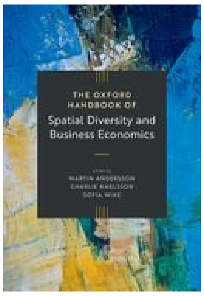


**The Oxford Handbook of Spatial Diversity and Business Economics**

Martin Andersson (ed.) et al.

https://doi.org/10.1093/oxfordhb/9780198866190.001.0001

**Published online:** 22 July 2025 **Published in print:** 04 September 2025 **Online ISBN:** 9780191898396 **Print ISBN:** 9780198866190

Search in this book

CHAPTER

# 15 More Ways to Rome: The Agglomeration and Firm Productivity Relationship through a Configurational Lens

Daniël Speldekamp, Joris Knoben, Frank van Oort

https://doi.org/10.1093/oxfordhb/9780198866190.013.0023 Pages 372–393

**Published:** 22 July 2025

## Abstract

A recent stream of studies demonstrates that agglomeration heterogeneously affects firms. There are many heterogeneous and interacting conditions in the firm-agglomeration relation, ranging from variations in size, age, degree of innovativeness, automation, resources used, networks exploited, and managerial styles applied in firms, to sectoral and occupational composition, and regional contexts of agglomeration (either specialization, diversity, or a mixture of these). To deal with this plethora of conditions this chapter suggests a reorientation for future agglomeration studies by employing configurational theory and analysis preceding or alongside more traditional regression techniques. In this chapter, we employ a configurational toolset (fsQCA) that explores the potentials of heterogeneous firm- and agglomeration-level conditions related to firm productivity in direct comparison to a previously employed econometrical analysis. This comparison shows that a well-defined configurational analysis is able to accurately select the conditions that matter significantly in the econometrical analysis. It is argued that when research aims to explore relevant dimensions in the firm-agglomeration relation ('what relates significantly to productivity?'), the advantages of QCA are found in its sensitivity to detect interacting conditions and dimensions even when they are empirically rare. QCA does not (yet) directly contribute to identifying causal effects compared to some econometric methodological applications ('what causes productivity?'), but we show that the exploratory function of QCA is ideal for mapping complex prior knowledge, suggesting which (combinations of) ways lead to Rome.

**Keywords:** agglomeration, specialization, diversity, firm productivity, QCA analysis, identification

**Subject:** Business Strategy, Innovation, Business and Management

**Series:** Oxford Handbooks

**Collection:** Oxford Handbooks Online

## 15.1 Introduction

DESPITE differences between many formal modelling approaches to regional growth and innovation, and specifically the role of firms therein, much common ground can be found (McCann and Van Oort, 2019). In each of the different approaches, from the new economic geography, urban economics, geography of innovation, and managerial economics to institutional-evolutionary economic geography, 'the role of agglomerations is regarded as being a crucial element of regional performance, and the common element is the issue of local knowledge generation, accumulation and spillovers' (ibid., p. 16). A pluralistic economic geography discipline with diverging narratives notwithstanding, the common outcome is that of more productive and faster growing firms in agglomerations (Martin, 2021).

In this broad field of study, the microeconomic foundations of these external knowledge effects have gained ever more attention over time. Intuitively, it seems clear that the higher the average level of human capital (knowledge) or the more spatially concentrated economic agents are, the more 'luck' these agents will have with their meetings and cooperation (formal and informal), increasing the diffusion and growth of knowledge (Rauch 1993, p. 381). Face-to-face meeting opportunities, the internal knowledge base of firms (absorptive capacities), the size of firms, and compositional effects in larger agglomerations are all argued to be related to productive firm-agglomeration interactions ( Knoben et al., 2016).

