Unpacking Skill Supply and Wages*

Joseph Richardson†

February 2025

Abstract:

This paper investigates how graduate wages respond to changes in the supply of graduates with similar majors, grouped together by hierarchical clustering. Using major fixed-effect and instrumental variable regressions on American Community Survey data, we estimate statistically insignificant effects. Even our most conservative specification rules out elasticities stronger than -0.4. This inelasticity is inconsistent with previous research that finds stronger effects using broader education categories (e.g., Katz and Murphy, 1992). We reconcile this by demonstrating that the previous results derive from spurious time series regressions. Finally, we explain the inelasticity by showing that different education groups have significantly overlapping skillsets.

JEL Codes: I23, J24, J30

Keywords: College wage premium, Skill supply

*I thank Marco Alfano, Byunghoon Kang, Gerald McQuade, and participants at the Lancaster PhD seminar, alongside my supervisors Maria Navarro Paniagua and Ian Walker, for helpful comments and feedback.

† Department of Economics, Lancaster University Management School, LA1 4YX. Email: j.m.richardson@lancaster.ac.uk

1 Introduction

How skill prices respond to a change in the supply of education is a classic question in labour economics, dating back to Tinbergen (1974). Empirical research has traditionally examined how the college wage premium responds to changes in the supply of graduate labour, estimating an elasticity of substitution between skilled and unskilled labour (prominent examples include Katz and Murphy, 1992; Card and Lemieux, 2001; Autor et al, 2020; and Blundell et al, 2022). These estimates have been used to subsequently estimate the extent of skill-biased technological change (Goldin and Katz, 2008) and to convert microeconomic returns to education into macroeconomic ones (examples include Gethin, 2023; Hendricks and Schoellman, 2023; and Bils et al, 2024).

A limitation of the traditional aggregate approach is that graduates' skillsets vary considerably between majors. Graduates from lower return majors have labour market outcomes closer to high school graduates than graduates of higher paying majors. This suggests that more granular measures of education would better capture actual skill differences. Figure 1 demonstrates this heterogeneity in earnings by major group. Moreover, previous research suggests that part of these differences are causal[1] and they are driven, at least in part, by specific technical skills rather than general ones (Kinsler and Pavan, 2015; Deming and Noray, 2020).

In this paper, we investigate how a graduate's wage responds to an increase in the supply of graduate labour from similar majors. One may expect that one computer science graduate is substitutable for another computer science graduate, but not necessarily for graduates in fine arts. Investigating this question is complicated by having to define groups of majors with similar skillsets. We address this issue by adopting hierarchical agglomerative clustering from statistical learning methods. Hierarchical agglomerative clustering allows us to group together majors with similar skillsets in a data-driven manner using information on graduates' occupations. For example, if Economics and Finance majors are disproportionately likely to be financial analysts, they will likely be grouped together (as indeed they are in our analysis below).

Using major fixed-effect regressions on American Community Survey data from 2009-2019, we show that graduate wages are inelastic with respect to changes in the

[1] Large differences are apparent in cross-sectional studies controlling for a rich set of pre-college ability measures (Britton et al, 2022; Andrews et al, 2024) and are similar in magnitude to comparable regression discontinuity estimates (Hastings et al, 2013; Kirkeboen et al, 2016; Bleemer and Mehta, 2022).

supply of graduates from similar majors. Even our most conservative specification rules out elasticities stronger than -0.4.

This inelasticity is not an artifact of labour supply and major choice responding to labour demand shocks. All specifications contain major time trends to control for secular demand trends, following Katz and Murphy (1992), and our data are inconsistent with a labour demand interpretation for two reasons. Firstly, we would expect labour demand shocks to be highly correlated within skill clusters i.e., the demand for Economics graduates increases with the demand for Finance. Therefore, confounding labour demand shocks would cause the supply of graduates from one major to be highly correlated with others within its cluster. However, this is not the case and our data is much more consistent with changes in supply representing idiosyncratic year-to-year fluctuations. Second, to err on the side of caution, we also show that our results are robust to alternative specifications that are less susceptible to specific sources of demand shocks. For instance, we construct an instrumental variable by calculating the supply of native graduates in each cluster if all natives worked the typical hours of their major-age-sex cell in a base period of 2009. This instrument removes any variation from potentially endogenous migration and labour supply decisions and only uses variation from new graduates and pre-determined changes from lifecycle labour supply factors. In addition, to deal with potential endogeneity in the supply of new graduates, we also estimate our specification in first differences. As graduates typically select their major at least one year before graduating, all labour demand shocks in the first-differences error term transpire after students made their study decisions.

Our findings are also not attributable to measurement error in the graduate supply variable. We address the measurement error problem by randomly splitting our sample into two halves, calculating our supply measures in both halves separately, and then using one as an instrumental variable for the other. As both measures are equal to the true population value plus an independent measurement error, this provides consistent estimates that form the basis of our most conservative specifications.

We also investigate whether there is heterogeneity by major. Specifically, we investigate whether the elasticities vary between high and low wage majors. We find that for majors below the mean graduate wage, we can rule out even very small elasticities, such as -0.1. This inelasticity is consistent with some majors having low returns relative to non-graduates (Andrews et al, 2024), meaning there should be less of a wage premium to erode. However, while still statistically insignificant, our confidence intervals are wider for majors above the mean graduate wage. These confidence intervals include effects as strong as -0.27 and -0.49 in the OLS and IV specifications respectively. Therefore, we cannot rule out all economically significant effects for this group, but we can exclude some effects consistent with previous research.

Our inelastic estimates appear inconsistent with earlier research, suggesting that US graduates and non-graduates are not very substitutable (e.g., Katz and Murphy, 1992; Card and Lemieux, 2001; Goldin and Katz, 2008; Autor et al, 2008; Acemoglu and Autor, 2011; Autor et al, 2020). These papers estimate elasticities of the college wage premium with respect to relative supply that centre around -0.6[2]. As the returns to college are heavily dependent on major-specific skills (Kirkeboen et al, 2014; Britton et al, 2022; Andrews et al, 2024), which suggests that most of the skills taught in a degree are not general to all degrees, one might expect us to find an effect of a similar magnitude. The inconsistency can be explained by the methodology of the earlier studies which, with one partial exception[3] (Card and Lemieux, 2001), regress a non-stationary time series of the college wage premium on a non-stationary time series of relative supply. Since Granger and Newbold (1974), we have known that non-stationary time series can lead to a spurious regression problem. Using an augmented Dickey-Fuller test on Autor et al (2020)'s replication data, we demonstrate that the time series in question contain unit roots and are thus non-stationary. Furthermore, an Engle-Granger test suggests that the time series are not cointegrated either. Estimating analogous specifications after first differencing out the unit roots instead produces null results. This non-stationarity also parsimoniously explains why some applications of the Katz and Murphy (1992) model to other time periods have produced statistically significant results with the wrong sign (Beaudry and Green, 2005; Bowlus et al, 2023).

We rationalise the inelasticity of graduate wages to the supply of graduates by presenting descriptive evidence that the two groups have substantial skill overlaps. Nearly half of graduates work in occupations where non-graduates are in a majority and many more work in occupations where non-graduates constitute at least a substantial minority. Likewise, most graduates are not employed in an occupation where either their

[2] This would be an elasticity of substitution between graduates and non-graduates between 1.5 and 2 (-1 divided by a figure around -0.6), assuming a constant elasticity of substitution production function. The college wage premium here is defined as graduate wage divided by non-graduate wage. Similarly, relative supply is hours worked by graduates divided by hours worked by non-graduates. This does create subtle differences in interpreting the elasticities. For example, the relative supply variable moves by more than 1% in response to a 1% increase in the graduate share of labour due to the denominator also decreasing. Similarly, the college wage premium also includes effects on non-graduate wages. These effects move in opposite direction.

[3] Although Card and Lemieux (2001) use the time series in some specifications, they also focus on estimating age-group specific elasticities that are identified using panel variation. These are estimated to be around -0.2, which is much more consistent with our findings.

major or similar ones predominate, suggesting significant substitutability between graduates of different majors.

This paper makes several contributions. First, it provides the first general estimates of how graduate wages respond to changes in the supply of graduates from similar majors – a natural extension of previous research given mounting evidence on the skill differences between majors (e.g., Britton et al, 2022; Andrews et al, 2024). The only similar study that we are aware of is Qvist et al (2021), which examines the effects of Aalborg University becoming the second university in Denmark to offer courses in electrical and construction engineering. The overwhelming majority of their estimates on the effects of this expansion on engineering wages are null, and they are universally so for engineers who are at least one-year post-graduation.

Beyond estimating the substitutability of different majors, we also contribute new evidence on the broader topic of estimating the elasticity of substitution between skilled and unskilled workers[4]. By revisiting the Katz and Murphy (1992) framework with appropriate time series techniques, we show that these groups are in fact highly substitutable. This finding aligns with recent studies arguing that low elasticities of substitution (large wage effects) are incompatible with recent data. For instance, Blundell et al (2022) estimates a precise null relationship between the graduate wage premium and graduate supply using panel data on UK regions. Furthermore, they argue that European data more generally appears inconsistent with low substitutability. Moreover, Bils et al (2024) argue that worldwide elasticities of substitution between skilled and unskilled labour below 4 are implausible. They show that combining elasticities below 4 with the dynamics of GDP and Mincerian returns to education between 1960 and 2010 would require TFP to decline for low skill workers over that period. At elasticities around 1.5, such as those often found in Katz and Murphy (1992) style regressions, the required TFP drop for low skill workers becomes extreme at over 90%.

