 years of college education or more). A college-educated (non-college) worker is classified as under-employed (over-employed) when working in an occupation that is majority non-college (college). A college-educated (non-college) workers working in an occupation that is majority college (non-college) is considered as well-matched.

E Mincer regression

Estimation of parameters over time: Using Census data, we first focus on college workers, and estimate Mincer regressions of individuals' real weekly earnings on observed characteristics (race, gender, age, age^2), state dummies, and a well-matched dummy that equals 1 if the worker is well-matched, 0 otherwise.

For a given year t, we estimate an equation of the form:

$$lnw_i^{college} = \alpha + \beta_t \text{Well-Matched}_i + \gamma_t X_i + \delta \text{State}_i + \epsilon_i \tag{A.10}$$

where i refers to college worker i. X_i denotes the vector of observed individual characteristics (race, gender, age, age^2), and State$_i$ the state dummy. We repeat the estimation for all years t in the Census sample and get a time-series for the estimated coefficient. We focus on coefficient β, which turns out to be significant at a 1% level, for all years t. Figure 8 reports the value of $exp(\beta_t)$.

We repeat the estimation of equation (A.10) for non-college real weakly earnings $lnw_{it}^{non-college}$, for each given year t.

Estimation of parameters across states: For each given state k, we estimate an equation of the form:

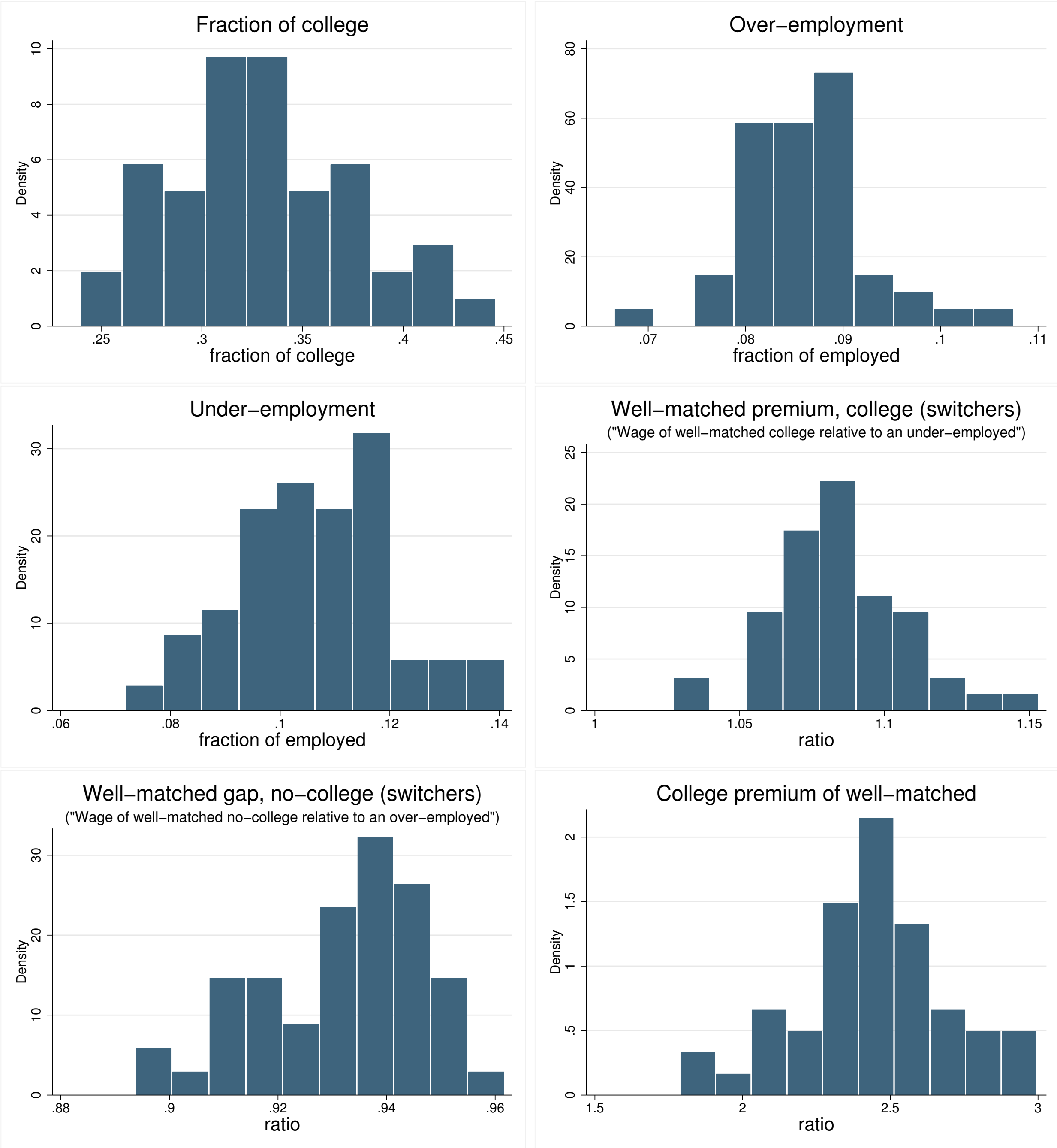
$$lnw_{it}^{college} = \alpha + \beta_k \text{Well-Matched}_{it} + \gamma_k X_{it} + \delta \text{State}_{it} + \epsilon_{it} \tag{A.11}$$

where i refers to college worker i, in year t. X_{it} denotes the vector of observed individual characteristics (race, gender, age, age^2), and State$_{it}$ the state dummy. We repeat the estimation for all states k and get a set of estimated coefficients for each state k. We focus on coefficient β, which turns out to be significant at a 1% level, for all states k. Figure A6 reports the value of $exp(\beta_k)$.

We repeat the estimation of equation (A.11) for non-college real weakly earnings $lnw_{it}^{non-college}$, for each given state k.

