THE UNIVERSITY OF CHICAGO

Alternative Pathways to Success: To what extent does Vocational Education contribute to regional economic growth and innovation?

By

Roy Xiao Hongze

July 2022

A paper submitted in partial fulfilment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Professor Dan A. Black

Preceptors: Can Mert Kokerer & Claire Watson[1]

[1] I would also like to extend my sincere gratitude to Professor Yanyan Sheng and Professor Christopher Clapp, both from the University of Chicago, as well as to all my coursemates for the patient guidance and helpful advice they all provided me over the course of writing this paper.

Abstract

The economic development and competitiveness of nations are increasingly thought to be underpinned by the growth and dynamism of regional economic hubs. Within the Regional Innovation Systems approach, education institutions such as universities feature prominently as generators of research, innovative technologies and human capital. However, the role of vocational education institutions in development is significantly under-researched. At the postsecondary level in the US, they mainly comprise community colleges and vocational/technical schools. In this study, I sought to uncover the possible contributions of VEIs to regional growth and innovation by conducting a quantitative panel data analysis. Collating publicly-available data from nearly 800 counties between 2010 to 2015, I assessed the suitability of various models for estimating the relationship between the population share of adults with associate's degrees as their highest education attainment, GDP growth and utility patent generation at the county-level. Results of my exploratory statistical analysis offer weak support to the hypothesis of a positive effect of increasing the population share of vocationally-trained workers on GDP at the county-level, but no significant effect was found on patenting activity. Overall, this study showcases the value of future quantitative research into the channels through which vocational education contribute to regional economic development.

Research Question

For my MA thesis, I propose to study the possible contributions of vocational education institutions (VEIs) to regional economic development in the United States. In particular, this exploratory statistical research will seek to examine the relationship between the human capital output from regional post-secondary vocational education and training (VET) programs, and economic growth as well as innovation within the US. Do US counties with a greater share of adults in the population that hold associate's degrees conferred by post-secondary VEIs record higher Gross Domestic Product (GDP) levels and greater patenting activity than others?

Motivation & Literature Review

This research topic lies within broader literature on the importance of education and human capital to regional economic development. Increasingly, the development and competitiveness of nations is thought to be underpinned by the growth and dynamism of regional economic hubs.[2] Attention of academics has in recent years been drawn to the Regional Innovation Systems (RIS) analysis approach, developed in the 1980s as a response the linear and static conceptualisations of the innovation process in earlier research. RIS studies instead focus on how innovation arises from interactions between industry, institutions and various actors in diverse networks, which support the growth and technological dynamism of regional economies.[3] Higher education institutions (HEIs) feature prominently as generators of research and innovative technologies, and as "system builders" that collaborate with industry and government to "generate mutually-beneficial innovation processes", as explained by Caniels and van den Bosch.[4] Universities and other such "knowledge infrastructures" are also identified by Charles as a key resource endowment and "active participants in the construction of regional competitive advantage".[5] Such institutions generate and contribute to the accumulation of commodifiable knowledge, human capital and social capital within a region, supporting industrial growth as well as facilitating interactive learning processes that bolster innovation.[6]

In particular, human capital as a resource and factor input has consistently been found to be key to the sustained vitality of regional economies. In studies on endogenous growth theory, researches frequently posit that increasing the quantity and quality of skilled labour through investment in higher education can boost productivity, innovation and thus economic growth.[7] Aghion et al. developed a comprehensive theory on the benefits of investing in higher

---

[2] Bruce Katz and Julie Wagner, "The Rise of Innovation Districts: A New Geography Of Innovation In America," *Brookings Metropolitan Policy Program* (May 2014): 1-2. Retrieved from https://www.brookings.edu/essay/ rise-of-innovation-districts/.
Michael E. Porter, "Clusters and the New Economics of Competition," *Harvard Business Review* (Dec 1998), https://hbr.org/1998/11/clusters-and-the-new-economics-of-competition.
[3] Henry Etzkowitz and Loet Leydesdorff, "The dynamics of innovation: from National Systems and 'Mode 2' to a Triple Helix of university-industry-government relations," *Elsevier Research Policy* 29 (2000): 111-7.
[4] Marjolein C.J. Caniels and Herman van den Bosch, "The role of Higher Education Institutions in building regional innovation systems," *Papers in Regional Science* 90, no. 2 (June 2011): 272-4.
[5] David Charles, "Universities as key knowledge infrastructures in regional innovation systems," *The European Journal of Social Science Research* 19, no. 1 (2006): 117-8.
[6] Ibid., 118-21.
[7] Robert J. Barro, "Human Capital and Growth," *American Economic Review* 91, no. 2 (May 2001): 12-7.

versus basic education for countries and regional economies based on their relative distance to the "technology frontier".[8] Based on the distinction that "high-brow" education fosters technological innovation while "low-brow" education fosters imitation, they empirically determined that technological and productivity growth generally increases across US states as their overall supply of educated workers grows. However, as the technological capabilities of individual states advance, growth comes to depend much more on innovation and thus on the supply of skilled human capital generated by HEIs.[9] Abel and Deitz conducted empirical research into the relation between HEIs and human capital across US regional economies. They found that while the direct relationship between the output of graduates from universities and the stock of skilled workers in the labour market is positive but weak, likely due to migration effects, the research and development activities of HEIs also generate spillovers that attracts human capital into the region.[10] Outside of the US, researchers have observed similar dynamics at play as well. Within the context of China over the 1990s to 2000s, empirical research conducted by Chi and Qian on the relation between the human capital output of HEIs and regional innovation activities identified a positive and significant correlation. Accounting for spatial dependence and spillover effects of innovation in their regression models, their study found that a one percentage point increase in the fraction of workers with tertiary education within a region was associated with a 9 percent increase in regional patent applications in 2006.[11] These studies are among many in the literature applying statistical methods to test the theories linking education to innovation and growth, and in general case studies from across

Jess Benhabib and Mark M. Spiegel, "The role of human capital in economic development: evidence from aggregate cross-country data," *Journal of Monetary Economics* 34 no. 2 (1994): 143–174.

[8] Philippe Aghion, L. Boustan, C. Hoxby and J. Vandenbussche, "The Causal Impact of Education on Economic Growth: Evidence from US," *Harvard University Department of Economics Publications* (Mar 2009): 1-3.

[9] Ibid., 7-15.

[10] Jaison R. Abel and Richard Deitz, "Do Colleges and Universities Increase Their Region's Human Capital?" *Federal Reserve Bank of New York Staff Reports* 401 (Oct 2009): 1-33.

[11] Wei Chi and Xiaoye Qian, "The role of education in regional innovation activities: spatial evidence from China," *Journal of the Asia Pacific Economy* 15, no. 4 (Nov 2010): 400-10.

the world seem to back up the importance of HEIs and education in general to regional development.

However, one group of postsecondary education institutions that remains under-researched are VEIs, which in the US comprise public community colleges, public vocational/technical schools and apprenticeship programs overseen by companies at the post-secondary level.[12] Fields of occupational study offered by postsecondary VEIs today range from consumer services, to engineering and manufacturing competencies, as well as public and social services.[13] Dortch summarises in her report for Congress in 2014 that the overall goal of VET (also termed Career and Technical Education or CTE) in the US is to prepare students for employment immediately after graduation. A successful VET education helps graduates reach "the achievement of industry-recognised credentials", thus close collaboration between industry and educators is often required in curriculum development.[14] The modern VET system was established following the passing of the 1917 Smith-Hughes Act supporting the teaching of agriculture, home economics and industrial trades in schools. This policy was the fruit of decades of advocacy by educators and industrialists for the integration of vocational subjects into the general curriculum and the establishment of dedicated VEIs. History however shows that public perceptions and support for VET were easily affected by events of each era, gaining prestige for example during the interwar period when VEIs supported the industrial and defence needs of the nation, but falling behind academic education in political importance when science and technology needs dominated during the Cold War.[15] Over the years, the association of VET with academically-less talented students and as a means of assimilating immigrants has also

---

[12] Cassandria Dortch, "Career and Technical Education (CTE): A Primer," *Congressional Research Service: CRS Report for Congress* R42748 (Feb 2014): 2-3.
[13] "Classification of Instructional Programs: 2000," Career and Technical Education (CTE) Statistics, National Centre for Education Statistics, last modified 2002, https://nces.ed.gov/surveys/ctes/tables/postsec_tax.asp.
[14] Dortch, "Career and Technical Education (CTE): A Primer," 2-3.
[15] William R. Ogden, "Vocational Education: A Historical Perspective," The High School Journal 73, no. 4 (Apr-May 1990): 250-1.

endured, despite ample research showing that the composition of the backgrounds and abilities of VET participants has been similar to other education tracks.[16] Not only does this help explain persistently poor employer attitudes towards vocational students, often viewed as second best to students trained in academic institutions,[17] but it may also explain the lack of research into the importance and contribution of VET to the US economy until recently.

While much more academic literature has been published on VET in education systems from the 1980s onwards, few researchers have directly addressed the question of how and how much VEIs contribute to economic development in the US. Evaluation of VET and its outcomes is itself still a relatively new area of economic and policy analysis, with methodologies constantly evolving and being drawn from other fields of research. Fretwell's 2003 paper proposing a comprehensive framework for assessing the internal inputs and external outputs of VEIs pinpoints the importance of evaluating their socio-economic impacts, asserting that like other forms of education, VET could generate higher worker productivity and enhance workers' ability to adapt and innovate.[18] Various methods for measuring the economic impact of VEIs were proposed, such as growth accounting studies and productivity studies, but within the US context there seems to have been little follow-up by other researchers on this topic. Instead, the bulk of research undertaken by education researchers and government agencies has focused on the labour market outcomes and private returns to graduates from VET programs. Early studies done by researchers at the National Centre for research in Vocational Education and the Bureau of Labour Statistics in the 1980s,[19] as well as more recent research conducted by the National Centre for Education Statistics (NCES), the Congressional Research Service and College Board Research have focused primarily on the characteristics and

[16] Nathan M. Semple, "Vocational Education: The Missing Link?" Peabody Journal of Education 63, no. 2 (Winter 1986): 75-8.
[17] Ibid., 84-5.
[18] David Fretwell, "A Framework for Evaluating Vocational Education and Training (VET)," *European Journal of Education* 38, no. 2 (June 2003): 177-190.
[19] Reviewed in Semple, "The Missing Link?", 82-3.

employment prospects of students with VET backgrounds, on trends in VEI enrolment and program composition.[20] The possible benefits of VET to the economy are frequently mentioned, notably its contribution in alleviating high unemployment and addressing inequality,[21] but comprehensive qualitative or quantitative evaluation of such outcomes seems to be lacking.