Our results also have implications for theories explaining why graduates and non-graduates may be highly substitutable in some contexts. One prominent explanation is endogenous technological change where firms endogenously adopt existing skill-biased management practices and technologies when skilled workers are abundant (Blundell et al, 2022). Another is directed technical change (Acemoglu, 2002), where an abundance of skilled workers increases the returns to developing skill-biased technologies. However,

[4] Assuming a constant elasticity of substitution production function, the elasticity of substitution is equal to the negative inverse of the elasticity of the college wage premium with respect to the relative supply of graduates compared to non-graduates.

both have limitations in explaining our results. The limitation of Blundell et al (2022)'s endogenous technology adoption explanation is that it was developed to reconcile high substitutability in countries like the UK with low substitutability in the US, which is at the forefront of the technological frontier. The idea being that countries behind the frontier can choose which existing frontier technologies to opt to invest in, and that the returns to skill-biased technologies are increasing in skill supply. However, this explanation becomes less convincing[5] when it turns out that there is high substitutability in the US as well. Meanwhile, the directed technical change explanation suffers from the short time-period (2009 to 2019) studied in our major-specific analysis), which is likely too short for many new technologies to be both invented and adopted. Instead, our results are most consistent with the education groups having significantly overlapping skillsets.

Finally, we also make a modest methodological contribution by suggesting an approach to addressing classical measurement error in values estimated from a sample. As our graduate supply variables are estimates of a population variable, they will necessarily contain some classical measurement error. Fortunately, the inconsistency introduced by classical measurement error can be addressed if we have an instrumental variable uncorrelated with that error. For many variables calculated from a sample, such as graduate supply or employment rates, randomly splitting the sample into two will enable the calculation of two statistically independent estimates for the parameter of interested. This method has similarities with the established psychometric technique of validating a measure's reliability through sample splitting (e.g., de Vet et al, 2017; Pronk et al, 2022), but we explicitly incorporate this idea into our estimation procedure to adjust our confidence intervals for unreliability. To the best of our knowledge, we are the first to construct an instrument to deal with measurement error in this way.

2 Data

We use data from eleven waves of the American Community Survey between 2009 and 2019. The American Community Survey provides a representative cross-sectional 1% sample of US population containing information on income, employment, education, and other demographic characteristics. We start in 2009, the year that bachelor's degree major[6] is first recorded, and end in 2019 to avoid any issues arising from labour market

[5] As not all individual US firms will be at the technological frontier, we cannot rule out this explanation having some impact.

[6] Notably, this means that our sample does not contain law or medicine as a subject, because those are only available at the postgraduate level in the United States. Unfortunately, information on the content of any graduate study is unavailable.

changes induced by COVID-19. Additionally, the 2020 sample faced significant quality control issues (US Census Bureau, 2021). In all, these data provide 28,102,181 individual observations with 7,769,718 having a bachelor's degree.

We use these data to calculate annual estimates of the proportion of total US hours worked by graduates of each major. We also calculate average wage series, weighted by hours worked, for each major by five-year age-group[7]. The underlying hourly wages used in the averages are calculated by dividing a worker's reported labour income by their total hours of work. To mitigate the influence of misreported income and working hours sometimes causing implausible wage values, we trim the wage values at the 1st and 99th percentiles for their sex.

2.1 Grouping Majors

As our data contains 176 different major categories and many of these are closely related, such as History and US History, we require a method for grouping majors together into sensible categories. Pre-existing categorisations are not necessarily suitable for our purposes as they may poorly reflect the skills graduates use on the labour market. For instance, the higher-level categorisation provided by the US Census Bureau groups together Economics, one of the highest earning majors, with Sociology, one of the lowest earning majors (see Figure 1) under the category social science. We address this categorisation issue in a data-driven way by using a clustering algorithm to group together similar majors.

To use clustering, we require data on what skills graduates of different majors possess. The obvious variable highly correlated with the skills a worker uses is their occupation. For instance, being employed as an engineer will signal that a worker's quantitative and engineering skills are likely to be relatively high. Therefore, we calculate the proportion of graduates working within each four-digit SOC occupation for each major and standardise these proportions into z-scores. As clustering approaches measure similarity using distance metrics and Euclidean distance performs poorly in high dimensions (Beyer et al, 1999), we reduce the dimensionality of the occupation data from 459 to 10 variables using principal component analysis. Another advantage of principal component analysis, in this setting, comes from the first few principal components being little affected by measurement error (Hellton and Thoresen, 2014), as the underlying proportions will be measured with some error for rarer majors and occupations.

Using this information on the occupations of graduates from each major, we group together majors using hierarchical agglomerative clustering, a method first introduced in

[7] The age-groups used are under 25, and 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-60, 60-64, and 65+.

Sokal and Michener (1958). We choose hierarchical agglomerative clustering over other methods because it is easily interpretable and is also more replicable than the popular k-means approach, as it does not require specifying random or arbitrarily chosen start points. In hierarchical agglomerative clustering, each major initially constitutes its own cluster, and the two most similar clusters are merged sequentially based on a linkage rule. This process continues until a stopping point is chosen. Figure 2 provides a stylised illustration with six majors instead of 176. First, subjects 5 and 6 are merged as they are the closest together, leaving us with five clusters, with the height of the upwards line in the dendrogram representing the distance between the two merged clusters. Then, subjects 1 and 2 become the two most similar clusters and are also merged. Afterwards, the cluster containing subjects 1 and 2 is merged with subject 3. This process continues until we tell the algorithm to stop.

Implementing hierarchical agglomerative clustering requires the researcher to choose both the linkage method and the stopping point. For the linkage method, we choose average linkage, as it is robust to potential outliers and clusters of unequal sizes (Ferreira and Hitchcock, 2009). Average linkage works by taking the location of each cluster in n-dimensional space (in this case 10) as the average position of its constituent members and then merging the two clusters closest in Euclidean space. We also consider results using Ward's linkage, which minimises the variance within the new set of clusters and produces more evenly sized clusters, in Appendix B. Meanwhile, we use the Calinski-Harabasz (Calinski and Harabasz, 1974) index to guide our choice of stopping point. The Calinski-Harabasz index relies on the intuition that an ideal clustering solution will produce tight and distinct clusters; this means that the members of a cluster are similar to one other but not to members of the other clusters. As higher values of the index indicate a better clustering solution, choosing around 30 clusters appears to be "optimal" (Appendix Table A1). However, no perfect ground truth is available and thus we also test the sensitivity of our results to choosing a different number of clusters. Notably, the effective number of clusters is always significantly lower than the headline number. For example, with 30 clusters, 80% of graduates are placed within the four largest clusters.

The clusters produced by this approach appear intuitively sensible (Appendix Table A2). The largest cluster primarily contains humanities subjects and certain social sciences. Reassuringly, we can also see that History and US History are both placed within the same cluster. Other groupings include Economics being paired largely with other business school subjects, mathematics majors being paired with computing majors, and education-focused majors being grouped together.

3 Empirical Strategy

Our primary specification is a two-way fixed effects model to estimate the elasticity of a graduate's wages with respect to the supply of graduates with similar degrees. Formally, our main regression is as follows:

$$\ln(\overline{wage_{mact}}) = \alpha_{ma} + \beta \ln\left(\frac{ClusterHours_{ct}}{TotalHours_t}\right) + \boldsymbol{\mu X}_{mact} + \tau_t + year_t * \gamma_m + \varepsilon_{mact} \quad (1)$$

where m represents a major, a represents a five-year age group, c represents a skill cluster, and t is the year. $\overline{wage_{mact}}$ represents the average hourly wage of graduates in group $mact$. Meanwhile, $\frac{ClusterHours_{ct}}{TotalHours_t}$ is the share of total hours worked by graduates in skill cluster c. Additionally, $\boldsymbol{X}_{mact}$ contains basic demographic covariates to control for the proportion of the group who are female, Black, Asian, or Hispanic.

We also include the necessary fixed effects and time trends required for causal identification. α_{ma} represents a set of major-by-age group fixed effects that control for time-invariant differences in unobserved productivity between major-age group categories. We split by age-group to control for the effects of labour market experience on wages. This is important because some of the variation in $\frac{ClusterHours_{ct}}{TotalHours_t}$ will come from new graduates and retirements, which are mechanically related with the group's average labour market experience. Another advantage of this disaggregation is that it may slightly increase precision and statistical power compared to total aggregation (Egerod and Hollenbach, 2024). Meanwhile, τ_t are a set of time fixed effects that capture any common shocks affecting all graduates. Finally, $year_t * \gamma_m$ are a set of major-specific linear time trends capturing any pre-existing trends in wages for each major that arise from the gradual adoption of new technologies. These time trends have similarities to the linear time trend controlling for secular labour demand trends in Katz and Murphy (1992).

Each observation is weighted by the number of individuals in that major-age category during the base year of 2009. This weighting makes the results representative of the US graduate labour market and reduces the influence of any measurement error in $\frac{ClusterHours_{ct}}{TotalHours_t}$, as the measured values will be more accurate when they are based upon more observations.

We also cluster the standard errors at the skill cluster level because the treatment is assigned at that level. Furthermore, as the typical regression only contains 30 clusters of unequal size, all tables report 95% wild cluster bootstrap (Cameron et al, 2008) confidence intervals using Webb (2023) weights to achieve the correct coverage levels.

These regressions are also run using only graduate observations, without any category for non-graduates, for three reasons. First, the non-graduate category likely contains more heterogeneity in education content than any specific major, making them less suitable for testing whether the type of education supplied matters for skill prices. Second, economic theory suggests that the average wages of graduates and non-graduates may change by different amounts when the supply of graduates changes, violating common shocks. In a Becker (1962) style human capital model, as the graduate share of the population increases due to higher returns or lower costs, the highest ability workers who were previously non-graduates become graduates. This shift reduces the average unobserved ability of both graduates and non-graduates, but not necessarily by the same amount. Consequently, assuming common time shocks for graduates and non-graduates would be inappropriate. Excluding non-graduates means that our time fixed effects will capture any bias from this channel[8] and mean that our regressions measure the effect of reallocating graduates from one major to another. Third, due to most Americans being non-graduates, including them would also harm precision in our weighted specification as most of the variation would come from a single time series.