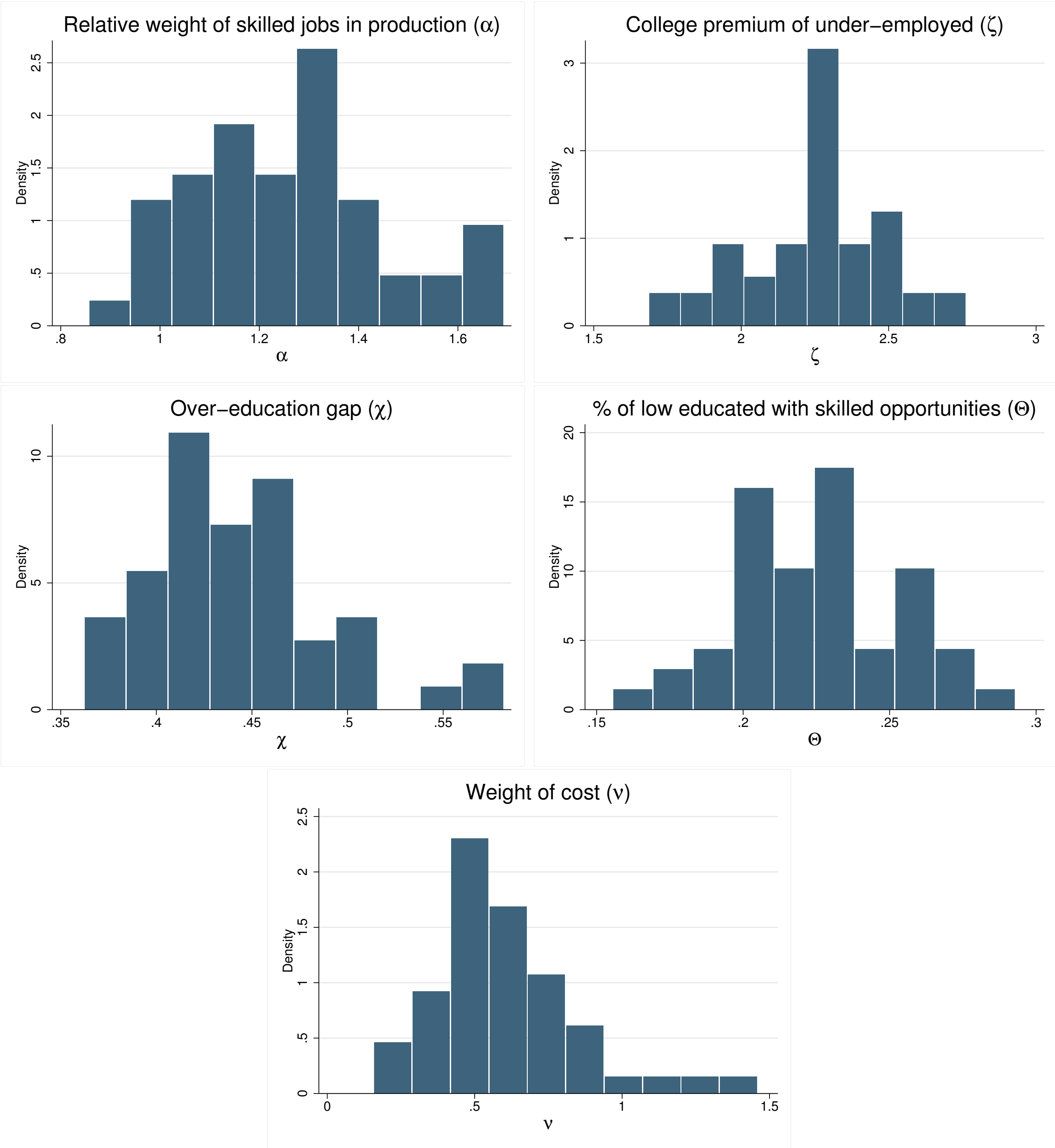
F Cross-State results

Figure A1: Histogram of Cross-State Calibration Targets



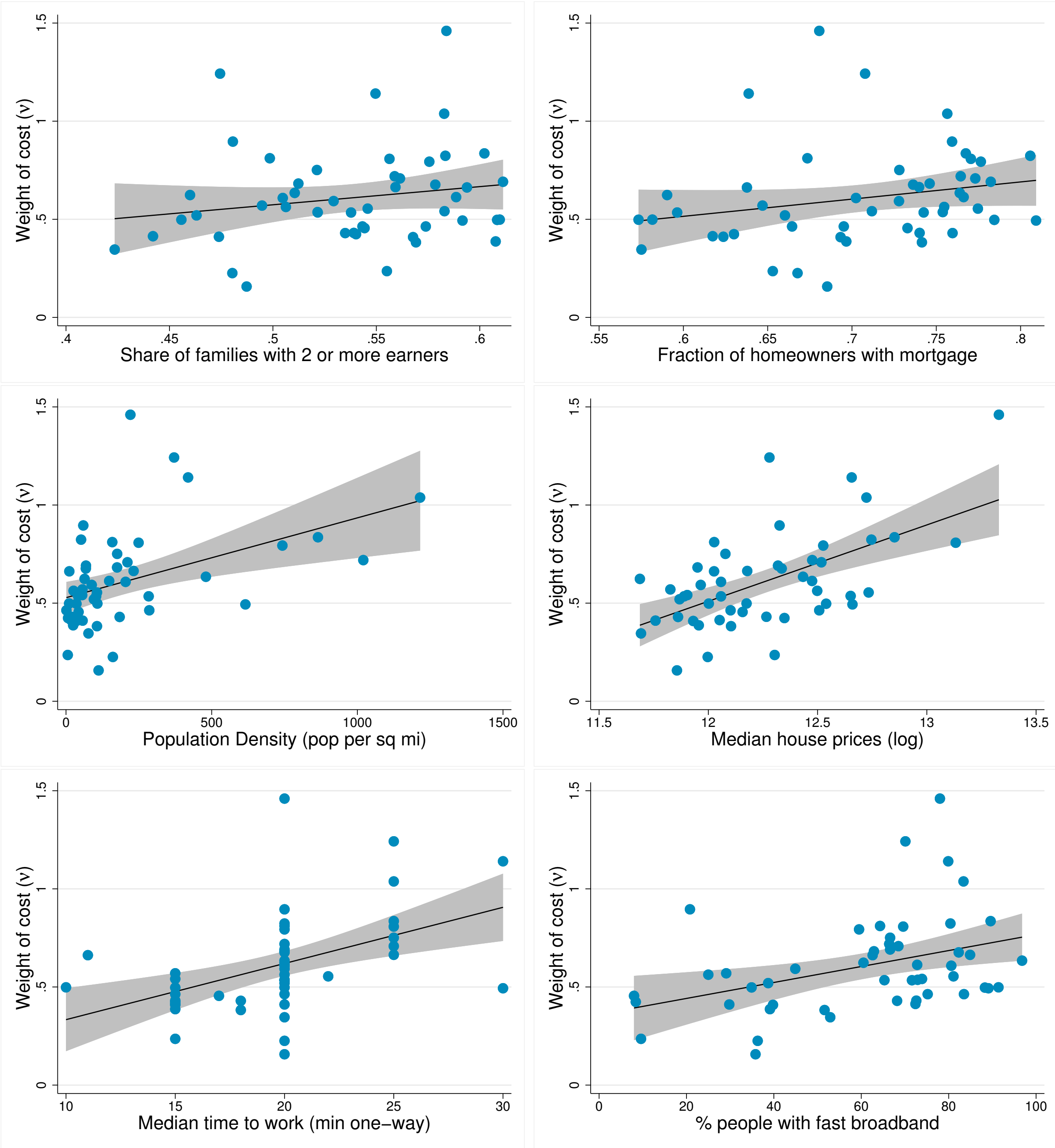
Note: Calculated from US Census data (1970, 1980, 1990, 2000-2017), and CPS monthly data on occupational switchers (1982-2017). A worker is considered as under-employed (over-employed) if she has a college (non-college) education and working in 4-digit occupations that are majority non-college (college). Wage from Census weekly earnings.

