Decomposing regional patenting rates: how the composition factor confounds the rate factor

Timothy R. Wojan\textsuperscript{a*}, Kathryn R. Dotzel\textsuperscript{b} and Sarah A. Low\textsuperscript{a}

\textsuperscript{a}Economic Research Service, Washington, DC, USA; \textsuperscript{b}Department of Agricultural, Environmental, and Development Economics, The Ohio State University, Columbus, OH, USA

(Received 14 April 2015; accepted 14 September 2015)

Patents per capita is a widely used innovation indicator. Rural areas generally perform very poorly using this metric, suggesting that inventive activity that leads to patents is an urban phenomenon. However, newly available inventor-disambiguated patenting data demonstrate that inventions per inventor are roughly equal across urban and rural areas. A critical assessment of the patents-per-capita measure questions its construct validity. An alternative measure is constructed that empirically identifies a plausible ‘inventive class’ and does not confound the patenting rate with irrelevant information. This allows the decomposition of overall patenting rates into a compositional factor and a rate factor which leads to a more meaningful regional comparison of patenting productivity.

\textbf{Keywords:} patents; innovation indicators; decomposition; standardization; urban–rural comparisons

\section*{Introduction}

Conventional wisdom holds that innovation and invention are predominantly urban phenomena – concentrated most heavily in global city agglomerations – and rare or idiosyncratic in rural areas (Carlino & Kerr, 2014; World Bank, 2009). Patents per capita rates are highest in urban areas; however, recently available data that track individual inventors and their locations provide evidence that seemingly contradicts this wisdom: patenting rates per inventor in rural areas are roughly equal to those of urban areas. These prolific rural inventors raise important questions about the geography of invention: Does the productivity of individual inventors inform the patent production capacity of a region? If the selection of successful inventors biases the measure of regional patenting productivity, then what is the appropriate pool of potential inventors and auxiliaries who support the patenting process? Is the convention of using population as this pool defensible?

Making sense of these seemingly incongruous data compelled a critical examination of population as the default denominator for computing patenting rates. Despite the importance attached to patents per capita as a primary indicator of a region’s innovative capacity (Carlino, Chatterjee, & Hunt, 2007; Furman, Porter, & Stern, 2002; Galindo-Rueda, 2013; Krammer, 2009; Organisation for Economic Co-operation and Development (OECD), 2009, 2010), we were unable to find any studies that confirm the validity of the construct.\textsuperscript{1} It is somewhat ironic that a primary indicator of a region’s

\textsuperscript{*}Corresponding author. Email: twojan@ers.usda.gov

© 2015 The Author(s). Published by Taylor & Francis.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
ability to codify new ways of thinking relies on a convention of convenience. We hope to demonstrate that a meaningful comparison of cross-sectional or longitudinal patenting rates requires defining a subpopulation that plausibly contributes to patenting.

We begin by evaluating the default metrics for regional innovation/invention—patents per capita and patents per inventor—to motivate our assessment of population as a denominator and the need to search for an alternative. The fact that the patents per capita data comport with a dominant mental map of what innovation data should look like demands an explanation of why. The portfolio of places argument in the World Bank’s *Reshaping Economic Geography* (2009) provides a rational explanation that relegates lower order places to filling more routine production, service and logistical roles. This compels the disturbing follow-on question: Why is there any patenting in rural areas?

This leads us to a hybrid rational/inductive identification of the regional inventive economy that provides an alternative basis for assessing the relative patenting productivity of a region by allowing us to compute patenting rates on the subset of the population who might plausibly contribute to patent production. The competing measures are compared axiomatically and empirically to assess their relative construct validity. A method for decomposing the population-denominated patenting rate into a compositional factor and a rate factor pertaining to the inventive class provides new insights on patent indicators, stimulating further debate on this important topic.

**Inventor-disambiguated patent data**

This analysis uses a novel database covering all utility patents granted by the US Patent and Trademark Office from 1975 to 2010. It was constructed from a new data product, supported in part by the National Science Foundation (NSF), which uses a Bayesian-supervised learning approach uniquely to identify all inventors that appear on utility patents (Lai, D’Amour, Yu, Sun, & Fleming, 2013). This means that inventors can be located and tracked across space and time. Using the US Geographical Survey (USGS) Geographic Names Information System, we assign each inventor to a county based upon the city and state of the inventor’s address provided at the date of patent application. For patents with more than one inventor, we assign each author an equal fraction of that patent. County identifiers associated with individual patents allow us to construct a consistent dynamic profile of rural patenting, rural inventors and rural technologies (Toole and Low, 2013).

County-level patent data may suffer from false precision, as place of invention is defined by the inventor’s place of residence, which may differ from the county where much of the inventive work took place. This is especially likely in large urban agglomerations. To address this problem, our analysis uses commuting zone (CZ) geography, a widely used multi-county areal unit. CZs are based on county-to-county commuting flows and are defined by a hierarchical cluster algorithm that identifies counties with strong commuting ties using no minimum population threshold (Parker, 2012; Tolbert & Sizer, 1996). The strategy should resolve the great majority of place of work versus place of residence discrepancies.

**What the data show**

Patents per capita purports that a region’s entire population—very young to very old, white collar and blue collar, employees in industrial and service sectors—possess equal
patenting capacity, while use of patents per inventor restricts the plausible patenting population to those who have achieved patenting success.

Patenting rates per inventor, computed as the total number of patents produced in years 2000–05 divided by the average annual number of inventors that contributed to patent production during the same period, are displayed as the third dimension in Figure 1.

