Appendix C: Innovation Index Definitions, Calculations and Models

Innovation Measures and Variable Definitions

Listed below are the concepts and variables used to construct the Portfolio Innovation Index. At the time of the analysis, 2006 was the latest year available (lya) for most time series. In all cases, researchers used the latest year available for all concepts and data series for this analysis and report. In the future, the innovation website will be updated on a regular basis with the most current data available for each concept or data.

The concepts and variables are grouped by sub-index:

Input: Human Capital (HC) Sub-Index

Annual Average Population Growth Rate for Ages 25-44 – *The rate of growth in county population for the 25-44 age group from 1997 to 2006 for this analysis.* The latest year available—*lya*—for this report was 2006. **Source: Census.**

$$popgroma_i = \frac{\ln(MAPOP_{t=lya}) - \ln(MAPOP_{t=1997})}{lya - 1997}$$

MAPOP = Mid-Aged Population (ages 25-44)

Educational Attainment – Higher educational levels in a population contribute to innovation by providing needed skills and knowledge. Higher educated workers are also more mobile both geographically and occupationally. **Source: Census.**

This indicator is broken into two sub-indicators:

i. ***Some college and associate's degrees*** – Even some college education can contribute to an increased level of skills and knowledge and contribute to a region's capacity to innovate. This measure may be more relevant than the bachelor's degree in more rural areas. This is measured as *percent of the population ages 25 to 64 with some college or an associate's degree.* The 2000 Census is the latest year available for these data.

$$PERASSOC_{i,t=lya} = \frac{ASSOC}{POPlya}$$

ASSOC = Number in Population with Some College or Associate's Degree, ages 25 to 64, lya

POPlya = Population in 2000—the lya for ASSOC–ages 25 to 64

ii. ***Bachelor's degree or higher*** *–Percent of the population ages 25 to 64 with a bachelor's degree or higher*. The 2000 Census is the latest year available for these data.

$$PERBACH_{i,t=lya} = \frac{BACH}{POPlya}$$

BACH = Number in Population with a Bachelor's Degree or Higher, ages 25 to 64, lya

POPlya = Population in 2000—the lya for BACH—ages 25 to 64

Technology-Based Knowledge Occupation Cluster – The technology-based knowledge occupation clusters (KOC) replace the creative class occupational component in earlier version of the innovation index. The KOC includes the following clusters: information technology; engineering; health care and medical science practitioners and scientists; math/stats/accounting; natural science and environmental management; and postsecondary education and knowledge creation. In contrast to the occupations that compose the creative class, the KOC does not include artists, musicians, or designers. The KOC indicator presents a share of technology-based cluster jobs for the year 2007, the latest year available. **Source: EMSI.**

$$KOC_{i,t=lya} = \frac{KOEMPlya}{TOTEMPlya}$$

KOEMPlya = Number of Technology-Based Knowledge Occupation Employment, lya, 2007 in this case*

TOTEMPlya = Total Employment (EMSI Definition), lya, 2007 in this case

* Requires aggregation of the six technology-based knowledge occupation clusters.

High-Tech Employment Share – Firms requiring a highly skilled and specialized workforce contribute to innovation in a county by providing a resource for workers, other firms and other industries. (This metric measures the point-in-time innovative capacity of the region as opposed to the growth of innovative capacity in the productivity and employment index.) High-tech employment, derived from a NAICS-based definition by Moody's, measures an aggregation of employment in key sectors (e.g., telecommunications, Internet providers, scientific laboratories) as the *average high-tech employment share of total employment from 1997 to 2006, the latest year available.* **Source: Moody's.**

$$avghtshare_i = \frac{\sum_{1997}^{lya} HTE_t}{\sum_{1997}^{lya} MdyTotEmp_t}$$

HTE = High Tech Employment in Year t

MdyTotEmp = Moody's Estimated Total Employment in Year t

Input: Economic Dynamics (ED) Sub-Index

The second input index measures *local* resources available to county entrepreneurs and businesses that encourage innovation close to home, or are limiting in the absence of such resources. For instance, direct investment from venture capital firms may provide the infusion of funding to quickly expand an operation. Likewise, Internet availability enables firms and entrepreneurs to learn new techniques based on best practices or communicate more effectively with researchers and innovators located in other geographic regions.

Average Venture Capital – Venture capital provides a source of funds to launch new ideas or expand innovative companies. Because the absolute volume of VC can vary widely, it is adjusted to reflect the relative size of a county's economy. **Sources: Decision Data Resources and Moody's.**

$$avgVCGDP_i = \frac{\sum_{2000}^{lya} VC_t}{\sum_{2000}^{lya} cuGDP_t}$$

VC = Total Venture Capital in Year t

cuGDP = Current-dollar County GDP in Year t

Average Private R&D – Research and development funds provides the resources for companies to launch new ideas or expand innovative companies. Because the absolute volume of R&D can vary widely, it is adjusted to reflect the relative size of a county's labor force. **Sources: Decision Data Resources and BEA.**

$$avgRDpCOMP_i = \frac{\sum_{1998}^{lya} RD_t}{\sum_{1998}^{lya} COMP_t}$$

RD = Total Research & Development Funds in Year t

COMP = Total Worker Compensation in Year t

Broadband Density and Penetration – Innovation and knowledge are linked to widespread Internet usage for individuals and businesses. This indicator measures the relative density of broadband providers available to residents in a given county, which serves as a proxy for broadband penetration that would be better measured by the number of business and residential broadband customers. This indicator is presented as the *population-weighted mean of broadband service providers available per county translated from population-weighted* ZIP *code data from 2000 to 2007, the latest year available.* Thus, this indicator transforms the FCC ZIP-code level data by weighting the number of broadband providers by population and aggregating it to county-level data. **Sources: IBRC, FCC, and Census.**