Figure A2: Histogram of Cross-State Parameter values



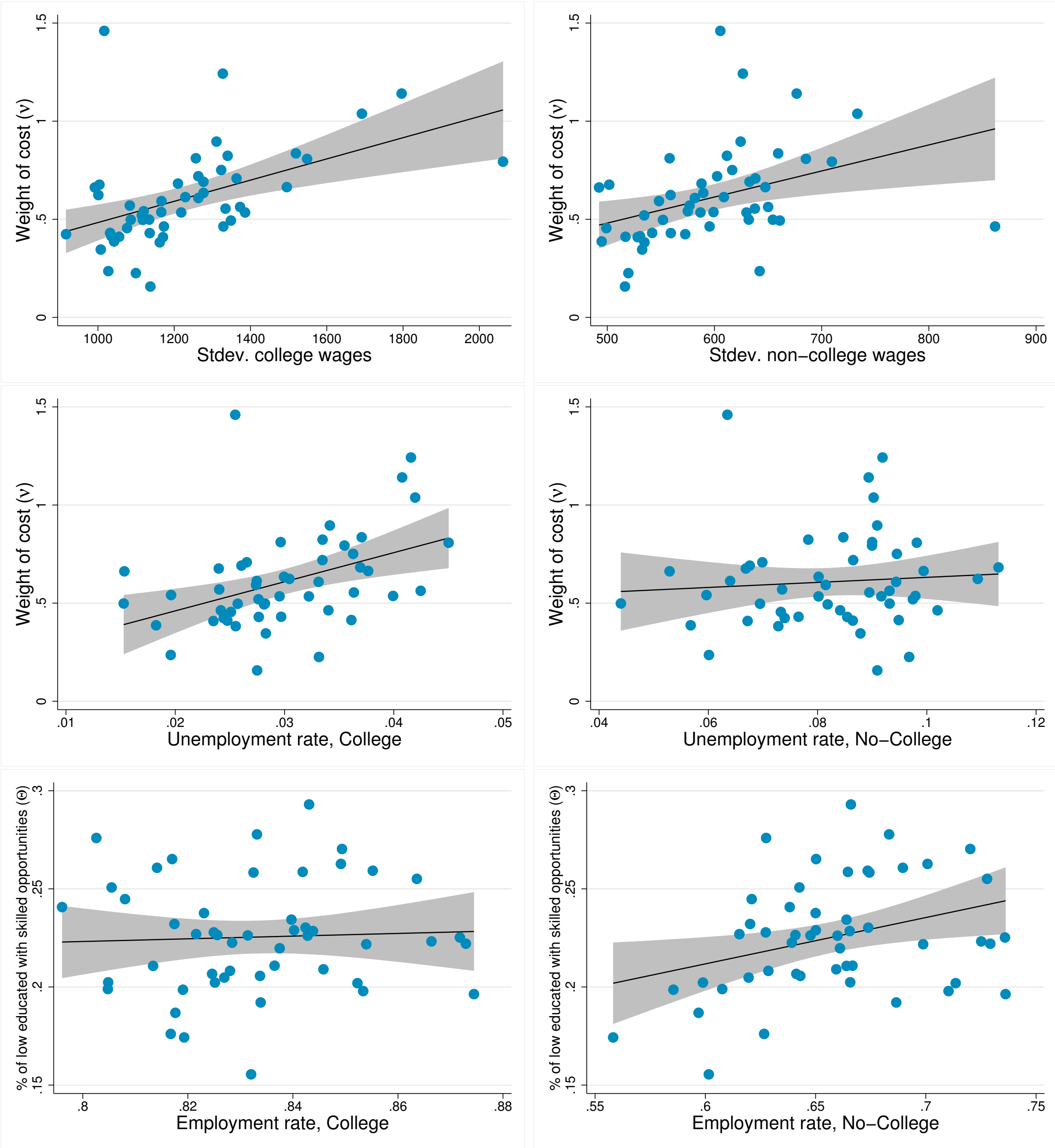
Estimated to fit empirical targets.

Figure A3: Association of ν with External Variables



Note: The scatter plots show the correlations of ν with the share of families with 2 or more earners (0.18), the fraction of homeowners with a mortgage (0.23), population density (0.44), the log of the median house prices (0.58), median time to work (0.47) and the percentage of people with access to fast broadband (0.38). When all variables are included jointly in a regression the log of median house prices and the median time to work are statistically significant at 5%.

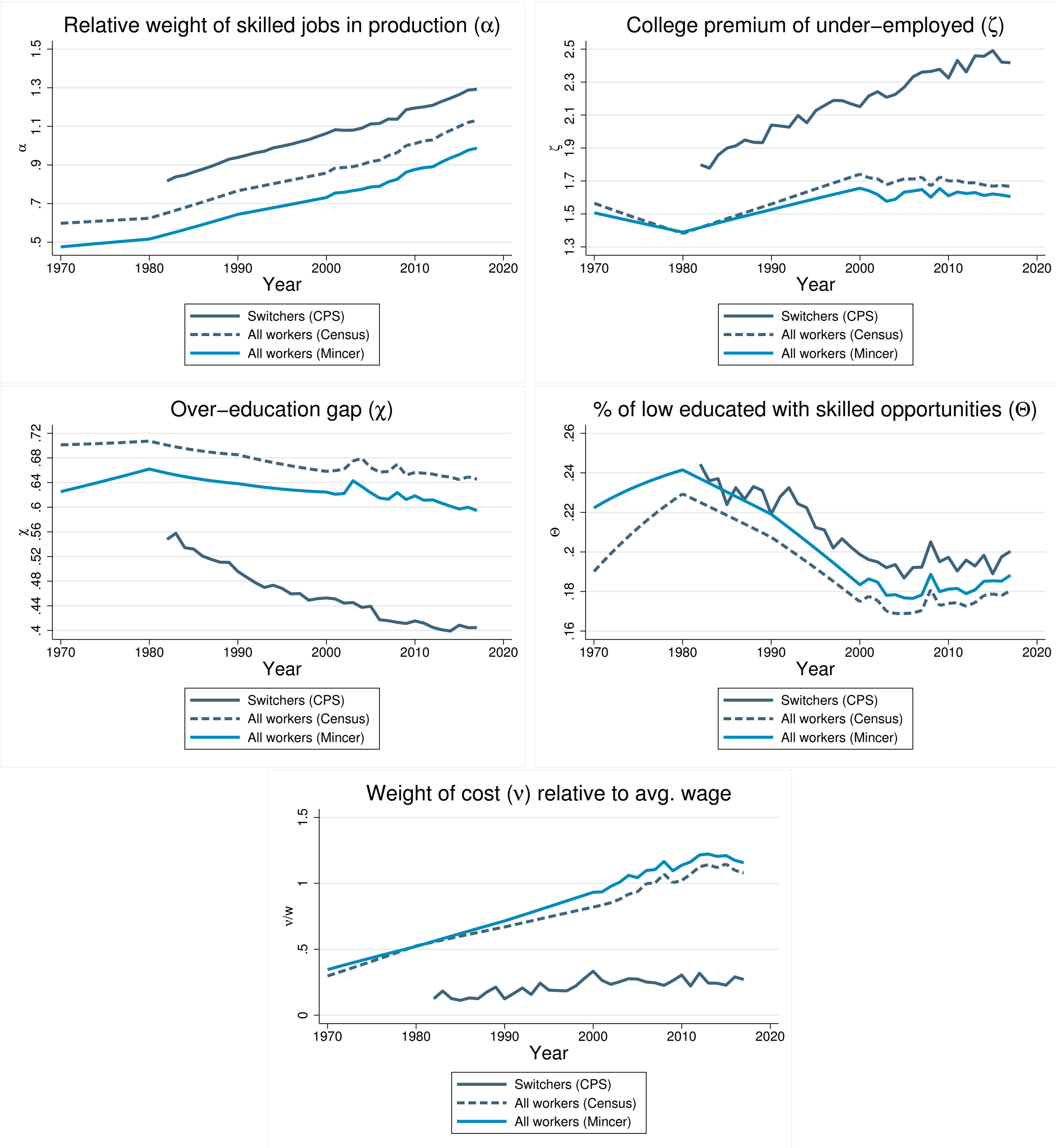
Figure A4: Association of ν and Θ with other Economic Variables