In the context of other countries however, we observe much more in-depth research into the trends in and impacts of VET. There is likewise a noticeable focus on the private returns and outcomes of workers that graduated from VEIs. Brunello and Rocco[22] as well as Choi et al.[23] separately examined data from the OECD's Program for the International Assessment of Adult Competencies (PIAAC), assessing graduates' skill levels and labour market outcomes across numerous countries. VEIs play a much larger role in the education system of European countries, and we see correspondingly more research into the status and outcomes of various VET systems. Brauns compared the systems of VET between Germany and France, assessing their relations to industry and the institutional challenges they faced during the 1980s.[24] He argues that the dual system of vocational training in Germany that is sensitive to industry needs produces a highly skilled workforce that immensely benefits German companies, while the absence of a consolidated VET system in France leaves vocational students at a disadvantage and deprives firms of workers with specialised training.[25] Budría and Telhado-Pereira examined the formation of human capital in VEIs and labour market outcomes of VET graduates in Portugal, finding that VET helps workers transition faster to employment and that

[20] Karen Levesque et al., *Vocational Education in the United States: Toward the Year 2000*, National Centre for Education Statistics, US Department of Education (2000), 147-72.
Dortch, *Career and Technical Education (CTE): A Primer*, 9-12.
Jennifer Ma and Sandy Baum, "Trends in Community Colleges: Enrollment, Prices, Student Debt, and Completion," *Research Brief, College Board Research* (Apr 2016): 1-21.
[21] Dortch, *Career and Technical Education (CTE): A Primer*, 1.
[22] Giorgio Brunello & Lorenzo Rocco, "The effects of vocational education on adult skills and wages: What can we learn from PIAAC?" *OECD Social, Employment and Migration Working Papers No. 168* (2015): 4-6.
[23] Su Jung Choi, Jin Chul Jeon and Seoung Nam Kim, "Impact of vocational education and training on adult skills and employment," *International Journal of Educational Development* 66 (2019): 129-38.
[24] Hildegard Brauns, "Vocational Education in Germany and France," International Journal of Sociology 28, no. 4 (Winter 1999): 57-98.
[25] Ibid., 84-7.

greater participation in VET helps "reduce the extent of skills mismatches" in the labour market.[26] Increasingly though, more studies are being published on how VEIs directly contribute to regional economic development. A seminal paper published by Lund and Karlsen examined the VET system in Norway and studied the co-evolution of VEIs with regional manufacturing industries in the face of technological change. The researchers argue for the importance of VET by differentiating between the "synthetic" and "analytical knowledge base" that develop within regional economies. Universities contribute to the latter through their research activities and training of scientists, while the practical skills and knowledge that comprise the former are generated from workplace training as well as the training of skilled workers in VEIs. A strong synthetic knowledge base gives rise to a "doing, using and interaction (DUI) mode of innovation" that improves industry products and processes, yet its importance has been neglected within HEI-centric RIS literature.[27]

Lund and Karlsen examined two manufacturing regions to obtain insights into how VEIs function as "sources of knowledge in adapting to new technologies".[28] In the Raufoss and Kongsberg regions, vocational colleges and universities are located within the main manufacturing clusters, from which the skilled workforce needed for industry are drawn. Through surveys and field research, the researchers identified numerous examples of VEIs responding flexibly and rapidly to the shifting knowledge demands of a wide variety of industries, developing new VET programs to support new factories and production processes, as well as forming new institutional configurations such as apprenticeship centres and industrial partnerships to meet employer demands. These VEIs are also seen to constantly

[26] Santiago Budria and Pedro Telhado-Pereira, "The contribution of vocational training to employment, job-related skills and productivity: evidence from Madeira," *International Journal of Training and Development* 13, no. 1 (2009): 69-71.
[27] Henrik Brynthe Lund and Asbjorn Karlsen, "The importance of vocational education institutions in manufacturing regions: adding content to a broad definition of regional innovation systems," *Industry and Innovation* 27, no. 6 (2020): 662-3.
[28] Ibid., 661.

upgrade their existing programs to help students meet the evolving knowledge requirements of the modern economy, enhancing their role as knowledge institutions and providers of skilled workers well-equipped for jobs of the future.[29] In doing so, VEIs in these regions boost the productivity and innovativeness of regional firms, contributing directly to industrial productivity and regional growth. This study breaks new ground in its analysis of the role VEIs play within RIS and uncovers numerous channels through which VET contributes to industrial competitiveness and regional economic development, but the lack of quantitative analysis on the magnitude of the benefits these dynamic VEIs bring hinders the persuasiveness of the researchers' claims. Yet within the US context, there is a stark lack of such in-depth studies on VEIs and their contributions to regional development, offering an avenue for further research.

Uncovering factors contributing to growth and innovation has always been a priority of policymakers in their search for strategies to raise the standards of living and competitive advantage. Research in the US on the economics of innovation and education picked up momentum in recent years following growing realization that the country is suffering a structural decline in economic competitiveness and is falling behind in the global innovation race.[30] The Economist Intelligence Unit assesses that a "skills gap" has emerged in the workforce, defined as a growing disparity between the skills workers possess and those needed in modern industries, necessitating widespread changes in the education system.[31] Upgrading VET seems to be a solution to this urgent problem, and businesses have identified VEIs such as community colleges as a "natural fit" for their needs. More research into the channels through which VET contribute to development is thus urgently needed to develop effective policies that link education to the needs of regional economies today.

[29] Ibid., 669-74.
[30] Robert D. Atkinson and Stephen J. Ezell, *Innovation Economics: The Race for Global Advantage* (New Haven: Yale University Press, 2012), 20-32.
[31] The Economist Intelligence Unit, *Closing the Skills Gap – companies and colleges collaborating for change* (London: The Economist Intelligence Unit Limited, 2014), 2-10.

Hypothesis, Data and Research Design

Research Hypothesis

Our hypothesis is that US counties with greater supply of vocationally-trained human capital (VHK) exhibited higher regional GDP levels and greater patenting activity over the period of 2010 to 2015. This hypothesis is informed by aforementioned studies in the literature linking human capital production to economic growth and innovation,[32] as well as studies defining VEIs as generators of human capital in the form of skilled workers.[33] The approach we will be taking to test this hypothesis is panel data analysis, as we are interested in the relationship between VHK, growth and innovation within the US across space and over time. Inspired by studies conducted by economists such as Aghion and Vandenbussche on the links between education, human capital and growth,[34] we believe that the contributions of human capital to economic development should be investigated on a longitudinal basis and that it is important to account for unobserved heterogeneity across regions when assessing the impact of education at a more granular level. In our estimation model we will be controlling for other factors that likely contribute to county-level growth and innovation in our models, such as the level of employment and wages, measures of business vitality, as well as the supply of university-educated human capital (UHK).

Data Sources, Rationale & Limitations

To measure the supply of VHK within counties, we use annual data on the share of adults that hold associate's degrees awarded by VEIs within the population of counties (between 0% to 100%), dubbing this key variable as VHKRATE. Estimates of this variable for

[32] Barro, "Human Capital and Growth," 12-7.
Benhabib and Spiegel, "The role of human capital in economic development," 143–174.
[33] Semple, "Vocational Education: The Missing Link?", 82-3.
Levesque et al., *Vocational Education in the United States: Toward the Year 2000*, 147-72.
[34] Philippe Aghion et al., "Causal Impact of Education on Economic Growth,": 28-38.
Jerome Vandenbussche, Philippe Aghion and Costas Meghir, "Growth, Distance to Frontier and Composition of Human Capital," *Journal of Economic Growth* 11, no. 2 (Jun 2006): 21-34.

each county studied are derived by dividing data on the estimated number of adults with associate's degrees as their highest education attainment, by data on the estimated total population within that county and multiplying by 100. Associate's degrees are defined by the NCES as "[degrees] granted for the successful completion of a subbaccalaurate program of study, usually requiring at least 2 but less than 4 full-time academic years of college study."[35] Postsecondary VEIs perform the important role of awarding industry-recognised credentials to graduating students that certify their mastery of "competencies, skills, and/or knowledge that [are] recognized as necessary or desired for a particular occupation by the relevant industry," with associate's degrees awarded by public community colleges and private vocational schools one prominent form of industry-recognised credentials.[36] Though associate's degrees comprise about 40% of all VET certificates awarded in 2011-12, with the remainder consisting various other subbaccalaurate certificates,[37] in this paper we define VHK as only comprising individuals with associates' degrees as their highest education attainment.

This choice is mostly due to limitations of the data source we are drawing on for our research. County-level estimates of the number of individuals above 25 years old with various levels of education attainment are obtained from the American Community Survey (ACS) 1-Year Data, published annually by the US Census Bureau.[38] Beginning from 2010, the survey produces annual estimates of the number of adults with various levels of education attainment in a subset of around 800 to 850 counties from various states in the US. The levels of education attainment examined range from each year of the K-12 system to the bachelor's, Master's and PhD degrees offered by HEIs. However, the estimates available in this source are not as granular or specific for VET-related degrees and certificates, with the only relevant levels of

[35] "Postsecondary Taxonomy," Career and Technical Education (CTE) Statistics, National Centre for Education Statistics, accessed June, 2022, https://nces.ed.gov/surveys/ctes/tables/glossary_college.asp.
[36] Dortch, *Career and Technical Education (CTE): A Primer*, 4-5.
[37] Ibid., 12.
[38] United States Census Bureau, *American Community Survey – Education Attainment Data* (Nov 23, 2021), distributed by U.S. Department of Commerce, https://www.census.gov/programs-surveys/acs/data.html.

educational attainment being ‘Some college, less than 1 year’, ‘Some college, 1 or more years, no degree’ and ‘Associate’s degree’. However, the broad categories of “some college” makes it difficult to separate individuals that attended postsecondary VEIs from those who dropped out of university or enrolled in shorter academic programs. We also lack specification of the types of degrees or certificates that are awarded to individuals within these estimation categories. We can be more certain that the annual estimates of the number of adults that hold associate’s degrees available in the ACS represent a significant portion of the human capital VEIs generate, thus we stick to a narrower quantitative definition of VHK.

This does however mean our measure of VHK suffers from limitations of incomplete representation of the full range of skilled human capital VEIs generate over time, as well as the assumption of homogeneity between adults holding associate’s degrees due to our inability to distinguish the specific qualifications they obtained and the fields they specialise in. In an ideal situation we would be able to obtain county-level data on the full range of postsecondary VET certifications awarded to graduates or held by working adults. This could possibly be done by recording the number of certifications awarded by VEIs located in each county we examine based on program of study, but the process is extremely resource intensive and could be left to a more in-depth future study. There are also broader issues with the use of educational attainment data as a proxy for human capital and in studying its relationship with growth. Notably, Bils and Klenow found empirical evidence of possible reverse causation with growth affecting individuals’ choices in education, thus the positive effect economists commonly observe between schooling and growth may be erroneously interpreted.[39] This and other issues impose limitations to the kinds of conclusions we can draw from our research, but due to the poor availability of alternative measurements of human capital generated by education

[39] Mark Bils and Peter J. Klenow, “Does Schooling Cause Growth?” *American Economic Review* 90, no. 5 (Dec 2000): 1166-71.

institutions at the county-level in the public domain, we persist with this form of data for estimating the effect of VHK while remaining acutely aware of the limitations of our analysis. We also do not attempt to uncover the exact extent to which VHK directly causes growth and innovation within counties in our research, aiming instead to present an exploratory statistical analysis of their relationship after controlling for covariates and confounders as best as we can. Future researchers can attempt to apply causal inference methods such as the instrumental variable approach to more confidently assess the causal effect of VHK on regional economic development and innovation with greater confidence.