CZs are identified as either ‘urban containing a global city’, ‘urban without a global city’ or ‘predominantly rural’. The image is disconcerting for those well versed in the geography of invention (or as confounded with the geography of innovation), as the anticipated red peaks of invention in well-known global cities and the ivory valleys of rare rural and small city invention are replaced by an eerily uniform distribution of inventiveness. The few notable exceptions are limited to predominantly rural CZs and urban without a global city CZs. Two exceptions with compelling stories are the counties of Renville and Redwood in Minnesota and the Boise, Idaho CZ. Renville–Redwood counties stand out as a singular case where the inventive process is prominent and exalted in a rural community – it is the home of the Minnesota Inventors Hall of Fame and actively supports student inventors. Boise provides a more conventional story of the geographic concentration of patents – since the 1970s it has developed as a satellite of Silicon Valley (Mayer, 2009). A transplanted Hewlett-Packard facility, homegrown Micron Technology, and a growing number of high-tech firms have relied heavily on quality-of-life amenities to attract highly skilled labour. Aside from a small number of exceptions, this map suggests that the projection of an individual measure of productivity to a geographical area may poorly represent that area’s inventive capacity.

Indeed, the selection of inventors as the denominator may raise valid questions about models of the inventive process, which are supported by strong priors that inventor productivity is higher in global cities. But ‘patents per inventor’ does not inform the inventive capability of a place. Having selected successful inventors for the metric, it may represent nothing more than a quantification of anecdotal evidence of the rare rural

Figure 1. Patents per inventor by commuting zone, 2000–05.
inventor. Patenting rates calculated on a per capita basis applied to these data restore confidence in our priors on the geography of invention (Figure 2).

In Figure 2, the Silicon Valley CZ containing San Jose is returned to its point of prominence, along with other recognized high-tech centres such as San Francisco and Austin, Texas. A few of the top patenting predominantly rural CZs are home to land grant research universities such as the University of Idaho and Washington State University in Moscow/Pullman and Oklahoma State University in Stillwater. Although the map does not support a strict dominance of global city CZs over other urban CZs, it does clearly indicate that patenting in rural CZs is muted. However, the challenge to one denominator should extend to the other even if the map of the metric comports with our priors.

The validity of population as the denominator for computing patenting rates is best challenged by a thought experiment of industrial clustering run amok. Say that alongside nanotechnology cities, biotechnology cities and software cities we had retirement cities and tourism cities. Comparing patenting rates across these two groups based on population would be wholly uninteresting. Yet, in the real world, some places may have a substantial portion of their population supported by economic activities that have exceedingly low patenting rates. So should the patenting rate of Paris take a hit just because it also happens to be the leading tourism destination in the world? Tourism in Paris arguably supports a more vibrant café economy and such places may be an essential component of the invention ecosystem. If tourism employment is generally associated with higher patenting rates, we will have learned something and opened up new questions about the inventive process. If not, we should stop penalizing the patenting rates of beautiful places.

This same plea applies to much less notable places characterized by a concentration of essential economic activities with exceedingly low patenting rates. The production of food and fibre in rural areas is the most obvious example. The World Bank’s report on
Reshaping Economic Geography provides a convenient oversimplification that allocates the whole of the innovation economy, where invention is presumed to take place, to the largest cities:

Research over the last generation indicates that different forms of human settlement facilitate agglomeration economies for different forms of production. A somewhat-oversimplified (but not altogether incorrect) generalization would be that market towns facilitate scale economies in marketing and distributing agricultural produce, medium-size cities provide localization economies for manufacturing industries, and the largest cities provide diverse facilities and foster innovation in business, government, and education services. (World Bank 2009, p. 128)

A more nuanced view of the report would acknowledge that parts of the innovation economy may locate in some rural areas and medium-sized cities. From this perspective, both of the above maps distort the true inventive capacity of places.

A hybrid rational/inductive approach for identifying the inventive class

The essential problem is how to define an exemplary population for evaluating the inventive productiveness of a region. In the absence of a single compelling alternative, total population has been uncritically accepted as the valid default measure. Since indicators using population as the denominator comport with where conventional wisdom expects inventive productiveness to be highest, there has been no reason to challenge its adequacy. However, the thought experiment above demonstrates why this is a weak construct for assessing inventive productiveness. Since ‘it is not obvious what the appropriate set of occupations should be’ (Carlino et al. p. 404) the absence of a perfect denominator has dispelled the need to look for a much better denominator. Our goal is to begin this search for a better denominator that represents an inventive class that does not dilute the contribution of the part of the economy that plausibly contributes to patenting.5

A ‘blue sky’ approach to the problem that recognized occupation-specific human capital as the essential input into the patenting process would require new data that link successful inventors to their occupations. These data could be used to calculate the patenting probability of each detailed occupation on a national scale. Then the expected number of patents produced by a given occupation in any region would be equal to the product of the occupation’s patenting probability and the region’s occupational structure. The residual would be either a patenting surplus or a deficit. Using regional characteristics to

| CZ type                        | N   | Population | Number of inventors | Number of patents | Patents per capita | Patents per inventor |
|-------------------------------|-----|------------|---------------------|-------------------|--------------------|----------------------|
| Urban containing a global city| 30  | 115,586,003| 83,544              | 343,310           | 0.00297            | 4.1093               |
| Urban without a global city   | 282 | 140,403,620| 54,082              | 221,345           | 0.00158            | 4.0928               |
| Predominantly rural           | 391 | 25,097,444 | 3,254               | 13,336            | 0.00053            | 4.0983               |

Note: ‘Inventors’ are averaged for 2000–05 and used as the denominator to compute ‘Patents per inventor’.

