Triangulating regional economies: Realizing the promise of digital data

Maryann Feldman *, Nichola Lowe

University of North Carolina, Chapel Hill, USA

A R T I C L E   I N F O

A B S T R A C T

Innovative data sources offer new ways of studying spatial and temporal industrial and regional development. Our approach is to study the development of an entrepreneurial regional economy through a comprehensive analysis of its constituent firms and institutions over time. Our study region is defined by the location of large multinationals recruited to North Carolina's Research Triangle Park and the adjacent area. We have built a database of 4200 technology-intensive entrepreneurial firms that draws on over 30 distinct data sources and includes details on company founders, annual firm employment and engagement with the entrepreneurial ecosystem. By outlining our approach in this paper, our primary objective is to create a transferable framework for analyzing regional dynamics in other locations.

Published by Elsevier B.V.

1. Introduction

Digital data provides a means to understand the functioning of innovative industries and regional economies. Aggregate industrial data produced by a variety of government agencies and dependent on rigid and inflexible geographic units, while useful for certain analyses, obscures many research questions relevant to studying the evolution of regional economies. Consideration of the role of firms—as dynamic actors with evolving capabilities and temporally delineated strategic investments in a region—is frequently absent because of the restrictions to sharing and collecting micro-level data. Moreover, comparative analysis typically assumes static or time-invariant institutional features and therefore is unable to capture the processes by which organizations and institutions within a region change and adapt over time. Thus, we are left to assume that regional institutional features are deterministic and static rather than socially constructed. Moreover, outcomes we observe are the result of a complex set of actors and specific interactions that occur in a temporal and varying regional setting. An unfortunate disconnect exists between the theoretical definition of region as integrated contiguous space and the political and Census geography for which data are readily available. Insufficient micro-level data has inhibited understanding of the underlying dynamic processes within regions that lead to, and sustain, innovation and entrepreneurship. Fundamental questions about overlapping and evolving pathways for innovation and regional economic growth remain obscure.

A more nuanced understanding of the sources of spatial and temporal variation within a regional economy is possible by combining data from a variety of electronic sources to gain a more complete picture of constituent organizations, their relationships and the ways in which inventors and entrepreneurs transverse institutional and geographic space. Digital data sources easily permit the construction of detailed firm records useful to studying regional industrial and institutional dynamics. This paper describes our efforts to combine and integrate public and private data sources to create detailed time-series micro-level data. Our framework is designed to study the development of the region by studying extinct, existing and emergent firms and institutions, triangulating data to gain a more complete understanding of the development of firms. The data are organized in a relational database that can support multiple forms of analysis, including mapping and network visualization, statistical modeling or more qualitative approaches. Focusing in-depth on one region removes unobserved heterogeneity that falls to the error term in cross-sectional intra-regional comparisons and permits an examination of the dynamics of innovative economies, specifically how organizations and institutions work together, adapt and improvise to define a functioning regional economic and social system. By recognizing sources of variation within a regional economy, we not only gain a more sophisticated framework for identifying concurrent pathways to nurture the development of innovative firms but also a better understanding of why certain regional attributes predominate and contribute to regional advantage. As such, we are in a better position to situ-

* Corresponding author.
    E-mail address: maryann.feldman@gmail.com (M. Feldman).

ate economic actors within an institutional environment, but also examine how their actions and interactions contribute to institutional transformation over time.

The primary purpose of this paper is to describe our approach for combining and creating digital data to study regional economies and industries. Our illustrative project integrates annual data from third-party data sources and tools into a digital infrastructure that is organized around both established and entrepreneurial firms in North Carolina's Research Triangle region. This project extends efforts to bring new electronic sources of private sector data and new analytic tools to the study of dynamic regional economies (Feldman et al., 2012). We describe our data collection efforts that define the region by examining the location of firms and tracking their development and program participation through annual events that are collected from a variety of sources. In partnership with UNC's Renaissance Computing Institute (RENCI) we have designed and implemented a relational database organized around records for over 4200 technology-intensive firms.

Our database design is organized around three components that are discussed in turn. The first is a relational database that tracks firms over time, using a variety of third party data sources. The second component is a data archive that preserves and catalogues studies that collected survey data and other single occurrence variables so that other scholars can re-examine these from their disciplinary perspective. The third component adds context through a digital archive of public documents, reports, and oral histories. Combining these elements, we create a time series data platform linking individuals, firms and institutions while providing historical context to their interactions. In the process, we have created a digital template that would enable scholars to conduct data intensive studies to address fundamental questions about innovation, regional growth and economic development. Though our particular research focus is the Research Triangle region located around North Carolina's Research Triangle Park, in this paper we describe a replicable methodology. We offer our project as an example of how digital data sources can be integrated to better understand regional economies.