With our county-level VHK data on hand, we then transform the variable into a share variable VHKRATE by dividing VHK by the estimated total population within each county in a particular year (POP). Population estimates are obtained from the Economic Profile by County dataset (code CAINC30) published by the Bureau of Economic Analysis (BEA) as part of their Regional Economic Accounts public database, which covers a range of economic indicators for over 3000 counties in the US over the years 1969 to 2020.[40] Focusing on around 800 counties for which we have a full set of VHK estimates between the years 2010 to 2015, we divide each county's estimated VHK by their POP for each year to obtain a set of VHKRATE estimates that we will be using as the variable of interest in our panel analysis. The decision to use county-level VHKRATE over raw VHK data is due to concerns about significant multicollinearity between VHK and the control variables we are including in our estimation model. This issue will be explained in further detail below. Our choice to restrict analysis to these six years is partly due to our desire to understand the effects of VHK over more recent years, and partly due to restrictions in the availability of data on innovation which is elaborated in further detail below. Some descriptive statistics of VHKRATE follow: mean

[40] Bureau of Economic Analysis, *Regional Economic Accounts – CAINC30: Economic Profile by County* (Dec 8, 2021), distributed by US Department of Commerce, https://apps.bea.gov/regional/downloadzip.cfm.

of 5.64% with standard deviation of 1.43% across our panel, with minimum share of 1.31% and maximum share of 11.4% among our county samples.

The dependent variables we are interested in studying are annual real GDP and the counts of utility patents filed within a county per year (PATENT). Real GDP measures the value of all final goods and services produced within a county adjusted for price inflation since a benchmark year. The estimated annual real GDP figures, in thousands of chained 2012 dollars,[41] for our sample of around 800 counties between 2010 to 2015 are also obtained from the BEA Regional Economic Accounts database, where county-level GDP data is available for each year between 2000 to 2020 (code CAGDP1).[42] County-level innovation statistics are obtained from the US Patent and Trademark Office website, which provides detailed information on the number of utility patents filed within every US county between 2000 to 2015.[43] Measuring innovation is a challenging task, given its often intangible nature as well as reliance on network effects and collaboration within firms as well as institutions. Patents filed by resident inventors to protect new inventions, information and improvements on existing technologies is one of the most common ways researchers measure the level of innovation within a region, given their tangible nature and their suitability as an indicator of successful innovation.[44] The absence of patent data after 2015 at the time of writing this paper restricts

[41] "Chain-type estimates provide the best available method for comparing the level of a given series at two points in time. Chained-dollar estimates are obtained by multiplying the chain-type quantity index for an aggregate by its value in current dollars in the reference year (currently 2012) and dividing by 100. For analysis of changes over time in an aggregate or in a component, the percentage changes calculated from the chained-dollar estimates and the chain-type quantity indexes are the same. Thus, chained-dollar estimates can be used to compute "real" (i.e., inflation-adjusted) rates of growth."
From "National Economic Accounts," Bureau of Economic Analysis, accessed June, 2022, https://apps.bea.gov/iTable/definitions.cfm?did=1&reqId=19.
[42] Bureau of Economic Analysis, *Regional Economic Accounts – CAGDP1: GDP Summary by County and MSA* (Dec 8, 2021), distributed by US Department of Commerce, https://apps.bea.gov/regional/downloadzip.cfm.
[43] United States Patent and Trademark Office, *Calendar Year Patent Statistics (January 1 to December 31): Reports by U.S. Metropolitan Area, Micropolitan Area, and County* (2015), distributed by US Department of Commerce, https://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports_cbsa.htm.
[44] Zoltan J. Acs, Luc Anselin and Attila Varga, "Patents and innovation counts as measures of regional production of new knowledge," *Research Policy* 31 (2002): 1069-1085.
Rifat Atun, Ian Harvey and Joff Wild, "Innovation, Patents and Economic Growth," *International Journal of Innovation Management* 11, no. 2 (June 2007): 279-90.

our analysis to the period before this year. We also account for the nationwide economic shock that hit the US in the years 2008 and 2009 due to the financial crisis. As the effect of this negative shock on regional economies likely obscures the possible effect changes in VHKRATE had on GDP and PATENT, we begin our analysis from the year 2010, where we believe most counties across the US began recovering economically. These considerations therefore restrict our analysis to the six years between 2010 and 2015. In the future when the USPTO publishes patent data for more recent years, this research can be extended to examine the relationship among VHK, growth and innovation over a longer time period.

To discern as accurately as possible the magnitude and significant of the effect VHK might have on regional growth and innovation, we sought to identify a set of appropriate control variables and include them in our panel regression models. Each control variable has to have an effect on or at least be related to both our outcome variables, be independent from or at least have very low correlation with our main variable VHKRATE, and have data available at the county level over the time period of 2010 to 2015. We reference existing studies on the relationship among human capital, growth and innovation within the literature in our process of identifying important covariates of VHKRATE that should be included. Within the Solow growth model with human capital specified by Mankiw, Romer and Weil,[45] as well as the multi-state endogenous growth model specified by Aghion, Boustan, Hoxby and Vandenbussche,[46] labour, technology, the levels of physical capital and human capital (measured by labour skills or education level), as well as the level of investment in education and capital all feature as crucial factor inputs contributing to GDP and income growth. Macroeconomic factors affecting innovation and patenting activity have received somewhat less discussion, with research that touch upon it listing scientific human capital, research and

[45] N. Gregory Mankiw, David Romer and David N. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* 107, no. 2 (May 1992): 415-21.
[46] Philippe Aghion et al., "Causal Impact of Education on Economic Growth,": 30-6.

development, institutions, business vitality, financial development and trade as key factors influencing the innovativeness of countries.[47] As the focus of our research is at the county level, there are limits to our ability to obtain data corresponding to, or are suitable proxies for the above covariates of VHKRATE from publicly-available sources. We ultimately settled on a short list of control variables that we believe contribute to both GDP growth and patenting activity at the county level. They include: the population share of adults above 25 years old with bachelor's degree as highest education attainment (UHKRATE), obtained from the ACS 1-Year data series; estimated total county population (POP), estimated total employment/number of jobs per capita within counties (EMPLOYPC), as well as average wages and salaries within counties (WAGE), all from the BEA Economic Profile by County dataset; new business applications per capita (BIZFORMPC) from the US Census Bureau Business Formation Statistics dataset[48]; as well as the total number of establishments of all sizes (TOTEST) and total number of large establishments with over 500 employees (BIGEST), from the US Census Bureau Statistics of US Businesses (SUSB) annual data series.[49]

We recognize that this set of control variables does not cover the full range of factors affecting regional growth and innovation, but we believe that there is strong economic justification for the inclusion of each of these variables in our panel regression models. UHKRATE represents the main alternative form of human capital to VHK generated by postsecondary education institutions within the US; POP, EMPLOYPC and WAGE reflect the state and growth of the local labour force, as well as their contribution to the county economy; while BIZFORMPC, TOTEST and BIGEST together provide us a comprehensive measure of

[47] Florence Jaumotte and Nigel Pain, "From Ideas to Development: The Determinants of R&D and Patenting," *OECD Economics Department Working Papers* no. 457 (Dec 2005): 26-33.
Sakshi Malik, "Macroeconomic Determinants of Innovation: Evidence from Asian Countries," *Global Business Review* (Jan 2020): 7-10.
[48] United States Census Bureau, *Business Formation Statistics – Annual County Data 2005-2021* (June 23, 2022), distributed by U.S. Department of Commerce, https://www.census.gov/econ/bfs/index.html.
[49] United States Census Bureau, *Statistics of U.S. Businesses (SUSB)* (July 12, 2022), distributed by U.S. Department of Commerce, https://www.census.gov/programs-surveys/susb/data/datasets.html.

the vitality of industries and markets within each county under study. The links between these variables and GDP growth are more immediately evident than their links to innovation and patenting activity, and we expect our list of explanatory variables to explain a larger proportion of the variance in GDP data than PATENT data at the county level. The absence of good data on R&D expenditures, quality of scientific institutions and financial development at the county level means this is a limitation we are unable to overcome at the time of this research, and is another direction future researchers can take.

To our set of explanatory and dependent variables, we apply logarithmic transformation to each of them to transform these right-skewed variables into log variables that are approximately normal. In the case of the PATENT variable, we apply an inverse hyperbolic transformation due to the substantial presence of zeros in the dataset.[50] Most of the counties we have data on do not vary greatly in terms of population size, GDP, business and innovation activity, but there are several counties such as Los Angeles County and New York County whose economic size is so large that their presence skews the distribution of all the variables we collated in our panel dataset. Benefits of taking logarithms are that it is a monotonic transformation, in that it preserves the original order of the data for each of our variables, and that a log-log regression model allows for the convenient interpretation of coefficients as elasticity, i.e. in percentage change terms.[51]

Another important step we took is to transform several variables from their original form obtained from their respective datasets, into a share or per capita variable through dividing

[50] Ghislain B.D. Aihounton and Arne Henningsen, "Units of measurement and the inverse hyperbolic sine transformation," *The Econometrics Journal* 24, no. 2 (May 2021): 334-351.
[51] "In instances where both the dependent variable and independent variable(s) are log-transformed variables, the interpretation is a combination of the linear-log and log-linear cases above. In other words, the interpretation is given as an expected percentage change in Y when X increases by some percentage. Such relationships, where both Y and X are log-transformed, are commonly referred to as elastic in econometrics, and the coefficient of log X is referred to as an elasticity."
From Kenneth Benoit, "Linear Regression Models with Logarithmic Transformations," *Methodology Institute, London School of Economics* (Mar 2011): 2-4.

each of them by POP (VHK into VHKRATE, UHK into UHKRATE, EMPLOY into EMPLOYPC, BIZFORM into BIZFORMPC). This was done to address the high degree of multicollinearity within our set of explanatory variables.[52] Figure 1 is a matrix that displays the measured correlation between each pairing of our explanatory and dependent variables. We observe a high degree of correlation between LNGDP, SINHPATENT and each of our explanatory variables as expected, but worryingly we also observe high levels of correlation within our set of untransformed explanatory variables, with near perfect correlation observed between some pairs. As we are interested in measuring as accurately as possible the magnitude and direction of the effect of VHKRATE on LNGDP and SINHPATENT, the high degree of correlation between VHKRATE and the controls we include is a significant problem. This inference is backed by our estimate of the variance inflation factors (VIF) within our set of explanatory variables, as seen in the left panel of Figure 2 below, where the VIFs for our entire set of explanatory variables (after applying logarithmic transformation) are too large.[53]

Transforming our main variable of interest and several control variables where appropriate helps deal with this issue, as seen in Figure 1 where the correlations of our explanatory variables in share or per capita form (VHKRATE, UHKRATE, EMPLOYPC and BIZFORMPC at the bottom of the matrix) against our remaining control variables are much lower than the correlations under their original forms. In particular, our VHKRATE share variable has close to zero correlation with the entire set of control variables, and as seen in

[52] Multicollinearity is an issue within regression models as high correlation between supposedly independent variables inflate the standard errors of estimated coefficients, causing difficulty in distinguishing the individual effects of a particular explanatory variable on the dependent variable
From "Enough if Enough! Handling Multicollinearity in Regression Analysis," Minitab Blog, last modified April, 2013, https://blog.minitab.com/en/understanding-statistics/handling-multicollinearity-in-regression-analysis.
[53] The Variance Inflation Factor (VIF) method "quantifies how much the variance is inflated" for each predictor in a multiple regression model due to "the existence of correlation among the predictor variables in the model." A VIF of 1 means no correlation between the ith predictor and the remaining j predictors, while as a general rule of thumb VIFs greater than 4 suggests some degree of multicollinearity, and VIFs exceeding 10 is a sign of serious multicollinearity.
From "STAT 462 Applied Regression Analysis: 10.7 – Detecting Multicollinearity Using Variance Inflation Factors," *PennState Eberly College of Science*, 2018, https://online.stat.psu.edu/stat462/node/180/.