Note: The scatter plots show the correlations of ν with the standard deviation of workers with college (0.84), standard deviation of workers with non-college (0.37), unemployment rate of college graduates (0.42) and with the unemployment rate of workers without college (0.08), as well as the correlation of Θ with the employment rate of college graduates (0.05) and employment rate of non-college graduates (0.34).

G Comparison Alternative Calibrations

Figure A5: Time-Series Parameter Values



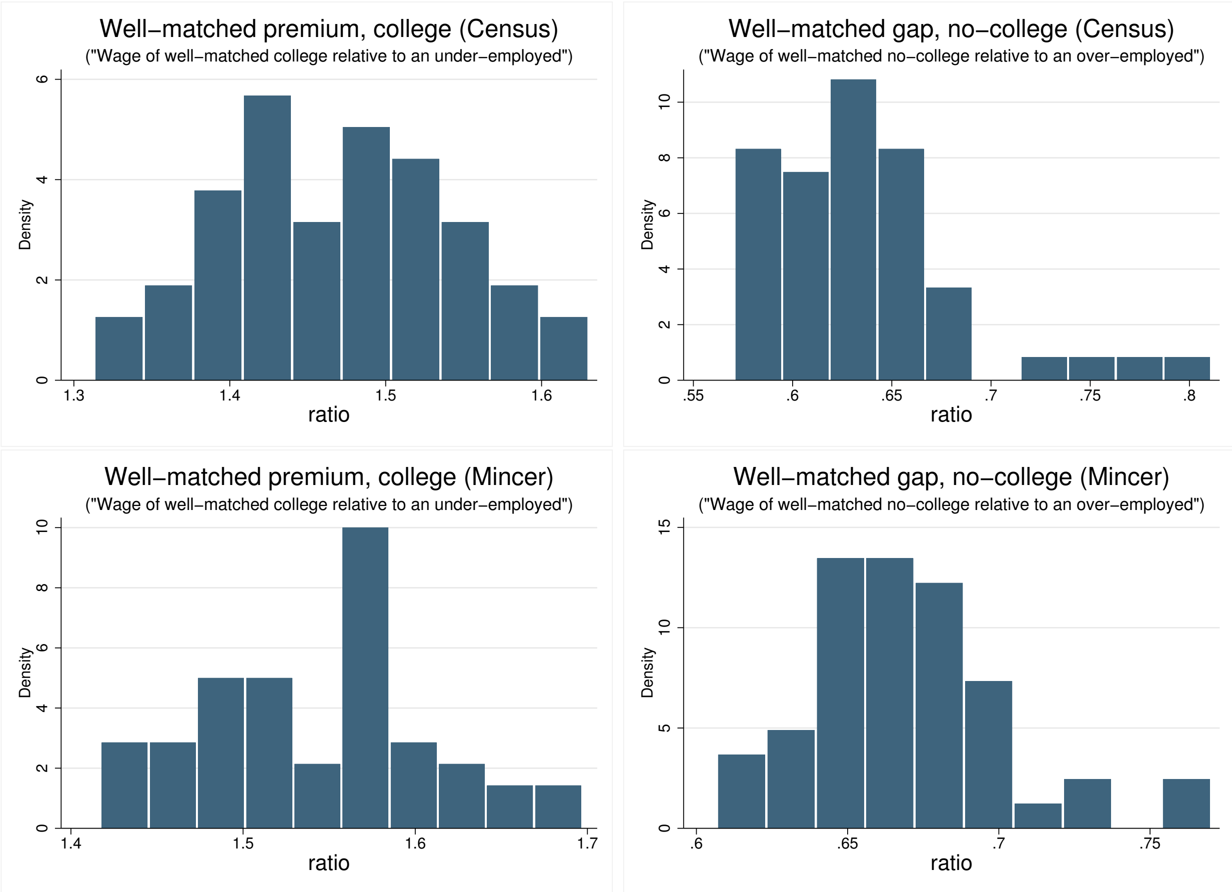
Estimated to fit empirical targets.