Despite this consensus on the outcomes of agglomeration forces—the increase in productivity as firms
p. 373 agglomerate—the exact circumstances at both the firm ↳ and agglomeration levels of analysis are subject to heterogeneous contextualization (Andersson et al., 2019; Stavropoulos et al., 2020). Heterogeneous agglomeration forces are found in seemingly competing explanations of urbanization- (i.e. diversity) and localization- (i.e. specialization) based local economies (Beaudry and Schiffauerova, 2009; De Groot et al., 2016), in knowledge intensities of and spillovers between firms (Knoben et al., 2016), and in compositional effects of firms according to size, age, sectoral membership, and technology adoption (Van Oort et al., 2012; Faggio et al., 2017; Steijn et al., 2022). In addition, pluralistic approaches in related disciplines researching the agglomeration-firm productivity relation have led to an ambiguity of conceptualizations and measurement methodologies that hampers the external validity and generalizability of many research findings (Van Oort, 2015; Speldekamp et al., 2020a). This in itself is not surprising given that any method employed to research relations with observational data (in our case of firms and geographical contexts), measured and applied by researchers, is highly dependent on theory and local institutional knowledge (Cunningham, 2021, pp. 8–9). Even when similar methods are utilized, competing theories exist because of the complexity and possible interactions of (multilevel) conditions that can be focused on. For instance, there is disagreement on which types of firms benefit from agglomeration economies. Is it the large and strong firms with the greatest ability to absorb external knowledge (McCann and Folta, 2011), or the small, weak firms with the most to gain and least to lose (Grillitsch and Nilsson, 2017)? Alternatively, could it be average firms occupying a 'Goldilocks position'—leaking less knowledge than their strong counterparts, and having a greater absorptive capacity than weak firms (Frenken et al., 2015; Knoben et al., 2016; Hervas-Oliver et al., 2018)? Theoretical and empirical backing exists for each of these arguments: there are more ways (firm size circumstances) that lead to (correlate with or even cause) Rome (productivity, growth, or innovation).

The fact that more than one way leads to Rome in firm-agglomeration analysis also fuels the application of pluriform methods. There is a need for establishing common ground in these methods, to safeguard that diverging outcomes are not attributable to the methodologies used. This chapter explores the common ground and potential crossover between two methodologies that have been separately used to research the firm-agglomeration performance relationship, namely that of econometric-based multilevel modelling (see for a recent overview of this method in economic and managerial applications Oshchepkov and Shirokanova, 2022) and fuzzy-set Qualitative Comparative Analysis (fsQCA) ( Fainshmidt et al., 2020; Fiss, 2011 provide an overview for international business and organization research). Both types of analysis have been presented as ways forward in the firm-agglomeration discussion (respectively by Van Oort et al., 2012; Knoben et al., 2016; Speldekamp, 2020b). Typically, multilevel modelling can deal with relatively large datasets that distinguish firm from agglomeration (contextual) levels, while fsQCA is suitable even for small datasets that are comprised with information on the same multilevel aspects. Both methods cannot easily deal with causal inference at the lowest level (Oshchepkov and Shirokanova, 2022, pp. 2–3; Speldekamp, 2021, pp. 140–141), but this does not
p. 374 mean it ↳ is impossible. Oshchepkov and Shirokanova (2022, p. 7) falsify the common critique of multilevel models being unable to use instrumental variables, regression discontinuity, or difference-in-difference estimation. In principle such techniques can be incorporated, yet the same large challenges exist as in other econometric approaches (identification strategies, appropriate instruments selection, pre-trend development, and matching of treated and untreated units), and there is a lack of good software capable of accommodating this for multilevel analysis. Similarly, QCA-models currently lack proper applications that focus on causal inference (see Fiss et al. 2013). In this chapter, both methods are therefore used for exploring the relevant dimensions in the firm-agglomeration relation ('what relates significantly to firm productivity?'), exploiting the advantages of both methods (for multilevel analysis: dealing with large datasets with firm and context level variances treated simultaneously; for QCA: detecting different pathways to an outcome and uncovering interacting conditions and dimensions, even when such patterns are empirically rare). The comparison shows that a well-defined configurational analysis is able to accurately select the conditions that also matter significantly in the multilevel analysis. This means that in an explorative sense, the two methods can be seen as complementary—an outcome that has not been shown before in this field, and only rarely in adjacent scholarships (e.g. Torugsa and Arundel, 2017 in their paper on public administration agencies).

We will compare the two methodologies using an identical dataset to the one presented in the study by Knoben et al. (2016), which consists of a representative sample of Dutch firms. That paper took firm-level and agglomeration-level heterogeneity into account simultaneously and focused on the interactions between these two levels of analysis in explaining the effect of agglomeration on firm performance. The central argument was that while some firms will benefit from agglomeration, others may be harmed by it. To assess these claims, the authors estimated multilevel models on firms' productivity with nonlinear interaction effects between the firm-level conditions (size, internal knowledge base, and face-to-face contacts) and those at the agglomeration-level (urbanization, localization, and knowledge intensity). The results showed that the effects of different dimensions of agglomeration on firm performance are strongly and nonlinearly moderated by firm characteristics. Moreover, the moderation effect was not uniform across the different agglomeration dimensions. In this chapter we re-analyse the dataset using fsQCA. We want to determine to what extent the methodologies produce the same explorative outcomes, how the methodologies can complement one another, and what the weak and strong aspects of the methodologies are when compared vis-à-vis each other.