3.1 Potential Endogeneity

3.1.1 Labour Demand Shocks

The most obvious source of potential endogeneity in equation (1) is that major time trends could insufficiently control for any major-specific labour demand shocks that are correlated with the supply of similar graduate labour. Fortunately, careful consideration of how these shocks would manifest themselves suggests a potential falsification test, and suggest alternative specifications exist that can handle these shocks, albeit sometimes at the cost of precision. The falsification tests stems from the idea that when graduates of one major face a positive labour demand shock, other graduates with similar skillsets should also face a positive labour demand shock. Therefore, if there are large shifts in major take-up, or in the labour supplied by existing graduates, in response to labour demand shocks, we would observe a strong correlation between the labour supplied by one major and the supply from other majors within the same skill cluster. Yet, Table 1 shows that we can rule out the supply of a major growing by more than 0.02% when the

[8] One could still imagine a Roy (1951) type self-selection model where students have different positively correlated unobserved abilities for each subject, with higher returns to education in subjects where their ability is higher, causing some bias of a similar nature. Therefore, the marginal student moving between subjects will likely have a lower subject-specific ability draw in both their new and old subject than the typical graduate. Alternatively, if their outside option was not university, they would simply have lower ability. However, this mechanism would bias our estimates downwards, so cannot change the story told by our results.

supply of graduates from similar majors increases by 1%, in the OLS model. For the IV specification, the corresponding upper bound is 0.2%. These results suggest that our identifying variation primarily represent idiosyncratic year-to-year shifts around the linear time trend that are uncorrelated with any labour demand shocks.

An alternative approach to dealing with labour demand shocks is to consider how labour supplied could react to demand shocks and adjust our econometric approach to exclude any potentially endogenous variation. Changes in the labour supplied by graduates of a major can only come from the production of new graduates, changes in the hours worked of existing workers, immigration, or graduates exiting the labour market. To address students potentially considering the demand for different majors when making their study decisions, we estimate a first-differenced specification. As graduates typically choose their major at least one year prior to graduating, this ensures that labour demand shocks in the error term occur after study decisions are made.

Meanwhile, to handle any potentially endogenous variation from hours worked, immigration, or labour force exits, we construct an instrumental variable. This instrument calculates the log share of total hours worked by US-born workers from each skill cluster under a counterfactual where everyone works the average hours of their major-age-sex combination in a base period of 2009. As a result, all the variation comes from changes in the size of each major-age-sex cell. Therefore, the changes in graduate supply induced by the instrument come from the production of new graduates and pre-determined shifts in labour supply due to lifecycle factors. As the instrument is calculated using only US-born individuals and because the hours are fixed at 2009 levels, it is unaffected by endogenous migration and shifts in hours of work. We define this instrument formally in equation (2):

$$ln\left(\frac{PredictedClusterHours_{ct}}{PredictedTotalHours_t}\right) = ln\ \left(\frac{\sum_{x \in MAS} \mathbb{1}_{x \in C_c}(x)(\overline{Hours_{x2009}} * NativePopShare_{xt})}{\sum_{x \in MAS}(\overline{Hours_{x2009}} * NativePopShare_{xt})}\right) \quad (2)$$

where i represents major, c skill cluster, t the year, and MAS is the set of all possible major-age-sex combinations in the data. $\mathbb{1}_{x \in C}$ is an indicator function equal to one if x is a member of the set C_c, the set of all major-age-sex cells in skill cluster c, and zero otherwise. Meanwhile, $\overline{Hours_{x2009}}$ is the average hours worked by a member of the major-age-sex cell x in 2009 and $\overline{Hours_{2009}}$ is the average hours worked by US-born adults in 2009. Finally, $NativePopShare_{xt}$ is the share of the US-born population that lies within major-age-sex cell x at time t.

3.1.2 Measurement Error

The other potential source of inconsistency in our specification comes from classical measurement error in the $\frac{ClusterHours_{ct}}{TotalHours_t}$ variable. At first sight, it may appear that this should be addressed by our large 1% sample of the US providing precise estimates. However, this variable is highly persistent as new graduates remain in the labour market for a long time and exits due to retirement or death tend to be permanent[9]. Adding fixed effects and time trends will therefore subtract out a large portion of the true variation in $\frac{ClusterHours_{ct}}{TotalHours_t}$, but none of the serially uncorrelated measurement errors, potentially creating a measurement error problem.

We resolve this issue by running a robustness test where we randomly split the underlying data into two samples and use the calculation of $\frac{ClusterHours_{ct}}{TotalHours_t}$ in one sample as an instrument for its measurement in the other sample. As both measurements are statistically independent measures of the true value, this approach yields consistent estimates. However, the attached confidence intervals may be relatively conservative, as psychometricians have long known that the correlation between two halves of a measure underestimates the reliability of the full measure (Spearman, 1910; Brown, 1910). Therefore, there could be room for efficiency improvements.

Owing to the specific nature of the measurement error problem in this context, we also add a second instrument which is the interaction of our first instrument with the natural logarithm of the skill cluster's size in 2009. This is because the measurement error, as a percentage of $\frac{ClusterHours_{ct}}{TotalHours_t}$, will be smaller for larger clusters and we take the logarithm to handle skewness in the distribution of cluster sizes.

4 Results

Table 2 presents the results from estimating equation (1) using the OLS and IV approaches, alongside results from an analogous first-differences specification. All estimates are statistically insignificant but the confidence intervals are small enough to exclude many values of potential economic significance. The OLS 95% confidence intervals exclude elasticities of graduate wages with respect to the supply of similar graduates stronger than -0.1, and elasticities stronger than -0.2 are excluded in the IV specification. In contrast, previous studies such as Katz and Murphy (1992) or Autor et

[9] It is possible that this would allow the relative supply variable to be trend stationary. However, as the sample period is just over 50 years and the typical career is almost as long, it will behave like a non-stationary series in the finite sample available. Indeed, spurious regression is still a problem in such a time series with sufficiently persistent stationary processes (Granger et al, 2001).

al (2020), estimate an elasticity around -0.6 for the relative wages of US graduates to non-graduates with respect to relative supplies. Additionally, the fixed effect and first-difference specifications produce very similar results, suggesting that major choices responding to labour demand shocks is not a significant source of bias.

These null results are robust to other reasonable groupings of majors. Figure 3 demonstrates that our estimates remain statistically insignificant regardless of how many clusters we group the majors into. However, the precision of the estimates does vary, with precision increasing with the number of clusters. Therefore, we can rule out very small elasticities when major is measured at a very fine level, but only larger ones at the broadest levels. Notably, the OLS and IV estimates are very similar even in the aggregations that produce tight confidence intervals, which also suggests that the OLS results contain little bias from labour demand shocks. Readers can decide what level of aggregation they think is most interesting, but there is a case for thinking that both finer and coarser levels are interesting, although dropping below 30 clusters may be undesirable[10]. We do not estimate elasticities levels much broader than the 20 clusters specification in Figure 3 – for example, close to 90% of the observations would lie within a single cluster when using only 10 clusters. We also present figures similar to figure 3 in Appendix B when taking alternative approaches to clustering that cover changing the occupation definition, the number of principal components, and using Ward's linkage to create more evenly sized clusters. The precise null results remain regardless of our choices.

Although we detect a null pattern overall, it could mask some heterogeneous effects between skill clusters. Specifically, one might expect a stronger elasticity between wages and supply for more skilled majors. Indeed, detecting an effect from some of the lower wage majors would be inconsistent with evidence that whatever skills they teach have little labour market value (Britton et al, 2022; Andrews et al, 2024). Therefore, we split the sample into majors which earn about the mean graduate wage and those that earn below it. This split reveals that our results are primarily driven by a very precise null result in the lower wage group, where elasticities stronger than -0.1 lie outside both the above and below mean confidence intervals. For higher wage clusters, the results remain null, but the IV results could still be consistent with elasticities of up to -0.5 in the expected direction, with the OLS results being more precise.

[10] The big change between 20 and 30 is that the humanities and social sciences cluster is merged with the education cluster. We do not think that merging a purely vocational cluster with one mostly containing general degrees is desirable for the purposes of our analysis and that this reflects the algorithm running out of sensible merging options with a low number of clusters.

4.1.1 Errors-in-variables Model

To address any classical measurement error in the graduate supply variable, we randomly split our sample in two to calculate the supply variable for both samples and use one as an instrument for the other. By construction, these two variables will measure the population value with an independent measurement error, meaning an instrumental variables approach will produce consistent estimates. We also add an additional instrument interacting the initial instrument with the natural logarithm of a cluster's size in 2009, to reflect the fact that the measurements are more precise in larger clusters.

Table 4 presents the results from our errors-in-variables analysis. Columns with an internal instrument use the procedure described above, while ones with external instruments use the value of the IV from equation (2) calculated in the other half of the sample. Meanwhile, as the choice of which sample to use to select as an instrument or an endogenous variable is arbitrary, we present both options separately as model 1 and model 2. All regression models in this procedure have strong instruments and produce null results, as one would expect from the earlier results, but with wider confidence intervals. Nonetheless, these confidence intervals still consistently exclude elasticities stronger than -0.39. That -0.39 is also likely to be conservative given that the estimates when we swap the samples used for the instrument and endogenous variable also produce similar results. These estimates are also robust to using limited information maximum likelihood estimation (Appendix Table A3), which is less susceptible to weak instrument bias in overidentified models (Blomquist and Dahlberg, 1999).

5 Explaining our Results

The presence of large differences in the returns to a degree by major (Kirkeboen et al, 2016; Bleemer and Mehta, 2022; Britton et al, 2022; Andrews et al, 2024) and an elasticity of substitution between skilled and unskilled labour of around 1.5 to 2 in the US (Katz and Murphy, 1992; Autor et al, 2008; Autor et al, 2020) appear mutually inconsistent with our estimates. If the returns to a degree mostly reflect subject-specific skills and an expansion in graduates erodes the graduate wage premium, then that result should be driven by a decrease in the return to those subject-specific skills[11].

We argue that the discrepancy can be understood by revisiting the evidence for a link between the college wage premium (graduate wage divided by non-graduate wage) and the relative supply of graduates to non-graduates. Specifically, we argue that these

[11] Advanced skills being highly complementary to unskilled workers could also explain the apparent inconsistency, although it is unclear why a physicist would complement a high school dropout but not an English graduate.

results were obtained by regressing non-stationary unit root time series upon one another, resulting in the type of spurious regression warned about in Granger and Newbold (1974). The basic time series setup, introduced by Katz and Murphy (1992) and used in studies such as Autor et al (2008) and Autor et al (2020), is presented in equation (3):

$$\ln\left(\frac{w_{St}}{w_{Ut}}\right) = \alpha_0 - \frac{1}{\sigma_{SU}}\ln\left(\frac{L_{St}}{L_{Ut}}\right) + \alpha_1 t + \varepsilon_t \tag{3}$$

where $\frac{w_{St}}{w_{Ut}}$ is the ratio between the average wage of graduates and the average wage of non-graduates in the United States, $\frac{L_{St}}{L_{Ut}}$ is the ratio between total hours worked by graduates and total hours worked by non-graduates, and t is a time trend to control for secular labour demand trends. The aim of this setup is to estimate the parameter σ_{SU}, which is the elasticity of substitution between skilled and unskilled workers under a constant elasticity of substitution production function.

