HARVARD | BUSINESS | SCHOOL

Spatial organization of firms and location choices through the value chain

Juan Alcácer
Mercedes Delgado

Working Paper

13-025

August 1, 2013

Spatial organization of firms and location choices through the value chain*

Juan Alcácer
Harvard Business School

Mercedes Delgado
Fox School of Business and Institute for Strategy and Competitiveness

Version May 28, 2013

Abstract

We explore the impact of geographically bounded, intra-firm linkages (internal agglomerations) and geographically bounded, inter-firm linkages (external agglomerations) on firms' location strategies. Using data from the Census Bureau's Longitudinal Business Database, we analyze the locations of new establishments of biopharmaceutical firms in the U.S. in 1993–2005. We consider all activities in the value chain and allow location choices to vary by R&D, manufacturing, and sales. Our findings suggest that internal agglomerations have a positive impact on location. The effects of internal agglomerations vary by activity, and they arise both within an activity (e.g. among plants) and across activities (e.g. between sales and manufacturing). Our results also suggest that previous estimates of the effect of external agglomerations may be overestimated because the existing literature abstracted from internal agglomerations.

Keywords: Location choices, agglomeration economies, value chain, organization theory.

* For helpful comments we are grateful to Joanne Oxley, Scott Stern, Ian Cockburn, Jeff Furman, Pierre Azoulay, Mat Marx, David Ridley, Ram Mudambi, and participants in the seminars at Rotman School of Management, University of Toronto, MIT Sloan, Carlson School of Management, University of Minnesota, the NBER Productivity seminar, U.S. Census Bureau Center for Economic Studies, Fox School of Business/Kauffman Foundation Workshop on Empirical Entrepreneurship, Strategic Management Society (SMS), and Academy of Management annual meetings. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Mercedes Delgado gratefully acknowledges support from the Center for International Business Education and Research (CIBER) at Temple University.

1. INTRODUCTION

The determinants of firms' location choices are the subject of a large body of research spanning multiple disciplinary fields. Most research in economics (Hanson 2001, Delgado et al. 2010) and strategy (Shaver and Flyer 2001, Alcacer and Chung 2013) focuses on elements of the external environment, specifically agglomeration economies, as drivers of firm location choices. Taken as a whole, this research suggests that physical proximity *between* firms boosts productivity and therefore creates incentives for collocation.

Much less is known about the role of internal agglomerations—geographically bounded, intra-firm linkages that positively impact performance—on location choices. Neglecting the role of internal drivers in firms' location decisions is a particularly surprising omission in the strategy literature, which recognizes that links between activities across the value chain are important levers for developing competitive advantage (Porter 1996) and for innovating (Cohen and Levinthal 1990). Understanding internal agglomerations is also important because their effect on location choices is likely to be present in any multi-unit firm, which has been the most common organizational form for firms since the 1940s (Chandler 1962).

To improve our understanding of location choices, this study investigates if internal agglomerations influence the locations of distinct activities in the value chain (manufacturing, R&D, and sales). Specifically, we ask two related questions: *Does the location of a firm's existing facilities affect its subsequent location choices? If so, does that effect vary by activity in the value chain?* We argue that internal agglomerations do exist, and that they prompt firms to collocate activities across the value chain. More specifically, we suggest that the three external agglomeration economies identified by Marshall (1920)—access to knowledge spillovers, specialized labor, and specialized suppliers—are as likely to boost firm performance when they occur within firms as when they occur across firms. Internal agglomerations also arise because geographical proximity enhances firms' ability to control (Giroud 2013, Kalnins and Lafontaine 2013) and coordinate (Chandler 1962, Henderson and Ono 2008) activities across the value chain, which results in better firm performance. Additionally, we suggest that these internal agglomerations vary by activity.

The presence of internal agglomerations does not preclude the effect of external ones. In fact, we advocate for a comprehensive framework where *both* internal and external agglomerations drive location choices. We conceptualize external agglomerations as centrifugal forces that may drive firms to disperse their activities geographically, and internal agglomerations as centripetal forces that may drive within-firm collocation, either across activities (e.g. manufacturing and R&D) or within activities (e.g. multiple R&D labs). Regardless of the direction these forces push firm locations, failing to consider internal

agglomerations leads to an omitted variable problem—and may well have biased previous work estimating external agglomeration's effect on location choice. This possible bias prompts a related research question: *Does the effect of external agglomerations on location choice vary when internal agglomerations are properly accounted for?*

We explore these questions using data from the Longitudinal Business Database of the Census Bureau. Our dataset encompasses location choices for new establishments of biopharmaceutical firms for distinct activities of the value chain (R&D, production, and sales) from 1993 to 2005. We examine internal agglomerations taking into account the spatial organization of *all* activities in the value chain of a firm. Failing to consider all activities leads to an omitted variable problem—pervasive in the literature—in which the estimated collocation levels between, say, sales and manufacturing may be biased by failing to control for the existing locations of another activity, such as R&D. Empirically, our paper is the first to systematically explore collocation decisions for all firm activities, rather than for subsets of activities. Conceptually, we offer a comprehensive framework for location choices that encompasses both internal and external drivers—a framework that, according to our results, better explains actual location choices. This new framework encompasses a diverse array of literatures, which we curated and synthesized to identify five primary sources of internal agglomerations and the theoretical mechanisms on which they are based.

We find that internal agglomerations are, in fact, an important driver of location choices that has been overlooked in the literature. The impact of internal agglomerations varies by activity in the value chain: although present in all activities, it is larger for R&D and manufacturing than for sales. Our results document unexpected patterns in the data, such as asymmetries of internal agglomeration between sales and manufacturing, whose explanation requires further conceptual development. Furthermore, our results suggest that previous estimates of the effect of external agglomerations may actually overstate their influence, because the impacts of internal agglomerations were overlooked.

2. WHAT DRIVES LOCATION CHOICES?

Most research on firms' location choices is based on external drivers. This approach, although fruitful, offers a limited portrait of actual location choices by omitting or minimizing the role of internal agglomerations (geographically bounded, intra-firm linkages). Internal agglomerations are likely to play an important role in the location choices of multi-unit firms, the most common organizational form since the 40s (Chandler 1977). Multi-unit firms are bound together through intra-firm linkages that impact performance (Porter 1996). To the extent that these interdependencies are weakened by distance, firms are likely to consider their existing physical footprint as they decide where to expand.

Recent studies on external drivers implicitly recognize the existence of internal agglomerations by taking steps to rule them out of empirical tests. For example, Alcacer and Chung (2007) focused on first-time entrants into the U.S. because "*…incumbent firms have prior investments that may affect subsequent location choices and create dependence among observations by the same firm."* Other studies have focused on the role of external drivers on the location of start-ups (e.g. among others, Glaeser and Kerr (2009) and Delgado et al. (2010)), a choice that will not be affected by internal agglomerations. Instead, our focus is to study the location decisions of existing firms as a function of *both* external drivers *and* internal agglomerations.

2.1. Internal Agglomerations: Bringing Firm Activities Together?

Internal agglomerations arise because geographical proximity between within-firm activities improves firm performance. Previous research suggests diverse mechanisms of internal agglomerations; some of these are analogous to those identified by Marshall (1920) for external agglomeration economies: improved access to knowledge spillovers, specialized labor, and specialized suppliers.

For example, Cohen and Levinthal (1990) highlight that the continuous exchange of information among activities, such as R&D and manufacturing, allows firms to develop the absorptive capacity needed to acquire external knowledge and innovate upon it. To the extent that physical proximity increases communication and knowledge sharing, collocating activities will allow for productivity gains. Indeed, knowledge spillovers between R&D and manufacturing have received special interest in the literature. Tecu (2011) used patent data as an output measurement to find that R&D in chemicals was 2.5 times more productive in locations where manufacturing was also present (after controlling for external drivers of innovation such as proximity to universities and peer firms). Examining the reverse effect, in the case of R&D on manufacturing, Adams and Jaffe (1996) demonstrated that the positive effect of parent-firm R&D on plant productivity diminishes with geographic distance. In both cases, within-firm knowledge spillovers were cited as the internal agglomeration mechanism behind superior performance. Despite these insights, few studies have examined how these types of spillover might work simultaneously, or explored their asymmetry.

Economies of scale and scope in internal labor markets are another source of internal agglomerations. Fixed costs from investments to attract, retain, and motivate workers, such as day care facilities, gyms, or cafeterias, are easier to spread over a geographically concentrated labor force. Large-scale operations also enable firms to develop specialized labor that can be shared across activities (e.g. sharing R&D personnel across diverse research projects). Although, many relevant studies find evidence of such scope economies in R&D (Helfat 1997, Henderson and Cockburn 1996) and generally in performance (Hamilton et al. 2003, Tate and Yang 2011), they abstract from the spatial organization of firms. We argue that

geographic proximity among same-firm units helps firms achieve economies of scope by making labor easier to share or redeploy. For example, Di Minin and Bianchi (2011) find that physical proximity of R&D personnel to IP lawyers increases the chances that an innovation will be patented, explaining why R&D is homebound in global industries. Proximity of specialized labor will also facilitate personnel rotation across activities, which was identified as a key practice behind the productivity gains of Japanese firms in the 80s (Clark and Fujimoto 1987, Mansfiled 1988).

A third mechanism of internal agglomerations is the access to intermediate inputs. Efficiency reasons for vertical integration are discussed in Grossman and Hart (1986), but there is little empirical evidence on the magnitude and type of these efficiencies. To the extent that firms have designed their operations around flows of intermediate inputs, we expect that collocating units that produce these inputs will decrease transportation and coordination costs, and ultimately impact performance.