Figure 2 its VIF falls significantly to just 18.5 when compared to the new set of controls. This is however still higher than the desired value of less than 10 and indicates a persistent degree of multicollinearity. To further address this, we will first run our regression models using our set of transformed explanatory variables, then apply Principal Components transformation to our set of control variables to generate several component variables. These principal components together account for a substantial portion of the variation in the data but are each completely independent from VHKRATE and between themselves, with the VIF values for the set containing VHKRATE and five component variables all found to be close to 1 as shown in Figure 2. This strategy effectively eliminates the presence of multicollinearity, at the expense of losing a minor portion of the variance explained by our control variables.[54]

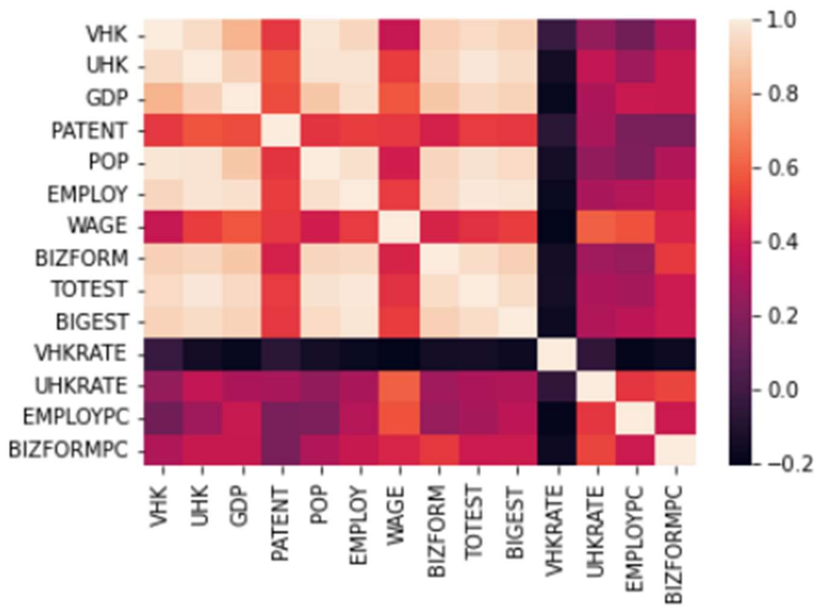

Figure 1. Correlation Matrix of untransformed explanatory and dependent variables, as well as transformed share variables, with the choropleth legend on the right indicating the estimated correlation value between any pair of variables

[54] William F. Massy, "Principal Components Regression in Exploratory Statistical Research," *Journal of the American Statistical Association* 60, no. 309 (Mar 1965): 234-56.
Kenneth Leung, "Principal Components Regression – Clearly Explained and Implemented," Towards Data Science, last modified April 6, 2022, https://towardsdatascience.com/principal-component-regression-clearly-explained-and-implemented-608471530a2f.

Original Set	VIF Value	Modified Set	VIF Value	PCA Set	VIF Value
LNVHK	1476.59	VHKRATE	18.53	VHKRATE	1.00479
LNUHK	1809.03	UHKRATE	20.27	Comp_0	1.00205
LNPOP	6786.88	LNPOP	9693.42	Comp_1	1.000109
LNEMPLOY	12070.64	EMPLOYPC	72.63	Comp_2	1.00198
LNWAGE	1614.77	LNWAGE	2398.27	Comp_3	1.000223
LNBIZFORM	754.19	BIZFORMPC	10.20	Comp_4	1.000424
LNTOTEST	4119.51	LNTOTEST	3649.47		
LNBIGEST	1151.93	LNBIGEST	973.55		

Figure 2. Table of Variance Inflation Factors, assessing multicollinearity within our original set of explanatory variables, transformed set of share and log variables, and our set of VHKRATE and component variables derived via Principal Component transformation

Panel Regression Methodologies

As stated, the goal of our research is to define as precisely as possible the existence and magnitude of any relationship between the supply of VHK, GDP growth and patenting activity within US counties. With our panel dataset on hand, our strategy will be to choose and fit the regression model that is the most suitable for the data that we have. We will be applying and evaluating three different panel regression approaches, with the aim of identifying the model that gives us the most valid and unbiased estimate of the relationship between VHKRATE and GDP, as well as between VHKRATE and PATENT. The models described in (1), (2) and (3) below have LNGDP as their outcome variable, but the same model specifications apply for estimating the relationship between VHKRATE and SINHPATENT.

$$lnGDP_j = \beta_0 + \beta_1 VHKRATE_j + \underline{\beta} Controls_j + \varepsilon_j \qquad (1)$$

We specify a Pooled Ordinary Least Squares (OLS) model in (1), whereby the panel structure of the data is ignored and we fit a multivariate regression model to a flattened dataset with j total observations, without consideration for county and time characteristics. We are only interested in the coefficient β_1 of VHKRATE and its p-value in this model, which tells us the direction, magnitude and statistical significance of the relationship between VHKRATE and LNGDP, while $\underline{\beta}$ represents a vector of the coefficients for our set of control variables with respect to LNGDP. For the inferences from this model to be valid, the assumption of exogeneity must hold, in that there must be zero correlation between our explanatory variables and any unobserved independent variables 'hidden' in the error term that influence LNGDP or SINHPATENT. If there is endogeneity, we will likely find the residual errors of the model to be heteroskedastic, in that their variance will not be constant across all values of our explanatory variables.[55] Furthermore, if residual errors are correlated with the dependent variable or there is autocorrelation, the pooled model will generate biased coefficient estimates.[56] We employ a number of tests to analyse the residual errors of this pooled model and test its goodness-of-fit, namely the White and Breusch-Pagan tests for heteroskedasticity, as well as the Durbin-Watson test of autocorrelation, all of which are well-established within econometrics literature.[57]

However, even if this model is found to be valid and appropriate, we concede that we cannot accurately determine the possible causal effect of VHKRATE on LNGDP or SINHPATENT using this approach. Our Panel OLS model only allow us to disentangle the structure of the correlation between VHKRATE and LNGDP or SINHPATENT after controlling for other factors correlated with the dependent variable. For causation to be determined through linear regression methods without bias, we must either be able to measure and include all confounding variables affecting both VHKRATE and LNGDP or SINHPATENT, or obtain a valid instrumental variable correlated with the outcome only

[55] Sachin Date, "The Pooled OLS Regression Model For Panel Data Sets," Time Series Analysis, Regression and Forecasting, last modified Jan 2022, https://timeseriesreasoning.com/contents/pooled-ols-regression-models-for-panel-data-sets/.
Jason W. Osborne and Elaine Waters, "Four assumptions of multiple regression that researchers should always test," *Practical Assessment, Research, and Evaluation* 8, no. 2 (2002): 4-5.
[56] Bernhard Brugger, "A Guide to Panel Data Regression: Theoretics and Implementation with Python," Towards Data Science, last modified Jan 6, 2021, https://towardsdatascience.com/a-guide-to-panel-data-regression-theoretics-and-implementation-with-python-4c84c5055cf8.
[57] Hun Myong Park, "Practical Guides to Panel Data Modelling: A Step by Step Analysis using Stata," *Public Management & Policy Analysis Program*, International University of Japan (Oct 2011): 12.

through our VHKRATE variable and uncorrelated with unobservable confounders.[58] Obtaining a valid and strong instrument that is highly correlated to VHKRATE but uncorrelated to LNGDP is not impossible but extremely difficult, likely requiring new in-depth data on how vocational education is taught and managed at the county level to be obtained through surveys or from unpublished government statistics. This process is beyond the scope of this research project and thus measuring correlation and drawing inferences on the possible relationship between VHK, GDP and PATENT is our focus when applying this and other regression methodologies.

$$lnGDP_{i,t} = \beta_0 + \beta_1 VHKRATE_{i,t} + \underline{\beta} Controls_{i,t} + \alpha_i + \gamma_t + \varepsilon_{i,t} \qquad (2)$$

We then specify a two-way fixed effects (FE) model, which is a popular method for estimating the effect of explanatory variables on economic outcomes within panel datasets. FE regression methods are designed to deal with characteristics or factors that are not directly observable or measurable but heterogeneously affect different observations or groups, but have to be controlled for as ignoring their effects may result in inaccurate model inferences.[59] In practice, this means that the intercept of the model is allowed to vary freely across groups by including dummy variables, such that we treat the individual effects of unobserved independent variables as constant across time or entities. Crucially, since the entity or time effects are embedded in the intercept of the model, the FE method allows for endogeneity, that is nonzero correlation between unobserved entity attributes and explanatory variables.[60] In our model described in (2), the county FE parameter α_i controls for unobservable time-invariant heterogeneity, that is unobservable attributes affecting growth or innovation that differ between

[58] Matheus F. Alves, "08 – Instrumental Variables," Causal Inference for the Brave and True, last modified 2022, https://matheusfacure.github.io/python-causality-handbook/08-Instrumental-Variables.html.
[59] Sachin Date, "The Fixed Effects Regression Model For Panel Data Sets," Time Series Analysis, Regression and Forecasting, last modified Jan 2022, https://timeseriesreasoning.com/contents/the-fixed-effects-regression-model-for-panel-data-sets/.
[60] Park, "Panel Data Modelling," 7-8.

the i counties in our panel dataset but are constant across time. The time FE parameter γ_t controls for county-invariant heterogeneity, that is unobservable attributes that are constant across counties but differ across our t time periods. Including these parameters means that we only need to control for attributes or factors that vary across counties and over time in our regression model, and by observation we know that our variable of interest VHKRATE and selected controls meet this criterion.

In essence, we are taking advantage of the panel structure of our data to deal with unobservable confounders and get closer to an unbiased estimation of the causal effect of raising VHKRATE on GDP or PATENT, as long as all remaining confounders that vary across space and time are controlled for.[61] The interpretation of the key coefficient B1 is thus that for a given county in our panel dataset, as VHKRATE varies across time by one unit (i.e. a one percent change in the share of workers within the population with associate degree as their highest education attainment), LNGDP increases or decreases on average by a factor of e^{β_1}. In addition, we will also be applying clustered standard errors in our two-way FE model, to account for temporal serial correlation in the error terms within each county we are measuring.[62] In our panel dataset we make repeated observations of around 800 counties under study over a common time period, which means that the observations for each county over time are likely not independently and identically distributed.