Table A1: Counterfactual Exercises (Alternative Calibrations)

	All workers (Mincer)			***All workers (Census)***		
	Mincer (1970-2017)			*Census (1970-2017)*		
	o	u	CP	o	u	CP
Variation	-0.05	0.09	0.36	-0.05	0.09	0.38*
Percentage explained by						
α	-132%	-33%	109%	-78%	-33%	142%
n	103%	143%	-71%	94%	141%	-55%
χ	8%	-3%	14%	12%	-4%	23%
ζ	-5%	4%	9%	-4%	5%	9%
ν	237%	-33%	318%	71%	31%	76%
Θ	28%	-4%	20%	9%	-2%	6%
α, n	-40%	38%	-69%	25%	25%	-48%
χ, ζ, ν, Θ	237%	-33%	318%	31%	31%	117%
Technology (α, χ)	-132%	-33%	127%	-78%	-33%	173%
Education (n, ζ)	95%	150%	-51%	86%	150%	-36%
Mismatch (ν, Θ)	237%	-33%	318%	79%	31%	86%

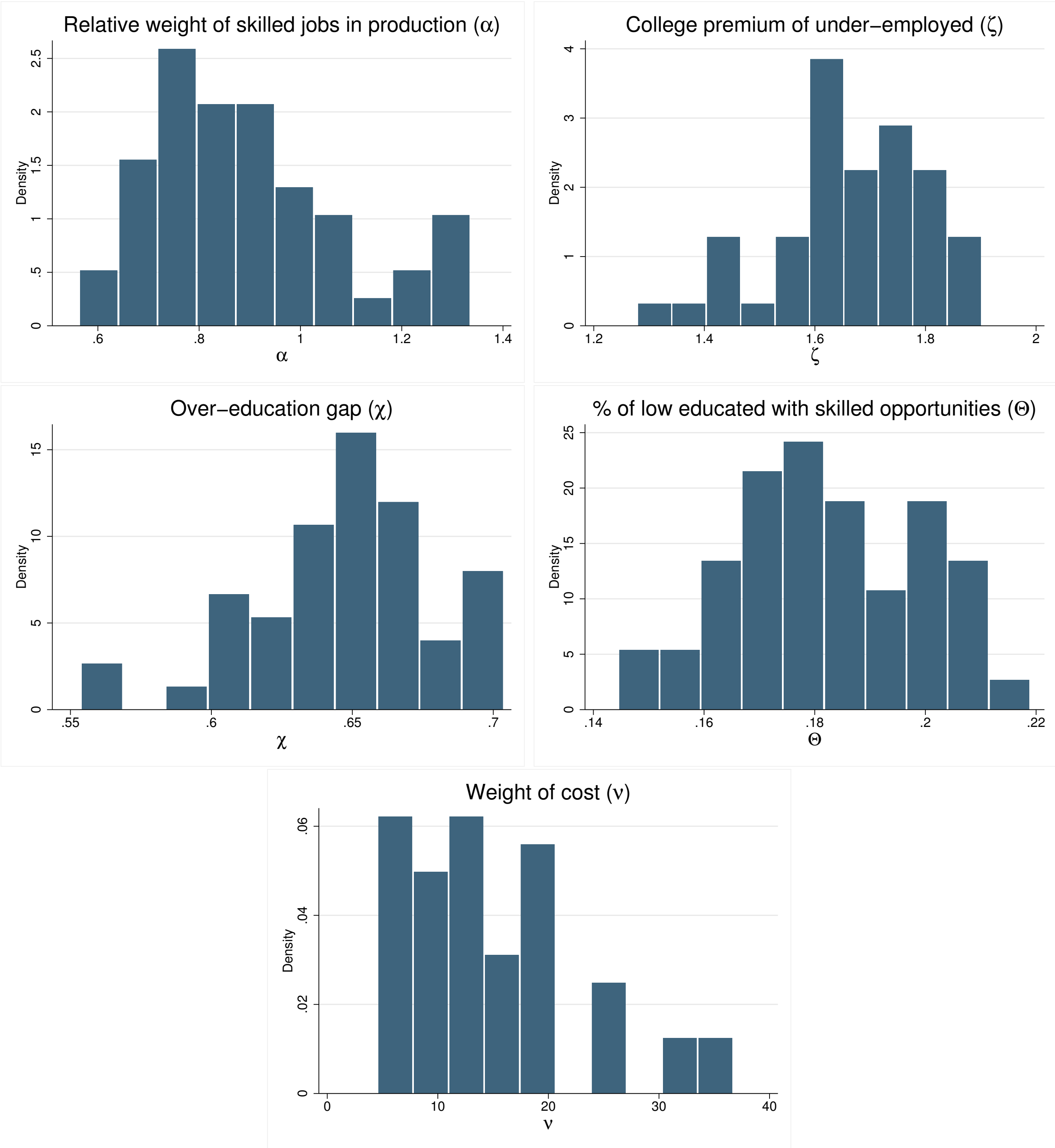
Own calculations based on model simulations, varying each parameter or combination of parameters while keeping the remaining parameters constant, at their value in the beginning of the sample.

Figure A6: Histogram of Cross-State Calibration Targets (Alternative Calibration)



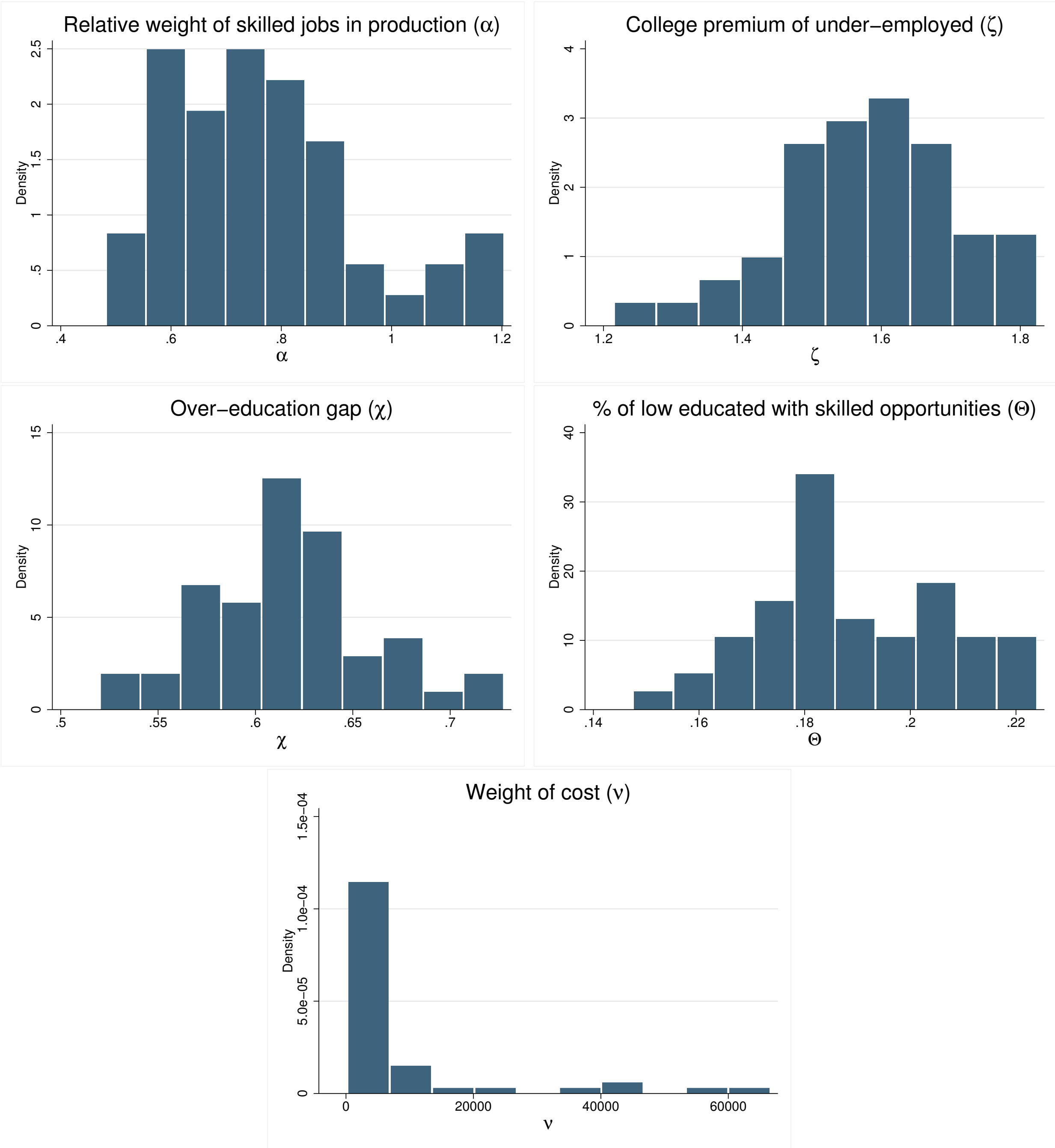
Note: Calculated from US Census IPUMS data (1970, 1980, 1990, 2000-2017, bottom panels) on all workers, and using Mincer regressions. A worker is considered as under-employed (over-employed) if she has a college (non-college) education and working in 4-digit occupations that are majority non-college (college).