Reshaping Economic Geography
explain that residual would provide the best estimates of factors that contribute or detract from patenting productivity.

Since these data are not currently available, our ability to compute the patenting productivities of various occupations is severely limited. The best we can hope to do with currently available data is to compare the regional occupational structure with the number of patents produced to try to isolate those occupations that are consistently associated with patenting activity. One possible empirical strategy is to begin with no priors on which occupations should be included and merely select those occupations that demonstrate a consistent statistical association with patenting using some stepwise or other data-mining protocol. However, the cost of specification errors from that approach is easily reduced by retaining all occupations for which there is a strong a priori justification for inclusion in an inventive class. The collection of occupations used by the NSF to track developments in the innovation and inventive economy would appear to meet a strong a priori justification threshold, comprised of science, engineering and technical (SET) occupations.  

The primary justification for expanding the analysis beyond the core SET occupations is the existence of non-core occupations that are directly involved in patenting. For example, ‘Art and design workers’ are not included in the SET occupations, but 40% of authors on design patents also appear on utility patents (Nichols, 2013). Carlino et al. (2007, p. 404) additionally caution that ‘a substantial amount of invention occurs when users of a product or process [outside of the class of knowledge workers] modify it to suit their particular needs’. A second group of occupations that motivates an expansion of the analysis beyond the core SET occupations are occupations that may characterize inventive ecosystems, even if they are not directly involved in patenting. While the inclusion of this second set of non-core occupations could reduce the direct connection between the resulting patenting rate using an inventive class denominator and patenting productivity, any attempt to differentiate these two non-core groups would be arbitrary.

At the root of our analysis to identify non-core SET occupations involved in patenting is a simple linear regression of aggregated 2000–05 CZ-level patent totals (Patents) on the share of the CZ’s workforce in year 2000 Census occupations and 2003 Rural–Urban Continuum Code dummy variables (RU):

$$\text{Patents} = \beta_0 + \beta_1 \text{OccSh}_1 + \cdots + \beta_{12} \text{OccSh}_{12} + \text{SET} + \text{ExclOccSh} + \text{RU} + \varepsilon \quad (1)$$

To mitigate the effects of collinearity between occupation shares, we select 20 shares to include in each of 10,000 separate regressions. Share of employment in each of eight core SET occupations (SET) is included in each regression due to their strong a priori connection to invention and 12 additional non-core occupations are selected randomly. To reduce omitted variable bias, the share of the workforce in the excluded occupations (ExclOccSh) is included in each regression, with ‘Cashiers’ chosen as the omitted occupation due to their relative ubiquity across CZs and presumably low inventive potential.

Following each regression, we update a collection of count variables that record instances of inclusion for each occupation share as well as whether each occupation share coefficient is positive and significant at the 10% level. These measures allow us to calculate the percentage of time a particular occupation share effect is positive and significant in the iterative regression analysis. To account for differences in composition of inventive class in metropolitan and non-metropolitan areas – suggested by the lack of businesses and organizations directed to formal R&D investment in rural areas – we separately analyze metro and non-metro CZs (Freshwater, 2012).
Our inventive subset inclusion criteria are as follows. Occupations associated with coefficients that are positive and significant in at least 75% of their regressions in the metro or non-metro analysis are characterized as inventive. We additionally consider occupations associated with positive and significant coefficients in at least 50% of their regressions in the metro and non-metro analysis in our inventive subset to account inventive processes that appear to be widespread, if not clearly defined. However, no occupations meet these criteria. Of the 84 non-core SET occupations included in the analysis (excluding Cashiers), we identify eight as ‘inventive’ (Table 2). All 16 occupations are used to define the inventive population in both urban and rural CZs.

The weak association between employment shares in the majority of SET occupations and patenting is somewhat surprising. ‘Engineers’, ‘Computer specialists’ and ‘Social scientists and related workers’ are the only three SET occupations that would be selected if relying on the inductive procedure alone. The occupation that was most consistently associated with patenting in both urban and rural areas is ‘Advertising, marketing, promotions, public relations and sales managers’. Support for including designers in the inventive class was also derived from both urban and rural inventing. College professors are only consistently associated with patenting in rural areas.

The occupations outside the creative class (Florida, 2002) that are consistently associated with patenting raise interesting questions about the rural patenting process – a process that has been dismissed as largely random and idiosyncratic. Some explanations are easier to come by than others. For example, customer service representative employment is highly correlated with the presence of headquarters establishments, which in turn is highly correlated with patenting. For the other occupations outside of the creative class there are clear indications from the data that they tend to concentrate in high

### Table 2. Inventive occupations.

| Occupation                                                                 | Percentage positive and significant (10% level) |
|----------------------------------------------------------------------------|-----------------------------------------------|
| **Science, engineering and technical (SET)**                               |                                              |
| Architects, surveyors and cartographers                                    | 0.00                                          |
| Computer specialists                                                      | 9.24                                          |
| Drafters, engineering and mapping technicians                               | 0.00                                          |
| Engineers                                                                  | 53.05                                         |
| Life and physical scientists                                              | 0.00                                          |
| Life, physical and social science technicians                              | 0.00                                          |
| Mathematical science occupations                                          | 0.00                                          |
| Social scientists and related workers                                     | 97.65                                         |
| **Identified**                                                             |                                              |
| Advertising, marketing, promotions, public relations and sales managers    | 100.00                                        |
| Art and design workers                                                     | 76.03                                         |
| Assemblers and fabricators                                                 | 0.00                                          |
| Customer service representatives                                           | 0.00                                          |
| Entertainers and performers                                                | 0.20                                          |
| Metal workers and plastic workers                                          | 0.00                                          |
| Postsecondary teachers                                                     | 0.00                                          |
| Printing workers                                                           | 0.00                                          |