Though the two-way FE model seems intuitively the most suitable method for modelling the relationship between VHKRATE and GDP or PATENT, there are limitations to its application and the conclusions that can be drawn from the regression results. The FE

[61] Matheus F. Alves, "14 – Panel Data and Fixed Effects," *Causal Inference for the Brave and True*, last modified 2022, https://matheusfacure.github.io/python-causality-handbook/14-Panel-Data-and-Fixed-Effects.html.
[62] David McKenzie, "When should you cluster standard errors? New wisdom from the econometrics oracle," *World Bank Blogs*, last modified Oct 16, 2017, https://blogs.worldbank.org/impactevaluations/when-should-you-cluster-standard-errors-new-wisdom-econometrics-oracle.

method assumes independence between the entities under study, which implies independence of the time or county-invariant attributes affecting entities, such that each entity's error term and FE parameters have no correlation with the parameters of other entities.[63] However, it is well-established in macroeconomics literature that economic spillovers and migration effects that influence the development of regional economies in close geographical proximity, and thus the growth and innovation trends of counties located in the same states or bordering each other are unlikely to be entirely independent.[64] We also assume that the effects of the time-invariant attributes we control for on the growth and innovation activity of individual counties are constant over time, and that the effects of the county-invariant factors we control for are constant across counties. This however may be unrealistic in the case of complex economic phenomenon such as GDP growth and patenting activity, and by omitting controls for such attributes in our model our inferences may continue to suffer from omitted variable bias.[65]

Furthermore, Torres-Reyna asserts that FE models "[do] not work well with data for which within-cluster variation is minimal or for slow changing variables over time."[66] Given the consistent but slow rate of increase we generally observe in the VHKRATE data of counties, we might be concerned that the possible effects of VHK on growth and innovation may not be accurately discernable using our model and data covering just a short period of six years. Also, the inferences we make from our FE model cannot be generalized to counties and time periods that are not included for analysis within our panel dataset. The set of counties we are analyzing

---

[63] Oscar Torres-Reyna, "Panel Data Analysis: Fixed and Random Effects using Stata", Princeton University, last modified Dec 2007, https://www.princeton.edu/~otorres/Panel101.pdf.
Joos Korstanje, "Assumptions of linear regression," Towards Data Science, last modified Jun 21, 2021, https://medium.com/towards-data-science/assumptions-of-linear-regression-fdb71ebeaa8b.
[64] Philippe Aghion et al., "Causal Impact of Education on Economic Growth,": 13-4.
Chi and Qian, "Education in regional innovation activities," 399.
[65] Terrence D. Hill, Andrew P. Davis and J. Micah Roos, "Limitations of Fixed-Effects Models for Panel Data," *Sociological Perspectives* (Jun 2020): 12-3. Accessed from https://www.researchgate.net/publication/334000163_Limitations_of_Fixed-Effects_Models_for_Panel_Data.
[66] Torres-Reyna, *Panel Data Analysis*, 10.
Hill, Davis and Roos, "Limitations of Fixed-Effects Models," 10-1.

are all relatively large in terms of population and economic size and can be said to be representative of regional economies across the US. Yet, due to this inherent limitation of FE models, the external validity of our inferences is limited and we must be careful about generalizing our findings to the broader population of counties and across other periods in time.[67] Finally, we cannot claim that our model as specified in (2) includes all observable confounding variables that vary across counties and time, thus our estimation of the causal effect likely remains biased. Pairing our model with the instrumental variable approach will help us definitively address this, whereby we estimate the effect of VHKRATE on LNGDP or SINHPATENT through an instrument that is uncorrelated with both the outcome variable and the confounders affecting VHKRATE.[68] However, confidently determining a valid and strong instrumental variable is a difficult process, thus extending this research via a 2-stage least squares panel regression study is a great avenue for future research.

$$lnGDP_{i,\,t} = \beta_0 + \beta_1 VHKRATE_{i,\,t} + \underline{\beta} Controls_{i,\,t} + \delta_i + \gamma_t + \mu_{i,\,t} \qquad (3)$$

An alternative to the FE model is the random effects (RE) model, in which the effects of unobservable attributes or factors on entities are modelled as random variables with unique variances to be estimated. The key difference between the FE and RE methods is the assumption of zero correlation between unobserved attributes and explanatory variables included in the model, such that there is unobserved heterogeneity but no endogeneity issue. In practice, the effects of unobserved attributes on an entity are modelled as a component of its error term in RE models, such that the differential effects of these attributes between entities is accounted for by entity-specific error terms, while the regression intercept remains constant

[67] Date, “Fixed Effects Regression Model,” *Time Series Analysis, Regression and Forecasting*.
[68] Philippe Aghion et al., “Causal Impact of Education on Economic Growth,”: 20-22.
Pierre Wilner Jeanty and Frederick J. Hitzhusen, “Analysing the Effects of Conflicts on Food Security in Developing Countries: An Instrumental Variable Panel Data Approach,” *American Agricultural Economics Association* (July 2006): 5-13.

across them.[69] In (3) we specify a mixed-effects model, whereby county effects are modelled as random while time effects are modelled as fixed. The county-specific random effects parameter δ_i is a normally distributed random variable with a mean of zero and constant variance to be estimated. It is constant across time and has zero correlation with the overall error term $\mu_{i,\,t}$, while we model time effects as fixed using parameter γ_t. This method does not require the inclusion of dummies and thus reduces the number of parameters to be estimated, while also allowing for the inclusion of time-invariant explanatory variables, but if the exogeneity assumption fails to hold the RE model will produce biased and inconsistent estimates.[70] To decide between modelling county effects as fixed or random variables, a popular approach is to apply the Hausman specification test to test for endogeneity within our model. If the null hypothesis of zero correlation between our set of explanatory variables and the error term is rejected, modelling county effects as a fixed parameter is preferred.[71] However, a more intuitive explanation for favoring one of these models can likely be found without relying exclusively on a test of dubious validity, as statisticians have found that the Hausman test tends to 'over-reject' its null hypothesis under certain model conditions.[72]

[69] Sachin Date, "The Random Effects Regression Model For Panel Data Sets," *Time Series Analysis, Regression and Forecasting*, last modified Jan 2022, https://timeseriesreasoning.com/contents/the-random-effects-regression-model-for-panel-data-sets/.
[70] Jeanty and Hitzhusen, "An Instrumental Variable Panel Data Approach," 7.
[71] Park, "Panel Data Modelling," 9.
[72] Teodora Sheytanova, "The Accuracy of the Hausman Test in Panel Data: a Monte Carlo Study," *Advanced level thesis 1*, Örebro University School of Business (2014): 43-5.

Results & Analysis

VHKRATE on LNGDP

The first section of our analysis focuses on the relationship between the VHKRATE and LNGDP in the US at the county level over the period of 2010 to 2015. We begin with a simple Pooled OLS regression of VHKRATE against LNGDP without controls ('UNIVP'), and compare the results against a Pooled regression with our set of control variables included ('POOLED'). Then, we run two FE regressions, the first including just a county FE parameter ('COUNTY FE') and the second including both county and time FE parameters ('2WAY FE'). Finally, we run a random effects regression with only a county RE parameter included ('COUNTY RE'), as well as a mixed-effects regression with both county RE and time FE parameters ('MIXED'). The reason for running the county FE and county RE models in addition to the three models specified in the Methodology section above is to allow us to observe more clearly how these different methods of controlling for time-invariant heterogeneity across counties affect the results we obtain on the relationship between VHKRATE and LNGDP.

We observe some interesting results from our regression analysis presented in Figure 3 below. The coefficient of VHKRATE in the 'UNIVP' model is -0.1560 and is statistically significant at the 99 percent confidence level, but its coefficient in the 'POOLED' model with all control variables included is -0.0002 and is not statistically significant. Assuming the Pooled OLS model is appropriate for testing our research hypothesis, this likely indicates that some, if not all, of our control variables are significantly associated with both VHKRATE and LNGDP. Failing to control for these economic factors in our 'UNIVP' regression biased the parameter estimate of VHKRATE, thus the effect of VHKRATE on LNGDP shown in the univariate model reflects in fact the statistically significant effect of the control variables on LNGDP, with their p-values all found to be close to zero. The overall R-squared value of the 'POOLED' model is very high at 0.9881, indicating that this model we specified explains a significant proportion of the variation in LNGDP.

Variable	UNIVP	POOLED	COUNTY FE	2WAY FE	COUNTY RE	MIXED
	Coefficients	Coefficients	Coefficients	Coefficients	Coefficients	Coefficients
VHKRATE	-0.1560*** (0.0109)	-0.0002 (0.0013)	-0.0007 (0.0012)	0.0016 (0.0010)	-0.0050*** (0.0010)	0.0017* (0.0009)
UHKRATE		-0.0106*** (0.0007)	-0.0025*** (0.0008)	-0.0008 (0.0006)	-0.0045*** (0.0007)	-0.0023*** (0.0007)
LNPOP		0.7551*** (0.0142)	0.8649*** (0.1107)	1.0526*** (0.0939)	0.8361*** (0.0246)	0.8888*** (0.0232)
EMPLOYPC		1.3656*** (0.0261)	1.4025*** (0.1243)	1.5881*** (0.1207)	1.5602*** (0.0449)	1.5263*** (0.0418)
LNWAGE		1.0850*** (0.0150)	0.2423*** (0.2423)	0.6934*** (0.0740)	0.4087*** (0.0195)	0.9121*** (0.0247)
BIZFORMPC		-13.525*** (0.7176)	-5.1718*** (1.7581)	-3.3003** (1.4433)	-11.563*** (1.1731)	-7.9354*** (1.0998)
LNTOTEST		0.1966*** (0.0125)	0.2376*** (0.0811)	0.0298 (0.0725)	0.2644*** (0.0223)	0.0711*** (0.0219)
LNBIGEST		0.0501*** (0.0084)	-0.0919*** (0.0354)	0.0371 (0.0373)	-0.0394*** (0.0146)	0.0419*** (0.0138)
Intercept	16.805*** (0.0631)	-7.3528*** (0.1440)	0.7070 (0.8784)	-5.6393*** (1.2495)	-1.2782*** (0.1984)	-6.2197*** (0.2477)
Dep Var	LNGDP	LNGDP	LNGDP	LNGDP	LNGDP	LNGDP
Estimator	PooledOLS	PooledOLS	PanelOLS	PanelOLS	RandomEff	RandomEff
Cov. Est.	Unadj	Unadj	Clustered	Clustered	Unadj	Unadj
N	4734	4734	4734	4734	4734	4734
Overall R^2	0.0418	0.9881	0.9713	0.9788	0.9817	0.9882
Within R^2	-4.0594	0.0639	0.4853	0.2448	0.4591	0.5129
F-statistic	206.36	4.9e+04	463.95	300.78	8813.3	6211.3
P-value (F-stat)	0.000	0.000	0.000	0.000	0.000	0.000

Legend: *: significant at the 10%; **: significant at 5%; ***: significant at 1%
Unadj: no adjustments made to standard errors
Standard errors in parentheses