There are good theoretical reasons to suspect a unit root in both the $\ln\left(\frac{w_{St}}{w_{Ut}}\right)$ time series and the $\ln\left(\frac{L_{St}}{L_{Ut}}\right)$ series. For the former, permanent technology shocks affecting the relative demand for skilled and unskilled labour could result in a unit root process. While the latter series is heavily driven by the production of new graduates, retirements, and deaths; all of which are highly persistent processes.

Indeed, using the replication data from Autor et al (2020), we show that an augmented Dickey-Fuller cannot reject a null hypothesis of a unit root for either time series. This Dickey-Fuller test allows for the series to have a constant term and a time trend, like in the regression setup. We also cannot reject a unit root in the pre or post 1992 periods, which Autor et al (2020) argues are structurally distinct. The results of these tests are presented in Panel A of Table 5, with the p-values ranging between 0.412 and 0.993.

Nevertheless, if the time series are cointegrated, the resulting regressions will not be spurious. Unfortunately, applying an Engle-Granger test for cointegration using the MacKinnon (2010) critical values, reveals that we also can not reject a null hypothesis of no cointegration. These test results are presented in Panel B of Table 5 and show that the test statistic is well below the critical value even at the 10% level.

As the time series appear to contain unit roots and are not cointegrated, we need to first difference the time series to obtain stationary series. Panel A of Table 6 presents the results from estimating variants of equation (3) in levels, while panel B is in first differences. Additionally, column (2) of panel A estimates an analogous specification to Table A1 of Autor et al (2020), excluding four data points prior to 1963 when annual data was not available[12]. Regardless of the sample period or specification of the time trends, first-differencing produces statistically insignificant estimates that are meaningfully different from the original results. The new confidence intervals also typically contain values consistent with our results at the finer skill cluster level. Overall, this suggests that changes in the relative supply of college graduates have not had large effects on the college wage premium.

Significant substitution between skill groups is also what we should expect after examining the structure of the US labour market, as college graduates and non-graduates have highly overlapping skills. Most graduates work in occupations where large shares, or even a majority, of workers do not possess a degree. Figure 4 plots the density of the occupational graduate share for the occupations of US graduates. Although some graduates work in graduate-dominated professions, there are clearly many occupations where the skills acquired during a degree are either unnecessary or can be acquired via alternative means.

A similar story also holds when looking at our skill clusters. Figure 5 demonstrates that most graduates work in occupations where their skill cluster is a clear minority (Figure 5), although there are some occupations where a specific skill cluster predominates. This obviously limits the extent to which we should see effects. However, this graph would likely look slightly less stark if we had data on postgraduate majors. Therefore, we do not know if a graduate studied specific subjects directly linked to occupational licensing requirements, such as law and medicine, that are only available at the postgraduate level in the US, where we might expect supply effects.

6 Conclusion

In this paper, we provide a new perspective on how graduate wages respond to changes in the supply of graduates. One aspect of this approach is to consider the significant heterogeneity in skills between majors. We demonstrate that graduate wages appear inelastic to the supply of graduate labour from similar majors, with us being able to rule out elasticities stronger than -0.4 in even our most conservative specification. This result appears inconsistent with prior research (e.g., Katz and Murphy, 1992; Goldin and Katz, 2008; Autor et al, 2008; Autor et al, 2020) arguing that the undifferentiated returns to college are highly responsive to the supply of graduates. We reconcile this by demonstrating that prior results stem from regressing non-stationary time series upon one another and that any response attenuates substantially after correcting for this. Furthermore, we argue that substantial substitutability between graduate and non-

[12] First differencing with those four observations would be impossible.

graduate labour should be expected given the makeup of the labour market. There is significant occupational overlap between graduates and non-graduates and between graduates from different majors, implying sharp limits on how different the skills typically used by the different groups are.

These results have significant implications for education policy and our understanding of labour markets. Firstly, the inelasticity of wages to an increase in the supply of graduates of similar majors imply substantial benefits to reallocating students from low return majors to higher return majors or possibly alternative career paths[13]. Low return majors are unlikely to see meaningful wage improvements when their numbers significantly shrink, whereas the wage advantage of higher return majors is likely to be reasonably robust. For instance, Bleemer and Mehta (2022) estimate a 46% return to studying Economics over a second choice major and this advantage is unlikely to be significantly eroded by realistic reallocations of students. Meanwhile, our re-analysis of the US elasticity of substitution between skilled and unskilled workers has three major implications. First, it suggests that skill-biased technological change has played a larger role in wage inequality than previously thought, as there was little countervailing effect from increasing higher education. Second, it also limits the extent to which education can be used as a policy lever to reduce inequality. And third, by shifting our beliefs about the elasticity of substitution between high and low skilled labour upwards, it suggests that the macroeconomic returns to education are likely to be slightly greater in general (Gethin, 2023).

Nevertheless, there are still several stones left unturned in this paper. For example, our research suggests that existing education data can be a weak guide to the specific skills that an individual worker uses, even at the major level, but we do not investigate alternatives. One related area of research shows that occupational licensing requirements, which govern the market supply of a very narrow set of skills, appear to affect wages (Kleiner and Krueger, 2013; Pizzola and Tabarrok, 2017; Dodini, 2023). Therefore, with the appropriate data and methodology, one could investigate how responsive the wage of an occupation like doctors responds to the supply of licensed professionals.

Finally, there may also be room to improve on the estimates of the novel parameter estimated in this dataset. If a researcher can access similar data with sufficiently precise measures of graduate supply over longer time periods, they could obtain even more precise estimates of the novel parameter estimated here. These results would be particularly interesting for the higher wage majors.

---

[13] It is possible that some low return majors may lead to lower economic returns than alternative training or work after including fees, opportunity costs, and government subsidies.

References

Acemoglu, D., 2002. Directed Technical Change. The Review of Economic Studies 69, 781–809. https://doi.org/10.1111/1467-937X.00226

Acemoglu, D., Autor, D., 2011. Skills, Tasks and Technologies: Implications for Employment and Earnings*, in: Card, D., Ashenfelter, O. (Eds.), Handbook of Labor Economics. Elsevier, pp. 1043–1171. https://doi.org/10.1016/S0169-7218(11)02410-5

Andrews, R.J., Imberman, S.A., Lovenheim, M.F., Stange, K., 2024. The Returns to College Major Choice: Average and Distributional Effects, Career Trajectories, and Earnings Variability. The Review of Economics and Statistics 1–45. https://doi.org/10.1162/rest_a_01503

Autor, D., Goldin, C., Katz, L.F., 2020. Extending the Race between Education and Technology. AEA Papers and Proceedings 110, 347–351. https://doi.org/10.1257/pandp.20201061

Autor, D.H., Katz, L.F., Kearney, M.S., 2008. Trends in U.S. Wage Inequality: Revising the Revisionists. The Review of Economics and Statistics 90, 300–323. https://doi.org/10.1162/rest.90.2.300

Beaudry, P., Green, D.A., 2005. Changes in U.S. Wages, 1976–2000: Ongoing Skill Bias or Major Technological Change? Journal of Labor Economics 23, 609–648. https://doi.org/10.1086/430288

Becker, G.S., 1962. Investment in Human Capital: A Theoretical Analysis. The Journal of Political Economy LXX, 9–49.

Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U., 1999. When Is "Nearest Neighbor" Meaningful?, in: Beeri, C., Buneman, P. (Eds.), Database Theory — ICDT'99. Springer, Berlin, Heidelberg, pp. 217–235. https://doi.org/10.1007/3-540-49257-7_15

Bils, M., Kaymak, B., Wu, K.-J., 2024. Labor Substitutability among Schooling Groups. American Economic Journal: Macroeconomics 16, 1–34. https://doi.org/10.1257/mac.20220288

Bleemer, Z., Mehta, A., 2022. Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major. American Economic Journal: Applied Economics 14, 1–22. https://doi.org/10.1257/app.20200447

Blomquist, S., Dahlberg, M., 1999. Small sample properties of LIML and jackknife IV estimators: experiments with weak instruments. Journal of Applied Econometrics 14, 69–88. https://doi.org/10.1002/(SICI)1099-1255(199901/02)14:1<69::AID-

JAE521>3.0.CO;2-7

Blundell, R., Green, D.A., Jin, W., 2022. The U.K. as a Technological Follower: Higher Education Expansion and the College Wage Premium. The Review of Economic Studies 89, 142–180. https://doi.org/10.1093/restud/rdab034

Bowlus, A., Lochner, L., Robinson, C., Suleymanoglu, E., 2023. Wages, Skills, and Skill-Biased Technical Change: The Canonical Model Revisited. Journal of Human Resources 58, 1783–1819. https://doi.org/10.3368/jhr.0617-8889R1

Britton, J., van der Erve, L., Belfield, C., Vignoles, A., Dickson, M., Zhu, Y., Walker, I., Dearden, L., Sibieta, L., Buscha, F., 2022. How much does degree choice matter? Labour Economics 79, 102268. https://doi.org/10.1016/j.labeco.2022.102268

Brown, W., 1910. Some Experimental Results in the Correlation of Mental Abilities. British Journal of Psychology, 1904-1920 3, 296–322. https://doi.org/10.1111/j.2044-8295.1910.tb00207.x

Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics 3, 1–27. https://doi.org/10.1080/03610927408827101

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-Based Improvements for Inference with Clustered Errors. The Review of Economics and Statistics 90, 414–427. https://doi.org/10.1162/rest.90.3.414