Other related mechanisms behind internal agglomerations arise because value-chain activities within the firm are governed by organizational hierarchy rather that by market mechanisms. Within hierarchies, activity control and coordination are paramount to performance. In terms of control, Kalnins and Lafontaine (2013) found that the distance to headquarters is associated with lower revenue for establishments in the hotel industry, a finding that they attribute to the difficulty headquarters have controlling and monitoring their establishments from a distance. Similarly, Giroud (2013) argues that better control and monitoring explain his finding that a reduction in travel time between headquarters and plants increases plant investment by around 9% and total factor productivity by 1.4%.

Proximity between headquarters and plants has also been associated with the desire to improve performance by improving coordination. Henderson and Ono (2008) found that, after controlling for external drivers such as access to specific services in metropolitan areas, firms prefer to keep headquarters close to their manufacturing bases to facilitate coordination. Coordination as a driver of collocation extends beyond headquarters and plants. Chandler (1962, 1977) identified coordination as the fundamental force that kept different parts of a multi-unit firm together, and suggested that coordination costs increase with distance. Ketokivi (2006) explores under what circumstances collocation of R&D and manufacturing decrease coordination costs.

Despite differences in theoretical frameworks, methods, and the mechanisms proposed, the papers above share a common feature: They each suggest, and find evidence to support, that collocating activities has a positive impact on performance. Therefore, we expect that *firms are more likely to locate new establishments in places where they already have operations.* That collocation would be triggered by firms' attempts to achieve better performance through any of the mechanisms that drive internal

agglomerations: knowledge and information flows, development of specialized labor and inputs, control and monitoring, and coordination.

We note that previous research suggests, conceptually and empirically, that the effect of internal agglomerations may change by activity. For example, Kleinbaum et al. (2008) used email frequency as a proxy for coordination needs in a large technology firm and found heterogeneity in information flows, with most information exchanges happening within activity, especially in R&D, and with above-average information exchanges between sales and R&D and between sales and supporting services. Van Den Bulte and Moenaert (1998) found that information flows among R&D personnel increased when dispersed R&D personnel were geographically concentrated, suggesting there are informational benefits from R&D collocation.

What accounts for these differences? A plausible explanation is that internal agglomerations vary according to activity. For example, knowledge generated by R&D may be tacit and hard to transfer, making collocation *within* activities more desirable for R&D establishments; coordination across plants may be more important than among sales establishments, etc. Similar arguments can be used to hypothesize that agglomeration economies vary *between* activities. For example, the need to collocate R&D and manufacturing may be greater than the need to collocate R&D and sales because R&D and manufacturing require more coordination and knowledge transfers than R&D and sales. Due to the dearth of relevant theoretical models and the piecemeal nature of empirical findings in this area, we do not develop a set of testable hypotheses per activity or pair of activities here. Nevertheless, we do expect that *internal agglomerations vary by activity in the value chain*.

2.2. External Drivers of Location: Driving Firm Activities Apart?

Two sets of external drivers have been commonly studied in the literature: unique location endowments and agglomeration economies. Ricardo (1817) was the first to posit that a stochastic distribution of natural resources across the geographic space drives economic exchanges, an idea that is at the core of economic geography and international economics today. Some studies have since expanded Ricardo's view of endowments to encompass a location's institutional features, such as IP regimes, labor regulation, and the unique technological knowledge present in universities. This literature primarily predicts that firms will flock to locations where they can tap abundant inputs at low costs (e.g. labor in China), rare and unique resource (e.g. access to a port), or location-specific incentive programs (e.g. tax policies).

Marshall (1920) introduced a second and related external driver to explain location choices: agglomeration economies. He argued that collocated firms would enjoy higher productivity because geographic concentrations of firms would attract larger pools of specialized labor and suppliers, and

would also facilitate the flow of knowledge from one firm to another. These agglomeration economies exert a multiplier effect on firm productivity by increasing the benefits that firms would otherwise receive only from a location's physical and institutional endowments, a concept that has been adopted and expanded by researchers in various fields, including Porter (1998) in strategy, Jacobs in urban economics (1984), and Krugman (1991) in international economics. While some papers have identified negative effects from external agglomerations in the form of incentives to avoid collocating with other firms (see e.g. Shaver and Flyer 2001 and Alcacer and Chung 2007), most empirical and theoretical models suggest that the effect of agglomeration economies on firm performance and on location choices will be positive (see e.g. Henderson 2003, Gleaser and Kerr 2009, Delgado et al. 2010).

Previous research also suggests that the strength of external location drivers will vary by value chain activity. For example, Audrestch and Feldman (1996) found that innovation in the United States is more concentrated than manufacturing, while Alcacer (2006) found that competition and external agglomerations exert different effects on inter-firm collocation for R&D, manufacturing, and sales in the wireless handset industry.

From this literature on the external drivers of location choices we can establish our null hypothesis: that only the potential for external agglomeration influences location choices. Additionally, we can expect that firms will adopt a basic level of geographic dispersion for different value chain activities because activities may be attracted to different features in the external environment.[1]

Summarizing the last two sections, we acknowledge that intra- and inter-firm linkages are important for multi-unit firms' performance, argue that location choices are the result of *both* external and internal drivers, and conceptualize the location decision as a function of these two sets of forces. The first set is internal agglomerations, centripetal forces that drive within-firm collocation; the second set is external agglomerations, centrifugal forces that drive firms to disperse activities geographically in search of the best external environment. Empirically, these forces may work in the same direction, as when, for example, a firm is already located in the best external environment: internal *and* external agglomerations would induce collocation. However, it is also plausible that these forces present firms with a tradeoff: Stay in a suboptimal location to preserve inter-firm links, or move some activities to a better external environment.

[1] For example, large service-oriented cities offer the best external environment for headquarters, but plants tend to locate in smaller cities or rural areas where wages are lower (Henderson and Ono 2008). Analogously, locations with top universities and a large number of Ph.D's may be ideal for R&D labs but not for manufacturing (Ketokivi and Ali-Yrkkö 2009, Tecu 2011).

3. EMPIRICAL DESIGN

Our empirical design examines the location decisions of new establishments of biopharmaceutical firms during 1993–2005. The biopharmaceutical industry is an ideal setting to examine firm location choices through the value chain, for several reasons. First, previous research has shown that biopharmaceutical firms tend to cluster, suggesting that external agglomerations exist (see e.g. Aharonson et al. 2008, Feldman 2003, Feldman and Schreuder 1996, Furman et al. 2005, Zucker et al. 1998). Second, Pisano (1996) found high levels of interdependencies among R&D and manufacturing activities in the value chain, suggesting the presence of potential internal agglomerations. Third, the industry experienced an important period of growth during our sample period, characterized by two critical transformations. The first was the shift from drug discovery, normally triggered by basic research in universities, to more applied development and the scaling-up of manufacturing. The second was the development of new activities for commercialization and support, which allows us to examine multiple activities within the value chain.

3.1. Sample

Our primary data source is the establishment-level Longitudinal Business Database (LBD) of the Census Bureau. The LBD provides annual observations of the universe of U.S. establishments with payroll, including each establishment's date of entry, physical location, industry code, and number of employees.

Our first challenge was to identify firms in biopharmaceuticals. The traditional approach has been to select firms in a particular set of SIC codes. For example, Toole (2003) and Cortright and Mayer (2002) defined biopharmaceutical firms as those with establishments in SIC-2830 (manufacturing of medicinal chemicals and botanical products (SIC-2833), pharmaceutical preparations (SIC-2834), diagnostic substances (SIC-2835), and biological products (SIC-2836)) and in SIC-8731 (commercial R&D in physical, engineering, and life sciences research). Another approach is to identify firms using firm directories (e.g. Zucker et al. 1998). Both approaches have pros and cons. SIC-based sampling captures all firms, well known or not, but it may include firms whose scope reaches beyond biopharmaceuticals.[2] Directory-based sampling guarantees that firms are in biopharmaceuticals, but it does not cover all firms.

We followed a hybrid approach to define our sample. We matched LBD data to BioScan (1992), a detailed (but not exhaustive) directory of worldwide biopharmaceutical firms also used by Zucker et al. (1998), to obtain frequencies of establishments' SIC codes for the BioScan-matched firms (see Table 1.a).

[2] This is particularly the case with firms in SIC-8731, an aggregated industry that includes labs that may not be associated with biopharmaceuticals.

Three features in Table 1.a deserve to be highlighted. First, approximately 75% of BioScan-matched firms had at least one establishment in SIC-2830 or SIC-8731, suggesting that membership in these SIC codes is an appropriate criteria to identify whether a firm in the LBD data is in biopharmaceuticals. Second, the frequency of SIC codes allowed us to identify activities in the biopharmaceutical value chain (referred to hereafter as the 'bio value chain'): R&D (associated with SIC-8731), manufacturing (associated with SIC-2830), and sales (associated with SIC-5120).[3] Third, there were establishments with SIC codes outside the bio value chain. We aggregated these other firm activities into other value chains (referred to hereafter as 'other value chains') that include 'other R&D' (SICs 8732 to 8734), 'other manufacturing' (SICs 20 to 39; except 2830) and 'other sales' (SICs 50 to 59, except 5120). The rest of firm activities were classified as support for any value chain. The main support activities included business services (including headquarters), medical labs, and financial, insurance, and real estate services (e.g. holding companies).

The resulting final sample spanned around 600 multi-unit firms in the LBD database, with at least one establishment in either SIC-2830 or in SIC-8731 as of 1992.[4] Among these, our analysis focuses on 157 firms that opened a new biopharmaceutical establishment during our sample period, 1993–2005.