Figure 3. Regression results of VHKRATE against LNGDP, using three different types of panel regression models

However, results of our goodness-of-fit tests displayed in Figure 4 show that Pooled OLS models are not suitable for the dataset we have at hand. Both the White test and the Breusch-Pagan test assess for heteroskedasticity of residual errors of a regression model, and applied to the 'POOLED' model we obtain very large test statistics for both tests which are rejected at the 99 percent confidence level, indicating significant heteroskedasticity and thus endogeneity. Applying the Durbin-Watson test of autocorrelation onto the 'POOLED' model, we obtain a value of 0.527 which indicates strong positive autocorrelation of residual errors. Rejection of these tests mean that the key assumptions of exogeneity, homoskedasticity and non-autocorrelation underpinning the Pooled OLS model are violated, indicating that either FE or RE models are more suitable for exploring the data at hand.[73]

Goodness-of-Fit test	Test Statistics and Significance			
	LM-Stat	LM p-value	F-statistic	F p-value
White Test	557.065	5.045e-90	14.213	1.406e-96
Breusch-Pagan Test	312.029	1.131e-62	41.676	7.057e-65
	Test Result			
Durbin-Watson Test	0.52657			

Figure 4. Results of three goodness-of-fit tests for the Pooled OLS model of VHKRATE against LNGDP

Moving then to the results of the models with FE parameters included, in Figure 3 we observe that controlling for county FE does not significantly alter the coefficient of VHKRATE, which remains negative and statistically insignificant under 'COUNTY FE'. However, in the '2WAY FE' model with both county and time FE included, the coefficient of VHKRATE is positive at 0.0016 with a p-value of 0.1087. To determine the fit of these FE models we examine their estimated within R-squared values, instead of the overall R-squared value as FE models examine specifically within-cluster variation and not cross-cluster variation,[74] and we find that it decreases to 0.485 when county FE are included and further decreases to 0.2448 when both county and time FE are included. Though this indicates that our model as specified in (2) does not fully explain the variation in LNGDP within counties over time in our panel dataset, the model is not entirely misspecified and the explanatory variables we picked for analysis do account for some of the changes observed annually in county-level growth. Controlling for county RE in our regression models however give us very different results. The coefficient of

[73] Park, "Panel Data Modelling," 7.
[74] Torres-Reyna, *Panel Data Analysis*, 10.

VHKRATE in 'COUNTY RE' is still negative, but its magnitude is larger at -0.005 compared to its value in the 'COUNTY FE' model and is importantly found to be statistically significant at the 99 percent confidence level. With the inclusion of time FE alongside county RE, the coefficient of VHKRATE in 'MIXED' becomes positive at 0.0017 and is statistically significant at the 90 percent confidence level. The magnitude of the effect of VHKRATE on LNGDP estimated with our mixed effects model is similar to that estimated using our two-way FE model, but smaller coefficient standard errors mean that the T-statistic of the VHKRATE coefficient is smaller under the mixed effects model than under the two-way FE model.

Hausman Test Results: County Fixed Effects Model vs County Random Effects Model	
Chi-Squared	639.963
Degrees of freedom	9
p-value	5.507e-132

Figure 5. Results of Hausman specification test comparing results of two-way Fixed Effects model versus mixed effects model

Applying the Hausmann specification test and comparing 'COUNTY FE' against 'COUNTY RE', the results as presented in Figure 5 tell us that the null hypothesis of zero correlation between unobserved entity effects and the explanatory variables in these models is rejected at the 99 percent confidence level. More importantly, it is easy to argue based on intuition that the assumption of zero correlation between our observed explanatory variables and all unobservable covariates necessary for the application of RE to be valid is unlikely to hold. Unobservable, time-invariant county-level attributes affecting growth such as geography, county or state-level policies and trade relations with other countries that remained unchanged between 2010 to 2015 are likely to display some degree of correlation with our set of explanatory variables encompassing population, wages, education, and business vitality. Thus, we determine that modelling unobserved time-invariant heterogeneity using county RE is not appropriate for the panel dataset at hand and for our research question, and we thus rely on the coefficient estimates derived from the two-way FE model as described in (2).

While we could conclude our analysis of the relationship between VHKRATE and LNGDP here, there is yet a further step that we can take to assess the robustness of our regression results. As previously mentioned, a key concern to address when specifying any regression model is the issue of multicollinearity within the set of explanatory variables we select for analysis. Though the transformation of key explanatory variables into population shares successfully reduced the degree of correlation between VHKRATE and our control variables, a small degree of multicollinearity persists as seen in the VIF results presented in Figure 2 and it may continue to bias our estimate of the coefficient of VHKRATE. One solution is to apply Principal Components transformation to our set of controls, a mathematical process of extracting the main 'components' from a bunch of covariates and recombining them into several component variables that are entirely independent from each other.[75] By first standardizing our control variables and then transforming them into five distinct component variables, we create a new set of control variables that as seen in Figure 6 cumulatively explain 99.6 percent of the variance within the original set of controls, and that as seen in Figure 2 are completely independent from VHKRATE and each other with VIF values close to 1. We then run a two-way FE regression of VHKRATE against LNGDP while controlling for the five component variables we created and compare the results with our original two-way FE model. In Figure 7 we observe that the estimated coefficient of VHKRATE within the 'PCA' model is larger compared to that obtained from the original two-way FE model at 0.0018 with a lower p-value of 0.0889, meaning that our hypothesis is in fact not rejected at the 90 percent confidence level. However, we remain careful about relying on the results obtained from this

[75] Kevin Dunn, "6.6. Principal Component Regression (PCR)," Process Improvement Using Data, last modified May 4, 2022, https://learnche.org/pid/latent-variable-modelling/principal-components-regression.

model. Though it benefits from the elimination of multicollinearity within our set of explanatory variables, the loss of some portion of the variance within our original control variables likely biases our coefficient estimate to a small extent. What this method does offer is a robustness check of the results obtained using the two-way FE model as described in (2), showing that the persistent degree of multicollinearity likely does not significantly bias our estimate of the relationship between VHKRATE and LNGDP in the two-way FE model.

Principal Components Transformation: Variance Analysis	
Eigenvalues/Explained Variance:	5.97; 0.0975; 0.0519; 0.0341; 0.0216
Explained Variance Ratio:	0.963; 0.0157; 0.00838; 0.00551; 0.00348
Cumulative Explained Variance Ratio:	0.963; 0.979; 0.987; 0.992; 0.996

Figure 6. Variance analysis of the five Principle Component variables created using our set of control variables after standardization

Figure 7 thus focuses our attention on the relevant regression results for analysis. With two-way FE regression evaluated to be the most suitable method for our research question and panel dataset, we present the regression results of a univariate two-way FE model ('UNIVFE'), a two-way FE model with controls included ('2WAY FE'), as well as a two-way FE model with principal components included as controls ('PCA'). We observe that no significant effect of VHKRATE on LNGDP is identified within 'UNIVFE', with the coefficient found to be close to zero. When we estimate this effect after controlling for each county's population, wage, education and business vitality levels, results from '2WAY FE' imply that a one unit increase in the VHKRATE of a county is on average estimated to increase its annual LNGDP by a factor of $e^{0.0016}$ = 1.0016 in the same year. Such a large contrast in the estimated coefficient value for our key variable is due to the inclusion of highly significant predictors of GDP growth in our full two-way FE model, notably LNPOP, EMPLOYPC, LNWAGE AND BIZFORMPC, that help uncover a more accurate estimation of the true effect of VHK on growth. Although strictly speaking the hypothesis of a positive and significant effect of VHKRATE on LNGDP is still rejected at the 90 percent confidence level, we argue that the results of our regression analysis offers weak support for the theory that accumulation of VHK contributes to county-level economic growth over time.

Variable	UNIVFE	2WAY FE	PCA
	Coefficients	Coefficients	Coefficients
VHKRATE	-0.0003 (0.0016)	0.0016 (0.0010)	0.0018* (0.0010)
UHKRATE		-0.0008 (0.0006)	
LNPOP		1.0526*** (0.0939)	
EMPLOYPC		1.5881*** (0.1207)	
LNWAGE		0.6934*** (0.0740)	
BIZFORMPC		-3.3003** (1.4433)	
LNTOTEST		0.0298 (0.0725)	
LNBIGEST		0.0371 (0.0373)	
Comp_0			0.4455*** (0.0332)
Comp_1			-0.2792*** (0.0312)
Comp_2			-0.1289*** (0.0273)
Comp_3			0.2646** (0.1320)
Comp_4			0.7852*** (0.1110)
Intercept	15.928*** (0.0089)	-5.6393*** (1.2495)	15.916*** (0.0059)
Dep Var	LNGDP	LNGDP	LNGDP
Estimator	PanelOLS	PanelOLS	PanelOLS
N	4734	4734	4734
Overall R^2	0.0002	0.9788	0.9589
Within R^2	-0.0013	0.2448	0.4254
F-statistic	0.0781	300.78	235.06
P-value (F-stat)	0.7799	0.000	0.000

Legend: *: significant at the 10%; **: significant at 5%; ***: significant at 1%
Standard errors in parentheses

Figure 7. Two-way Fixed Effects regression results of VHKRATE against LNGDP, with only VHKRATE, VHKRATE plus controls, and VHKRATE plus principal components included respectively

Results from our additional ‘PCA’ model supports this conclusion, as the direction and size of the effect of VHKRATE is found to be similar while also being statistically significant at the 90 percent confidence level, but we exercise caution in relying on these results due to the limitations of this model. It is interesting to examine which control variables are found to have a statistically significant effect on LNGDP. Changes in population, employment, wages and business formation rates predictably have significant effects on county-level GDP, but interestingly UHKRATE – the population share of adults with bachelors’ degrees – turns out to have no significant effect. If education attainment data is a valid proxy for measuring the output of human capital from educational institutions, this raises questions about our understanding of how postsecondary education is linked to regional economic growth.

VHKRATE on SINHPATENT

The second section of our analysis focuses on the relationship between VHKRATE and SINHPATENT in the US at the county level over the period of 2010 to 2015. Similarly, we begin by pooling all observations together and examining the coefficient of VHKRATE within a univariate Pooled OLS model (‘UNIVPP’) and in another Pooled model with controls included (‘POOLEDP’). Then, we estimate two separate two-way FE regression models, one with no controls included (‘UNIVFEP’) and another with control variables included (‘2WAY FEP’). Just as we had earlier determined that the application of county RE in our models of VHK against GDP is unlikely to be valid, we also believe that it is unlikely for the assumption of zero correlation between observed and unobservable covariates to hold for our models regressing VHK against PATENT, thus we exclude analysis of RE and mixed-effects models. As a robustness check, we also regress VHKRATE and the five Principal Component variables generated from our control variables against SINHPATENT in a two-way FE model (‘PCA FEP’), so as to assess the results after eliminating multicollinearity.