Card, D., Lemieux, T., 2001. Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis*. The Quarterly Journal of Economics 116, 705–746. https://doi.org/10.1162/00335530151144140

de Vet, H.C.W., Mokkink, L.B., Mosmuller, D.G., Terwee, C.B., 2017. Spearman–Brown prophecy formula and Cronbach's alpha: different faces of reliability and opportunities for new applications. Journal of Clinical Epidemiology 85, 45–49. https://doi.org/10.1016/j.jclinepi.2017.01.013

Deming, D.J., Noray, K., 2020. Earnings Dynamics, Changing Job Skills, and STEM Careers*. The Quarterly Journal of Economics 135, 1965–2005. https://doi.org/10.1093/qje/qjaa021

Dickey, D.A., Fuller, W.A., 1979. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. Journal of the American Statistical Association 74, 427–431. https://doi.org/10.2307/2286348

Dodini, S., 2023. The spillover effects of labor regulations on the structure of earnings and employment: Evidence from occupational licensing. Journal of Public Economics 225, 104947. https://doi.org/10.1016/j.jpubeco.2023.104947

Egerod, B., Hollenbach, F.M., 2024. How many is enough? Sample Size in Staggered Difference-in-Differences Designs. https://doi.org/10.31219/osf.io/ac5ru

Ferreira, L., Hitchcock, D.B., 2009. A Comparison of Hierarchical Methods for Clustering Functional Data. Communications in Statistics - Simulation and Computation 38, 1925–1949. https://doi.org/10.1080/03610910903168603

Gethin, A., 2023. Distributional Growth Accounting: Education and the Reduction of Global Poverty, 1980-2022.

Goldin, C., Katz, L.F., 2008. The Race between Education and Technology. Harvard University Press. https://doi.org/10.2307/j.ctvjf9x5x

Granger IV, C.W.J., Hyung, N., Jeon, Y., 2001. Spurious regressions with stationary series. Applied Economics 33, 899–904. https://doi.org/10.1080/00036840121734

Granger, C.W.J., Newbold, P., 1974. Spurious regressions in econometrics. Journal of Econometrics 2, 111–120. https://doi.org/10.1016/0304-4076(74)90034-7

Hastings, J.S., Neilson, C.A., Zimmerman, S.D., 2013. Are Some Degrees Worth More than Others? Evidence from college admission cutoffs in Chile. NBER Working Paper Series. https://doi.org/10.3386/w19241

Hellton, K.H., Thoresen, M., 2014. The Impact of Measurement Error on Principal Component Analysis. Scandinavian Journal of Statistics 41, 1051–1063. https://doi.org/10.1111/sjos.12083

Hendricks, L., Schoellman, T., 2023. Skilled Labor Productivity and Cross-Country Income Differences. American Economic Journal: Macroeconomics 15, 240–268. https://doi.org/10.1257/mac.20200256

Katz, L.F., Murphy, K.M., 1992. Changes in Relative Wages, 1963-1987: Supply and Demand Factors. The Quarterly Journal of Economics 107, 35–78. https://doi.org/10.2307/2118323

Kinsler, J., Pavan, R., 2015. The Specificity of General Human Capital: Evidence from College Major Choice. Journal of Labor Economics 33, 933–972. https://doi.org/10.1086/681206

Kirkeboen, L.J., Leuven, E., Mogstad, M., 2016. Field of Study, Earnings, and Self-Selection*. The Quarterly Journal of Economics 131, 1057–1111. https://doi.org/10.1093/qje/qjw019

Kleiner, M.M., Krueger, A.B., 2013. Analyzing the Extent and Influence of Occupational Licensing on the Labor Market. Journal of Labor Economics 31, S173–S202. https://doi.org/10.1086/669060

MacKinnon, J.G., 2010. Critical values for cointegration tests (Working Paper No. 1227). Queen's Economics Department Working Paper.

Pizzola, B., Tabarrok, A., 2017. Occupational licensing causes a wage premium: Evidence from a natural experiment in Colorado's funeral services industry. International

Review of Law and Economics 50, 50–59. https://doi.org/10.1016/j.irle.2017.04.005

Pronk, T., Molenaar, D., Wiers, R.W., Murre, J., 2022. Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment. Psychon Bull Rev 29, 44–54. https://doi.org/10.3758/s13423-021-01948-3

Qvist, H.-P.Y., Holm, A., Munk, M.D., 2021. Demand and Supply Effects and Returns to College Education: Evidence from a Natural Experiment with Engineers in Denmark. The Scandinavian Journal of Economics 123, 676–704. https://doi.org/10.1111/sjoe.12400

Roy, A.D., 1951. Some thoughs on the distribution of earnings. Oxford Economic Papers 3, 135–146. https://doi.org/10.1093/oxfordjournals.oep.a041827

Sokal, R.R., Michener, C.D., 1958. A Statistical Method for Evaluating Systematic Relationships. University of Kansas Science Bulletin 38, 1409–1438.

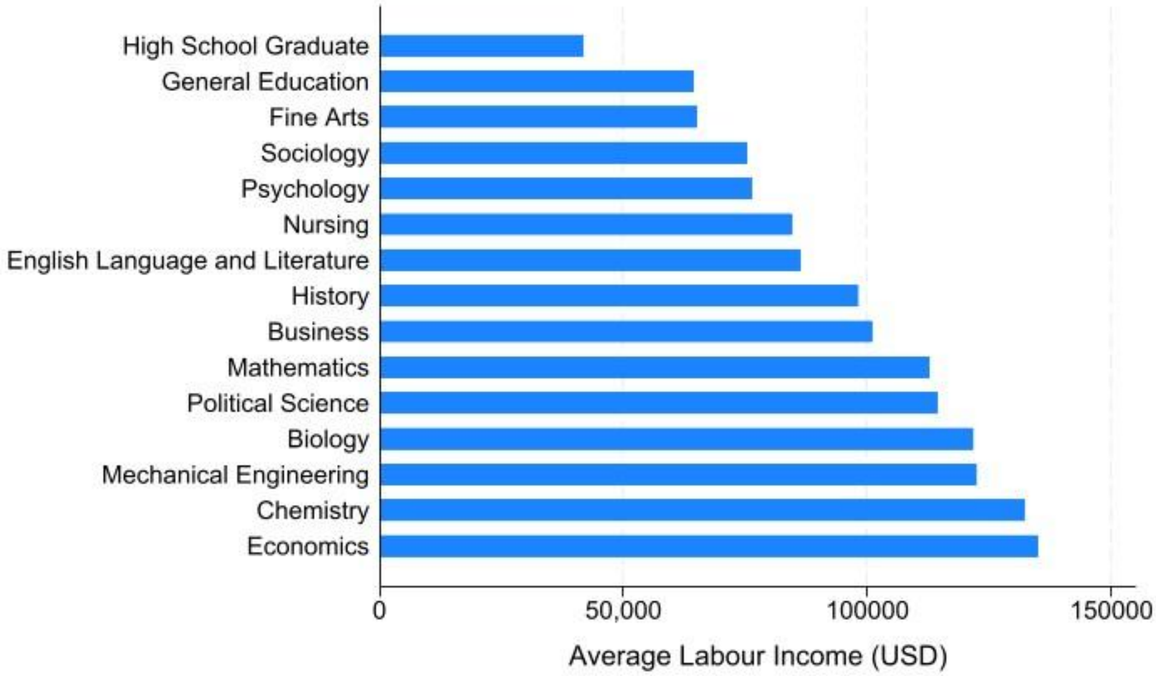
Spearman, C., 1910. Correlation Calculated from Faulty Data. British Journal of Psychology, 1904-1920 3, 271–295. https://doi.org/10.1111/j.2044-8295.1910.tb00206.x

Tinbergen, J., 1974. Substitution of Graduate by Other Labour. Kyklos 27, 217–226. https://doi.org/10.1111/j.1467-6435.1974.tb01903.x

US Census Bureau, 2021. Pandemic Impact on 2020 American Community Survey 1-Year Data [WWW Document]. Census.gov. URL https://www.census.gov/newsroom/blogs/random-samplings/2021/10/pandemic-impact-on-2020-acs-1-year-data.html (accessed 2.18.25).

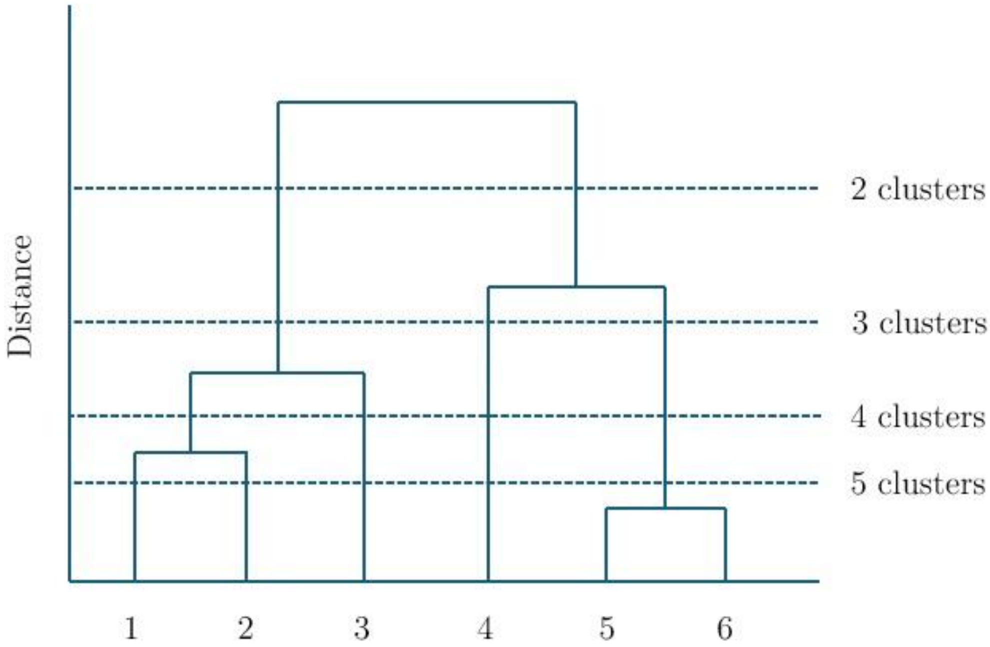
Webb, M.D., 2023. Reworking wild bootstrap-based inference for clustered errors. Canadian Journal of Economics/Revue canadienne d’économique 56, 839–858. https://doi.org/10.1111/caje.12661