3.2. Econometric Specification: A Location Choice Model

Because internal agglomerations elude direct observation and their effects on firm performance are endogenous to the choice of location, we infer the existence of internal agglomerations from location choices. Thus our empirical analysis examines the relationship between internal and external agglomeration drivers and the (continental) U.S. locations of new establishments opened by biopharmaceutical firms.[5] The baseline econometric specification is as follows:

$$Location_{firt} = \beta_1 Ln(X_{frt-1}^{Internal}) + \gamma_r + \gamma_{fit} + \varepsilon_{firt} \qquad (1)$$

where $Location_{firt}$ is equal to 1 if firm *f* chooses Economic Area (EA)[6] *r* to open a *new establishment* in focal bio activity *i* (R&D, manufacturing, or sales) at time *t*; $X_{frt-1}^{internal}$ is a set of variables that captures firm *f*'s geographic footprint of activities (i.e. for all activity *i* in bio value chain or other value chains) in

[3] SIC-5120 is Drugs, Drug Proprietaries, and Druggists' Sundries, including wholesale distribution of prescription drugs, proprietary (patent) drugs, druggists' sundries, and toiletries.
[4] Of these, 112 firms also appeared in BioScan.
[5] Following traditional location studies, we do not consider expansions through acquisitions because they often do not represent location choices. The goal of many acquisitions is to buy some technology or other assets of the firms, such as patents.
[6] The Bureau of Economic Analysis (BEA) defined 179 economic areas (EAs) spanning the U.S. that reflect meaningful economic regions and ensure comprehensive regional coverage (Johnson and Kort 2004). We excluded EAs in Alaska and Hawaii to minimize concerns about transportation costs.

EA *r* at time *t-1*; γ_r is a vector of EA fixed effects; γ_{fit} is a vector of firm–bio–activity–year fixed effects; and ε_{firt} is the error term.

We estimated equation (1) using conditional (fixed-effects) logit models of the location decisions for each activity in the bio value chain.[7] The fixed effects (or groups) in the model are firm–activity–year (γ_{fit}), and within a group the firm chooses among multiple regions (EAs) to locate the focal activity *i*.[8] The fixed effects will capture unobserved firm–activity attributes that affect the location decision (e.g. the size of a firm's bio-R&D effort; the type of bio R&D a firm pursued (applied or science-oriented), etc.). Because our sample may include multiple location choices from the same firm (i.e. firms may open multiple establishments over time), and because a firm's subsequent location decisions are likely to be correlated, we clustered the standard errors by firm.

3.2.1. Internal agglomerations

Internal agglomerations are intra-firm linkages that are geographically bounded (i.e. facilitated with proximity among units of the firm). To capture the role of internal agglomerations in a firm's location strategies, we developed various firm–region–year variables that account for the spatial organization of all the firm's activities. Specifically we grouped same-firm employment by value chain (bio value chain and other value chains) and by activity (bio R&D, bio manufacturing, bio sales, and support) across regions. The main internal agglomeration variables $X^{Internal}_{frt-1}$ are defined as follows:

$$Firm\,Bio\ Value\text{-}Chain\,Employ_{frt-1} = \sum_{k \in Bio-VC} emp_{fkrt-1} \tag{2a}$$

$$Firm\,Other\ Value\text{-}Chains\,Employ_{frt-1} = \sum_{k \in Other-VCs} emp_{fkrt-1} \tag{3a}$$

$$Firm\,Support\,Employ_{firt-1} = \sum_{k \in Support} emp_{fkrt-1} \tag{4a}$$

where *k* indexes all firm activities (4-digit SIC codes) that map into the bio value chain (equation 2a), into other value chains (equation 3a), and into support (equation 4a); and emp_{fkrt-1} is the employment in firm–EA–activity *k* the year prior to expansion. Recall from Section 3.1 that the range of industry SIC codes used for bio value chain, other value chains, and support were mutually exclusive. To examine within- and across-activity internal economies, the *Firm Bio Value-Chain Employ* variable is also broken down

[7] There were more than 1,200 location choices for new establishments in bio value chain activities: over 700 entries in R&D, 92 in manufacturing, and 408 in sales.

[8] The choice set corresponds to the group of 130 EAs selected for expansion in our sample in 1993–2005. The selected EAs vary by biopharmaceutical activity. Note that firms may open establishments in multiple EAs in a given bio-activity year (i.e. there could be multiple positive outcomes within a group).

into its component activities (*Firm Bio R&D Employ*, *Firm Bio Mfg Employ*, and *Firm Bio Sales Employ*). Table 1.b shows the descriptive statistics for each of these variables.

While our analysis focuses on levels of same-firm employment, we also develop internal agglomeration variables that measure the extent of employment specialization. The goal is to assess whether firms locate in regions where firm activities are over-represented. These variables will further confirm that firm location choices are not random but instead driven by internal agglomerations.

Our measure of firm-level regional specialization in the bio value chain (*Firm Bio Value-Chain Spec)* captures whether a firm's employment in these activities is over-represented in a particular location (given the size of the firm and its portfolio of activities and locations), and it is defined as:

$$Firm\,Bio\ Value\text{-}Chain\,Spec_{frt-1} = \frac{\sum_{k \in Bio-VC} emp_{fkrt-1}}{emp_{frt-1}} * \left[\frac{\sum_{k \in Bio-VC} emp_{fkt-1}}{emp_{ft-1}} \right]^{-1} \qquad (5)$$

where *k* indexes all firm activities (4-digit SIC codes) that map into the bio value chain, and the employment variables are as follows: emp_{fkrt-1} and emp_{frt-1} are employment in firm–activity–EA and in firm–EA at *t-1*; emp_{fkt-1} aggregates employment by firm–activity; and emp_{ft-1} is firm total employment. For example, if a firm locates 100% of its bio value-chain employment in location *r* and has 50% of its total employment in the bio value chain, then this variable will be 2 in that location (and zero in all other locations).[9] Similarly we compute *Firm Other Value-Chains Spec*, *Firm Support Spec* and the breakdown of equation 5 in its components: *Firm Bio R&D Spec*, *Firm Bio Mfg Spec*, and *Firm Bio Sales Spec.*

3.2.2. External agglomerations

Our baseline specification controls for external location drivers using Economic Areas (EA) fixed effects. These region dummies capture external economies of agglomerations (e.g. the specialization of the region in biopharmaceuticals) as well as regional endowments and policies that may influence the extent of external benefits (e.g. physical endowments, policies that favor manufacturing activities, and wages).[10]

Alternatively, in the supplementary analysis we define variables that capture potential external agglomerations in each region the year before an establishment expanded. We built on Delgado et al. (2012) to compute our main variable of external economies of agglomeration. External agglomeration benefits may arise from the specialization of a region in a cluster of related biopharmaceutical activities.

[9] The specialization variables may be more meaningful for larger, diversified firms.
[10] The extent of external benefits will depend on a firm's ability to interact with other firms in the region. In this paper, we abstract from the firm-level mechanisms that facilitate external benefits.

To capture this, we defined the specialization of region *r* in biopharmaceuticals using the employment location quotient, i.e. the share of regional employment in biopharmaceuticals compared to the share of U.S. employment in biopharmaceuticals. To compute this variable, we excluded the focal firm's own biopharmaceutical employment at the region to avoid cofounding the external and internal agglomerations. To be consistent with our definition of the bio value chain at the firm level, we define the biopharmaceutical cluster as including the same industry codes. More formally:

$$Region\ Bio\ Value\text{-}Chain\ Spec_{frt-1} = \frac{\sum_{k \in Bio-VC}(emp_{irt-1} - emp_{firt-1})}{emp_{rt-1}} * \left[\frac{\sum_{k \in Bio-VC} emp_{kUSt-1}}{emp_{USt-1}} \right]^{-1} \quad (6)$$

where *k* indexes all activities that map into the bio value chain; emp_{irt-1} and emp_{iUSt-1} are employment in bio activity *i* at time *t-1* for EA *r* and the U.S., respectively; emp_{firt-1} is the focal firm's own employment in bio activity *i* at *r*; and emp_{rt-1} and emp_{USt-1} are total employment for *r* and for the United States.

A value of *Region Bio Value-Chain Spec* greater than 1 indicates that biopharmaceuticals was over-represented (in terms of employment) in EA *r*. For example, Table A1 in the Appendix shows EAs with high bio value-chain specialization (and a high share of U.S. biopharmaceuticals employment) in the base year (1992), including New York-Newark-Bridgeport, NY-NJ-CT-PA, and Raleigh-Durham-Cary, NC (with a *Region Bio Value-Chain Spec* of 2.12 and 1.95, respectively).

Finally, we also control for the role of region size in firm location using the EA employment outside the bio value chain at time *t-1* (*Region employment (Outside Bio Value Chain)*). This variable will capture the potential benefits of locating in larger regions, including access to customers and urbanization economies.

4. EMPIRICAL ANALYSIS

Our analysis starts with a descriptive examination of the location-choice patterns of *new* biopharmaceutical establishments. We then explore the existence and heterogeneity of internal agglomerations and their differential effect by activity (Section 4.1) and the changes on the effect of external agglomerations when internal agglomerations are factored in (Section 4.2). While our main analysis focuses on the location of new establishments, we also consider the implicit location choice associated with growing an existing establishment. We argue that growth in existing establishments may be associated with internal agglomerations, and failing to consider these events as location choices could underestimate the relevance of internal agglomerations. For this reason, we complete our analysis by examining location patters for *all* types of establishments, new and existing (Section 4.3).