(1) $$bb\_lya_i = \frac{WT_{t=lya}}{POPEST_{t=lya}}$$

WT = Broadband Weighting Factor for the lya, in this case, 2007

POPEST = Broadband Population Estimate for the lya, in this case, 2007

(2) $$\boldsymbol{bbd_i = \frac{\ln\left(\frac{WT_{t=lya}}{POPEST_{t=lya}}\right) - \ln\left(\frac{WT_{t=2000}}{POPEST_{t=2000}}\right)}{lya - 2000}}$$

WT = Broadband Weighting Factor for Year t

POPEST = Broadband Population Estimate for Year t

Establishment Churn – Innovative and efficient companies replace outdated firms, or those firms unable to modernize techniques and processes. Average churn measures the creative destruction of a constantly transforming economy by taking *total establishment births and deaths, and expansions and contractions, relative to the total number of firms in a county for all years available.* **Source: Census.**

$$\boldsymbol{avgchurn_i = \frac{\sum_{1999}^{lya}(Birth + Death + Expansion + Contraction)_t}{\sum_{1999}^{lya}(Death + Expansion + Contraction + Constant)_t}}$$

Birth = Establishment Births in Year t

Deaths = Establishment Deaths in Year t

Expansion = Establishment Expansions in Year t

Contraction = Establishment Contractions in Year t

Constant = Establishment Constants in Year t

Establishment Sizes –

1. *Average Small Establishments per 10,000 Workers* from 1997 to 2006, the latest year available. **Sources: Census and BEA.**

$$\boldsymbol{smestpw_i = \frac{\sum_{1997}^{lya} SMALL_t}{\sum_{1997}^{lya} TOTEMP_t}}$$

SMALL = Small Establishments with Less than 20 Employees for Year t from County Business Patterns

TOTEMP = BEA Total Employment in Ten Thousands for Year t

2. *Average Large Establishments per 10,000 Workers* from 1997 to 2006, the latest year available. **Sources: Census and BEA.**

$$\boldsymbol{lgestpw_i = \frac{\sum_{1997}^{lya} LARGE_t}{\sum_{1997}^{lya} TOTEMP_t}}$$

LARGE = Large Establishments with More than 500 Employees for Year t from County Business Patterns

TOTEMP = BEA Total Employment in Ten Thousands for Year t

State Context (SC)

The third index measures state innovation resources available to entrepreneurs and businesses. These resources may not necessarily be used by all businesses, but their proximity and availability provide access to innovation capacity.

S&E Graduates from State Institutions per 1,000 – The number of graduates from science and engineering programs within a given state increases the supply of individuals trained to meet growing demands on the skilled labor force. This measure is the *S&E graduates in the state (or states if a region crosses state boundaries) per 1,000 members of the population.* **Source: Census and National Science Foundation.**

$$SEGRADS_i = \frac{SEGRAD_{lya}}{TOTPOP_{lya}}$$

SEGRAD = Number of Science and Engineering Graduates—Bachelor's and Advanced Degrees—for the latest year available

TOTPOP = Total Population in Thousands for the latest year available

R&D spending per capita – *Total per capita spending by universities and private firms by state (or states if a region crosses state boundaries).* **Source: Census and National Science Foundation**

$$TOTRD_i = \frac{STATERD_{lya}}{TOTPOP_{lya}}$$

STATERD = Research and Development Expenditures by Universities and Private Firms—for the latest year available

TOTPOP = Total Population for the latest year available

Output: Productivity & Employment (PE) Sub-Index

These output indicators measure economic improvement, regional desirability, or are the direct outcomes of innovation. They suggest the extent to which local and regional economies are moving up the value chain, creating an attractive environment for living or are direct consequences of innovation. Innovative economies will attract people seeking particular jobs, exhibit growth in productivity, and contribute new products to the marketplace.

Job Growth – *Change in BEA employment divided by the change in population from 1997 to 2006, the latest year available.* The conditional nature of the equation provides for the fact that a county or region may have growing employment but a declining population, which would be considered a positive outcome. **Sources: BEA and Census.**

$$\boldsymbol{jobpop_i} = \begin{Bmatrix} \text{IF}\left(\boldsymbol{TOTEMP_{t=lya}} - \boldsymbol{TOTEMP_{t=1997}}\right) > (0)\ \text{ANDIF} \\ \left(\boldsymbol{TOTPOP_{t=lya}} - \boldsymbol{TOTPOP_{t=lya}}\right) < (0) \\ \text{THEN}\left(\text{ABS}\left(\frac{\left(\boldsymbol{TOTEMP_{t=lya}} - \boldsymbol{TOTEMP_{t=1997}}\right)}{\left(\boldsymbol{TOTPOP_{t=lya}} - \boldsymbol{TOTPOP_{t=1997}}\right)}\right)\right) \\ \text{ELSE}\left(\frac{\left(\boldsymbol{TOTEMP_{t=lya}} - \boldsymbol{TOTEMP_{t=1997}}\right)}{\left(\boldsymbol{TOTPOP_{t=lya}} - \boldsymbol{TOTPOP_{t=1997}}\right)}\right) \end{Bmatrix}$$