Source: Toole and Low 2013; US Census Bureau.
patenting areas, but what connects them to patenting is not so clear. CZs that rank highly in their share of metal and plastic workers, fabricators, and printing workers also tend to rank highly in terms of total number of patents. For example, the CZ containing Warsaw, Indiana – called the ‘Orthopaedic Capital of the World’ – has the second highest number of patents produced of any rural CZ, the second highest share of metal and plastic workers, and the sixth highest share of printing workers. The regular co-location identified in the data does not yet have a conceptual explanation, but we are confident that these occupations are strongly associated with rural patenting.

Comparing patenting rates denominated by population and inventive class

The topography of the geography of invention in the patents per capita map is retained in the map of patents per inventive class member in Figure 3, but with a higher base plateau throughout and numerous eruptions of predominantly rural CZs. The map directly challenges the characterization of rural inventing as idiosyncratic and muted.

To examine the validity of the alternative measures of inventive activity more fully, we adopt the axiomatic approach to indicators made famous by Sen’s (1976) assessment of alternative poverty measures. Axiomatically, both patents per capita and patents per inventive class member will show an increase in patenting rate with a decline in the relevant denominator, ceteris paribus. In the case of patents per capita, a decline in population tells us nothing about the change in the inventive productiveness of a region. In the patents per inventive class member case, however, a decline in a region’s inventive class, ceteris paribus, signifies an increase in its inventive productiveness, as fewer potential inventors and their auxiliaries are producing the same number of patents.

Table 3 provides the information needed to compare patenting rates and population growth across global city CZs. From the axiomatic critique, we would expect those global city CZs experiencing relative population decline to fall in rank moving from the per capita measure to the alternative inventive class measure. Alternatively, CZs
Table 3. Patenting statistics for commuting zones containing global cities.

| Patents per capita rank | Global city commuting zone | Patents per capita | Patents per inventive class member | Patents per inventive class member rank | Change in rank | Population growth, 1975–2000 | Inventive class employment share |
|-------------------------|---------------------------|-------------------|-----------------------------------|----------------------------------------|---------------|-----------------------------|-------------------------------|
| 1                       | San Jose, CA              | 0.02165           | 0.19267                           | 1                                      | 0             | 45.90%                      | 0.11239                       |
| 2                       | San Francisco, CA         | 0.00614           | 0.07456                           | 2                                      | 0             | 38.44%                      | 0.08230                       |
| 3                       | Portland, OR              | 0.00538           | 0.06863                           | 4                                      | -1            | 62.25%                      | 0.07845                       |
| 4                       | Minneapolis, MN           | 0.00523           | 0.05485                           | 8                                      | -4            | 40.58%                      | 0.09532                       |
| 5                       | Raleigh, NC               | 0.00518           | 0.05774                           | 6                                      | -1            | 83.77%                      | 0.08971                       |
| 6                       | Seattle, WA               | 0.00514           | 0.06373                           | 5                                      | 1             | 65.93%                      | 0.08059                       |
| 7                       | San Diego, CA             | 0.00491           | 0.07277                           | 3                                      | 4             | 74.86%                      | 0.06754                       |
| 8                       | Boston, MA                | 0.00467           | 0.05597                           | 7                                      | 1             | 11.20%                      | 0.08344                       |
| 9                       | Denver, CO                | 0.00315           | 0.03520                           | 15                                     | -6            | 69.69%                      | 0.08940                       |
| 10                      | Detroit, MI               | 0.00312           | 0.03814                           | 10                                     | 0             | 1.50%                       | 0.08190                       |
| 11                      | Dallas, TX                | 0.00309           | 0.04121                           | 9                                      | 2             | 79.46%                      | 0.07498                       |
| 12                      | Philadelphia, PA          | 0.00246           | 0.03799                           | 11                                     | 1             | 3.09%                       | 0.06471                       |
| 13                      | Phoenix, AZ               | 0.00241           | 0.03791                           | 12                                     | 1             | 142.65%                     | 0.06350                       |
| 14                      | Houston, TX               | 0.00233           | 0.03737                           | 13                                     | 1             | 81.15%                      | 0.06246                       |
| 15                      | Cincinnati, OH            | 0.00232           | 0.03161                           | 17                                     | -2            | 17.49%                      | 0.07344                       |
| 16                      | Cleveland, OH             | 0.00232           | 0.03241                           | 16                                     | 0             | -3.02%                      | 0.07159                       |
| 17                      | Milwaukee, WI             | 0.00228           | 0.02702                           | 19                                     | -2            | 8.64%                       | 0.08445                       |
| 18                      | Chicago, IL               | 0.00202           | 0.02851                           | 18                                     | 0             | 15.42%                      | 0.07073                       |
| 19                      | Los Angeles, CA           | 0.00201           | 0.03523                           | 14                                     | 5             | 55.89%                      | 0.05703                       |
| 20                      | Atlanta, GA               | 0.00198           | 0.02598                           | 20                                     | 0             | 102.81%                     | 0.07617                       |
| 21                      | Washington, DC            | 0.00188           | 0.02070                           | 25                                     | -4            | 39.83%                      | 0.09097                       |
| 22                      | Baltimore, MD             | 0.00156           | 0.02256                           | 21                                     | 1             | 16.34%                      | 0.06907                       |
| 23                      | St. Louis, MO             | 0.00147           | 0.02162                           | 23                                     | 0             | 8.66%                       | 0.06805                       |
| 24                      | Kansas City, MO           | 0.00147           | 0.02070                           | 26                                     | -2            | 26.07%                      | 0.07089                       |
| 25                      | Columbus, OH              | 0.00130           | 0.01743                           | 28                                     | -3            | 27.95%                      | 0.07464                       |
| 26                      | Orlando, FL               | 0.00127           | 0.02237                           | 22                                     | 4             | 140.73%                     | 0.05669                       |
| 27                      | Charlotte, NC             | 0.00098           | 0.01447                           | 30                                     | -3            | 58.68%                      | 0.06770                       |
| 28                      | Tampa, FL                 | 0.00095           | 0.01741                           | 29                                     | -1            | 68.60%                      | 0.05455                       |
| 29                      | Miami, FL                 | 0.00088           | 0.02070                           | 24                                     | 5             | 63.80%                      | 0.04271                       |
| 30                      | New York, NY              | 0.00087           | 0.01863                           | 27                                     | 3             | 6.86%                       | 0.04675                       |