Table 2 provides a simple tabulation of the location choices of new biopharmaceutical establishments based on the intensity of external and internal agglomerations in the chosen location. The 1,226 location choices during 1993–2005 are divided into six categories based on whether they chose locations with low or high external agglomerations and low, medium, or high internal agglomerations the year prior to the expansion. In Table 2, *external agglomerations* in a location are defined as *high* if the location choice is among the top-10 EAs by share of U.S. biopharmaceuticals that also have high specialization. These top-10 locations account for around 50% of total U.S. biopharmaceuticals employment (e.g. see Table A1 for the list of Top-10 biopharmaceutical EAs in 1992). The rest of the EAs are broadly defined as locations with *low* external agglomerations. Internal agglomerations in a location are defined as high if the location is the firm's base EA for biopharmaceuticals, medium if the location has some firm employment (but is not the base EA); and low if the location is new to the firm (no preexisting firm presence).

Table 2 shows that a majority of location choices for new establishments (63%) occur in EAs with same-firm presence (internal agglomerations), with the rest (37%) in new locations (external drivers). Surprisingly, many location choices (69%) occur outside the top-10 biopharmaceutical EAs. For these cases of low-external locations, most expansions (38%) occur in locations with same-firm presence. There are also expansions in locations where both internal and external agglomerations are high (7%), but these are a small number, reducing the concern about firms choosing a top biopharmaceutical regional cluster for their first location and subsequently expanding into that same location. Importantly, very similar patterns occur if we drop the location choices for bio sales, which could be more geographically dispersed than manufacturing and R&D.[11] Overall, these findings suggest that *both* internal and external agglomerations matter for the location choices of biopharmaceutical firms.

4.1. Baseline Results: Internal Agglomerations Exist

To confirm the patterns identified in Table 2, we conducted a more systematic analysis of the location choices of new establishments. Table 3 shows the results of estimating equation 1 using a conditional logit specification. Recall that the coefficients are identified using across location variation in both external and internal agglomerations controlling for firm–activity-year fixed effects. Coefficients are transformed into odds-ratios to facilitate comparisons.

Model 1 introduces our main measure of external drivers: Economic Areas (EAs) fixed effects. These region dummies capture external agglomerations and endowments that drive the location choices of firms. Not surprisingly, these external drivers are important to explain location choices of new establishments,

[11] The main difference is that the percentages of low-internal and low-external locations decline slightly, to 25%, with an increase in the percent of location choices in EAs with same-firm presence.

as suggested by the fit of the model (log likelihood of -3,119, and pseudo R-squared of 0.22). Model 2 adds our main measures of internal agglomerations: same-firm employment in the bio value chain, other value chains, and support activities. Including these variables improves the model fit; the log likelihood goes up significantly (at the 1% level), suggesting the importance of considering internal agglomerations in location choices. The coefficients for internal agglomerations variables are all positive and significant: a 1% increase in *Firm Bio Value-Chain Employ* in a location leads to a 0.31% increase in the probability of choosing that location (i.e. *ln (1.365)/100*). The magnitude is statistically larger for firm bio value-chain activities than for other value chains and for support.[12] The relatively larger collocation with biopharmaceutical activities is consistent with the hypothesis that more meaningful interdependencies will occur among activities in the same value chain than among activities in different value chains.

Model 3 used alternative internal agglomerations variables that capture firm specialization (versus employment level) across regions. The goal is to assess whether firms locate in regions where the firms' activities are over-represented. The results are robust in statistical significance, but there are some changes in the magnitude of the effects: Internal agglomerations based on specialization tend to result in lower coefficients.

Table 4 shows the results for location choices by bio activity. Note that although internal agglomerations still have a positive and significant effect on location, the coefficients' magnitude varies by activity. Specifically, the importance of collocating a firm's own bio value chain activities is especially larger for manufacturing expansions (models 2 and 5), followed by R&D (models 1 and 4) and sales expansions (as suggested by the magnitude of the odds ratios for *Firm Bio Value-Chain Employ*). Indeed, for manufacturing expansions this is the only type of internal agglomerations that seems to matter. This suggests strong linkages across new and existing plants within the same value chain that may not be easy to transfer across different value chains.

In contrast, internal agglomerations in other value chains seem to matter only for the location of R&D and sales. These positive estimates for *Firm Other Value-Chains Employ* and *Firm Other Value-Chains Spec* suggest that R&D and sales activities may tap into complementary knowledge and tasks that reside in other value chains, or that there may be economies of scope associated with both activities. For example the same sales force can be used to market related products in different SIC codes. Finally, support activities only seem to matter for the location of new R&D establishments. The lack of significance for *Firm Support Employ and Firm Support Spec*, which includes employment in

[12] The difference in the estimated coefficients of *Firm Bio Value-Chain Employ* and *Firm Other Value-Chains Employ* is positive and significant at 1% (the Chi-square test is 13.97).

headquarters, in the location of manufacturing (Models 2 and 4) seems inconsistent with those studies that highlight the benefits for plant performance of proximity to the headquarters. One potential explanation for this discrepancy is that these studies examine the effect of the proximity between HQs and manufacturing abstracting from other activities of the value chain,[13] which reinforces the importance of exploring agglomeration economies considering all activities instead of isolated pairs.

Our analysis has shown strong evidence of internal agglomerations within the bio value chain. Now we turn to explore the extent of within- and across-activity collocation in biopharmaceuticals. To do so, in Table 5 we break down the variable *Firm Bio Value Chain* into its component activities (*Firm Bio R&D, Firm Bio Mfg*, and *Firm Bio Sales*), measured as employment and specialization.[14] Results suggest strong within-activity interdependencies for R&D and, especially, for manufacturing establishments. Manufacturing establishments are more likely to be located in EAs where there is a higher presence of same-firm manufacturing in biopharmaceuticals (positive and significant coefficient for *Firm Bio Mfg Employ* and *Firm Bio Mfg Spec* in models 2 and 5, respectively). Similarly, R&D establishments seem to collocate with same-firm bio R&D activity (models 1 and 4). This within-activity collocation is consistent with the idea that relevant information exchanges take place within-firm, same-activity (Kleinbaum et al. 2008; Van den Bulte and Moenaert 1998). For example, if R&D relies on the continuous exchange of tacit knowledge, we may expect bio R&D sites to collocate. Even if geographical dispersion may be desirable, firms need communication technology and sophisticated management practices to be able to break R&D or manufacturing activities across locations (see Fort 2011). Thus, same-activity collocation may reflect the lack of proper technology to overcome distance. In contrast, for sales establishments we find weak within-activity interdependencies. Sales activities that aim to cover geographically dispersed customers may account for low within-activity collocation.

We also find evidence of across-activity collocation for each type of bio activity. R&D and manufacturing benefit from across-activity collocation as well as the within-activity collocation. New R&D establishments are more likely to be located in EAs where there is a presence of same-firm manufacturing in biopharmaceuticals (positive and significant coefficient for *Firm Bio Mfg Employ* and *Firm Bio Mfg Spec* in models 1 and 4 respectively); and manufacturing establishments are also more likely to be located in EAs where there is presence of same-firm R&D in biopharmaceuticals (positive and significant coefficient for *Firm Bio R&D Employ* and *Firm Bio R&D Spec* in models 2 and 5,

[13] For example, if we re-estimate model 2 in Table 4 including only internal agglomerations in *Firm Support Employ*, the estimated coefficient of this variable becomes positive and significant (odd ratio of 1.33), suggesting positive co-location between new manufacturing facilities and support activity.

[14] Note that the findings in Table 5 are consistent with those in Table 4 (i.e. findings for *Firm Other Value Chains* and *Firm Support* are robust).

respectively). To further understand what could be driving this across activity collocation, we consider the sub-sample of firms that are diversifying into new bio activities (i.e. manufacturing firms opening their first R&D establishment or R&D firms opening their first manufacturing site). Firms that begin a new activity may need more coordination with related activities to successfully integrate the new activity with the rest of the value chain and therefore can inflate the role of internal agglomerations. If we exclude the observations that correspond to diversifying into new bio activities, the collocation of R&D and manufacturing becomes insignificant, but within-activity collocation remains relevant.

Note that new sales activity collocates with existing, same-firm manufacturing facilities, but new manufacturing activity does not collocate with sales. The asymmetry of internal agglomerations between sales and manufacturing is puzzling. The fact that sales offices are opened in existing manufacturing locations seems to be in part explained by firms beginning to commercialize their products (i.e. firm has no prior bio sales activity). The extent of collocation of sales with manufacturing may depend on the type of customer (other businesses of final consumers), but we cannot assess this with our data.

4.2. Agglomerations: External versus Internal

While our baseline econometric specification controls for external drivers using EAs fixed effects (Table 3), we alternatively use region–year external agglomeration variables in Table 6 to compare external and internal agglomerations. The external agglomeration variables are the specialization of the region in biopharmaceuticals (*Ln Region Bio Value-Chain Spec*) and the size of the region (*Ln Region Employment (outside Bio Value Chain*)). Model 1 only includes the external agglomeration variables. Both the specialization of the region in biopharmaceuticals and, particularly, the size of the region have a positive influence on the location choice of new biopharmaceutical establishments. Model 2 adds the internal agglomeration variables of same-firm employment in the bio value chain, other value chains, and support activities. The inclusion of these variables increases the fit of the model significantly (at the1% level). Importantly, the estimated coefficients of the external agglomeration variables decline, especially for region size (the odds ratio declines from 1.38 to 1.21 for regional specialization in the bio value chain; and from 2.42 to 1.88 for the region size). This suggests that the effect of external agglomerations in firm locations could be overestimated if we do not take into account the role of internal agglomerations.