TOTEMP = BEA Total Employment for Year t

TOTPOP = Population for Year t

Change in Share of High-Tech Employment – Firms requiring a highly skilled and specialized workforce are drawn to innovative areas. Growth in this sector suggests an increasing presence of innovation. High-tech employment, derived from a NAICS-based definition by Moody's, measures an aggregation of employment in key sectors (e.g., telecommunications, internet providers, scientific laboratories). The measure is calculated as *the average annual rate of change in the share of high-tech employment from 1997 to 2006, the latest year available.* **Source: Moody's.**

$$\boldsymbol{HTESd_i} = \frac{\ln\left(\frac{\boldsymbol{HTE_{t=lya}}}{\boldsymbol{MdyTotEmp_{t=lya}}}\right) - \ln\left(\frac{\boldsymbol{HTE_{t=1997}}}{\boldsymbol{MdyTotEmp_{t=1997}}}\right)}{\boldsymbol{lya - 1997}}$$

HTE = Moody's Definition of High-Tech Employment in Year t

MdyTotEmp = Moody's Estimated Total Employment in Year t

Change in Gross Domestic Product per Worker – GDP measures economic output and increases in GDP per worker measures increases in worker productivity. This measure is the *annual rate of change in current-dollar GDP per employee from 1997 to 2006, the latest year available.* **Source: BEA and Moody's.**

$$\boldsymbol{GDPWcud_i} = \frac{\ln\left(\frac{\boldsymbol{cuGDP_{t=lya}}}{\boldsymbol{TOTEMP_{t=lya}}}\right) - \ln\left(\frac{\boldsymbol{cuGDP_{t=1997}}}{\boldsymbol{TOTEMP_{t=1997}}}\right)}{\boldsymbol{lya - 1997}}$$

cuGDP = Current-Dollar County GDP from Moody's for Year t

TOTEMP = BEA Total Employment for Year t

Gross Domestic Product per Worker – GDP measures economic output per worker at a point in time. The measure is *current-dollar GDP per employee in 2006, the latest year available.* **Sources: BEA and Moody's.**

$$\boldsymbol{cuGDPW_i} = \frac{\boldsymbol{cuGDP_{t=lya}}}{\boldsymbol{TOTEMP_{t=lya}}}$$

cuGDP = Current-Dollar County GDP from Moody's for Year t

TOTEMP = BEA Total Employment for Year t

Average Patents per 1,000 Workers – New patented technologies provide an indicator of individuals' and firms' abilities to develop new technologies and remain competitive. The measure is *the number of utility patents issued per 1,000 workers for the entire time period, 1997 to 2006, the latest year available.* **Sources: Decision Data Resources and BEA.**

$$avgPatpw_i = \frac{\sum_{1997}^{lya} Patents_t}{\sum_{1997}^{lya} TOTEMP_t}$$

Patents = Total Patents Issued in Year t

TOTEMP = BEA Total Employment in Thousands for Year t

Output: Economic Well-Being (EWB) Sub-Index

Innovative economies improve the economic well-being of residents because they earn more and enjoy a higher standard of living. This is evident in lower poverty rates, greater job availability, and an economic base that increases the rewards to employees over time.

Average Poverty Rate – Innovative economies have greater employment opportunities with higher compensation, thus lowering rates of poverty. Reduced rates of poverty will tend to lag growth in employment opportunities. As a result, the last three years of the most recent data are used. In addition, a high poverty rate is a negative outcome so this measure is the inverse of the *average poverty rate from 2003 to 2005, the last three years available.* **Source: Census.**

$$avgpovR_i = \frac{\sum_{lya-2}^{lya} POV_t}{\sum_{lya-2}^{lya} POVUNIV_t}$$

POV = Total Impoverished Persons for Year t

POVUNIV = Total Population Estimate (Poverty Universe) for Year t

Average Unemployment Rates – The unemployment rate is the number of persons seeking employment as a percentage of the total labor force. The last three years of the most recent data for this series are used. In addition, a high unemployment rate is a negative outcome. As a result, this measure is the inverse of *average unemployment rate from 2005 to 2007, the last three years available*. **Source: BLS.**

$$avgunempR_i = \frac{\sum_{lya-2}^{lya} UNEMP_t}{\sum_{lya-2}^{lya} LF_t}$$

UNEMP = Number of Unemployed Persons for Year t

LF = Number of Persons in Labor Force for Year t

Average Net Migration – Total migration of all persons into a county or region serves as an indicator of whether an area is attractive to job seekers and families. Net migration is provided as *net-migration rate 2000—the year of the last Census—to 2007, the latest year available.* **Source: Census.**

$$netmigR_i = \frac{\sum_{2000}^{lya} NETMIG_t}{\sum_{2000}^{lya} TOTPOP_t}$$

NETMIG = Total Net Internal Migration for Year t

TOTPOP = Total Population for Year t

Per Capita Personal Income Growth – Personal income is the broadest measure of a person's income because it includes rental income, dividends and interest payments, in addition to salary, wages and benefits. As a result, it is probably the best measure of well-being. On the other hand, the measure is based on the location of residence, not the location of work. Thus, high personal income may or may not reflect the economic returns to innovation. This measure of well-being is the *average annual rate of change in per capita personal income from 1997 to 2006, the latest year available.* **Source: BEA.**

$$PCPId_i = \frac{\ln\left(\frac{INC_{t=lya}}{POP_{t=lya}}\right) - \ln\left(\frac{INC_{t=1997}}{POP_{t=1997}}\right)}{lya - 1997}$$

INC = BEA Personal Income for Year t

POP = BEA Population Estimate for Year t

Compensation – In contrast to personal income as reported by the BEA, compensation is measured based on the place of work. For this reason, there may be a more direct link between the employee returns to innovation and the activity itself. The sources of compensation can be the more traditional source of an employer, as well as "self-compensation" of proprietors. The compensation measures are average rates of change from the base year to the present.