2. Defining regions by studying firms[1]

The availability of new digital third party data sources calibrated to individual firm addresses provides a means to define a region. After all, places like Silicon Valley or Route 128 are not easily identified on any map; however, they are well known as organizing platforms for innovative economic actors and activity. In practice, innovative locations are defined by the location of firms. Silicon Valley, the prototype for an innovative region, is so named because it was the epicenter for the silicon-based semiconductor industry. Personal computer component makers followed semiconductors, subsequently followed by networking and Internet companies–all creating a revolution in electronics miniaturization and computers, fueled by entrepreneurial firms (Lécuyer, 2008). The local community adopted the name Silicon Valley, which was important to establishing regional identity. Geographically Silicon Valley now encompasses all of the Santa Clara Valley and the southern East Bay. As with any desirable real estate, the boundaries have expanded over time, following yet also influencing the location of entrepreneurial firms, who at times cannot afford prime real estate but want to identify with, and be part of, Silicon Valley.

The Research Triangle, our study region, offers another illustrative example. Research Triangle can refer to the Research Triangle Park (RTP), a 7000-acre industrial park with its own dedicated

zip code. But the RTP is a 1960 style low-density research campus of mostly large multinational firms, which until recently had real estate covenants that precluded entrepreneurial firms. Instead, most of the entrepreneurial activity in the region is located in the contiguous communities adjacent to the Park-places with names like Morrisville, Cary and Apex that are not well recognized on their own but are integral locations for economic activity. Often these communities are counted as part of the metropolitan areas of Raleigh or Durham, reflecting the tradeoff between industry detail and geographic specificity. Often finer geographic data are suppressed to maintain confidentiality as required for government data collections. Additionally, regions can be defined by government entities. For example, the state of North Carolina considers the Research Triangle Region to be a 13 county planning region that stretches to the Virginia border and provides data aggregated over this large and diverse region. In reality, there are important synergies between the Research Triangle and the cities of Greensboro and Winston Salem and firms gravitate towards those locations, which are outside the state-defined region. In addition, the larger 13 county region masks the micro-geography, as firms in specific sectors may agglomerate in closer proximity that is dilute when considered against activity in the larger region.

As these examples and other illustrate, we study innovative and entrepreneurial regions that are socially constructed in so far as they do not conform neatly to rigid political boundaries or government statistical units (Amin, 1999). In reality, the functional boundaries of regional economies are reinforced instead by the location of prominent institutions, often universities, government labs or large successful firms, and are influenced by existing transportation routes and land use patterns—factors that in turn influence and reflect firm location (Audretsch et al., 2005; Bathelt et al., 2010). The definition of region is also fluid and expands over time due to idiosyncratic and serendipitous economic and social events (Feldman et al., 2012). Spatial patterns follow a logic that motivates firms to locate near others with similar products, and markets, and close to employees with the requisite workforce skills. The location patterns of related firms define the region (Markusen, 1994). The definition of the region evolves as more firms form and grow, and are attracted to the region.

Examining federally-defined geographic units such as metropolitan areas or counties can also mask cross-border activity and often lead jurisdictions to act as though they are in competition when they could gain from collaboration (Bartik, 1991). Equally important, the geographic co-location of innovative, creative firms in small places such as multi-tenant buildings, neighborhoods, or industrial parks is often invisible when data are available only for larger, administrative units (Sassen, 1989). Firm address level data are important to understanding innovative geographies and allow us to define the region of study based on the behavior of firms rather than the other way around: what appears to be an agglomeration at the county level may indeed be several geographically (and often technologically) distinct groups of firms each with different social relationships and underlying patterns of development (Kohlhase and Ju, 2007).

The technology-intensive firms driving regional economies are themselves fluid and difficult to classify using standard industrial classification schemes. As firms struggle to survive, they often pivot to modify their products or services, but there is no incentive and limited opportunity to update their standard industrial classifications—the main mechanism primarily used to understand industrial activity. New industrial activities, such as 'clean tech' or even optical science defy more standard classifications (Lane, 2011; Clark, 2014). Therefore, by using keywords or combinations of phrases or membership organizations we are able to better assess a broad range of activities and technology applications (Feldman and Lendel, 2010). Both patents and product announcements provide

---

[1] There is a long academic tradition of studying regions by examining their constituent firms s advocated by Markusen (1994).