For this, the chapter is organized in the following way. First, we detail the research setting and data sourced from Knoben et al. (2016). We elucidate how fsQCA differs from multilevel regression modelling, and why it can improve our understanding of heterogeneity relating to agglomeration effects. We unpack the difficulties encountered when using such an approach, such as calibrating the data to accommodate fsQCA's set-theoretical foundation and controlling for industry differences. The final section provides recommendations on the p. 375 potential use of both methodologies ↳ (crossover) in the firm-agglomeration debate, suggesting that QCA can be used when fewer observations are available, and can precede multilevel or standard econometric modelling by giving valuable clues for interaction effects, possible heterogeneity, and complexity insights that can be dealt with in both methods. QCA informing a more quasi-experimental setup in multilevel or econometric modelling is, as said, the least but still extremely valuable crossover between the methods that can identify relevant heterogeneity.

## 15.2 Data and Methods

Our data is sourced from Knoben et al. (2016) and was originally collected by the Netherlands Environmental Assessment Agency through a 2005 survey.[1] It includes firms from 128 of the 467 Dutch municipalities, containing the 8 largest cities, 10 medium-sized cities, and their suburban and rural surroundings. Through this approach 'as much regional heterogeneity as possible' was captured. Although this suggests that extreme cases were prioritized, Knoben et al. (2016, p. 138) show that most of their regional measures do not differ significantly from the mean for all Dutch municipalities.[2] The dataset only includes firms in manufacturing and business services, because retail and customer-related services are impacted by the distribution of the population, and not the agglomeration dimensions under study. Random stratified sampling (of municipalities, industry, and establishment size) was used, informed by the LISA database containing all Dutch establishments where paid work takes place. In total, the survey was completed by 1,676 establishments, and any non-responders were replaced with sampled establishments from the same stratus. The final sample contains 2,009 establishments (with a response rate of about 7%) and is representative of all Dutch firms in the sampled regions in 2005 (Van Oort et al., 2012).

### 15.2.1 A Set-Theoretic Approach

Given our research aims, it is important to unpack the defining features of fsQCA and compare it to regression techniques. At its core, fsQCA considers cases, in this chapter firm-establishments, as combinations of conditions, uncovering their necessity and sufficiency for an outcome, in our case high firm-level productivity (Ragin, 2008, 2012). Conditions and the outcome are measured through sets, of which cases can be a member or non-member. To illustrate, a firm may be fully out of the set of being subject to geographic localization, fully in the set capturing geographic urbanization, and fully in the outcome set of high firm productivity. As fsQCA's adaptation of Boolean algebra uses continuous values between 0 (absence) and 1 (presence), it captures degrees of set membership and allows for ambiguity (denoted by a value of 0.5).[3] Necessity and sufficiency patterns are p. 376 uncovered by analysing sub- and superset relationships (Greckhamer et al., ↳ 2018). To exemplify, it may be that firm-establishments with a particular combination of conditions consistently achieve high productivity, rendering this configuration a subset of the outcome and thus sufficient. Alternatively, all high-productivity cases may have a particular combination of conditions and be a superset of the outcome and therefore necessary. Crucially, in its detailed analysis of such patterns, multiple conjunctions or configurations of sets (of conditions) can be associated with an outcome, and there is thus the potential for equifinality (Misangyi et al., 2017). Further, their inverse does not necessarily lead to the absence of the outcome, making fsQCA's understanding of complexity asymmetrical.