Regarding the internal agglomerations effects, our findings are robust to those reported in model 2 of Table 3 (using EA fixed effects). We can compare internal and external agglomerations in biopharmaceuticals by examining the estimated coefficients of *Firm Bio Value Chain* versus *Region Bio Value-Chain Spec*. There is no statistical difference between their estimated coefficients, suggesting that internal and external agglomerations are equally important in the bio value chain. These results are robust in model 3 using the alternative internal agglomeration variables based on firm specialization. Overall,

these finding reinforce that both external and internal agglomerations are important, though further research will be required to understand the complementarities and trade-offs between these two types of agglomerations and their dynamics.

4.3. Expanding in an Existing Location

So far our analysis has focused on location choices for new establishments. However, internal agglomerations may also stem from expansion in existing establishments. For example, when firms grow within the same establishment, coordination and control costs will be the lowest and interdependencies the highest. Thus, failing to recognize that an expansion in an existing establishment is equivalent to choosing again the same location could underestimate the relevance of internal agglomerations. Furthermore, if the locations where a firm is already present are rich with external agglomerations, abstracting from growth in existing establishments may also underestimate the importance of external agglomeration. Therefore, to further understand the role of internal agglomerations on location choices, we estimate our baseline models using the location decisions for both new *and* existing establishments.

We define the location for an existing establishment as equal to 1 if the firm chooses EA *r* to increase employment in an existing establishment in bio activity *i* at time *t* at a level that is larger than the median size of new establishments opened during 1993–2005 for the same bio activity.[15] There were 2,337 expansions in existing establishments in the bio value chain in our sample.[16] Perhaps not surprisingly, expansions via new establishments are a much rarer phenomenon than expansions via existing establishments. Finally we generate the variable $Location\ all_{firt}$, which is equal to 1 if firm *f* either opens a new establishment or increases employment in an existing establishment above the threshold in economic area *r*, in focal bio activity *i* (R&D, manufacturing, or sales) at time *t*.

Table 7 introduces results of estimating equation 1 for expansions in both existing and new establishments (e.g. with dependent variable equal to $Location\ all_{firt}$). Model 1 only includes the EAs fixed effects; model 2 adds our measures of internal agglomerations based on employment level; model 3 measures based on specialization. Including these variables improve the model fit (the log likelihood goes up significantly relative to model 1). As expected, the main internal economies occur within the firm bio value chain, and the estimated coefficient is significantly larger than when we only model the locations of

[15] The pseudo-median size of new establishments was computed using the range of percentiles 40 to 60 to avoid disclosure problems. The thresholds corresponded to values of 7, 45, and 21 employees for bio R&D, bio manufacturing, and bio sales, respectively. We considered alternative thresholds to define internal expansions that take into account the size of the expanding firms. The new thresholds are the median size of new biopharmaceutical establishments by bio activity and firm size (firms are coded as small (large) if their size is below (above) the median size of expanding firms in 1992). The number of expansions only increases slightly with this new criterion.
[16] Of these, there were 1,282 in R&D, 546 in manufacturing, and 509 in sales.

new establishments. The odds ratio of *Firm Bio Value Chain* is 2.520 in model 2 versus 1.365 for expansions with new establishments (model 2, Table 3). Consistent with our prior findings in Table 3, there are also internal agglomerations in other value chains and in support activities.[17]

In summary, our findings in Tables 2–7 offer relevant new insights. First, internal agglomerations play a role in the location of biopharmaceutical activity and they arise in the bio value chain, in other value chains, and in support activities. Second, evidence of internal agglomerations emerges both within and across activities. Third, the effect of internal agglomerations varies by activity (R&D, manufacturing, and sales). Fourth, although external agglomerations are also important, their effect may be over-estimated if we abstract from the spatial organization of the firm.

5. EXPLORING FIRM HETEROGENEITY AND ROBUSTNESS ANALYSIS

Before settling on these results, we conducted several robustness tests using different ways to capture firm heterogeneity, samples, and explanatory variables.

5.1. Exploring Firm Heterogeneity

While our model accounts for firm–activity–year fixed effects, it is possible that additional unobserved firm heterogeneity influences firms' location decisions. We addressed this concern in several ways. First, we used "weighted" internal agglomeration with firm-specific weights to account for the extent of relatedness between activities of the value chain. Second, we explored the geographical diversification and activity diversification of firms since location choices in these cases could be more idiosyncratic, as explained below.

Weighted Internal Agglomerations: Exploring Firm Heterogeneity in the Value Chain. Our main internal agglomeration variables used so far assume that all activities in a firm value chain are equally related. Here we develop a new set of internal agglomeration variables that account for firm-level heterogeneity in the extent of relatedness between value chain activities. For a firm, two activities (say bio R&D and bio manufacturing) may have meaningful and geographically bounded linkages (knowledge flows, input-output links, shared labor and skills, etc.) and could then benefit from collocation, while other pair of activities of the firm may be less related (say bio R&D and bio sales).[18] To account for this, we develop internal economies variables that weigh more the employment that seems most relevant for

[17] We considered two alternative samples for expansions in existing establishments. First, we dropped internal expansions of establishments that were created after 1992 (25% of the expansions) because new establishments are expected to grow in a given location across time. The results in Table 7 are very similar after dropping these observations. Second, as described above, we considered alternative thresholds to define internal expansions that take into account the size of the expanding firms; the same findings hold.

[18] Firm communication technologies may work better to coordinate certain activities (e.g. well defined input–output links in manufacturing) than others (e.g. tacit knowledge flows in R&D).

the focal bio activity *i*. Similarly to equations 2a-to-4a (Section 3.2.1), the set of weighted internal agglomeration variables are defined as follows:

$$Firm\,Bio\ Value\text{-}Chain\,Employ\ (weighted)_{firt-1} = \sum_{k \in Bio-VC} w_{fikt-1} * emp_{fkrt-1} \quad (2b)$$

$$Firm\,Other\ Value\text{-}Chains\,Employ\ (weighted)_{firt-1} = \sum_{k \in Other-VCs} w_{fikt-1} * emp_{fkrt-1} \quad (3b)$$

$$Firm\,Support\,Employ\ (weighted)_{firt-1} = \sum_{k \in Support} w_{fikt-1} * emp_{fkrt-1} \quad (4b)$$

where *i* is the focal bio activity of a new establishment subject to location choice; *k* indexes all activities that map into the bio value chain (equation 2b), into other value chains (equation 3b), and into support (equation 4b); emp_{fkrt-1} is the employment in firm–EA–activity *k* the year prior to expansion; and w_{fik} is the firm-specific relatedness between pairs of activities *i* and *k* (e.g. bio R&D and bio manufacturing in equation 3b). Building on Porter (2003),[19] we calculated w_{fikt-1} using the "locational correlation" of employment. This measure is the correlation coefficient between firm employment in activity *i* and firm employment in activity *k* across regions (EAs) where the firm had any positive employment:[20] $LC_{fikt-1} = Correlation(Employment_{firt-1}, Employment_{fkrt-1})$. The greater the collocation of a pair of firm activities across regions in the past (LC_{fik}), the greater the extent of geographically bounded linkages between those activities. For a given pair of activities, some firms may have developed management practices or communication technologies that allow them to break down the value chain across locations (Fort 2011). In such a case, even if the pair of activities has meaningful linkages, the optimal level of collocation might be reduced (resulting in a lower LC_{fik} for these firms).

The locational correlation weights were then transformed to values between 0 and 1 ($w=(1+LC)/2$) to compute the internal agglomeration variables.[21] For example, if the bio activity *i* of a new establishment was R&D, the *Firm Bio Value-Chain Employ* variable in equation 2b was computed as follows: $emp_{firt-1} + w_{fi2830t-1} emp_{f2830rt-1} + w_{fi5120t-1} emp_{f5120rt-1}$; where $w_{fi2830t-1}$ and $w_{fi5120t-1}$ capture the firm-specific relatedness of R&D with manufacturing and with sales.

Table 8 shows the results of estimating equation 1 using the weighted internal agglomeration variables. Model 1 includes the weighted employment in firm bio value chain, other value chains, and support activities, while model 2 shows the weighted specialization variables. Findings using the

[19] Delgado, Porter, and Stern (2013) use locational correlation of employment at the region–industry level to capture many types of interdependencies between pair of industries.

[20] We allowed the relatedness between a pair of activities (or industries) to change by year because a firm's portfolio of EAs and activities changes as the firm expands.

[21] We assumed that if a firm had only one EA, w=1 for all pairs of activities in the location. We also assumed that the firm's new bio activities (i.e. the firm diversified into new activities) are related to its existing activities (w=1). In the sensitivity analysis we computed a simpler measure of firm collocation patterns using a Jaccard index.

weighted internal agglomeration variables are robust in statistical significance to those using the un-weighted variables, but there are some changes in the magnitude of the effects. Weighted internal economies tend to result in slightly lower coefficients. For example, the estimated odds ratio for *Firm Bio Value-Chain Employ* is 1.286 when weighted (model 1, Table 8) and 1.365 when unweighted (model 2, Table 3). This suggests that internal agglomerations are firm-specific and that accounting for firm-specific interdependencies may offer a more accurate estimate of the average role of internal economies.

Firm Geographical Diversification. There may be some unobserved attributes that drive firms to be specialized in one location (EA). Firms that have a single location could be problematic since the initial external drivers that pulled them to that location may not be disentangled from any internal forces inducing them to expand in the same location. We dropped observations if a firm had a single EA at *t-1*. Very few firms in our sample fall into this group and our findings are robust to their exclusion.