i. ***Annual Wage and Salary Earnings per Worker*** - *Average annual rate of change in wage and salary earnings per worker from 1997 to 2006, the latest year available.* **Source: BEA.**

$$wspWd_i = \frac{\ln\left(\frac{WS_{t=lya}}{WSEMP_{t=lya}}\right) - \ln\left(\frac{WS_{t=1997}}{WSEMP_{t=1997}}\right)}{lya - 1997}$$

WS = BEA Wage & Salary Earnings for Year t

WSEMP = BEA Wage & Salary Employees for Year t

ii. ***Proprietors' Income per Proprietor*** - *Average annual rate of change in proprietors' income per proprietor from 1997 to 2006, the latest year available.* **Source: BEA.**

$$propincd_i = \frac{\ln\left(\frac{PRINC_{t=lya}}{PREMP_{t=lya}}\right) - \ln\left(\frac{PRINC_{t=1997}}{PREMP_{t=1997}}\right)}{lya - 1997}$$

PRINC = BEA Nonfarm Proprietors Income for Year t

PREMP = BEA Nonfarm Proprietors Employment for Year t

Portfolio Innovation Index Calculation

The five dashboard, aggregate indices are weighted as follows to produce the Portfolio Innovation Index:

- 0.3 - Human Capital (HC), an input measure
- 0.3 - Economic Dynamics (ED), an input measure
- 0.3 - Productivity and Employment (PE), an output measure
- 0.1 - Economic Well-Being (EWB), an output measure
- 0.0 - State Context (SC), a statewide measure not included in the PII

The initial calculation ("iteration 0") for county *j* is

$$\mathrm{PII}_{j,0} = 0.3(HC_j) + 0.3(ED_j) + 0.3(PE_j) + 0.1\,(EWB_j)$$

or alternatively

$$\mathrm{PII}_{j,0} = \sum_{i=1}^{4} \mathrm{A_i X}_{sj}$$

Where each X_{sj} represents a specific sub-index value and A_i represents the weight of the sub-index in the portfolio index.

Sub-indices are generally calculated as

$$X_{sj} = 100 * \sum_{i=1}^{n} \alpha_i \left(\frac{x_{ij}}{x_{iU}}\right) \ldots \alpha_n \left(\frac{x_{nj}}{x_{nU}}\right)$$

Where x_{ij} is the *i*-th variable (or measure) county j relative to the U.S. average for variable x_i. The ratio is weighted by a specified alpha for the *i*-th variable (given in column four of Table 14).

Several issues arise when attempting to construct PII_j as described above. First, negative values for several measures such as population growth rates or high-tech employment growth rates can dramatically reduce PII_j because their weighted ratios are negative. This can result in a given PII_j to be less than zero. To mitigate this effect, the entire range for a variable (or measure) was shifted upward by the absolute value of the minimum of the range. All data are shifted according to

$$x_{kj} = \{\min(x_{ij})\} + x_{ij} \text{ and } x_{kU} = \{\min(x_{ij})\} + x_{iU}$$

The value of the range-shifted variable is denoted by the subscript *k* for variable *i*. The sum of the range-shifted variables are added to unaltered x_i's for "iteration 1" of the PII for county or region *j*:

$$PII_{j,1} = 100 * \left[\left(\sum_{k=1}^{n} \alpha_i \left(\frac{x_{kj}}{x_{kU}}\right) \ldots \alpha_n \left(\frac{x_{nj}}{x_{nU}}\right)\right) + \left(\sum_{i=1}^{n} \alpha_i \left(\frac{x_{ij}}{x_{iU}}\right) \ldots \alpha_n \left(\frac{x_{nj}}{x_{nU}}\right)\right)\right]$$

While this procedure eliminates the negative values of the PII, there are extreme values for some variables or measures that can dominate or swamp the index value. The initial calculation can generate results for PII_j an order of magnitude larger than the U.S. average. Such results are usually produced by a single variable with an extreme ratio of x_{ij} to x_{iU}. Such extreme values limit the usefulness of the index. For instance, several counties in California have venture capital investments 18 times the U.S. average. To restrict the results of PII_j and narrow the distribution, limits are applied to each subset of weighted ratios using a conditional statement:

$$\theta_{ij} = \left\{\begin{array}{c} \text{IF } \left(\frac{x_{ij}}{x_{iU}}\right) > \left(\frac{x_{iU}+2\sigma_{ij}}{x_{iU}}\right) \text{THEN } \alpha_{x_i} \left(\frac{x_{iU}+2\sigma_{ij}}{x_{iU}}\right) \\ \text{ELSE IF } \left(\frac{x_{ij}}{x_{iU}}\right) < \left(\frac{x_{iU}-2\sigma_{ij}}{x_{iU}}\right) \text{THEN } \alpha_{X_1} \left(\frac{x_{iU}-2\sigma_{ij}}{x_{iU}}\right) \\ \text{ELSE } \alpha_{x_1} \left(\frac{x_{ij}}{x_{iU}}\right) \end{array}\right\} \text{(Equation 5)}$$

where θ_{ij} restricts the value of the index variable x_i to $\pm 2\sigma$ from the U.S. average for variable x_i . In other words, θ_{ij} creates a ceiling for PII_j by capping high ratios for a given variable x_i. The procedure also creates a floor for underperforming counties.

Even by applying limits of θ_{ij}, there are several extreme values of venture capital that generate an extremely large σ_{ij}. To systematically address this issue, values greater than $4\sigma_{ij}$ are omitted from the initial σ_{ij} calculation. This further restriction is only applied to distributions such as venture capital that have a substantial positive skew.

$$PII_{j,2} = 100 * \left[\left(\sum_{i=1}^{n} \theta_{ij}\right) + \left(\sum_{k=1}^{n} \theta_{kj}\right)\right]$$

Where $PII_{j,2}$ is the second iteration of the innovation index for county j. The PII_j in 86 counties were unaltered with the application of this equation. In 2,762 counties, the floor and ceiling thresholds of iteration two had the effect of increasing the county's innovation index (relative to the nation) and decreasing the index in 276 counties (relative to the nation).