Figure 1: **Full-time Labour Income by Major (2019)**



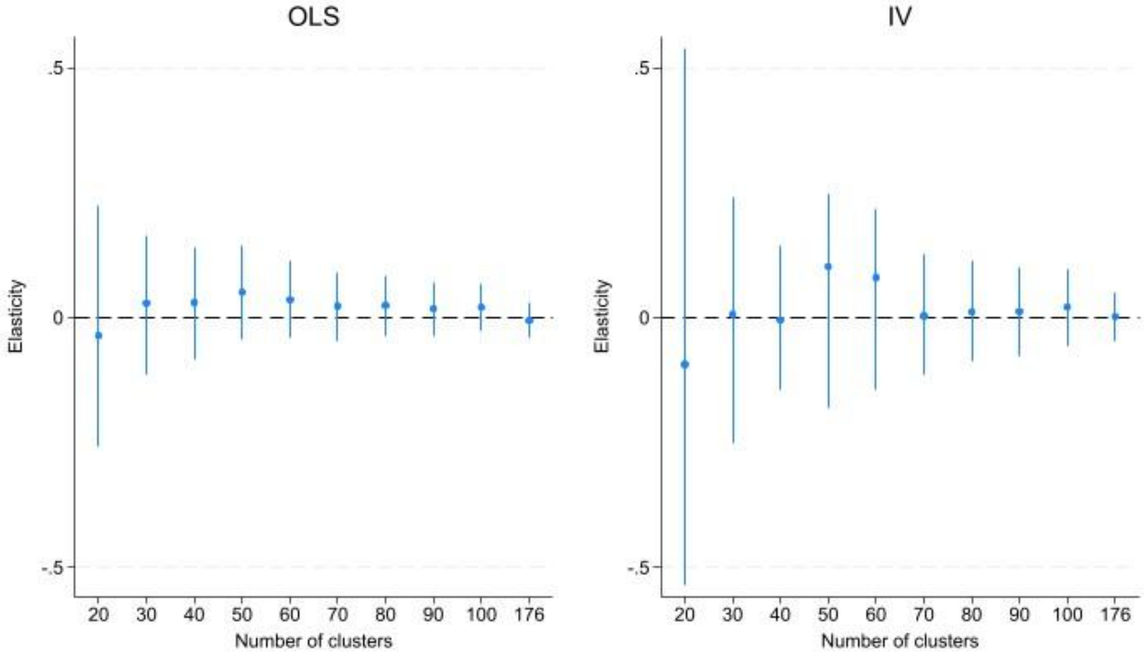
Notes: Data are from the 2019 American Community Survey. A full-time worker is classified as anyone working at least 1560 hours a year, which is 30 hours per week.

Figure 2: **Example Hierarchical Agglomerative Clustering Dendrogram**



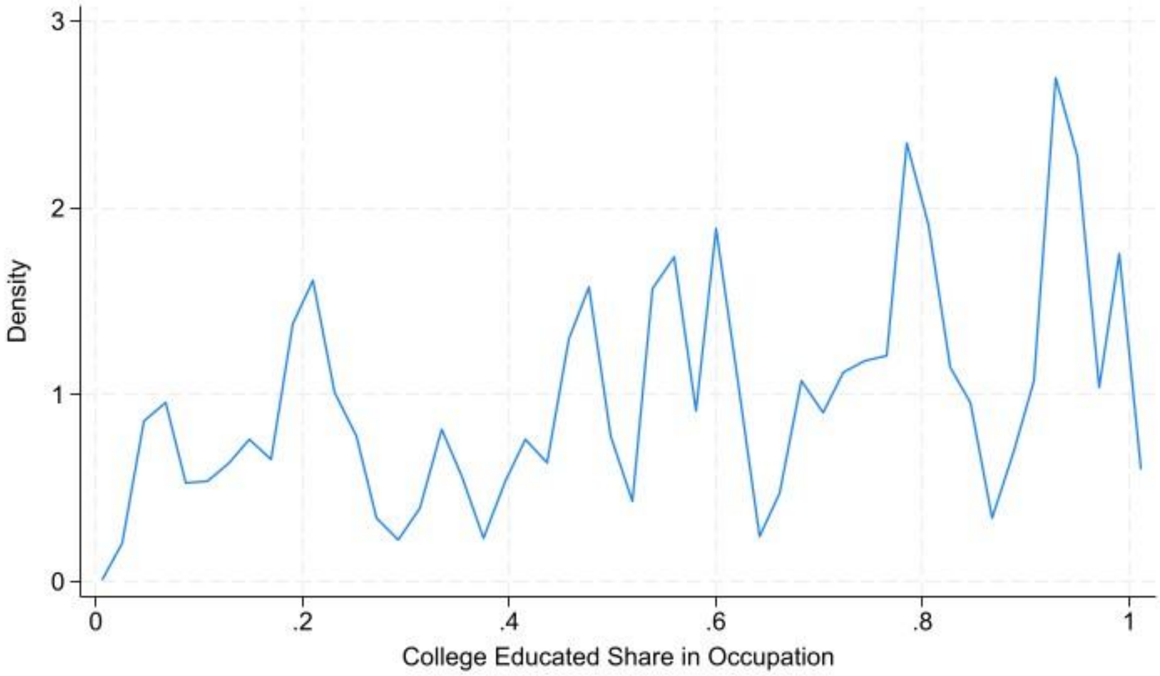
Notes: As one moves up the dendrogram, the most similar clusters are merged. Initially, this is subjects 5 and 6. This continues until a researcher decides to stop and cut the dendrogram.

Figure 3: **Elasticity of Graduate Wages to Supply of Similar Majors with Varying Numbers of Clusters**



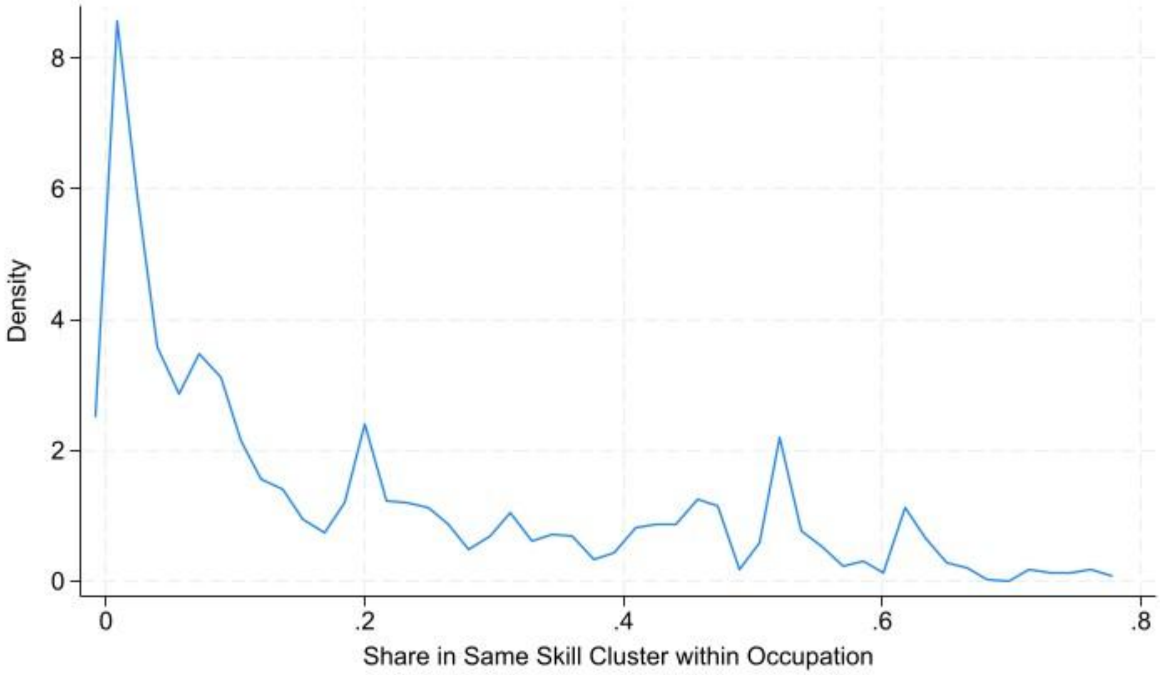
Notes: Point estimates come from estimating equation (1). The IV estimates use the instrument in equation (2) in a two-stage least squares setup. The 95% confidence intervals come from 5000 wild cluster bootstrap replications using Webb weights, clustered at the skill cluster level. The number of clusters is the number of clusters we tell the hierarchical clustering algorithm to sort the majors into.

Figure 4: **Graduate Share in Occupations Worked by Graduates**



Notes: Kernel density plot using an Epanechnikov kernel. Underlying data contains all college graduates in the American Community Survey from 2009-2019. College educated share is the proportion of workers in their four-digit occupation with at least a bachelor's degree.

Figure 5: **Same Skill Cluster Share in Occupations Worked by Graduates**



Notes: Kernel density plot using an Epanechnikov kernel. Underlying data contains all college graduates in the American Community Survey from 2009-2019. Share in the same skill cluster is the proportion of workers in their four-digit occupation who majored in a subject within their skill cluster.

Table 1: **OLS Regression of Supply of Hours Worked from a Major on the Supply of Hours Worked from Other Majors within the Same Skill Cluster**

Dependent variable:	$\ln\left(\frac{MajorHours_{ict}}{TotalHours_t}\right)$	
	(1)	(2)
	FE	IV-FE
$\ln\left(\frac{ClusterHours_{ct}}{TotalHours_t} - \frac{MajorHours_{ict}}{TotalHours_t}\right)$	-0.005	0.042
	(0.012)	(0.101)
	[-0.03, 0.02]	[-0.17, 0.22]
F-statistic		26.7
Major-by-age fixed effects	Yes	Yes
Year-by-age fixed effects	Yes	Yes
Controls	Yes	Yes
Time trend	Yes	Yes
R-squared	0.001	0.001
Observations	17,912	17,444

Notes: Major-age (five year) combinations, the unit of analysis, are weighted proportionally to the sum of their members. Standard errors clustered at the skill cluster level are in parentheses. 95% wild cluster bootstrap confidence intervals from 5000 replications using Webb weights are in square brackets. Controls cover the proportion of asian, hispanic, black, and female workers used to calculate the major-by-age group. Within R-squared values are reported. The instrument is defined as proportion of hours worked by natives from within the same skill cluster, excluding those from the same major, if everyone worked the hours typical of their age-major-sex combination in the base period of 2009.