Note: ‘Change in rank’ denotes the difference between patents per capita rank and patents per inventive class member rank.
experiencing rapid population growth should rise in rank moving from the population-denominated to the inventive class-denominated patenting rate. This is in fact what we see. The 15 slowest growing CZs had an average fall in rank of 0.73, which is matched by an equivalent rise for the 15 fastest growing CZs. However, regions with the largest increases in rank are reminiscent of the Paris example above: centres of media and entertainment (Los Angeles and New York) and tourism and retirement centres (Orlando and Miami) perform better under the inventive class-denominated patenting rate. Confounding the rate factor with the composition factor creates the largest distortion in Miami, the global city with the smallest share of employment in the inventive class. Detroit, which ranks 10th using either denominator to compute patenting rates, is 3.54 times more productive in producing patents than Miami using the per capita measure, but only 1.84 times more productive using the inventive class measure. The ordinal change in rank is modest, but the relative change in the quantitative measure of patenting rate can be very large.

The problem of the compositional factor confounding the rate factor has long been understood in demography – perhaps because the process of childbirth is so much better understood than the process of invention. Implicit in the patent per capita measure is a ‘black box’ model – in which input factors are transformed into output – where a region’s population defines the relevant measure of input. If the black box model were extended to demography, fertility of a region would simply be the number of live births in the region divided by its population. Fertility would decline with an improvement in life expectancy and would increase with a relative decline in the young or elderly populations. By defining regional fertility as a product of the general fertility rate and compositional factors such as the proportion of women in the population and the proportion of those women who are of child-bearing age, demographers and public health analysts have the ability to compare fertility across regions with different compositions and to analyze the fertility of a single region through time as its composition changes.

The same tools of standardization and decomposition are available for the study of patenting activity, as measured by patents per capita, if an inventive class is defined. The composition investigated here was derived rationally and inductively based on occupations that are either recognized by national statistics as being at the core of the inventive economy or demonstrated a strong statistical association with patenting in the preceding analysis. Patenting rate computed using the inventive class as the denominator provides our rate factor, whereas inventive class as a share of total population provides our compositional factor:

\[
\text{Patents per capita} = \frac{\text{Patents}}{\text{Inventive class}} \times \frac{\text{Inventive class}}{\text{Total population}}
\]

Standardization tells us what observed patenting rates across populations would be if their rate (compositional) factors were identical, while decomposition answers how much of the difference in the observed patenting rates across populations can be attributed to differences in their rate (compositional) factors. Thus, by standardizing and decomposing patenting rates, we can determine how much of the difference in the population-denominated patenting rate across populations is attributable to differences in patenting productivity of the inventive class and how much is attributable to differences in the size of the inventive class relative to the total population.

Suppose we have two populations, \( i \) and \( j \), and two factors, \( \alpha \) and \( \beta \). Following Das Gupta (1993), let the observed patenting rate of population \( k \) be expressed as:
\[ R_k = \alpha_k \beta_k \]

Then, for \( k \in \{i,j\} \), the \( \alpha \)-standardized rate for population \( k \) is:

\[ \frac{\alpha_i + \alpha_j}{2} \beta_k \]

while the \( \beta \)-standardized rate for population \( k \) is:

\[ \frac{\beta_i + \beta_j}{2} \alpha_k \]

Factor effects for \( \alpha \) and \( \beta \) can be defined as:

\[ \alpha \text{-effect} = \frac{\beta_i + \beta_j}{2} (x_i - x_j) \]

\[ \beta \text{-effect} = \frac{\alpha_i + \alpha_j}{2} (\beta_i - \beta_j) \]

or as the difference in \( \beta \)- and \( \alpha \)-standardized rates for populations \( i \) and \( j \), respectively. This results in the identity:

\[ R_i - R_j = \alpha \text{-effect} + \beta \text{-effect} \]