Variable	UNIVPP	POOLEDP	UNIVFEP	2WAY FEP	PCA FEP
	Coefficients	Coefficients	Coefficients	Coefficients	Coefficients
VHKRATE	-0.0595*** (0.0178)	0.0696*** (0.0097)	0.0011 (0.0084)	0.0041 (0.0089)	0.0040 (0.0092)
UHKRATE		0.1653*** (0.0051)		0.0067 (0.0067)	
LNPOP		0.6834*** (0.1055)		0.5960 (0.5971)	
EMPLOYPC		-0.7814*** (0.1936)		0.7720 (0.7829)	
LNWAGE		1.5944*** (0.1112)		0.8004** (0.3367)	
BIZFORMPC		-66.809*** (5.3255)		-2.2111 (21.571)	
LNTOTEST		0.9876*** (0.0924)		0.6594 (0.4580)	
LNBIGEST		-0.5369*** (0.0627)		-0.0586 (0.1695)	
Comp_0					0.4806*** (0.1300)
Comp_1					-0.2270 (0.2103)
Comp_2					-0.3267*** (0.0979)
Comp_3					0.4780 (0.4260)
Comp_4					1.0209*** (0.3008)
Intercept	4.4389*** (0.1035)	-27.337*** (1.0685)	4.0975*** (0.0471)	-17.342** (7.1499)	4.0811*** (0.0514)
Dep Var	SINHPATENT	SINHPATENT	SINHPATENT	SINHPATENT	SINHPATENT
Estimator	PooledOLS	PooledOLS	PanelOLS	PanelOLS	PanelOLS
N	4734	4734	4734	4734	4734
Overall R^2	0.0024	0.7461	-8.4e-05	0.6377	0.6265
Within R^2	-0.0322	-0.0647	0.0004	0.0650	0.0491
F-statistic	11.194	1735.6	0.0143	4.0640	5.4440
P-value (F-stat)	0.0008	0.000	0.9048	0.0001	0.000

Legend: *: significant at the 10%; **: significant at 5%; ***: significant at 1%
Standard errors in parentheses

Figure 8. Comparison of results of regressing VHKRATE against SINHPATENT using Pooled OLS models versus two-way Fixed Effects models

Results from our two Pooled OLS models as shown in Figure 8 raise the prospect of interesting findings. The estimated coefficient of VHKRATE within the 'UNIVPP' model is -0.0595 and is statistically significant at the 99 percent confidence level. After including controls however, its coefficient within the 'POOLEDP' model dramatically transforms into a positive value at 0.0696 that is also statistically significant at the 99 percent confidence level, with all control variables found to be statistically significant as well. Overall R-squared also rises from 0.0024 under 'UNIVPP' to 0.7461 under 'POOLEDP', thus were Pooled OLS the appropriate regression methodology, this model would be a useful tool for understanding how VHK positively affects innovation over time at the county level. However, results of our goodness-of-fit tests presented in Figure 9 again show us that the Pooled OLS model is not suitable for the dataset we have at hand. Both the White test and Breusch-Pagan tests for heteroskedasticity of residual errors are rejected at the 99 percent confidence level, while results from the Durbin-Watson test indicate strong positive autocorrelation of residual errors. Rejection of these tests mean that the key assumptions of exogeneity, homoskedasticity and non-autocorrelation underpinning the Pooled OLS model are violated when applied to our dataset, thus the two-way FE regression model is more suitable for this research question.

Goodness-of-Fit test	Test Statistics and Significance			
	LM-Stat	LM p-value	F-statistic	F p-value
White Test	556.160	7.665e-90	14.187	2.258e-96
Breusch-Pagan Test	199.104	9.870e-39	25.931	1.383e-39
	Test Result			
Durbin-Watson Test	0.57190			

Figure 9. Results of three goodness-of-fit tests for the Pooled OLS model of VHKRATE against SINHPATENT

Looking at the results of our various FE models, the effect of VHKRATE on SINHPATENT is by contrast found to be almost negligible. Under the 'UNIVFEP' model, the coefficient of VHKRATE is positive but statistically insignificant with a very large p-value, and likewise under the full '2WAY FEP' model the coefficient value is positive but not found to be statistically significant at the 90 percent confidence level. It is notable that all our control variables, with the exception of LNWAGE, are found to have no significant effect on the annual SINHPATENT generated within counties in the '2WAY FEP' model. The estimated within R-squared value for both FE models are also extremely small at 0.0004 and 0.065 respectively, implying that our models explains less than 10 percent of the variation observed in SINHPATENT. Applying Principal Components transformation to address multicollinearity, we find that the estimated values of the coefficient of VHKRATE are very similar between the '2WAY FEP' and 'PCA FEP' models, but the effect of VHKRATE on SINHPATENT is still found to be not statistically significant. Hence, the specification of these models in terms of the predictors chosen is likely poor, and a better model is needed to uncover the relationship between vocational education and innovation.

In sum, our results lead us to conclude that there is unlikely to be any direct relationship between the share of VHK within a county's population and its patenting activity over time. This may not surprise some researchers, as it can be argued that utility patents in the modern era are predominantly generated in large companies that hire university-educated workers and in universities undertaking commodifiable research, due to the substantial knowledge requirements and cutting-edge facilities required for innovation today.[76] Thus, VEIs are unlikely to be prominent generators of patents and VET graduates may predominantly contribute only indirectly to patenting activity through the companies that they work for. Interestingly though, our results indicate that neither UHKRATE, the share of university-educated workers, nor LNBIGEST, the number of large establishments within counties, significantly predict recorded patenting activity within counties, with only our wage variable found to have a positive and significant effect. This may be explained by the fact that our measure of UHK rate focuses on workers with bachelor's degrees as their highest education attainment, and thus does not reveal the true effect that HEIs have on innovation.

[76] John Sutton, *Technology and Market Structure: Theory and History*, Boston: MIT Press (1998), 9-18. David C. Mowery, Richard R. Nelson, Bhaven N. Sampat and Arvids A. Ziedonis, "The growth of patenting and licensing by U.S. universities: an assessment of the effects of the Bayh-Dole act of 1980," *Research Policy* 30 (2001): 116-8.

Summary & Evaluation

The role education and educational institutions play in spurring economic growth and innovation has been widely explored in macroeconomics literature and continues to be emphasized by economists and policymakers as a key pillar of development. Human capital generated through the education of young adults plays a big role in endogenous growth theory, where it is posited to combine with technological advancement to boost productivity growth. However, the bulk of existing research has focused on how higher education institutions such as universities generate human capital and contribute to the accumulation of knowledge. Far less attention has been directed towards vocational education institutions such as community colleges, technical institutes as well as industry-led apprenticeship programs. Much of the research that has been done on vocational education in the US has been focused on the labour market outcomes of and private returns to workers that were educated in vocational institutions, whilst in Europe in-depth research into the contributions of vocational education to regional economies has been growing in popularity. This research paper sought to bridge this gap in the literature by conducting an exploratory statistical analysis of the relationship between the share of adults in the population that hold associate's degrees as their highest education attainment, GDP growth and patenting activity at the county level.

Data from around 800 counties across the US between the years 2010 and 2015 were collated from various online databases published by government agencies such as the BEA and the Census Bureau. County-level annual education attainment data, GDP figures and counts of utility patents filed by local entities were compiled alongside population, wage, employment and business vitality data into a panel dataset. Several variables including our measure of vocational-trained human capital were transformed into a population share form to address the significant degree of multicollinearity detected within our set of our explanatory variables, and as an additional step we applied Principal Components transformation to our set of controls to generate independent component variables to control for in our regression model. We assessed the suitability of three different panel regression models for testing our hypothesis, namely a Panel OLS model that ignores the panel structure of the data, a two-way fixed effects regression model controlling for county and time fixed effects, as well as a mixed-effects regression model controlling for county random effects alongside time fixed effects. Evaluation of their respective goodness-of-fit and the validity of their core assumptions lead us to conclude that the two-way fixed effects model is the most appropriate regression approach for investigating the relationship between VHKRATE, LNGDP and SINHPATENT. Yet, there are critical limitations to the conclusions that can be drawn from this model. Assumptions of independence between the counties under study, of the constancy of the effects that the unobserved time-invariant and county-invariant attributes we are controlling for have on growth and innovation, as well as limitations to the external validity of the inferences we can draw from the regression results and to our confidence that the estimates of the causal effect are unbiased due to our inability to include all observable confounders that vary across counties and over time – these must all be taken into account when drawing conclusions from the results of our two-way fixed effects model.

Overall, our regression results tell us that the direct effect of vocational education on the GDP growth and patenting activity of counties is minimal at best. Focusing on the estimated value of the coefficient of VHKRATE and the statistical significance of the difference of this value from zero, our model regressing VHKRATE against LNGDP estimates that for the counties under study a one unit increase in the former variable on average increases the latter by a factor of 1.0016 in the same year. This causal effect however is strictly speaking not found to be statistically significant at the 90 percent confidence level, but we argue that the estimated p-value of 0.1087 indicates that there is weak evidence of a small and positive effect of raising the output of human capital from vocational institutions on growth. Our alternative two-way fixed effect model controlling for independent component variables confirms the robustness of this result. Interestingly, we observe that while changes in population, wage rates, employment and business formation have significant effects on county-level growth as one might expect, the other key source of human capital – the share of adults with bachelor's degrees in the population – is not found to have any significant effect on growth. These intriguing findings make a strong case for further in-depth research into the influence of vocational education on the growth and vitality of regional economies.

Our model regressing VHKRATE against SINHPATENT however reveals no significant causal effect of raising human capital output from vocational institutions on regional innovation activity. This may be due to misspecification of our model linking VHKRATE to innovation, but it may also be explained by the recognition that VEIs primarily focus on preparing students for skilled roles in industry. Thus, with patenting in the modern era becoming increasingly knowledge intensive and requiring cutting-edge technology, VEIs and VET graduates may only indirectly contribute to regional patenting activity, such as through the companies they work for. This does not however imply that VET does not contribute to regional innovation at all, and further research can be undertaken in the future to explore the alternative channels of innovation that VEI graduates excel at in different industries.

Bibliography

Primary Data Sources

Bureau of Economic Analysis. *Regional Economic Accounts – CAINC30: Economic Profile by County*. Dec 8, 2021. Distributed by US Department of Commerce. https://apps.bea.gov/regional/downloadzip.cfm.

Bureau of Economic Analysis. *Regional Economic Accounts – CAGDP1: GDP Summary by County and MSA*. Dec 8, 2021. Distributed by US Department of Commerce. https://apps.bea.gov/regional/downloadzip.cfm.

United States Census Bureau. *American Community Survey – Education Attainment Data*. Nov 23, 2021. Distributed by U.S. Department of Commerce. https://www.census.gov/programs-surveys/acs/data.html.

United States Census Bureau. *Business Formation Statistics – Annual County Data 2005-2021*. June 23, 2022. Distributed by U.S. Department of Commerce. https://www.census.gov/econ/bfs/index.html.

United States Census Bureau. *Statistics of U.S. Businesses (SUSB)*. July 12, 2022. Distributed by U.S. Department of Commerce. https://www.census.gov/programs-surveys/susb/data/datasets.html.

United States Patent and Trademark Office. *Calendar Year Patent Statistics (January 1 to December 31): Reports by U.S. Metropolitan Area, Micropolitan Area, and County*. 2015. Distributed by US Department of Commerce. https://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports_cbsa.htm.