Table 2: **Elasticity of Graduate Wages to Supply of Graduates with Similar Majors**

Dependent variable:	$\ln(\overline{wage_{iact}})$			
	(1)	(2)	(3)	(4)
	FE	IV-FE	FD	IV-FD
$\ln\left(\frac{ClusterHours_{ct}}{TotalHours_t}\right)$	0.029	0.006	0.049	-0.003
	(0.061)	(0.108)	(0.058)	(0.092)
	[-0.11, 0.16]	[-0.25, 0.24]	[-0.08, 0.18]	[-0.21, 0.19]
F-statistic		161.1		125.0
Major-by-age fixed effects	Yes	Yes	Yes	Yes
Year-by-age fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
R-squared	0.008	0.008	0.011	0.011
Observations	18,419	18,419	16,731	16,731

Notes: All regressions estimate equation (1) or a first-differenced analogue. Major-age (five year) combinations, the unit of analysis, are weighted proportionally to the sum of their members. Standard errors clustered at the skill cluster level are in parentheses. 95% wild cluster bootstrap confidence intervals from 5000 replications using Webb weights are in square brackets. Controls cover the proportion of asian, hispanic, black, and female workers used to calculate the major-by-age group. Within R-squared values are reported. The instrument is defined as proportion of hours worked by natives from within the same skill cluster if everyone worked the hours typical of their age-major-sex combination in the base period of 2009.

Table 3: **Elasticity of Graduate Wages to Supply of Graduates with Similar Majors by Average Wage of Major**

Dependent variable:	$\ln(\overline{wage_{iact}})$			
Major wage in 2009:	Below Mean		Above Mean	
	(1)	(2)	(3)	(4)
	FE	IV-FE	FE	IV-FE
$\ln\left(\frac{ClusterHours_{ct}}{TotalHours_t}\right)$	0.092	0.066	-0.051	-0.050
	(0.047)	(0.055)	(0.111)	(0.236)
	[-0.04, 0.20]	[-0.10, 0.18]	[-0.27, 0.23]	[-0.48, 0.59]
F-statistic		96.3		198.8
Major-by-age fixed effects	Yes	Yes	Yes	Yes
Year-by-age fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
R-squared	0.010	0.010	0.009	0.009
Observations	9,227	9,227	9,192	9,192

Notes: All regressions estimate equation (1). Major-age (five year) combinations, the unit of analysis, are weighted proportionally to the sum of their members. Standard errors clustered at the skill cluster level are in parentheses. 95% wild cluster bootstrap confidence intervals from 5000 replications using Webb weights are in square brackets. Controls cover the proportion of asian, hispanic, black, and female workers used to calculate the major-by-age group. Below mean average wage means that the average wage of graduates from that major was below the mean graduate's wage in 2009. Within R-squared values are reported. The instrument is defined as proportion of hours worked by natives from within the same skill cluster if everyone worked the hours typical of their age-major-sex combination in the base period of 2009.

Table 4: **Errors-in-Variables Model for the Elasticity of Graduate Wages to Supply of Graduates with Similar Majors**

Dependent variable:	$\ln(\overline{wage}_{iact})$			
	Sample Split 1		Sample Split 2	
	(1)	(2)	(3)	(4)
$\ln\left(\frac{ClusterHours_{ct}}{TotalHours_t}\right)$	0.027	0.029	-0.002	-0.054
	(0.132)	(0.160)	(0.113)	(0.113)
	[-0.31, 0.31]	[-0.39, 0.37]	[-0.28, 0.22]	[-0.38, 0.16]
F-statistic	22.6	27.9	19.6	25.6
Instruments	Internal	External	Internal	External
Major-by-age fixed effects	Yes	Yes	Yes	Yes
Year-by-age fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
R-squared	0.008	0.008	0.008	0.007
Observations	18,419	18,419	18,419	18,419

Notes: All regressions estimate equation (1). Major-age (five year) combinations, the unit of analysis, are weighted proportionally to the sum of their members. Standard errors clustered at the skill cluster level are in parentheses. 95% wild cluster bootstrap confidence intervals from 5000 replications using Webb weights are in square brackets. Controls cover the proportion of asian, hispanic, black, and female workers used to calculate the major-by-age group. Within R-squared values are reported. The external instrument is defined as proportion of hours worked by natives from within the same skill cluster if everyone worked the hours typical of their age-major-sex combination in the base period of 2009 from the other half of the sample. The internal instrument is the independent variable calculated in the other half of the sample. Either instrument is also interacted with the log of the cluster's size in 2009 to account for the structure of the measurement error. The distinction between Model 1 and Model 2 and is that Model 1 takes the value from half of the sample to be endogenous and the other to be the instrument and Model 2 flips this.

Table 5: **Autor et al (2020)'s Time Series Do Not Pass Diagnostic Tests**

Panel A: Augmented Dickey-Fuller Test with Time Trend

Series:	Sample	ADF Statistic	5% Critical Value	P-value
Relative Wage	1963-2017	-1.253	-3.496	0.899
Relative Supply	1963-2017	-1.357	-3.496	0.873
Relative Wage	1963-1991	-0.346	-3.588	0.988
Relative Supply	1963-1991	-0.138	-3.588	0.993
Relative Wage	1992-2017	-1.746	-3.596	0.730
Relative Supply	1992-2017	-2.340	-3.596	0.412

Panel B: Engle-Granger Test for Cointegration

Sample	ADF Statistic	5% Critical Value	10% Critical Value
1963-2017	-2.012	-3.961	-3.630
1963-1991	-2.615	-4.136	-3.758
1992-2017	-2.216	-4.164	-3.779

Notes: These statistics are calculated from the replication data provided by Autor et al (2020). The Engle-Granger test for cointegration critical values are calculated from the MacKinnon (2010) formula for N=2 with a time trend.

Table 6: **Autor et al (2020) Approach Before and After First Differencing**

Dependent variable:	$\ln\left({}^{w_{Skilled_t}}\!/\!_{w_{Unskilled_t}}\right)$				
	(1)	(2)	(3)	(4)	(5)
			Panel A: Levels		
$\ln\left({}^{L_{Skilled_t}}\!/\!_{L_{Unskilled_t}}\right)$	-0.254***	-0.644***	-0.688***	-0.602***	-0.183
	(0.048)	(0.057)	(0.070)	(0.079)	(0.170)
			Panel B: First Differences		
$\ln\left({}^{L_{Skilled_t}}\!/\!_{L_{Unskilled_t}}\right)$	-0.096	-0.121	-0.120	-0.162	-0.006
	(0.113)	(0.118)	(0.121)	(0.163)	(0.094)
Observations	54	54	54	28	26
Sample	1963-2017	1963-2017	1963-2017	1963-1991	1992-2017
Time trend	Linear	Linear spline	Quadratic	Linear	Linear

Notes: Heteroskedasticity-robust standard errors in parentheses. Data on college wage premiums and relative supplies are taken from Autor et al (2020). The linear spline fits a linear time trend between 1963 and 1992 and a different linear time trend between 1992 and 2017, as in many of Autor et al (2020)'s specifications. The first differenced regressions difference the time series and omit the constant to be analogous to the levels specification.

Table A1: **Calinski-Harabasz Index by Number of Clusters**

No. clusters	Calinski-Harabasz Pseudo-F
5	13.77
10	13.00
15	23.58
20	32.82
25	30.26
30	32.87
35	31.79
40	30.97
45	29.44
50	29.30

Notes: Calculates the Calinski-Harabasz (1974) statistic for a given number of clusters when clustering majors using average linkage on the first ten principal components of occupation data. A higher pseudo-F suggests a “better” clustering solution in the sense that the clusters are tighter and more distinct from one another.

Table A2: **Assignment of Majors into Clusters**

Major	Cluster
General agriculture	1
Agriculture production and management	1
Plant science and agronomy	2
Soil science	2
Forestry	2
Agricultural economics	3
Industrial and manufacturing engineering	3
Engineering and industrial management	3
Architecture	4
General engineering	4
Biological engineering	4
Architectural engineering	4
Chemical engineering	4
Civil engineering	4
Electrical engineering	4
Engineering mechanics, physics, and science	4
Environmental engineering	4
Geological and geophysical engineering	4
Materials engineering and materials science	4
Mechanical engineering	4
Metallurgical engineering	4
Mining and mineral engineering	4
Petroleum engineering	4
Miscellaneous engineering	4
Physical sciences	4
Geology and earth science	4
Geosciences	4
Aerospace engineering	5
Biomedical engineering	5
Astronomy and astrophysics	5
Atmospheric sciences and meteorology	5
Chemistry	5
Physics	5
Materials science	5
Animal sciences	6
Food science	6
Environmental science	6
Natural resources management	6

Hospitality management	20
Court reporting	21
Criminal justice and fire protection	21
Criminology	21
Naval architecture and marine engineering	22
Nuclear engineering	22
Oceanography	23
Engineering technologies	24
Electrical engineering technology	24
Industrial production technologies	24
Mechanical engineering related technologies	24
Miscellaneous engineering technologies	24
Transportation sciences and technologies	25
Construction services	26
Military technologies	27
Electrical and mechanic repairs and technologies	28
Cosmetology services and culinary arts	29
Precision production and industrial arts	30