Index values discussed and presented in this report are exclusively derived from the equation for $PII_{j,2}$.

Economic Growth Models for Empirical Index

Data and Variables

The county-level data are derived from numerous sources, including several proprietary sources. The public data sets include the U.S. Census Bureau (Population Estimates, County Business Patterns, TIGER/Line 2007), the U.S. Bureau of Economic Analysis, the U.S. Bureau of Labor Statistics, the Federal Communications Commission, and the U.S Department of Agriculture (Economic Research Service). Proprietary data includes estimates from Moody's economy.com, Innovation Economy 360, and Economic Modeling Specialists, Inc.

Whereas previous studies have focused primarily on states, we examine the innovative variability that occurs at the county level. Of considerable importance is the main dependent variable—GDP per worker. GDP is not estimated by the Bureau of Economic Analysis in geographic units smaller than MSAs, but models from Moody's provide estimates of county-level GDP. Based on these estimates, 97 percent of the counties in the United States experienced positive growth in GDP per worker measured in chained-dollars between 1997 and 2006.

Of the 3,111 counties included in this analysis, 2,924 had some level of utility patenting activity in the past decade. Far fewer counties, 530, benefited from recent venture capital investment and still fewer, 346 counties, received private research and development investments in the past decade. Together these two investment stream reached a combined total of 620 counties.[44] The relative concentration of these latter two investment streams in less than one-fifth of counties begs the question of the extent to which focused investment drives innovation and how much influence they have on growing an economy. If focused investments do drive innovation economies, one would expect to see a positive, statistically significant relationship between this variable and the dependent variable for counties with the presence of these major investments. Similarly, one would expect counties with patenting activity to exhibit stronger growth than others.

In addition to the inputs, we include measures of capacity and dynamics as well as a series of controls.[45] Improvements in economic conditions could be due to any of a number of circumstances and may have little to do with innovation itself. For instance, the presence of extractive industries that collect large rents on natural resources, such as gold mining or oil and gas drilling, could explain substantial growth in GDP but would hardly signify an improvement in innovation (although it may). Conversely, the cessation of significant natural resource extraction operations can adversely affect economic performance (Michaels 2007). Counties with high resource extraction were controlled for by using a modified USDA typology that identified resource-dependent counties as those with 15 percent or more of total compensation involved in resource extractive industries (USDA 2004).[46]

[44] Ideally, VC related to innovation should focus on seed capital and start-up funds as those are the investments that, while risky, would have the most immediate economic impact with relatively quick changes in productivity and profits. However, since VC is concentrated in relatively few counties, all stages of investment are considered for this analysis.

[45] Variable definitions are available in Table 48.

[46] The typology was modified to remove counties that relied heavily on low-value natural resources in such activities as sand and gravel quarrying.

The location of a county and its proximity to large urban centers is also potentially important. Previous research has identified a substantial difference between urban and rural growth due to knowledge spillovers (Jaffe et al. 1993). To acknowledge urban centers, we control for regions in metropolitan statistical areas as defined by the Office of Management and Budget (OMB) using a dummy variable, coding 1 for counties within MSAs. By definition, an MSA includes counties with substantial in-flows of workers to an urban core. Approximately one-third of U.S. counties are included in MSAs. Three-quarters of these MSA counties received investment flows. Metropolitan counties also show a statistically significant difference from non-metro counties in GDP per worker growth ($p<0.01$).

Economic growth is of primary interest, but current levels of GDP per worker can affect the extent to which growth occurs as neoclassical growth theories have indicated (Rupasingha et al. 2002). Thus, overall economic circumstances, in contrast to regional performance, were controlled for (Bergheim 2008).

Finally, regional considerations may influence growth rates. For example, regionalization effects implicit in random unidentified geographic characteristics such as weather or amenities and economic patterns like cost of living can also affect economic performance. Regionalization was specified using the U.S. Census Bureau's nine geographic divisions: Pacific, Mountain, West North Central, East North Central, West South Central, East South Central, Middle Atlantic, South Atlantic, and New England (Census 2008). Descriptive characteristics for variables included in the model are shown in Table 48 and Table 49.

Regression Results

The research team used the Ordinary Least Squares (OLS) procedure to test a series of competing standardized estimates that test hypotheses on our input variables. First we test the standardized input variables in the full extent as our preferred model, that is, unmodified from their original conception save for normalization. Second, we test the inputs using dummy variables as to determine if their mere presence was enough to influence economic growth. For the first specification standardized estimates are derived for three separate but related dependent variable growth measures from 1997 to 2006: GDP per worker (Model 1), PCPI (Model 3), and compensation per worker (Model 4). The second specification using dummy variables is shown only for GDP per worker (Model 2).

Outliers of the dependent variable ($\pm 2\sigma$) were removed in each specification. This procedure accounted for the removal of 135 extreme growth or loss counties in the GDP per worker specification. Mean GDP per worker growth was only marginally altered, but the procedure did remove all but 14 cases of negative growth (which were dispersed throughout the country). In addition to removing outliers on the dependent variables, exceedingly influential cases on the standardized independent variables were systematically omitted by calculating leverage (h_i).[47] The removal of leverage outliers varied based on dependent variables and model specification.