We also control for the diversification of firms into new locations. Entry into a new region (an EA where the firm had no prior employment) could be driven by unobserved firm attributes (e.g. firms with small portfolios of locations, R&D-oriented firms, decreasing internal agglomerations in existing facilities of the firm); and/or by changes in the external environment (e.g. radical innovations in the new location). We controlled for this by including a dummy equal to 1 if the firm had no employment in a location at *t-1* (*New Region*). As expected, firms are less likely to choose locations with no firm presence. All findings are robust to the inclusion of this dummy, reinforcing the importance of different types of internal agglomerations (within the focal value chain, across value chains, within activities, and across activities).

Firm Activity Diversification. We examined the activity diversification of biopharmaceutical firms. Smaller biotech firms may be more specialized (often in R&D), while big pharma firms tend to be more diversified, with multiple activities in the bio value chain as well as more diversification into other value chains (e.g. medical devices, downstream chemical products). The extent of firm diversification may capture an unobserved strategy choice of the firm when it was founded (Gans et al. 2013). Some firms' strategy is to specialize in some core activities and rely on other firms for complementary assets. This may be the case for many small biotech firms. In contrast, the strategy of other firms may be to choose a more vertically integrated value chain (big pharma firms), perhaps relying more on internal agglomerations.

We preliminarily explored how a firm's degree of diversification influences the extent of internal agglomerations by considering three distinct types of firms that open new bio establishments: *Bio Diversified*, *Bio Specialized*, and *Very Bio Specialized*. *Bio Diversified* firms are those that have two or more bio activities (SICs 8731-2830, or 2830-5120, or 8731-5120, or all) at *t-1*. The rest are bio-specialized firms, which have a single bio activity at *t-1* (SICs 8731 or 2830 or 5120) but could have

employment in other value chains or support activities.[22] Finally, *Very Bio Specialized* firms are a sub-set of bio-specialized firms that only have employment in a single bio activity at *t-1* (i.e. the firm only has employment in SIC-8731 or SIC-2830).

We estimate equation 1 separately for bio-diversified and bio-specialized firms, and examine internal agglomeration in bio value chains, other value chains, and support activities (using the same specification as in model 2 of Table 3). For both sub-samples we find similar patterns to those in our baseline results: internal agglomerations matter, they arise primarily within the bio value chain, and they arise to a lesser extent with other value chains. Perhaps surprisingly, the magnitude of internal agglomeration in the bio value chain seems smaller for bio-diversified firms than for bio-specialized firms. Finally, we also dropped observations if the firm at *t-1* was very bio-specialized. Although there is a non-negligible number of observations in this group (~7.5% of all location choices), results obtained by dropping them are similar in magnitude and significance to those in Table 3. Although the analysis suggests that internal agglomerations matter for specialized and diversified firms, more research will be required to understand how the extent of firms' vertical integration across the value chain impacts their ability to exploit internal and external agglomerations.

5.2. Alternative Samples and Variable Definitions

To address some additional econometric concerns, we estimated additional models using different samples and variable definitions. In terms of sampling, we applied specific criteria to identify biopharmaceutical firms from the LBD data. Because it may be possible that some of these firms are not actually in biopharmaceuticals, we re-estimated our models using only firms that were in the BioScan directory. Results using this subsample of firms are very similar to those obtained from the extended sample across specifications.

In terms of defining value chains, we tested whether our method of mapping bio value chain activities to SICs drives our findings on within- and across-activity collocation. Specifically, we defined new internal agglomeration variables considering a single value chain and examined collocation across broadly defined activities (i.e. without separating bio value chains from other value chains). This redefinition has the benefit of reducing sparcity for the explanatory variables as well as relaxing the definition of the core value chain of a biopharmaceutical firm. Findings with this new definition are reported in Table A2 and confirm prior results in Table 5 regarding within-activity and across-activity internal agglomerations.

[22] In our sample, many of the location choices for new bio establishments (more than 60%) correspond to *Bio Specialized* firms.

6. CONCLUSIONS

This paper examined the extent to which the geographical location of distinct activities in the value chain (manufacturing, R&D, or sales) is explained by internal agglomerations (i.e. geographically bounded intra-firm linkages) and external agglomerations. We proposed a conceptual framework where both internal and external agglomerations drive location choices. External agglomerations act as centrifugal forces that drive firms to disperse their activities geographically; internal agglomerations act as centripetal forces that drive within-firm collocation. Based on an exhaustive review of relevant papers across disciplines, we identified five mechanisms through which geographic proximity may lead to internal agglomerations: coordination, control, knowledge and information flows, economies of scope and scale in internal labor markets, and access to intermediate inputs. The latter three parallel the mechanisms identified as external agglomerations by Marshall (1920). Although the focus of our paper is empirical and we cannot unbundle these alternative mechanisms, we believe that our conceptual framework is a first step toward building theory that advances our understanding of internal agglomerations and their role on firm performance.

We tested our conceptual framework using data on the location of new establishments opened by biopharmaceutical firms in the continental U.S. during 1993–2005. Our findings offer several relevant insights. First, internal agglomerations have a relevant, positive impact on location. They are especially important within the focal value chain (bio value chain), and they also arise in related value chains of the firm. Second, the effects of internal agglomerations vary by activity, and they arise both within an activity (e.g. among plants) and across activities (e.g. between sales and manufacturing). Third, external agglomerations also have a relevant positive impact on location, but its effect may be overestimated if internal agglomerations are not considered.

These insights offer important contributions to the literature. They suggest that focusing on one side of agglomerations (external or internal) may produce biased estimates due to omitted variables. Both internal and external agglomerations play a role in the location of firms. Regarding internal agglomerations, failing to consider *all* activities in the value chain may lead to omitted relationships that can also bias results. For example, the location of firm R&D may be influenced not only by the presence of same-firm R&D and manufacturing, but also by the presence of same-firm sales and support activities. Furthermore, to better understand the links between activities (e.g. manufacturing and R&D) we need to examine the location decision of each activity.

Considering just one category of expansions (new establishments) may also underestimate the importance of internal agglomerations. For example, we show that if we also consider firms expansions by increasing activity levels in existing establishments, the estimated effect of internal agglomerations is

greater. After all, the default alternative to a new location is to stay in an existing location—an insight that is absent in most of the location literature.

All of these issues—the relevance of internal and external agglomerations, the distinct relationships between pairs of activities in the value chain, and the fact that effects may vary by the type of expansion—emphasize the need for a comprehensive framework, both at the theoretical level and at the empirical level, to understand the spatial organization of firms.

Several avenues for further research remain. First, the current empirical analysis focuses on location decisions of biopharmaceutical firms. One could extend the analysis to other industries with different degree of modularity (Baldwin and von Hippel 2011). Biopharmaceuticals has lower modularity than other industries, such as semiconductors and automotives, and this could affect the extent and types of agglomerations. Second, the effect of internal and external agglomerations may vary across time and our results may reflect a specific stage in the life cycle of biopharmaceutical firms. Industry maturity influences firms' location choices (e.g. Dumais et al. 2002, Duranton and Puga 2001). Relatedly, the life cycle of firms (younger versus older firms, smaller versus larger) may also influence location choices. For example, external agglomerations are especially important for startups (Delgado et al. 2010, Glaeser and Kerr 2009) and smaller firms (Henderson 2003, Rosenthal and Strange 2009), but as firms get larger, internal agglomerations may become more relevant. Future research that explores these dynamics would greatly enrich our understanding of the spatial organization and performance of firms. Third, in this paper we assumed that firms' location decisions are driven by the goal of maximizing profitability, and we abstracted from the role of location choices on firm performance. In related work, we plan to examine the effects of internal and external agglomerations on the subsequent performance of new facilities and the firm as a whole. Fourth, we have taken into account the spatial organization of firms with a novel approach, but we did not examine the business practices that shape firms' spatial organization. The location choice and performance of firms may depend on firm management practices that facilitate intra-firm and inter-firm interactions even with geographic separation (e.g. outsourcing practices, labor mobility practices, monitoring and control practices, IT investments; see e.g. Baldwin and von Hippel 2011, Choudhury 2010, Fort 2011). Finally, we did not look at overseas expansions due to data limitations. This omission is less important in our empirical setting since most American biopharmaceutical firms did not globalize during our study period, and the U.S. provided better biopharmaceutical clusters. Nonetheless, future research should consider the role of internal and external agglomerations when firms expand globally (e.g. Beugelsdijk et al. 2010).

REFERENCES

Aharonson, B., J. Baum, A. Plunket. 2008. Inventive and univentive clusters: The case of Canadian biotechnology. *Research Policy* **37** 1108–1131.

Adams, J., A. Jaffe. 1996. Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data. RAND Journal of Economics, The RAND Corporation, vol. 27(4), pages 700-721,

Alcacer, J. 2006. Location choices across the value chain: How activity and capability influence collocation. *Management Sci.* **52**(10) 1457–1471.

Alcacer, J., W. Chung. 2007. Location strategy and knowledge spillovers. *Management Sci.* **53**(5) 760–776.

Alcacer, J., W. Chung. 2013. Location strategies for agglomeration economies. Harvard Business School Working Paper Series 10–701.

Baldwin, C., E. von Hippel. 2011. Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization Science* **22**(6) 1399–1417.

Beugelsdijk, S., P. McCann, R. Mudambi. 2010. Place, space, and organization: economic geography and the multinational enterprise. *J. of Economic Geography* **10**(4) 485–493.

BioScan. 1992. *The Worldwide Biotech Industry Reporting Service*. AHC Media, Volume 6.

Cohen, W., D. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* **35**(1) Special Issue on Technology, Organizations, and Innovation: 128–152.