This is markedly different from regression techniques, which are correlational in nature, and utilize linear algebra. By default, isolated, net effects of factors—or, more formally, predictors—are key, and not their conjunction (Greckhamer et al., 2013). Thus, regression analysis assumes effects are independent from all values of the other factors and are non-competing—i.e. without the possibility of multiple paths to the outcome. This does not mean that conjunction and complexity are entirely neglected. Regression analysis can fruitfully accommodate multiple levels of causation (i.e. at both the firm-establishment and agglomeration-level, see Van Oort et al., 2012). However, regressions have an innately limited capacity to capture interrelationships, through for example moderation and/or interaction effects, becoming increasingly difficult to interpret when exceeding two or more factors (Misangyi et al., 2017). Despite having the potential to deepen our understanding of how firms are affected by agglomeration (Speldekamp et al., 2020b), more complex interrelationships, and the necessity of factors, are generally not be considered by researchers, including Knoben et al. (2016).

In addition to being useful in research settings characterized by complexity, fsQCA and related techniques have other, well-documented advantages. Although our data is representative of the larger population, probabilistic sampling is not required (Fiss, 2011). In addition, it can accommodate a low, medium, and large number of cases —all with a relatively large number of conditions that, as long as they can be transformed into sets, can be quantitative or qualitative in nature (Greckhamer et al., 2013). Although having incomplete data imposes the risk of leaving important pathways to an outcome undetected, it does not invalidate results from analyses with sufficient variation in terms of the absence and presence of sets. This is related to the treatment of limited diversity—a matter we turn to at the end of this methodology when detailing how fsQCA simplifies patterns in the data to arrive at conclusions (Fiss, 2011).

### 15.2.2 Set Calibration

To transform the conditions under study into continuous sets with scores between 0 and 1, we used three anchor points: full non-membership (0), the point of maximum ambiguity (0.5), and full membership (1) (Ragin, 2008). Following common practice, these were based on external and theoretical criteria whenever possible, and sample statistics in all other cases (Fiss, 2011; Chappin et al., 2015; Misangyi et al., 2017). Due p. 377 ↳ to methodological issues associated with set membership scores of exactly 0.50, we added a constant of 0.001 to scores below 1 (Hotho, 2014). All calibrations and analyses were performed using the fsQCA 3.1b program (Ragin et al., 2019). We sourced all operationalizations for the conditions and outcome under study from Knoben et al. (2016), which are distinguished by either being at the firm-establishment or agglomeration-level.

#### 15.2.2.1 Firm-Level Sets

The outcome of productivity and three conditions, namely size, internal knowledge base strength, and level of local connectedness, were measured at the firm-establishment level. This generates notable barriers to using fsQCA, as the dataset consists of firms in heterogeneous industrial activities. For each of these factors, industries differ in their mean values and their dispersion in absolute terms. Hence, any set calibration based on such data will result in analyses mainly capturing industry differences. In regression analysis, this is corrected using industry-fixed effects. However, fsQCA does not allow for this. Instead, we standardized all establishment-level conditions using Z-scores, by the four types of activity where most differences occur: labour intensive, capital intensive, and knowledge intensive industries, and firms in knowledge intensive services.[4] Thereby, prior to calibration, all mean values are zero, with the different case values denoting variation within the type of industry following the normal distribution curve, making them comparable over the full sample.

Below, we report the operationalizations prior to standardization, and the calibration based on Z-scores.

#### Outcome: Productivity

The outcome measure is productivity at the level of the establishment, i.e. the added value per full-time employee, where the added value is the gross turnover in 2004 (including taxes, subsidies, wages, profits) minus any purchases for that year (including all intermediate goods and services needed in the production process). This is a frequently used measure of performance (Ciccone, 2002; Beaudry and Swann, 2009; Rigby and Brown, 2015).

To best approximate the estimation technique used in the original paper, we calibrated the outcome set to denote above-average productivity. The calibration threshold for non-set membership was a Z-score of -1, i.e. one standard deviation below the mean. The crossover point was 0, the mean, and full set membership was set at a Z-score of 1.

#### Size, Knowledge Base Strength, and Local Connectedness

The other three measures at the establishment level capture firm size, the strength of the firm's internal knowledge base, and the firm's level of local connectedness. Each has the potential to affect firm productivity and shape the consequences of region-level conditions (Knoben et al., 2016). Size is operationalized by gross sales in 2004, which is ideal for research pertaining to multiple industries such as our study (Cohen and p. 378 Klepper, 1996). The strength of the knowledge base is denoted by the share of knowledge ↳ intensive jobs (of the total). For manufacturing establishments, these are occupations in R&D. For those in business services they are jobs in marketing and design-related consulting. This difference is due to the latter type of establishment relying more on professional knowledge (Illeris, 1996; Cassiman and Veugelers, 2006). Finally, local connectedness is the percentage of inter-organizational contacts, regardless of their purpose, occurring face-to-face ( Knoben and Oerlemans, 2006).