Summary results are provided in Table 50 and Table 51. The first model with full input values bears an adjusted R-squared of 0.189. Adding regional controls boosts the adjusted R-squared to 0.271 but generates several substantial changes in t-scores. In Model 1A, most of the human capital measures are significant and positive; however, the share of technology-based knowledge occupation workers is significant and negative

[47] The identification of leverage points and removal was based on $h_i > 3(k+1/n)$ where k is the number of variables and n is the number of remaining observations. For discussion of high leverage points, see Belsley et al. 2004; Seaver and Triantis 1995; Martin 1992; Rousseeuw and van Zomeren 1990.

implying that larger shares of knowledge occupations may not always generate increased economic growth, *ceteris paribus*. The economic dynamics measure of establishment churn appears to have a statistically significant negative impact. Further analysis of this measure shows churn to yield a positive relationship in metropolitan counties and a significant negative in non-metro counties. This implies that the notion of churn as an economic driver may be more appropriate for larger economies whereas it is detrimental to smaller, less diverse, or developing economies.

The innovation input measures, while all significant, yield mixed and conflicting results. VC investment is positive with a marginal beta coefficient, whereas patent and R&D investment generated negative relationships.[48] For each, the operationalization may prove limiting to the magnitude and direction of the relationships. There may also be a more systematic problem with the coding of the R&D variable, which relied heavily on the location of a firm's headquarters and not the location of R&D activity. Additionally, annualized estimates with appropriate lags may lead to better results.

In Model 1B, the addition of controls for regionalization suggests that on the whole several regions—East South Central and West South Central—grew substantially more quickly than others relative to the omitted New England region. The two South Central regions are of particular interest given that Hurricane Katrina struck the coastal counties in mid-2005 and may have led to an influx of federal development funds and disaster relief that boosted GDP at the tail end of the measured period (Bergheim 2008). Disaster relief funds are as of yet uncontrolled for.[49] For the other independent variables, the relationships are similar to those from Model 1A; however, bachelor's degrees, patenting activity, and broadband access rates all become insignificant presumably due to regional concentrations of these factors.[50]

The second series of models re-coded the variables for targeted investments (R&D and VC) and patent activity as dummy variables so that any county with the presence of the variable received a one, regardless of the concentration. Both specifications of this model, with and without regional controls, generate results similar to those from Model 1 with one noticeable exception. VC funds appear to be highly sensitive to the binary (or dummy variable) recoding and produce nearly significant negative results, whereas R&D investment is less sensitive to the re-coding. A possible explanation is that VC is only efficacious in larger, more sizeable amounts.

Alternate specifications of the dependent variable using PCPI and wage and salary growth (see Table 51) generate similar results. The obvious differences between the models are that with PCPI, establishment churn becomes positively significant suggesting that higher rates of churn are positive for residence-based income measures. Additionally, technology-based knowledge occupation cluster shares—the human capital measure deemed questionable in the GDP per worker growth model—generate a positive and statistically significant

---

[48] The idea that patent activity is negative is counterintuitive but may be consistent with an argument made by Heller and Eisenberg (1998) in the field of biomedical research, namely that increased patent activity in the long-term results in decreased innovation and the underutilization of resources as a result of more stringent intellectual property right enforcement mechanisms (see also Crepon and Duguet 1998). The negative relationship may also be the result of mis-specified patent data. As explained in the text of the main part of the report, patents may or may not be properly assigned to location of the research activity depending upon whether the assignee lives and works in the same county.

[49] We requested county-level data from the Department of Homeland Security's Federal and Emergency Management Agency but were referred to a website providing information at a level of little value in addressing the question at hand.

[50] Future avenues of research include testing for spatial autocorrelation. If spatial dependence is evident in the highly granular county-level data, then future research would employ spatial econometrics to analyze the most important influences of innovation.

coefficient in both alternate specifications. Regardless of the dependent variable, educational attainment, high-tech employment, and venture capital are consistently statistically significant and positive.

Table 48: Summary Statistics for Model Inputs

				All Counties				
Variable	Years	Label	U.S.	Mean	Median	Std	Min	Max
Mid-aged population growth rate, ages 25-44	1997-2006	popgroma	-0.2%	-0.7%	-0.7%	2.1%	-22.2%	9.4%
Percent of population ages 25-64 with some college or an associate's degree	2000	Perassoc	29.5%	29.1%	29.1%	6.2%	11.3%	47.2%
Percent of population ages 25-64 with a bachelor's degree	2000	Perbach	26.5%	18.0%	16.2%	8.2%	4.9%	64.0%
Average high-tech employment share	1997-2006	avghtshare	4.8%	2.9%	2.3%	2.5%	0.1%	51.2%
Average venture capital investment per $10,000 GDP	1997-2006	avgVCGDP	35.2	4.2	0.0	25.2	0.0	648.5
Average private research & development per $1,000 compensation	1997-2006	avgRDpCOMP	2.3	3.0	0.0	28.2	0.0	1081.7
Average patents per 10,000 workers	1997-2006	avgPatpw	18.8	4.0	1.8	7.3	0.0	101.2
Change in broadband density	2000-2007	bbd	16%	21%	21%	6%	0%	88%
Average establishment churn	1999-2004	avgchurn	0.80	0.74	0.74	0.06	0.42	0.96
Average small establishments per 10,000 workers	1997-2004	smestpw	364	412	400	101	36	1,176
Change in high-tech employment share	1997-2006	HTSd	-0.7%	0.0%	-0.2%	4.0%	-21.7%	33.1%
Gross domestic product per worker, current $	2006	cuGDPW	73,989	58,976	57,119	20,831	3,314	622,632
Technology-based knowledge occupations	1997-2007	KOC	1.0	0.6	0.6	0.3	0.1	3.9