Figure A7: Histogram of Cross-State Parameter Values (Census)



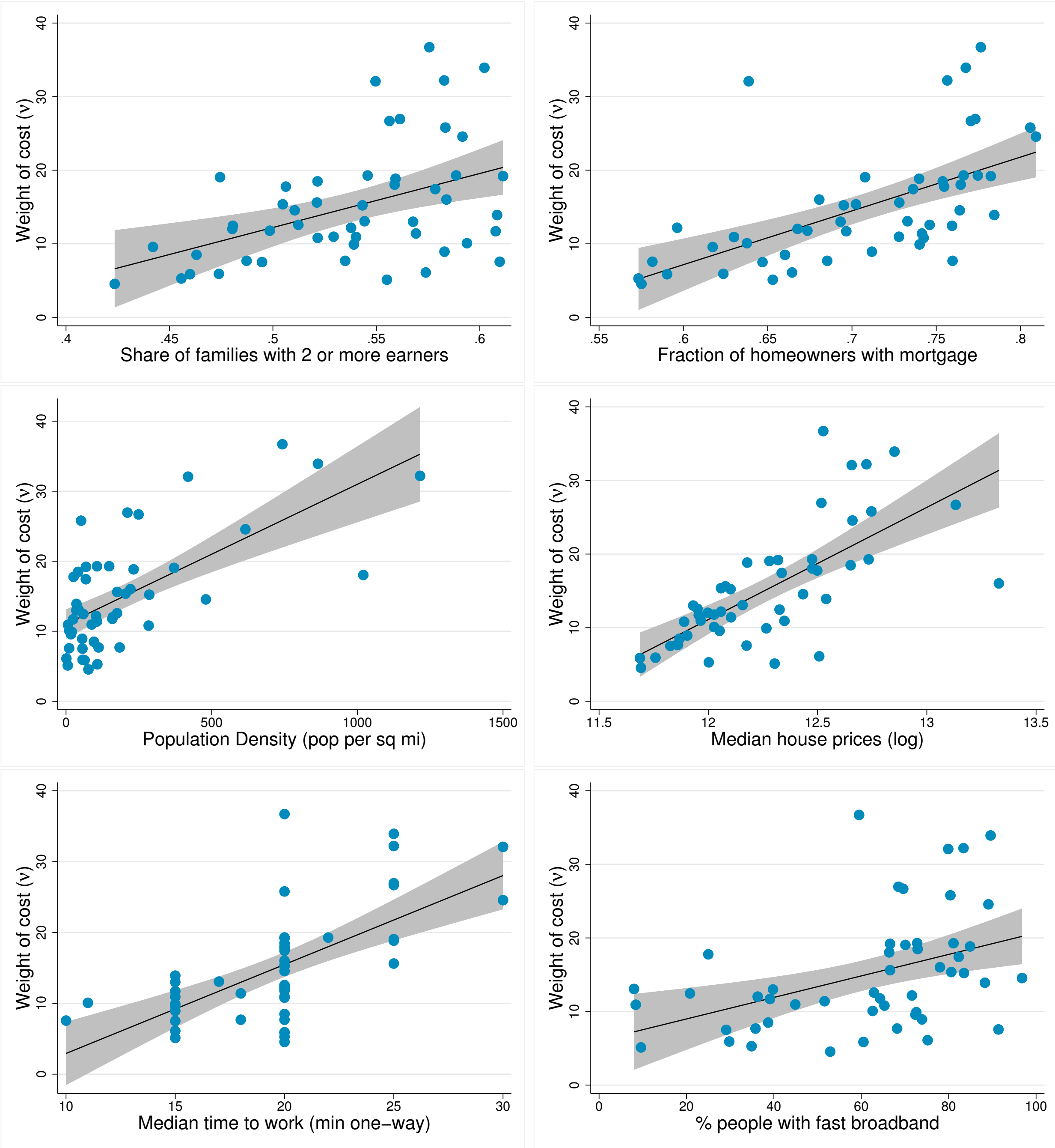
Estimated to fit empirical targets.

Figure A8: Histogram of Cross-State Parameter values (Mincer)