Secondary References

Abel, Jaison R. and Richard Deitz. "Do Colleges and Universities Increase Their Region's Human Capital?" *Federal Reserve Bank of New York Staff Reports* 401 (Oct 2009), 1-33.

Acs, Zoltan J., Luc Anselin and Attila Varga. "Patents and innovation counts as measures of regional production of new knowledge." *Research Policy* 31 (2002), 1069-1085.

Aghion, Philippe, Leah P. Boustan, Caroline Hoxby and Jerome Vandenbussche. "The Causal Impact of Education on Economic Growth: Evidence from the US." *Conference Draft, Brookings Papers on Economic Activity* (Mar 2009), 1-73.

Aihounton, Ghislain B.D. and Arne Henningsen. "Units of measurement and the inverse hyperbolic sine transformation." *The Econometrics Journal* 24, no. 2 (May 2021), 334-351.

Alves, Matheus F. Causal Inference for the Brave and True. Last modified 2022. https://matheusfacure.github.io/python-causality-handbook/08-Instrumental-Variables.html.

Angrist, Noam et al. “Measuring Human Capital.” *Social Science Research Network* (February 2019), 1-44.

Atkinson, Robert D. and Stephen J. Ezell. *Innovation Economics: The Race for Global Advantage*. New Haven: Yale University Press, 2012.

Atun, Rifat, Ian Harvey and Joff Wild. “Innovation, Patents and Economic Growth.” *International Journal of Innovation Management* 11, no. 2 (June 2007), 279-97.

Barro, Robert J. “Education and Economic Growth.” *Annals of Economics and Finance* 14, no. 2 (2013), 277-304.

Barro, Robert J. “Human Capital and Growth.” *American Economic Review* 91 no. 2 (May 2001), 12-7.

Benhabib, Jess and Mark Spiegel. “The role of human capital in economic development evidence from aggregate cross-country data.” *Journal of Monetary Economics* 34 (2) (1994), 143-174.

Benoit, Kenneth. “Linear Regression Models with Logarithmic Transformations.” *Methodology Institute, London School of Economics* (Mar 2011), 1-8.

Bils, Mark and Peter J. Klenow. “Does Schooling Cause Growth?” *American Economic Review* 90, no. 5 (Dec 2000), 1160-83.

Brauns, Hildegard. “Vocational Education in Germany and France.” *International Journal of Sociology* 28, no. 4 (Winter 1999), 57-98.

Brugger, Bernhard. “A Guide to Panel Data Regression: Theoretics and Implementation with Python.” Towards Data Science. Last modified Jan 6, 2021. https://towardsdatascience.com/a-guide-to-panel-data-regression-theoretics-and-implementation-with-python-4c84c5055cf8.

Brunello, Giorgio and Lorenzo Rocco, “The effects of vocational education on adult skills and wages: What can we learn from PIAAC?” *OECD Social, Employment and Migration Working Papers No. 168* (2015).

Budría, Santiago and Pedro Telhado-Pereira. “The contribution of vocational training to employment, job-related skills and productivity: evidence from Madeira.” *International Journal of Training and Development* 13, no. 1 (2009), 53-72.

Bureau of Economic Analysis. “National Economic Accounts.” Accessed June, 2022. https://apps.bea.gov/iTable/definitions.cfm?did=1&reqId=19.

Caniels, Marjolein C.J. and Herman van den Bosch. "The role of Higher Education Institutions in building regional innovation systems." *Papers in Regional Science* 90, no. 2 (June 2011), 271-86.

Charles, David. “Universities as key knowledge infrastructures in regional innovation systems.” *Innovation: The European Journal of Social Science Research* 19, no. 1 (Aug 2006), 117-30.

Choi, Su Jung, Jin Chul Jeon and Seoung Nam Kim. "Impact of vocational education and training on adult skills and employment." *International Journal of Educational Development* 66 (2019), 129-38.

Date, Sachin. "The Fixed Effects Regression Model For Panel Data Sets." Time Series Analysis, Regression and Forecasting. Last modified Jan 2022. https://timeseriesreasoning.com/contents/the-fixed-effects-regression-model-for-panel-data-sets/.

Date, Sachin. "The Pooled OLS Regression Model For Panel Data Sets." Time Series Analysis, Regression and Forecasting. Last modified Jan 2022. https://timeseriesreasoning.com/contents/pooled-ols-regression-models-for-panel-data-sets/.

Date, Sachin. "The Random Effects Regression Model For Panel Data Sets." Time Series Analysis, Regression and Forecasting. Last modified Jan 2022. https://timeseriesreasoning.com/contents/the-random-effects-regression-model-for-panel-data-sets/.

Dortch, Cassandria. "Career and Technical Education (CTE): A Primer." *Congressional Research Service: CRS Report for Congress R42748*, Feb 10, 2014. https://digital.library.unt.edu/ark:/67531/metadc282297/.

Dunn, Kevin. "6.6. Principal Component Regression (PCR)." Process Improvement Using Data. Last modified May 4, 2022. https://learnche.org/pid/latent-variable-modelling/principal-components-regression.

Etzkowitz, Henry and Loet Leydesdorff. "The dynamics of innovation: from National Systems and 'Mode 2' to a Triple Helix of university-industry-government relations." *Elsevier Research Policy* 29 (2000), 109-23.

Fretwell, David. "A Framework for Evaluating Vocational Education and Training (VET)." *European Journal of Education* 38, no. 2 (June 2003), 177-190.

Gennaioli, Nicola et al. "Human Capital and Regional Development." *The Quarterly Journal of Economics* 128, no.1 (2013), 105-64.

Hill, Terrence D., Andrew P. Davis and J. Micah Roos. "Limitations of Fixed-Effects Models for Panel Data." *Sociological Perspectives* (Jun 2020), 357-69.

Jaumotte, Florence and Nigel Pain. "From Ideas to Development: The Determinants of R&D and Patenting." *OECD Economics Department Working Papers* no. 457 (Dec 2005), 4-57.

Jeanty, Pierre Wilner and Frederick J. Hitzhusen. "Analysing the Effects of Conflicts on Food Security in Developing Countries: An Instrumental Variable Panel Data Approach," *American Agricultural Economics Association* (July 2006), 1-30.

Katz, Bruce and Julie Wagner. "The Rise of Innovation Districts: A New Geography Of Innovation In America." *Brookings Metropolitan Policy Program* (May 2014), 1-25.

Korstanje, Joos. “Assumptions of linear regression.” Towards Data Science. Last modified Jun 21, 2021. https://medium.com/towards-data-science/assumptions-of-linear-regression-fdb71ebeaa8b.

Leung, Kenneth. “Principal Components Regression – Clearly Explained and Implemented.” Towards Data Science. Last modified April 6, 2022. https://towardsdatascience.com/principal-component-regression-clearly-explained-and-implemented-608471530a2f.

Levesque Karen et al. *Vocational Education in the United States: Toward the Year 2000*. National Centre for Education Statistics, US Department of Education (2000).

Lim, Stephen S et al. “Measuring human capital: a systematic analysis of 195 countries and territories, 1990-2016.” *The Lancet* 392, no. 10154 (October 2018), 1217-34.

Lund, Henrik Brynthe and Asbjorn Karlsen. "The importance of vocational education institutions in manufacturing regions: adding content to a broad definition of regional innovation systems." *Industry and Innovation* 27, no. 6 (2020), 660-679.

Ma, Jennifer and Sandy Baum. “Trends in Community Colleges: Enrollment, Prices, Student Debt, and Completion.” *Research Brief, College Board Research* (Apr 2016), 1-23.

Malik, Sakshi. “Macroeconomic Determinants of Innovation: Evidence from Asian Countries.” *Global Business Review* (Jan 2020), 1-15.

Mankiw, N. Gregory, David Romer and David N. Weil. “A Contribution to the Empirics of Economic Growth.” *Quarterly Journal of Economics* 107, no. 2 (May 1992), 407-33.

Massy, William F. “Principal Components Regression in Exploratory Statistical Research.” *Journal of the American Statistical Association* 60, no. 309 (Mar 1965), 234-56.

McKenzie, David. “When should you cluster standard errors? New wisdom from the econometrics oracle.” World Bank Blogs. Last modified Oct 16, 2017. https://blogs.worldbank.org/impactevaluations/when-should-you-cluster-standard-errors-new-wisdom-econometrics-oracle.

Minitab Blog. “Enough if Enough! Handling Multicollinearity in Regression Analysis.” Last modified April 16, 2013. https://blog.minitab.com/en/understanding-statistics/handling-multicollinearity-in-regression-analysis.

Mowery, David C., Richard R. Nelson, Bhaven N. Sampat and Arvids A. Ziedonis. “The growth of patenting and licensing by U.S. universities: an assessment of the effects of the Bayh-Dole act of 1980.” *Research Policy* 30 (2001), 99-119.

National Centre for Education Statistics. “Classification of Instructional Programs: 2000.” Career and Technical Education (CTE) Statistics. Last modified 2002. https://nces.ed.gov/surveys/ctes/tables/postsec_tax.asp.

National Centre for Education Statistics. “Postsecondary Taxonomy.” Career and Technical Education (CTE) Statistics. Accessed June, 2022. https://nces.ed.gov/surveys/ctes/tables/glossary_college.asp.

Ogden, William R. "Vocational Education: A Historical Perspective." *The High School Journal* 73, no. 4 (Apr-May 1990), 245-51.

Osborne, Jason W. and Elaine Waters. "Four assumptions of multiple regression that researchers should always test." *Practical Assessment, Research, and Evaluation* 8, no. 2 (2002), 1-5.

Park, Hun Myong. "Practical Guides to Panel Data Modelling: A Step by Step Analysis using Stata." *Public Management & Policy Analysis Program, International University of Japan* (Oct 2011), 1-51.

Porter, Michael E. "Clusters and the New Economics of Competition." *Harvard Business Review*, Dec 1998. https://hbr.org/1998/11/clusters-and-the-new-economics-of-competition.

Semple, Nathan M. "Vocational Education: The Missing Link?" *Peabody Journal of Education* 63, no. 2 (Winter 1986), 70-102.

Sheytanova, Teodora. "The Accuracy of the Hausman Test in Panel Data: a Monte Carlo Study." *Advanced level thesis 1, Örebro University School of Business* (2014), 1-45.

Sutton, John. *Technology and Market Structure: Theory and History*. Boston: MIT Press, 1998.

The Economist Intelligence Unit, *Closing the Skills Gap – companies and colleges collaborating for change* (London: The Economist Intelligence Unit Limited, 2014).

Torres-Reyna, Oscar. "Panel Data Analysis: Fixed and Random Effects using Stata." Princeton University. Last modified Dec 2007. https://www.princeton.edu/~otorres/Panel101.pdf.

Vandenbussche, Jerome, Philippe Aghion and Costas Meghir. "Growth, Distance to Frontier and Composition of Human Capital." *Journal of Economic Growth* 11, no. 2 (Jun 2006), 1-40.

Wei Chi and Xiaoye Qian. "The role of education in regional innovation activities: spatial evidence from China." *Journal of the Asia Pacific Economy* 15, no. 4 (Nov 2010), 296-419.