Source: IBRC

Table 49: Summary Statistics for Dependent Variables

Variable	Label	U.S.	All Counties: Mean	Median	Std	Min	Max
Average annual rate of change in GDP per worker, current $, 1997 to 2006	GDPWcud	3.6%	3.4%	3.4%	1.9%	-25.2%	13.9%
Change in nominal wage and salary compensation per worker, 1997 to 2006	wspWd	3.8%	3.5%	3.5%	0.8%	-4.8%	9.8%
Change in nominal per capita personal income, 1997 to 2006	PCPId	4.1%	3.8%	3.7%	1.1%	-1.7%	15.7%

Source: IBRC

Table 50: Estimation Results for GDP per Worker Growth, 1997-2006

Label	Model 1A Std Beta (t-value)	Sig	Model 1B Stds Beta (t-value)	Sig	Model 2A Std Beta (t-value)	Sig	Model 2B Std Beta (t-value)	Sig
Intercept	0.0508 (12.18)	***	0.0509 (11.45)	***	0.051 (11.63)	***	0.0507 (11.24)	***
Perbach	0.0578 (2.21)	**	0.0357 (1.42)		0.0685 (2.56)	**	0.0375 (1.48)	*
Perassoc	0.0812 (4.19)	***	0.0993 (4.54)	***	0.0774 (4.01)	***	0.1044 (4.85)	***
KOC	-0.1012 (-3.99)	***	-0.0505 (-2.05)	**	-0.0904 (-3.43)	***	-0.043 (-1.7)	*
HTSd	0.1371 (8.02)	***	0.1389 (8.61)	***	0.13 (7.59)	***	0.1371 (8.5)	***
popgroma	0.0478 (2.05)	**	0.055 (2.51)	**	0.0464 (1.97)	**	0.056 (2.56)	**
avgchurn	-0.0917 (-3.87)	***	-0.1114 (-4.64)	***	-0.0781 (-3.27)	***	-0.1044 (-4.31)	***
smestpw	0.1128 (6.06)	***	0.157 (8.64)	***	0.1101 (5.85)	***	0.1507 (8.22)	***
avgPatpw	-0.0515 (-2.44)	**	-0.0182 (-0.89)					
patdumb: Average patents per 10,000 workers dummy					-0.0324 (-1.86)	*	-0.0067 (-0.4)	

Label	Model 1A Std Beta (t-value)	Model 1A Sig	Model 1B Stds Beta (t-value)	Model 1B Sig	Model 2A Std Beta (t-value)	Model 2A Sig	Model 2B Std Beta (t-value)	Model 2B Sig
avgRDpCOMP	-0.0342 (-1.89)	*	-0.0341 (-1.98)	**				
rddumb: Average private research & development per $1,000 compensation dummy					-0.0478 (-2.22)	**	-0.0172 (-0.84)	
avgVCGDP	0.0355 (1.81)	*	0.0349 (1.88)	*				
vcdumb: Average venture capital investment per $10,000 GDP dummy					-0.0261 (-1.15)		-0.0071 (-0.33)	
bbd	0.0281 (1.6)	*	-0.0109 (-0.63)		0.0285 (1.6)	*	-0.0125 (-0.72)	
oilgas: presence of fossil fuel extraction	0.0586 (3.38)	***	0.0108 (0.65)		0.0651 (3.75)	***	0.0144 (0.86)	
cuGDPW	0.3869 (20.63)	***	0.3632 (19.63)	***	0.3942 (20.63)	***	0.363 (19.4)	***
Metro: metropolitan statistical area county	0.1307 (6.27)	***	0.1345 (6.73)	***	0.1265 (5.99)	***	0.1288 (6.39)	***
ENC: East North Central Census division			0.0267 (0.57)				0.0097 (0.21)	
ESC: East South Central			0.0849 (2.05)	**			0.0741 (1.82)	*
MA: Middle Atlantic			0.011 (0.38)				0.0003 (0.01)	
MT: Mountain			0.0553 (1.54)	*			0.0458 (1.29)	
PAC: Pacific			0.0442 (1.44)	*			0.0347 (1.15)	
SA: South Atlantic			0.0214 (0.45)				0.0097 (0.21)	
WNC: West North Central			0.0764 (1.65)	*			0.0615 (1.35)	
WSC: West South Central			0.3386 (7.7)	***			0.3278 (7.54)	***

Label	Model 1A		Model 1B		Model 2A		Model 2B	
	Std Beta (t-value)	Sig	Stds Beta (t-value)	Sig	Std Beta (t-value)	Sig	Std Beta (t-value)	Sig
N	2,859		2,908		2,868		2,933	
Adj-R^2	0.1893		0.2711		0.1834		0.2661	

*** < .01; ** < .05; * < .15

Note: The variance inflation factor (VIF) is a method to test the severity of multicollinearity. Except for the regional dummy variables, no VIF score exceeded 2.5. Only "Perbach" and "avgchurn" exceeded 2 in model 1B and it should not surprise that the VIF for the regional variables ranged from 3.2 to 8.9. The VIF results for model 2b were similar, with KOC joining Perbach and avgchurn in the range of 2.0 to 2.5.
Source: IBRC