The calibration of these three measures was based on the highest tercile of the four types of industries, therefore capturing their strong presence. The anchor point for non-membership was the minimum recorded value in this tercile, the crossover point was the mean, and firms were fully in the set at the maximum recorded value.

### 15.2.2.2 Agglomeration-Level Sets

All agglomeration-level conditions are measured at the level of the municipality, given that externalities like knowledge spillovers decay rapidly over larger distances (Baldwin et al., 2010; Rigby and Brown, 2015). A three-year time lag is used (in line with Henderson, 2003). Whereas productivity was measured in 2004, the regional conditions pertain to 2002. Notably, the authors of the original study point out that the latter are time-invariant, which also limits endogeneity concerns (Knoben et al., 2016).

### Urbanization, Localization, and Knowledge Intensity

Three agglomeration-level measurements are used, denoting the level of urbanization, localization, and knowledge intensity. As urbanization denotes diverse concentrations of economic activity, it is best indicated by the presence of jobs (Jacobs, 1969). To correct for different municipal sizes, the total number of jobs per square kilometre was used, rather than an absolute count. For localization, an employment-based location quotient was used for the municipality and industry (by two-digit SIC code) in which an establishment is active (Van Soest et al., 2006; Beaudry and Schiffauerova, 2009). Finally, the level of knowledge intensity is the percentage of all establishments in a region that have generated technological innovations in the period between 2000 and 2002. Technological innovations are defined as 'the introduction of new or improved products, services, or processes for which the novelty or improvement lies in the application of new or recently developed technologies', closely matching Eurostat's Community Innovation Survey (Knoben et al., 2016, p. 141). This construct is highly correlated to similar patent or R&D-based measures, but less biased to particular types of knowledge activities (Hagedoorn and Cloodt, 2003).

Our set calibration of urbanization is based on the Statistics Netherlands' categorization (CBS, 2021). Establishments are not subject to urbanization if the job density is at or below 500, at the point of ambiguity at 1,000, and are fully in the set denoting urbanization at 1,500. For localization, these thresholds were 1, 1.5, and 2. This first value denotes that there is no specialization compared to the national average, whereas this is 1.5 and 2 times the national average for the other anchor points. For local knowledge intensity, there were no such easy guidelines, and we used sample statistics to capture the values for the 50th, 70th, and 90th percentiles.
p. 379 The corresponding value for non-membership ↳ was 57% of establishments in a region introducing technological innovations, whereas this was 61% and 65% for the crossover point and full membership.

Appendix Table A.15.1 shows a summary of our set operationalizations.

## 15.2.3 Uncovering Pathways

After preparing our data, the next step in uncovering the pathways to above-average productivity was to create a truth table. Such a table has $2^k$ rows, with *k* denoting the number of sets (Fiss, 2011). Each table row is a specific combination of these sets, which are either untrue (0) or true (1). Each row shows the number of cases matching this pattern, and their consistency—i.e. the extent to which cases correspond to the relationships expressed in a solution.

With fsQCA, the number of truth table rows are reduced by setting a minimum number of cases, and a minimum consistency value. When these are not met, rows are deleted as their particular patterns are either not represented often enough to be taken into consideration, or they do not consistently lead to the outcome. We set our minimum solution frequency at 5, in order to filter out exceptional cases from which no theoretical inferences can be made, and our consistency threshold was ≥ 0.85. Both are in line with recent recommendations made in the QCA scholarship (Rutten, 2022). In the reported analysis of the full sample, 95 cases met both thresholds.