Di Minin, A., & Bianchi, M. 2011. Safe Nests in Global Nets: Internalization and Appropriability of R&D in Wireless Telecom. ***Journal of International Business Studies***, 42(7): 910-934.

Delgado, M., M. E. Porter, S. Stern. 2012. Clusters, convergence, and economic performance. NBER Working Paper 18250.

Delgado M., M. E. Porter, S. Stern. 2010. Clusters and entrepreneurship. *J. of Econom. Geography* **10**(4) 495–518.

Delgado, M., M.E. Porter, S. Stern. 2013. Defining clusters of related industries. Working Paper.

Dumais, G., G. Ellison, E.L. Glaeser. 2002. Geographic concentration as a dynamic process. *Rev. of Economics and Statistics* **84**(2) 193–204.

Duranton, G., D. Puga. 2001. Nursery cities: urban diversity, process innovation, and the life cycle of products. *Amer. Econom. Rev.* **91**(5) 1454–1477.

Furman, J., M. Kyle, I. Cockburn, R. Henderson. 2005. Public & Private Spillovers, Location, and the Productivity of Pharmaceutical Research," *Annales d'Economie et de Statistique.*

Gans, J., F. Murray, S. Stern. 2013. *Choosing an Entrepreneurial Strategy*. Mimeo.

Giroud, X. 2013. Proximity and Investment: Evidence from Plant-Level Data, *Quarterly J. of Economics*, forthcoming.

Glaeser, E.L., W.R. Kerr. 2009. Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain? *J. of Economics and Management Strategy* **18**(3) 623–663.

Grossman, S.J., O.D. Hart. 1986. The costs and benefits of ownership: A theory of vertical and lateral integration. *J. of Political Economy* **94**(4) 691–719.

Feldman, M.P., D. Audretsch. 1999. Innovation in cities: Science-based diversity, specialization, and localized competition. *European Econom. Rev.* **43** 409–429.

Feldman, M.P. 2003. The locational dynamics of the US biotech industry: knowledge externalities and the anchor hypothesis. *Industry and Innovation* **10** 311–328.

Feldman, M.P., Y. Schreuder. 1996. Initial advantage: The origins of geographic concentration of the pharmaceutical industry in the Mid-Atlantic region. *Industrial and Corporate Change* **5** 839–862.

Fort, T. 2011. Breaking up is hard to do: Why firms fragment production across locations. Working Paper.

Hamilton, B., J. Nickerson, H. Owan. 2003. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *J. of Political Economy* **111**(3)

Hanson, G. 2001. Scale economies and the geographic concentration of industry J Econ Geogr (2001) 1(3): 255-276

Helfat, C. 1997. Know-how complementarities and knowledge transfer within firms: The case of R&D." *Strategic Management J.* **18**(5) 339–360.

Henderson, J.V. 2003. Marshall's Scale Economies. *J. of Urban Econom.* **53** 1–28.

Henderson, J.V., Y. Ono. 2008. Where do manufacturing firms locate their headquarters? *J. of Urban Economics* **63**(2) 431–450.

Jacobs, J. 1984. Cities and the Wealth of Nations: Principles of Economic Life. Random House, New York.

Kalnins, A., F. Lafontaine. 2013. Too far away? The effect of distance to headquarters on business establishment performance. *AEJ-Micro*, forthcoming.

Kenney, K., R. Florida. 1994. The organization and geography of Japanese R&D: Results from a survey of Japanese electronics and biotechnology firms. *Research Policy* **23**(3) 305–322.

Ketokivi, M. 2006. When does co-location of manufacturing and R&D matter? Discussion Papers 1051, *The Research Institute of the Finnish Economy*.

Kleinbaum, A. M., T. E. Stuart, M. Tushman. 2008. Communication (and coordination?) in a modern, complex organization. Harvard Business School Working Paper 09-004.

Krugman, P. 1991. Increasing returns and economic geography. *J. of Political Economy* **99** 483–99.

Marshall, A. 1920. *Principles of Economics; An Introductory Volume.* Macmillan and Co., London.

Pisano, G.P. 1996. *The Development Factory: Unlocking the Potential of Process Innovation.* Harvard Business School Press, Boston, MA.

Porter, M. E. 1996. What is strategy? *Harvard Business Review* November 61–78.

Porter, M. E. 1998. Clusters and Competition: New Agendas for Companies, Governments, and Institutions. In M.E. Porter (ed), *On Competition*. Harvard Business School Press, Boston, MA.: 197–299.

Porter, M.E. 2003. The Economic performance of regions. *Regional Studies* **37** 549–578.

Ricardo, D. 1817. *On the Principles of Political Economy and Taxation*. John Murray, Albemarle-Street Eds., London.

Rosenthal, S.S., W.C. Strange. 2004. Evidence on the nature and sources of agglomeration economies. In *Handbook of Regional and Urban Economics*, vol. 4, J.V. Henderson and J.F. Thisse (eds.), Amsterdam: Elsevier North-Holland.

Rosenthal, S.S., W.C. Strange. 2009. Small Establishments/Big Effects: Agglomeration, Industrial Organization, and Entrepreneurship. In *The Economics of Agglomeration*, E. Glaeser (ed). Cambridge, MA.

Tate, G., L. Yang. 2011. The bright side of corporate diversification evidence from internal labor markets. Working Paper.

Tecu, I. 2011. Where do firms patent? Measuring intra-firm spillovers for R&D. Working Paper.

Van den Bulte, C., R.K. Moenaert. 1998. The effects of R&D team co-location on communication patterns among R&D, marketing, and manufacturing. *Management Sci.* **44**(11) 1–18.

Zucker, L., M. Brewer, M. Darby. 1998. Intellectual capital and the birth of U.S. biotechnology enterprises. *Amer. Econom. Rev*. **88**(1) 290–306.

Table 1.a: Top-10 Industry Codes (SICs) of BioScan Firms in 1992

Value-Chain Activity	SIC Label	SIC code	No. firms	% of firms (out of 395)
Bio R&D	Commercial physical & biological research	**8731**	**168**	**43%**
Bio Mfg	Includes SICs 2833-to-2836	**2830***	**127**	**32%**
Bio Mfg	Diagnostic substances	2835	55	14%
Bio Mfg	Pharmaceutical preparations	2834	51	13%
Bio Mfg	Biological products	2836	47	12%
Bio Sales	Drugs, proprietaries, & sundries-wholesale	**5120**	**45**	**15%**
Support	Holding companies	6719	42	11%
Other Mfg	Industrial organic chemicals, n.e.c.	2869	40	10%
Other Sales	Medical & hospital equipment-wholesale	5169	37	9%
Support	Medical laboratories	8071	30	8%
Other Mfg	Surgical medical instruments	3841	29	8%

Note: 2830 (sum firms in 2833-to-2836). Sample of 395 BioScan firms, 112 of them are multi-unit firms. Source: LBD data and BioScan industry Report.

Table 1.b: Variables and Descriptive Statistics (Mean and Standard Deviation)

	Definition	Obs.=51k	
Location $_{firt}$	Dummy 1 if firm chooses Economic Area (EA) *r* to locate a new establishment in bio activity *i*	0.02 (0.15)	
Internal Agglomerations $_{t-1}$:		Employment* (1)	Specialization (2)
Firm Bio Value Chain	Firm bio value-chain (SICs 8731, 2830, 5120) employment or specialization in EA	21.84 (291.44)	0.17 (4.24)
Firm Bio R&D	Firm bio R&D (SIC 8731) employment or specialization in EA	7.11 (161.17)	0.14 (6.62)
Firm Bio Mfg	Firm bio mfg (SIC 2830) employment or specialization in EA	9.79 (199.05)	0.06 (1.29)
Firm Bio Sales	Firm bio sales (SIC 5120) employment or specialization in EA	4.94 (99.91)	0.26 (6.86)
Firm Other Value Chains	Firm other value-chains employment or specialization in EA	14.97 (137.41)	0.19 (2.31)
Firm Support	Firm support-activity employment or specialization in EA	4.13 (86.41)	0.27 (10.22)
External Agglomerations $_{t-1}$:			
Region Bio Value-Chain Spec	EA specialization in bio value chain (excluding firm own employment)	0.69 (0.90)	
Region Employment (outside Bio Value Chain)	EA employment (excluding bio value-chain employment)	729977.5 (1134466)	

Note: Number of observations is rounded to facilitate disclosure. Source: LBD data. *Column 1 reports the internal agglomerations variables based on employment level; and column 2 based on employment specialization. See Section 3 for detailed descriptions of these variables.

Table 2: Location Choices for New Bio Establishments, 1993–2005: Distribution by Intensity of Internal and External Agglomerations

			External Agglomerations $_{t-1}$: Low	External Agglomerations $_{t-1}$: High Top-10 EAs	
Internal	**Low :**	EAs with zero firm employ	372 (30%)	84 (7%)	**37%**
Agglomerations $_{t-1}$:	**Medium:**	EAs with any firm employ	419 (34%)	213(17%)	**52%**
	High:	EA with largest firm Bio employ	54 (4%)	84 (7%)	**11%**
			69%	**31%**	**100%**

Notes: In parentheses, % of all bio new establishments (out of 1,226). See Table A1 for a list of top-10 EAs. The distribution is very similar if bio sales expansions are excluded.