We begin by comparing the slow and fast growing global city CZs (Table 4), distinguishing between the two types of global cities based on whether the 1975–2000 growth rate falls above or below the global city median, 43.24%. The ‘high’ growth global city population has a rate factor of 0.0506 and a compositional factor of 0.0668, while the ‘low’ growth global city population has a rate factor of 0.0360 and a compositional

| Measures                              | Standardization of patenting rates | Decomposition of effects | Per cent distribution of effects |
|---------------------------------------|------------------------------------|--------------------------|---------------------------------|
| Rate factor-standardized patenting rate | 0.00289    0.00312                     | -0.00023 (CF effect)    | -28.9568                        |
| Compositional factor-standardized patenting rate | 0.00351    0.00250                     | 0.00101 (RF effect)     | 128.9568                        |
| Observed patenting rate               | 0.00338    0.00259                     |                          |                                 |

Notes: Rate factor is patents per inventive class member and compositional factor is inventive class as a share of total commuting zone population. ‘CF effect’ denotes ‘compositional factor effect’ and ‘RF effect’ denotes ‘rate factor effect’.
Source: Toole and Low 2013.
factor of 0.0720. Thus, the population-denominated patenting rate for fast growing global cities is 30% higher than that of slow growing global cities. Patenting rates for the two populations when we standardize by the rate factor – that is, assign an identical patents per inventive class member rate to each population and allow the inventive class population share to vary between populations as it does – are 0.00289 and 0.00312, respectively. Compositional factor standardized patenting rates – which can be similarly calculated by assigning the average of the high and low growth global city inventive class population shares to each population and allowing patents per inventive class member rates to vary as they do – are 0.00351 and 0.00250 for fast and slow growth global city populations.

Decomposition analysis shows us that most of the difference in slow and fast growing global city population-denominated patents per capita rates – 129% – is driven by differences in inventive class patenting productivity across populations. While the observed patenting rate for high growth global cities exceeds that of low growth global cities, if inventive class productivity were constant across the two populations, the patenting rate of slow growth global cities (0.00312) would exceed that of fast growth global cities (0.00289). The analysis tells us that high growth global cities are also marginally disadvantaged by their inventive class composition. If their composition were identical to that of low growth global cities, the gap between high and low growth global city patenting rates would widen (0.00351 versus 0.00250). Nonetheless, high growth global cities’ compositional disadvantage is more than offset by their productivity advantage.

Next, the standardization and decomposition analysis is extended to the inclusive set of CZs distinguished by level of urbanization: urban containing a global city (UCGC),

| Measures                                      | Urban containing global city | Urban without global city | Predominately rural |
|-----------------------------------------------|------------------------------|---------------------------|---------------------|
| Rate factor-standardized patenting rate        | 0.00199                      | 0.00155                   | 0.00121             |
| Compositional factor-standardized patenting rate | 0.00246                      | 0.00150                   | 0.00080             |
| Observed patenting rate                       | 0.00297                      | 0.00158                   | 0.00053             |

| Comparison of CZs                             | Effects                      | Percent distribution of effects |
|-----------------------------------------------|------------------------------|--------------------------------|
| UCGC versus UWGC                              | Rate factor: -0.00095        | Compositional factor: -0.00044 |
|                                               | Rate factor: 68.3978         | Compositional factor: 31.6022 |
| UCGC versus PR                                | Rate factor: -0.00166        | Compositional factor: -0.00078 |
|                                               | Rate factor: 68.0497         | Compositional factor: 31.9503 |
| UWGC versus PR                                | Rate factor: -0.00071        | Compositional factor: -0.00034 |
|                                               | Rate factor: 67.5856         | Compositional factor: 32.4144 |

Notes: Rate factor is patents per inventive class member and compositional factor is inventive class as a share of total commuting zone population.
UCGC, urban containing a global city; UWGC, urban without a global city; PR, predominantly rural.
Source: Toole and Low 2013.
urban without a global city (UWGC) and predominately rural (PR) (Table 5). Das Gupta’s (1993) three population methodology is adopted, which ensures that there is ‘only one standardized rate for a population when standardization is done with respect to the same factor’ (p. 97) and that the sum of the UCGC–UWGC effect and the UWGC-PR effect equals the UCGC-PR effect for both factors.

As seen in the last column of the ‘Decomposition of effects’ section of Table 5, when the three population methodology is applied to each CZ type pair, 32% of the difference in population-denominated patenting rates can be attributed to differences in the composition of the inventive population. Instead of global cities being roughly six times more productive (0.00297/0.00053) than predominately rural areas in producing patents, we see that when we hold inventive class population share constant, global cities are only three times as productive (0.00246/0.00080). That the new approach does not change the qualitative verdict that global cities are more inventive than predominately rural areas suggests that Figure 1 is not a good representation of the regional capacity for patentable innovation. However, the large difference between the compositional factor-standardized patenting rate and the observed patenting rate for predominately rural areas suggests that Figure 2 may also mislead. Contrary to being a matter of type where large cities support invention and smaller places generally do not, geography of invention is seemingly a matter of degree.

Discussion

Despite their widespread use, patents per capita have not been sanctified as any sort of official statistic for regional invention. Substantial efforts at the international level to harmonize patent statistics for cross-national comparisons provide strong evidence of the importance attached to these innovation indicators (Galindo-Rueda, 2013; OECD, 2009). Yet, the most recent review of the value of patent statistics is agnostic (National Research Council (NRC), 2014, pp. 5–9):

The panel makes no explicit recommendation here for NCSES [National Science Foundation’s National Center for Science and Engineering Statistics] to do more than continue to explore wider use of patent indicators and to engage in international cooperation on the development of indicators based on patent records to address user needs. There is no standard method for calculating indicators from patent data, and as noted earlier, analysis of these data without reservation can lead to incorrect inferences and misleading policy decisions. […] As NCSES continues to disseminate patent data as part of its [Science, Technology, and Innovation] STI indicators program, it would be valuable to users to have clear cautions regarding the use and misuse of these statistics for decision-making purposes.