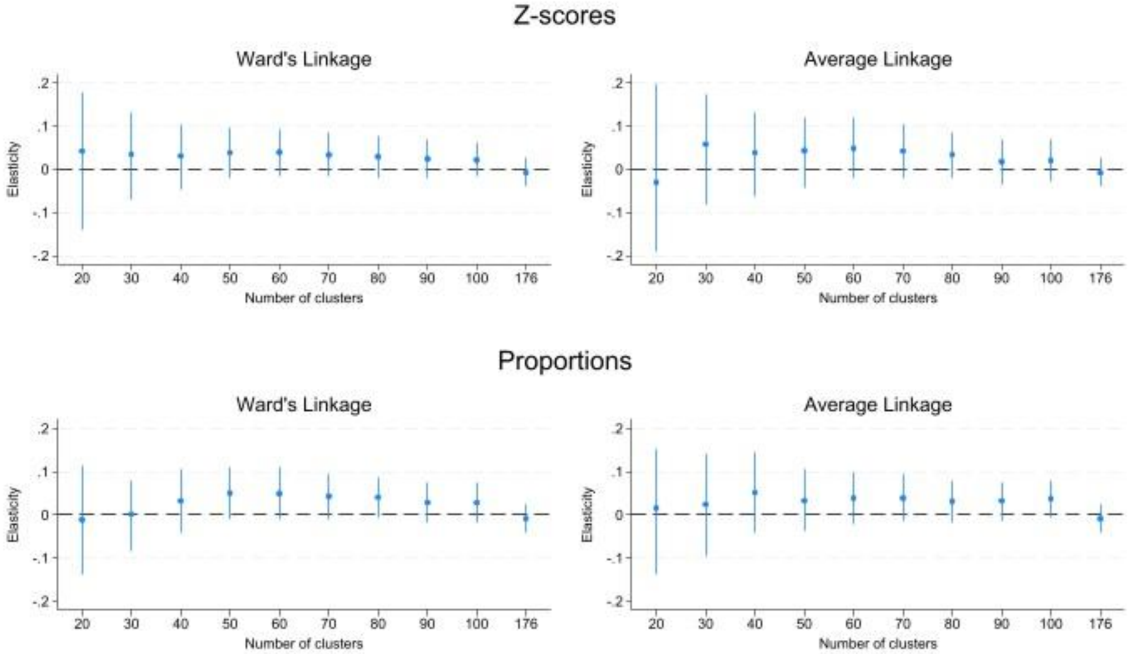
Table A3: **Errors-in-Variables Model for the Elasticity of Graduate Wages to Supply of Graduates with Similar Majors (LIML)**

Dependent variable:				
	Model 1		Model 2	
	(1)	(2)	(3)	(4)
$\ln\left(\frac{ClusterHours_{ct}}{TotalHours_t}\right)$	0.027	0.029	-0.002	-0.055
	(0.133)	(0.160)	(0.113)	(0.113)
	[-0.31, 0.31]	[-0.39, 0.37]	[-0.28, 0.22]	[-0.38, 0.16]
F-statistic	22.6	27.9	19.6	25.6
Instruments	Internal	External	Internal	External
Major-by-age fixed effects	Yes	Yes	Yes	Yes
Year-by-age fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
R-squared	0.008	0.008	0.008	0.007
Observations	18,419	18,419	18,419	18,419

Notes: All regressions estimate equation (1). Major-age (five year) combinations, the unit of analysis, are weighted proportionally to the sum of their members. Standard errors clustered at the skill cluster level are in parentheses. 95% wild cluster bootstrap confidence intervals from 5000 replications using Webb weights are in square brackets. Controls cover the proportion of asian, hispanic, black, and female workers used to calculate the major-by-age group. Within R-squared values are reported. The external instrument is defined as proportion of hours worked by natives from within the same skill cluster if everyone worked the hours typical of their age-major-sex combination in the base period of 2009 from the other half of the sample. The internal instrument is the independent variable calculated in the other half of the sample. Either instrument is also interacted with the log of the cluster's size in 2009 to account for the structure of the measurement error. The distinction between Model 1 and Model 2 and is that Model 1 takes the value from half of the sample to be endogenous and the other to be the instrument and Model 2 flips this.

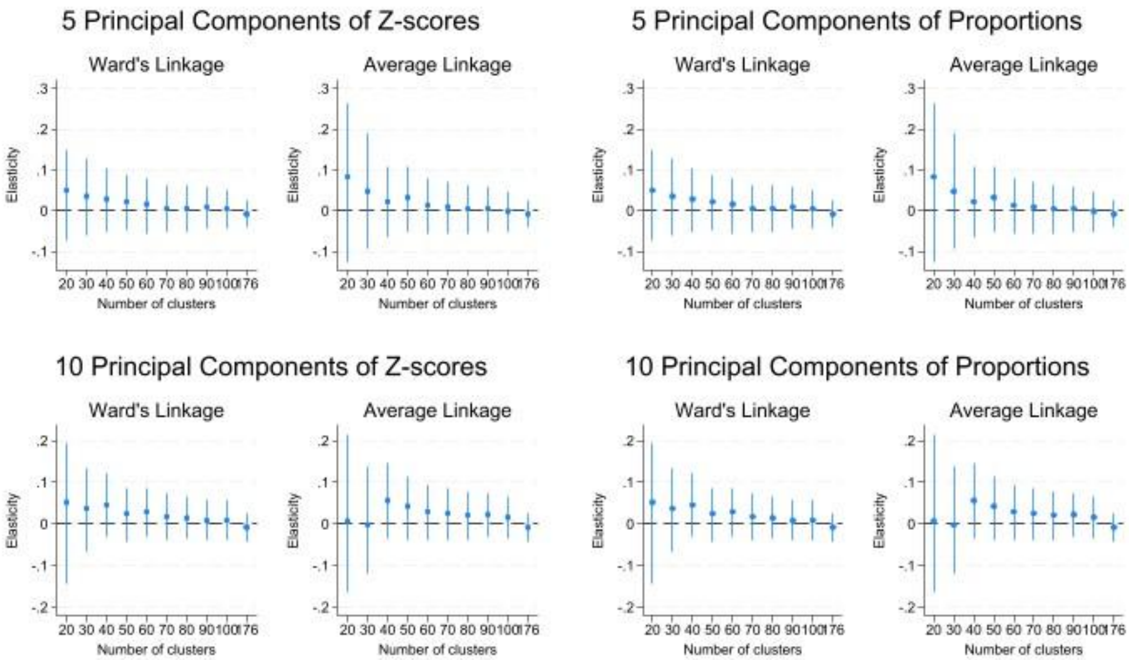
Appendix B

Figure B1: **Elasticity of Graduate Wages to Supply of Similar Majors When Clustering on 2-Digit Occupation**



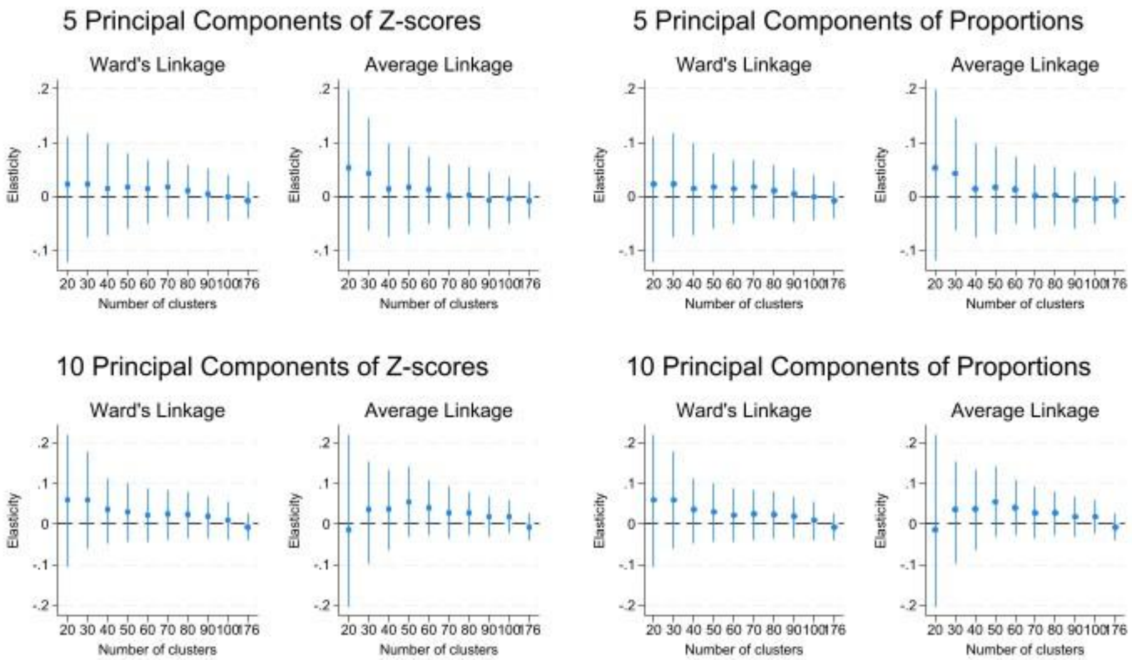
Notes: Point estimates come from estimating equation (1). The 95% confidence intervals come from 5000 wild cluster bootstrap replications using Webb weights, clustered at the skill cluster level. The number of clusters is the number of clusters we tell the hierarchical clustering algorithm to sort the majors into. Titles describe whether we z-score the data before clustering and whether we use average or ward's linkage.

Figure B2: **Elasticity of Graduate Wages to Supply of Similar Majors When Clustering on 3-Digit Occupation**



Notes: Point estimates come from estimating equation (1). The 95% confidence intervals come from 5000 wild cluster bootstrap replications using Webb weights, clustered at the skill cluster level. The number of clusters is the number of clusters we tell the hierarchical clustering algorithm to sort the majors into. Titles describe whether we z-score the data before taking principal components, the number of principal components we cluster on, and whether we use average or ward's linkage.

Figure B3: **Elasticity of Graduate Wages to Supply of Similar Majors When Clustering on 4-Digit Occupation**



Notes: Point estimates come from estimating equation (1). The 95% confidence intervals come from 5000 wild cluster bootstrap replications using Webb weights, clustered at the skill cluster level. The number of clusters is the number of clusters we tell the hierarchical clustering algorithm to sort the majors into. Titles describe whether we z-score the data before taking principal components, the number of principal components we cluster on, and whether we use average or ward's linkage.