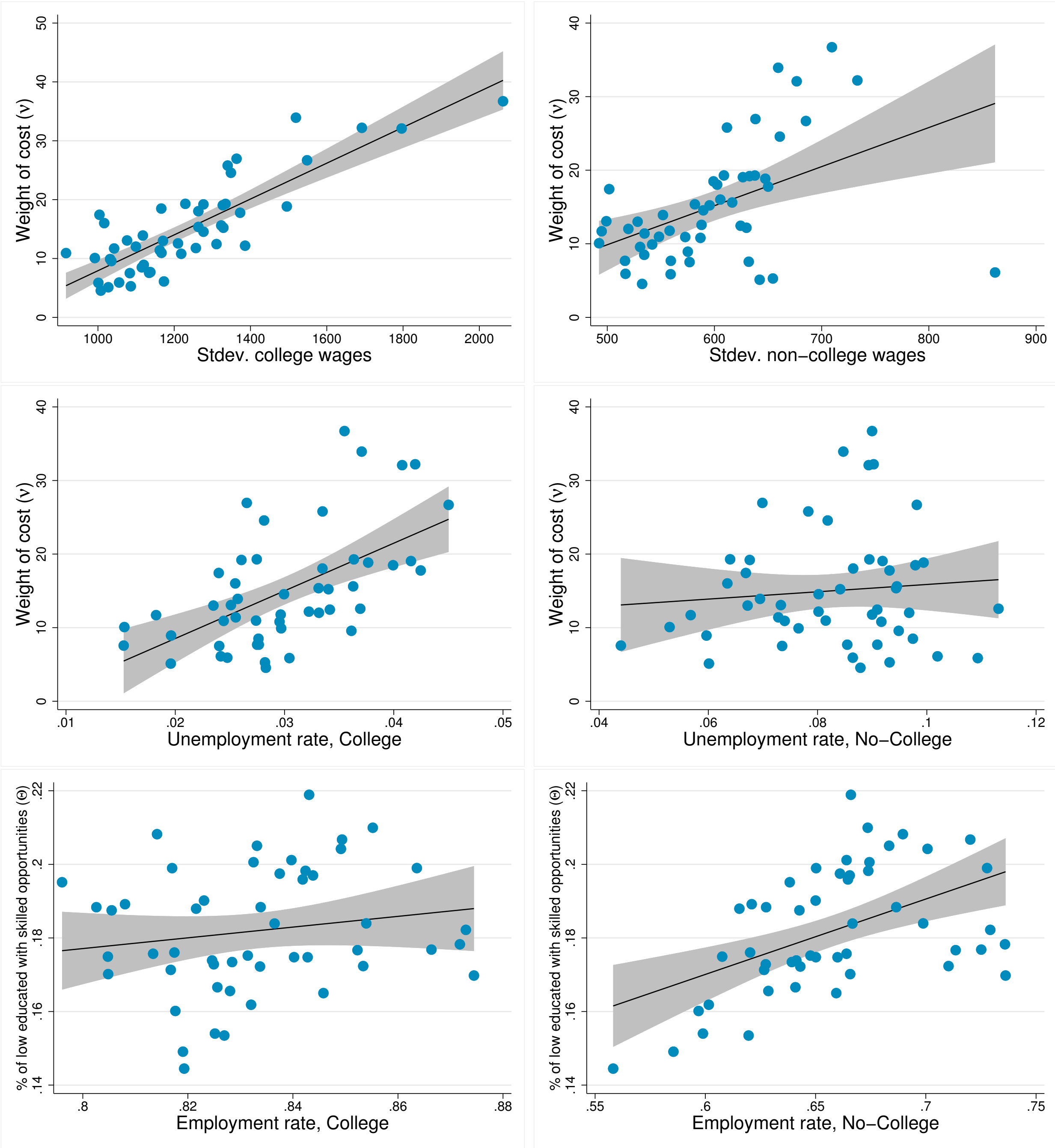
Estimated to fit empirical targets.

Figure A9: Association of ν with External Variables (Census)



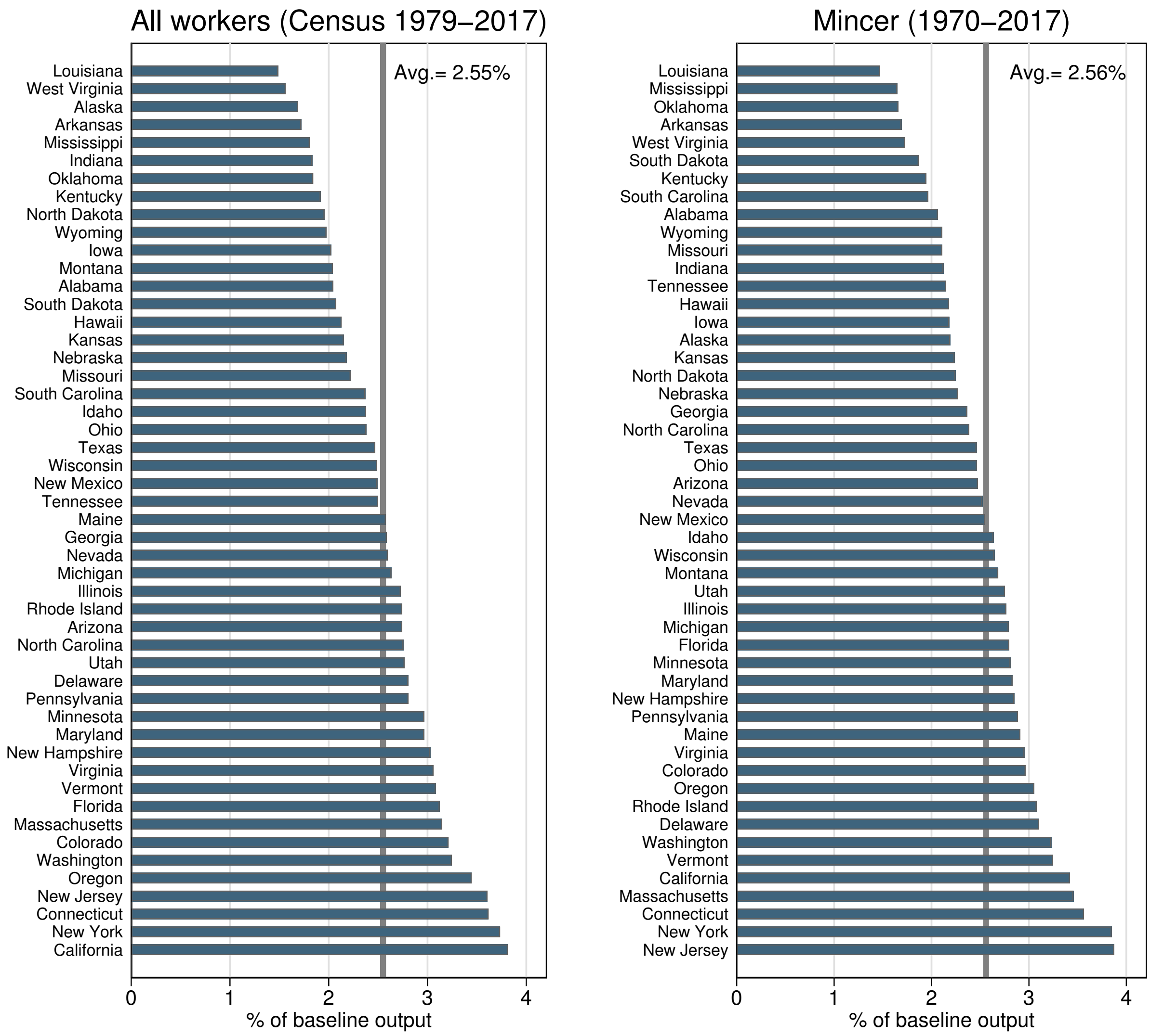
Note: The scatter plots show the correlations of ν with the share of families with 2 or more earners (0.451), the fraction of homeowners with a mortgage (0.58), population density (0.655), the log of the median house prices (0.653), median time to work (0.619) and the percentage of people with access to fast broadband (0.437). When all variables are included jointly in a regression the share of families with 2 or more earners, population density, the log of median house prices and the median time to work are all statistically significant at 5%.