Table 51: Estimation Results for Alternate Dependent Variable Measures

Label	Change in Per Capita Personal Income				Change in Wage and Salary Compensation			
	Model 3A		Model 3B		Model 4A		Model 4B	
	Std Beta (t-value)	Sig	Std Beta (t-value)	Sig	Std Beta (t-value)	Sig	Std Beta (t-value)	Sig
Intercept	0.0094 (3.62)	***	0.0185 (6.64)	***	0.0343 (15.5)	***	0.0427 (17.89)	***
Perbach	0.3164 (10.56)	***	0.286 (9.62)	***	0.1948 (7.39)	***	0.2208 (8.46)	***
Perassoc	-0.0339 (-1.71)	*	0.0561 (2.47)	**	0.0906 (4.57)	***	0.1287 (5.6)	***
KOC	0.0644 (2.55)	**	0.0547 (2.19)	**	0.1109 (4.21)	***	0.1056 (4.04)	***
HTSd	0.0367 (2.14)	**	0.0354 (2.14)	**	0.0789 (4.53)	***	0.0854 (5.05)	***
popgroma	-0.2318 (-9.98)	***	-0.2371 (-10.67)	***	0.1156 (4.8)	***	0.1039 (4.51)	***
avgchurn	0.2527 (10.79)	***	0.2019 (8.4)	***	-0.0012 (-0.05)		-0.0806 (-3.18)	***
smestpw	0.0923 (4.93)	***	0.1002 (5.4)	***	-0.0476 (-2.43)	**	-0.0227 (-1.15)	
avgpatpw	-0.0543 (-2.51)	**	-0.0432 (-2.04)	**	-0.0969 (-4.43)	***	-0.0582 (-2.66)	***
avgRDpCOMP	0.0131 (0.72)		0.0265 (1.5)	*	0.063 (3.4)	***	0.0732 (4.02)	***
avgVCGDP	0.0474 (2.43)	**	0.0386 (2.03)	**	0.1173 (5.82)	***	0.0751 (3.82)	***
bbd	-0.0534 (-3.03)	***	-0.0704 (-4)	***	-0.0454 (-2.54)	**	-0.0591 (-3.3)	***

Label	Change in Per Capita Personal Income				Change in Wage and Salary Compensation			
	Model 3A		Model 3B		Model 4A		Model 4B	
	Std Beta (t-value)	Sig	Std Beta (t-value)	Sig	Std Beta (t-value)	Sig	Std Beta (t-value)	Sig
oilgas: presence of fossil fuel extraction	0.0929 (5.43)	***	0.0766 (4.58)	***	0.137 (7.85)	***	0.1388 (7.99)	***
Pcpi	-0.3651 (-13.86)	***	-0.3093 (-11.7)	***				
Ws					-0.4309 (-17.59)	***	-0.374 (-14.9)	***
Metro: metropolitan statistical area county	0.1069 (4.92)	***	0.0942 (4.48)	***	0.0907 (4.21)	***	0.0758 (3.59)	***
ENC: East North Central Census division			-0.2721 (-5.72)	***			-0.2069 (-4.36)	***
ESC: East South Central			-0.0723 (-1.7)	*			0.0131 (0.31)	
MA: Middle Atlantic			-0.0484 (-1.67)	*			-0.0909 (-3.11)	***
MT: Mountain			-0.0771 (-2.1)	**			-0.0209 (-0.55)	
PAC: Pacific			-0.1367 (-4.46)	***			-0.0068 (-0.22)	
SA: South Atlantic			-0.1398 (-2.91)	***			-0.0144 (-0.3)	
WNC: West North Central			-0.2638 (-5.67)	***			-0.1393 (-2.92)	***
WSC: West South Central			-0.0029 (-0.06)				0.0371 (0.82)	
N	2,856		2,901		2,872		2,930	
Adj-R^2	0.1809		0.2351		0.158		0.1971	

*** < .01; ** < .05; * < .15

Source: IBRC

Table 52: Reweighted Betas from Preferred Empirical Model

Variable	**Standardized βxᵢ's** for significant, positive variables	**αxᵢ**: Reweighted **standardized βxᵢ's** (sum to 1.0)
Perbach: Percent of population, ages 25-64, with bachelor's degrees	0.058	0.12
Perassoc: Percent of population, ages 25-64, with some college or an associate's degree	0.081	0.16
popgroma: Population growth rate for ages 25-44	0.048	0.10
HTSd: Change in high-tech employment share	0.137	0.27
smestpw: Average small establishments per 10,000 workers	0.113	0.23
avgVCGDP: Average venture capital investment per $10,000 GDP	0.036	0.07
bbd: Change in broadband density	0.028	0.06

Note: Negative values (in red) in Table 50 were not included in the empirical index.
Source: IBRC

References

Belsley, D. A., D. Kuh, and R. E. Welsch. 2004. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Bergheim, S. 2008. The dependent variable: GDP growth. In *Long-Run Growth Forecasting,* 36-41. Berlin Heidelberg: Springer.

Crepon, B., and E. Duguet. 1997. Research and development, competition and innovation pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity. *Journal of Econometrics* 79 (2): 355-378.

Heller, M. A., and R. S. Eisenberg. 1998. Can patents deter innovation? The anticommons in biomedical research. *Science* 280 (5364): 698-701.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 103 (3): 577-598.

Martin, R. J. 1992. Leverage, influence and residuals in regression models when observations are correlated. *Communications in Statistics—Theory and Methods* 21 (5): 1183-1212.

Michaels, G. 2007. Long-term consequences of resource based specialization. Discussion Paper, Center for Economic Performance, London School of Economics.

Rousseeuw, P. J., and B. C. van Zomeren. 1990. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association* 85 (411): 633-651.

Rupasingha, A., S. Goetz, and D. Freshwater. 2002. Social and institutional factors as determinants of economic growth: Evidence from the United States counties. *Papers in Regional Science* 81: 139-155.

Seaver, B. L., and K. P. Triantis. 1995. The impact of outliers and leverage points for technical efficiency measurement using high breakdown procedures. *Management Science* 41 (6): 937-956.

U.S. Census Bureau. 2008. U.S. Census Regions and Divisions. www.census.gov/geo/www/us_regdiv.pdf.

USDA. 2004. Mining typology codes. Economic Research Service. www.ers.usda.gov/briefing/rurality/Typology/maps/Mining.htm.