Table 3: Location of New Establishments: Collocation Within and Outside the Bio Value Chain (Conditional Logit; Odd Ratios)

	Y_{firt}=Location Choices for New Establishment in Bio Activity *i*		
		Internal Vars. $_{frt-1}$: Employment	**Internal Vars. $_{frt-1}$: Specialization**
	(1)	(2)	(3)
Ln Firm Bio Value Chain		1.365**	1.195**
		(0.052)	(0.044)
Ln Firm Other Value Chain		1.126**	1.097**
		(0.038)	(0.017)
Ln Firm Support		1.118**	1.100**
		(0.042)	(0.021)
Region (EA) Fixed Effects	Yes	Yes	Yes
Groups: Firm–Bio–Activity–Year	380	380	380
Pseudo R-sq	0.22	0.26	0.25
Log-Likelihood	-3119	-2935	-2970
Firms	157	157	157
Obs	51k	51k	51k

Notes: Standard errors are clustered by firm. ** Significant at 1%. *Significant at 5% level.

Table 4: Location of New Establishments by Type of Bio Activity: Collocation Within and Outside the Bio Value Chain (Conditional Logit; Odds Ratios)

	Y_{firt}=Location Choices for New Establishment in Bio Activity *i*					
	Internal Vars. $_{frt-1}$: Employment			**Internal Vars. $_{frt-1}$: Specialization**		
	Bio R&D	Bio Mfg	Bio Sales	Bio R&D	Bio Mfg	Bio Sales
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Firm Bio Value Chain	1.286**	1.578**	1.253**	1.131*	1.347**	1.133*
	(0.093)	(0.130)	(0.065)	(0.071)	(0.063)	(0.060)
Ln Firm Other Value Chains	1.306**	1.046	1.093*	1.140**	1.038	1.074**
	(0.100)	(0.047)	(0.040)	(0.033)	(0.045)	(0.025)
Ln Firm Support	1.141*	1.070	1.035	1.109**	1.039	1.042
	(0.071)	(0.109)	(0.076)	(0.043)	(0.057)	(0.036)
EA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Groups: Firm-Bio-Activity-Year	192	72	116	192	72	116
Pseudo R-sq	0.26	0.21	0.22	0.26	0.18	0.21
Log-Likelihood	-1583	-257	-841	-1596	-267	-849
Firms	92	48	60	92	48	60
Obs.	23k	3k	9k	23k	3k	9k

Notes: Standard errors are clustered by firm. ** Significant at 1% level. *Significant at 5% level.

Table 5: Location of New Establishments: Collocation Within and Across Activities (Conditional Logit model; Odd ratios)

	Y_{firt}=Location Choices for New Establishment in Bio Activity *i*					
	Internal Vars $_{frt-1}$: Employment			Internal Vars $_{frt-1}$: Specialization		
	Bio R&D	Bio Mfg	Bio Sales	Bio R&D	Bio Mfg	Bio Sales
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Firm Bio R&D	1.282**	1.534**	1.105	1.181*	1.327**	1.095
	(0.101)	(0.224)	(0.112)	(0.102)	(0.113)	(0.066)
Ln Firm Bio Mfg	1.394**	1.529**	1.350**	1.582**	1.747**	1.542**
	(0.177)	(0.099)	(0.094)	(0.223)	(0.130)	(0.107)
Ln Firm Bio Sale	0.917	1.039	1.007	0.977	1.054	1.043
	(0.250)	(0.153)	(0.070)	(0.169)	(0.077)	(0.071)
Ln Firm Other Value Chains	1.303**	1.040	1.106**	1.141**	1.033	1.076**
	(0.105)	(0.075)	(0.039)	(0.034)	(0.038)	(0.026)
Ln Firm Support	1.131*	0.996	0.999	1.107**	1.013	1.021
	(0.070)	(0.089)	(0.073)	(0.043)	(0.050)	(0.037)
EA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Groups: Firm–Bio–Activity–Year	192	72	116	192	72	116
Pseudo R-sq	0.27	0.21	0.22	0.26	0.20	0.23
Log-Likelihood	-1581	-257	-841	-1591	-261	-833
Firms	92	48	60	92	48	60
Obs.	23k	3k	9k	23k	3k	9k

Notes: Standard errors are clustered by firm. ** Significant at 1% level. *Significant at 5% level.

Table 6: Location of New Establishments: Internal v. External Agglomerations (Conditional Logit model; Odds ratios)

	Y_{firt}=Location Choices for New Establishment, Bio Activity *i*		
		Internal Vars $_{frt-1}$:	
		Employment	**Specialization**
	(1)	(2)	(3)
Ln Firm Bio Value Chain		1.394**	1.221**
		(0.043)	(0.038)
Ln Firm Other Value Chains		1.107*	1.107*
		(0.044)	(0.018)
Ln Firm Support		1.142**	1.114**
		(0.040)	(0.017)
Ln Bio Value Chain Spec Region	1.377**	1.214*	1.256**
	(0.111)	(0.111)	(0.093)
Ln Employment Region	2.418**	1.878**	1.875**
(Outside Bio Value Chain)	(0.123)	(0.096)	(0.088)
Groups: Firm–Bio–Activity–Year	380	380	380
Pseudo R-sq	0.16	0.22	0.22
Log-Likelihood	-3330	-3092	-3124
Firms	157	157	157
Obs	51k	51k	51k

Notes: Standard errors are clustered by firm. ** Significant at 1%. *Significant at 5% level.

Table 7: Location of New and Existing Establishments (Conditional Logit model; Odd ratios)

	Y_{firt}=Location Choices for New and Existing Establishments, Bio Activity *i*		
		Internal Vars $_{frt-1}$:	
		Employment	**Specialization**
	(1)	(2)	(3)
Ln Firm Bio Value Chain		2.520**	1.814**
		(0.100)	(0.062)
Ln Firm Other Value Chains		1.098**	1.128**
		(0.033)	(0.017)
Ln Firm Support		1.082*	1.136**
		(0.035)	(0.015)
EA Fixed Effects	Yes	Yes	Yes
Firm–Bio–Activity–Year FEs	Yes	Yes	Yes
Pseudo R-sq	0.26	0.57	0.54
Log-Likelihood	-9511	-5556	-5864
Firms	335	335	335
Obs.	216k	216k	216k

Notes: Standard errors are clustered by firm. ** Significant at 1% level. *Significant at 5% level. Results are robust to using the weighted internal-agglomeration variables.

Table 8: Location of New Establishments: Collocation Within and Outside the Bio Value Chain. Weighted Internal Agglomeration Variables (Conditional Logit model; Odd ratios).

	Y_{firt}= Location Choices for New Establishment, Bio Activity *i*	
	Weighted Internal Vars $_{firt-1}$:	
	Employment	**Specialization**
	(1)	(2)
Ln Firm Bio Value Chain $_{firt-1}$	1.286**	1.249**
	(0.042)	(0.058)
Ln Firm Other Value Chains $_{firt-1}$	1.118**	1.099**
	(0.033)	(0.018)
Ln Firm Support $_{firt-1}$	1.109**	1.101*
	(0.036)	(0.021)
EA Fixed Effects	Yes	Yes
Groups: Firm–Bio–Activity–Year	380	380
Pseudo R-sq	0.26	0.25
Log-Likelihood	-2935	-2971
Firms	157	157
Obs.	51k	51k

Notes: Standard errors are clustered by firm. ** Significant at 1% level. *Significant at 5% level.

APPENDIX

Table A1: Top-10 EAs by Share of U.S. Bio Value-Chain Employment in 1992 (with High Employment Specialization)

EA Name	Share U.S. Bio Value-Chain Employment	Bio Value-Chain Specializaion	Bio Value-Chain Employment
New York-Newark-Bridgeport, NY-NJ-CT-PA	18.5%	2.12	101972
Chicago-Naperville-Michigan City, IL-IN-WI	6.7%	1.58	36661
Washington-Baltimore-Northern Virginia, DC-MD-VA-WV	4.7%	1.52	26010
Boston-Worcester-Manchester, MA-NH	4.7%	1.37	25678
Philadelphia-Camden-Vineland, PA-NJ-DE-MD	4.4%	1.61	24149
San Jose-San Francisco-Oakland, CA	4.3%	1.29	23858
Raleigh-Durham-Cary, NC	1.7%	1.95	9362
Indianapolis-Anderson-Columbus, IN	1.7%	1.36	9107
Knoxville-Sevierville-La Follette, TN	1.5%	4.20	8274
San Diego-Carlsbad-San Marcos, CA	1.5%	1.69	8262

Notes: Authors' calculations based on County Business Patterns data. The criterion for high bio value-chain specialization is a value greater than the percentile 80th of the Location Quotient (See equation 2 in Section 3).

Table A2: Location of New Establishments: Collocation Within and Across Activities with a Single Value Chain (Conditional Logit; Odd Ratios).

	Y_{firt}=Location Choices for New Establishment in Bio Activity *i*		
	Bio R&D	Bio Mfg	Bio Sales
Single Value Chain:	(1)	(2)	(3)
Ln Firm R&D Employ	1.304**	1.374**	1.162
	(0.104)	(0.166)	(0.113)
Ln Firm Mfg Employ	1.215*	1.471**	1.243**
	(0.111)	(0.104)	(0.069)
Ln Firm Sales Employ	1.214	1.077	1.062
	(0.136)	(0.124)	(0.061)
Ln Firm Support Employ	1.150*	1.034	1.023
	(0.075)	(0.085)	(0.069)
EA Fixed Effects	Yes	Yes	Yes
Groups: Firm–Bio–Activity–Year	192	72	116
Pseudo R-sq	0.26	0.21	0.22
Log-Likelihood	-1585	-258	-840
Firms	92	48	60
Obs.	23k	3k	9k

Notes: Standard errors are clustered by firm. ** Significant at 1%. *Significant at 5% level. Findings are robust to using the weighted internal-agglomeration variables.