The central purpose of this paper is to demonstrate that meaningful comparison of cross-sectional or longitudinal patenting rates requires defining a subpopulation that plausibly contributes to patenting. The hybrid rational/inductive identification of an inventive population or inventive class allows us to compute patenting rates on such a subpopulation, where patenting productivity is not confounded by population irrelevant to the patenting process. Separating the simple population-denominated patenting rate into a compositional factor and a rate factor introduces the concepts of standardization and decomposition that have been essential for meaningful cross-sectional and longitudinal comparisons of demographic phenomena. The importance of this method to the innovation literature is best expressed in the title of the NRC report Capturing Change in Science, Technology, and Innovation: Improving Indicators to Inform Policy.
The replacement for the population denominator in patenting rate statistics is neither obvious nor uncontentious. The principal objective of this research has been to demonstrate that differences in overall patenting rates are difficult to interpret as they represent differences in both rate factors and compositional factors. However, we do not claim that the compositional factor derived from our preliminary definition of the inventive class is definitive. We welcome debate on the best way to identify an exemplary population of those who might plausibly contribute to patenting. In the meantime, the methods demonstrated here are available to any locality wanting a more meaningful assessment of changes in patenting rates over time if there is agreement on the occupations that make up the local exemplary inventive population.

The main take away from this analysis regarding the geography of invention is that rural patenting rates denominated by a reasonable definition of the inventive class is one-third the patenting rates of global cities in the United States, on average, not the one-sixth suggested by the conventional per capita rate. At the individual CZ level, 5% of predominantly rural CZs have patenting rates higher than half of the global cities. Figure 3 demonstrates that the most productive patenting regions are in fact centred around small and medium-sized cities. The claim that patenting is overwhelmingly a global city phenomenon, based on evidence produced from conventional patenting rates, dichotomizes the innovation economy. That dichotomization is likely to contribute to suboptimal innovation policy, as it mischaracterizes the large potential contribution from medium city, small city, and rural inventing.

Shifting from an ‘inventive places of type’ to an ‘inventive places of degree’ perspective may hold little sway for many innovation researchers who will still claim that most inventive activity occurs in global cities. We are not worried that studying patenting and innovation in global cities will be reduced by the confirmation that a substantial amount of invention takes place elsewhere. What is more troubling are the simple linear notions of the mindset that contends that capturing the bulk of a phenomenon is all that should matter or that promoting a phenomenon only where it is most prevalent is the most efficient strategy. This linear view is best challenged by the fact that some of the most reliable patent producers today are located in a place thought better suited to growing apricots 60 years ago.

Acknowledgements
The authors thank Andy Toole for his wealth of knowledge regarding patenting and guidance throughout the research; the ERS GIS team for an introduction to three-dimensional maps; and Maryann Feldman for comments made on an earlier draft. The authors are solely responsible for any errors. The ideas expressed in this paper are those of the authors and are not attributable to the Economic Research Service or US Department of Agriculture.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This research was conducted at the Economic Research Service.
Notes

1. The only comment we were able to find questioning the validity of using population as a denominator addressed patenting rate indicators in developing countries: ‘The number of U.S. patents per capita is a common proxy used to measure the relative innovation efficiency of countries, but we believe that this computation underestimates the innovative capacity of developing countries, because it fails to detect the productivity of highly capable centers of excellence within countries with large populations’ (Morel, 2005, p. 401).

2. We choose to separate the Aleutians West Census Area from the Seattle, Washington CZ due to the geographic distance between the two regions.

3. Determining the criteria for the smallest CZ classification was straightforward: CZs that contain only non-metropolitan counties or only non-metropolitan and small ex-urban counties classified as part of a metropolitan statistical area (MSA) are labelled as ‘predominantly rural’. This classification corresponds to CZs containing no metro counties (i.e., all counties are assigned Rural–Urban Continuum codes 4–9) in 2003. CZs that contain cities included in the list of Global Cities constructed by Globalization and World Cities (GaWC) Research Network at Loughborough University are labelled as ‘urban containing a global city’ CZs. The criteria for global city status are determined by the availability of advanced producer services essential for the global coordination of activities by multinational corporations (Beaverstock, Taylor, & Smith, 1999). The remaining CZs make up the ‘urban without a global city’ category.

4. To increase the probability of patenting success in predominantly rural CZs during the period of interest, we define ‘patents per capita’ as the number of patents produced in 2000–05 divided by the year 2000 population.

5. Alternative denominators for computing regional patenting rates have been largely limited to employees, and research and development (R&D) expenditures or R&D employees. Replacing population with a measure of employment corrects for the distortion introduced by variation in the size of the dependent population across regions or through time (Meliciani, 2000; Porter, 2011). Using R&D expenditures and R&D employees attempts to define more narrowly patent productivity but runs into the problem that not all patents come from R&D laboratories contributing to the erroneous result that R&D is supposedly most productive where R&D laboratories are rare. The closest previous research to the current effort is to use the size of the science and engineering workforce in the denominator (Motoyama & Konczal, 2013).

6. One data limitation in the current analysis is that the publicly available occupational data are not disaggregated at a fine enough level to provide information on science and engineering management occupations that are included in the NSF SET occupations. A special tabulation of 2000 Census data to address this problem has been requested, but was not available at the time of writing.