Figure A10: Association of ν and Ω with Economic Variables (Census)



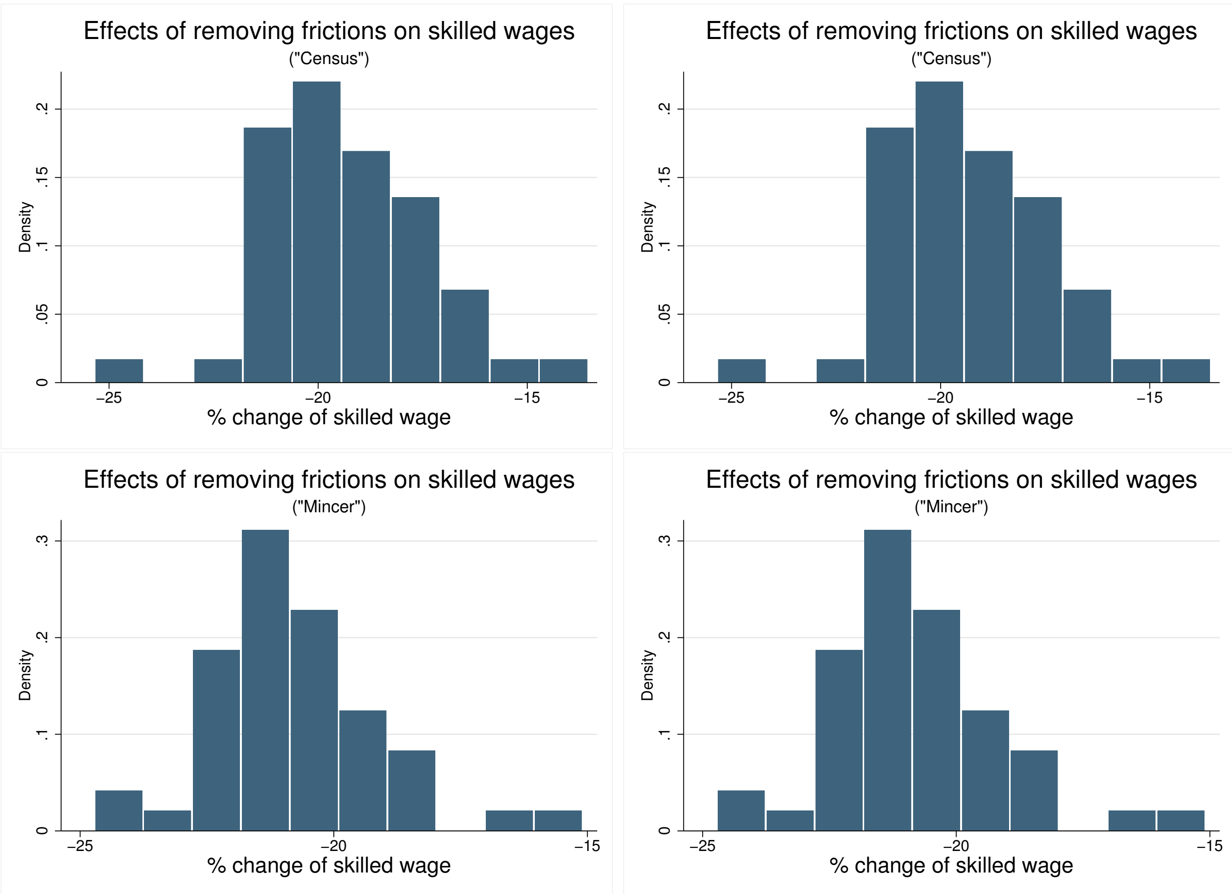
Note: The scatter plots show the correlations of ν with the standard deviation of workers with college (0.884), standard deviation of workers with non-college (0.498), unemployment rate of college graduates (0.506) and with the unemployment rate of workers without college (0.099), as well as the correlation of Θ with the employment rate of college graduates (0.110) and employment rate of non-college graduates (0.457).

Figure A11: Output costs of frictions ($\nu = 0$), Alternative Calibrations



Note: Model simulations. The graph plots the percentage variation of output relative to baseline when setting $\nu = 0$.

Figure A12: Effects of Removing Frictions of Skilled an Unskilled Wages, Alternative Calibrations



Note: Note: Model simulations. The graph plots the percentage variation of skilled and unskilled wages (w^h) and (w^l) relative to baseline when setting $\nu = 0$. The first panel is for "Switchers", and the second using the "Mincer" regressions.

Table A2: Regression of output cost of mismatch on observable variables

	Mincer		***Census***	
Variable	**coef. (t-stat)**	**[R-squared]**	**coef. (t-stat)**	**[R-squared]**
Panel A: Cross States				
n	8.33 (7.43)	[0.534]	8.17 (6.98)	[0.504]
o	9.30 (0.80)	[0.013]	7.40 (0.63)	[0.008]
u	23.71 (5.67)	[0.401]	24.38 (5.92)	[0.042]
$\frac{w_h}{w_l}$	1.34 (5.62)	[0.396]	1.64 (8.39)	[0.594]
$\frac{w_h}{\zeta w_l}$	**7.92** (24.04)	[0.923]	**6.969** (20.89)	[0.901]
$\frac{w_l}{\chi w_h}$	-6.31 (-3.21)	[0.176]	-7.59 (-6.40)	[0.461]
Obs.		(50)		(50)
Panel B:Simulated Data				
n	-1.61 (-13.79)	[0.127]	-2.50 (-62.26)	[0.099]
o	8.22 (6.22)	[0.029]	2.08 (7.34)	[0.001]
u	15.18 (17.70)	[0.194]	6.19 (32.97)	[0.030]
$\frac{w_h}{w_l}$	0.26 (4.42)	[0.014]	0.56 (44.79)	[0.054]
$\frac{w_h}{\zeta w_l}$	7.40 (64.95)	[0.764]	6.72 (349.55)	[0.777]
$\frac{w_l}{\chi w_h}$	1.341 (3.21)	[0.007]	-2.21 (-40.57)	[0.045]
Obs.		(1,302)		(35,119)

The table shows coefficient, t-statistic and R-squared of the univariate regression of the model-based output costs of mismatch on all observable variables in turn. n share of high-type workers in employment. u and o, share of under-employed and over-employed workers, respectively, in total employment. $\frac{w_h}{w_l}$ high-type wage premium of well-matched workers. $\frac{w_h}{\zeta w_l}$ well-matched premium of high-type workers relative to under-employed. $\frac{w_l}{\chi w_h}$ well-matched loss of low-type workers relative to over-employed.