7. As in the case of patents per capita, to increase the probability of patenting success in predominantly rural CZs during our period of interest, we define ‘patents per inventive class member’ as the number of patents produced in 2000–05 divided by year 2000 inventive class.

References

Beaverstock, J. V., Taylor, P. J., & Smith, R. G. (1999). A roster of world cities. Cities, 16, 445–458.

Carlino, G. A., Chatterjee, S., & Hunt, R. M. (2007). Urban density and the rate of invention. Journal of Urban Economics, 61, 389–419.

Carlino, G., & Kerr, W. R. (2014). Agglomeration and innovation (Working Paper No. 20367). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w20367.

Das Gupta, P. (1993). Standardization and decomposition of rates: A user’s manual current population reports special studies (p. 23–186). Washington, DC: Bureau of the Census.

Florida, R. (2002). The rise of the creative class: And how it’s transforming work, leisure, community, and everyday life. New York, NY: Basic Books.
Freshwater, D. (2012). *Rural innovation: Crucial, but rarely systemic*. University of Kentucky, Agricultural Economics Research Report. Retrieved July 29 2015, from http://ageconsearch.umn.edu/handle/139829

Furman, J. L., Porter, M. E., & Stern, S. (2002). The determinants of national innovative capacity. *Research Policy, 31*, 899–933.

Galindo-Rueda, F. (2013). The OECD measurement agenda for innovation. In F. Gault (Ed.), *Handbook of innovation indicators and measurement* (pp. 217–246). Northampton, MA: Edward Elgar Publishing, Inc.

Krammer, S. M. S. (2009). Drivers of national innovation in transition: Evidence from a panel of Eastern European countries. *Research Policy, 38*, 845–860.

Lai, R., D’Amour, A., Yu, A., Sun, Y., Fleming, L. (2013). Disambiguation and Co-authorship networks of the U.S. patent inventor database (1975–2010) (Harvard Dataverse Network [Distributor] V5, 2013, hdl.handle.net/1902.1/15705UNF:5:Rqsl3LsOEYLiHkk5jG/jRg==).

Mayer, H. (2009). *Bootstrapping high-tech: Evidence from three emerging high technology metropolitan areas*, Metro Economy Series for the Metropolitan Policy Program at Brookings. Washington: Brookings Institution. Retrieved September 19 2014, from http://www.brookings.edu/-/media/research/files/reports/2009/6/metro%20hightech%20mayer/06_metro_hightech_mayer.pdf.

Meliciani, V. (2000). The relationship between R&D, investment and patents: A panel data analysis. *Applied Economics, 32*, 1429–1437. doi:10.1080/00036840050151502.

Morel, C.M. (2005). Health innovation networks in developing countries address neglected diseases. *Science, 309*, 401–404.

Motoyama, Y., & Konczal, J. (2013). How can I create my favorite state ranking?: The hidden pitfalls of statistical indexes. *Journal of Applied Research in Economic Development, 10* Retrieved October 9, 2015 from http://journal.c2er.org/2013/09/how-can-i-create-my-favorite-state-ranking/

National Research Council. (2014). *Capturing Change in Science, Technology, and Innovation: Improving Indicators to Inform Policy*. Panel on Developing Science, Technology, and Innovation Indicators for the Future, R.E.Litan, A.W.Wyckoff, and K.H.Fealing, Editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Board on Science, Technology, and Economic Policy, Division of Policy and Global Affairs. Washington, DC: The National Academies Press.

Nichols, B. (2013). *Valuing the art of industrial design: A profile of the sector and its importance to manufacturing, technology, and innovation*. Research Report #56, Washington, DC: National Endowment for the Arts. Retrieved July 21 2015, from http://arts.gov/publications/valuing-art-industrial-design

Organisation for Economic Co-operation and Development (OECD). (2009). *OECD patent statistics manual*. Paris: OECD Publications. Retrieved September 10 2014, from http://www.oecdbookshop.org/oecd/display.asp?sf1=identifiers&st1=9789264056442

Organisation for Economic Co-operation and Development (OECD). (2010). *Territorial reviews: Venice, Italy 2010*. Paris: OECD Publications. doi:10.1787/9789264083523-en.

Parker, T. (2012). USDA economic research service commuting zones and labor market areas. Retrieved August 29 2014, from http://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas.aspx

Porter, M. E. (2011). *Cluster mapping project*. Cambridge, MA: Institute for Strategy and Competitiveness, Harvard Business School.

Rothwell, J., Lobo, J. Strumsky, D., & Muro, M. (2014). *Patenting prosperity: Invention and economic performance in the United States and its metropolitan areas*. Metropolitan Policy Program at Brookings Research report. Washington: Brookings Institution. Retrieved August 28 2014, from http://www.brookings.edu/research/reports/2013/02/patenting-prosperity-rothwell

Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica, 44*, 219–231.

Tolbert, C., & Sizer M. (1996). *US commuting zones and labor market areas: A 1990 update*. USDA Economic Research Service, Rural Economy Division. Staff Paper No. AGES-9614. Retrieved July 29 2015, from https://usa.ipums.org/usa/resources/volii/cmz90.pdf
Toole, A. A. & Low, S. A. (2013). Patenting in rural America: Inventors, teams, & technologies. Selected poster presentation at Annual Meeting of the Agricultural & Applied Economics Association, Washington, DC, August 4-6. Retrieved from http://ageconsearch.umn.edu/bitstream/149013/2/Toole%20Low%20AAEA%20Poster%202013-2.pdf

World Bank. (2009). World development report 2009: Reshaping economic geography. Washington: World Bank.