The final step was to further reduce the truth table to simplified combinations of sets of conditions associated with high firm productivity, using the Quine–McCluskey algorithm (see Ragin, 2017). This is critical, as researchers are otherwise left with highly complex and numerous pathways associated with an outcome. Some of the combinations of sets are logically redundant and can be combined (Fiss, 2007 offers a detailed discussion). Given the likelihood of empty truth table rows, or so-called logical remainders, this reduction process incorporates counterfactual analysis (Fiss, 2011). As Ragin and Sonnett (2005, p. 9) explain, this problem of limited diversity where empirical instances of particular combinations of sets are missing 'is the rule, not the exception, in the study of naturally occurring social phenomena'. This applies even to large-N datasets such as our own, although having more cases means that it is less prevalent. Following the norm, two distinct reduction processes were used that accommodate logical simplification in the presence of logical remainders (Saka-Helmhout et al., 2020). The first relies on all simplifying assumptions, both difficult and easy, and generates parsimonious solutions. The second relies only on easy counterfactuals, generating intermediate solutions. The distinction between difficult and easy counterfactuals is that the latter are consistent with substantive and theoretical knowledge, and are indicated in the analysis (Fiss, 2011).[5] In our case, intermediate solutions included just two assumptions, namely that the presence of large firm size, and a strong internal knowledge base likely contribute to high productivity. Assumptions on other sets were not easy, given that their effects differ in their interaction (Knoben et al., 2016), and are therefore only included in parsimonious solutions.

p. 380 With this approach, it is possible to discern between core and peripheral conditions (Fiss, 2011). Where core conditions are sets that are absent or present in both the intermediate and parsimonious solutions, peripheral conditions appear only in the intermediate solution. In other words, the evidence for causal relationships is the strongest for core conditions. Finally, it is important to note that in the fsQCA tables reported in the results section, ● represents the presence of core conditions, and ⊗ their absence. Peripheral conditions are denoted by smaller circles, i.e. · and ⊗. When conditions do not matter, the space is left blank. Coverage and consistency values are reported. Coverage is the proportion of cases captured by a pathway or solution, and thus indicates empirical relevance (Schneider and Wagemann, 2012; Park et al., 2020). Consistency is the correspondence between cases and the conditions in a solution, showing the 'strength' of a configuration to lead to an outcome. Furthermore, raw and unique coverage values are shown below each pathway, where raw coverage relates to the total share, and unique coverage only to the share covered exclusively by a particular pathway (Greckhamer et al., 2013). Finally, the overall solution coverage and consistency are those of the combination of all depicted pathways.

## 15.3 Results

In this results section, we first discuss the results of the fsQCA performed on the full sample of firms (see Table 15.1), using the set calibrations and procedures described in the above. However, we also discuss separate p. 381 analyses for the presence of high ↳ productivity in three sub-samples of firm size (small, medium, and large firms; see Tables 15.2 through 15.4).[6] This is to test whether there are nonlinear firm-agglomeration effects, which the literature suggests to exist (Hervas-Oliver et al., 2018). In all instances, results are compared to those obtained by multilevel regression analyses of the same data as reported in Knoben et al. (2016).

**Table 15.1** High firm productivity (full sample)

| | | 1a | 1b | 1c | 2 | 3 |
|---|---|---|---|---|---|---|
| *Firm-level* | Size | ● | ● | ● | | |
| | Internal knowledge base | | | | ⊗ | ● |
| | Face-to-face contacts | ⊗ | ⊗ | ⊗ | ● | ⊗ |
| *Agglomeration-level* | Urbanization | ⊗ | ⊗ | • | ● | ⊗ |
| | Localization | | ⊗ | • | ⊗ | ● |
| | Knowledge intensity | ⊗ | | • | ⊗ | ● |
| Raw coverage | | 0.10 | 0.10 | 0.02 | 0.03 | 0.05 |
| Unique coverage | | 0.01 | 0.02 | 0.00 | 0.02 | 0.04 |
| Consistency | | 0.90 | 0.93 | 0.99 | 0.86 | 0.84 |
| **Overall solution coverage** | | 0.18 | | | | |
| **Overall solution consistency** | | 0.87 | | | | |

*Notes:* ● = core condition (present); ⊗ = core condition (absent); • = peripheral condition (present); ⊗ = peripheral condition (absent); blank space = the causal conditions may be present or absent.

The first result from the Knoben et al. (2016) paper clearly visible in the fsQCA results is the large and positive effect of firm size on firm productivity. Paths 1a through 1c clearly reflect the importance of firm size for achieving above-average productivity. Pathways 1a and 1b show that large firms can achieve high productivity even in un-agglomerated environments (i.e. absence of the associated conditions) and without interacting intensively with firms in their environment (i.e. absence of face-to-face interactions). Pathway 1c, however, shows that large firms can achieve the same high productivity in agglomerated environments. These three pathways thus show that large firms can achieve high productivity in almost all environments—a central conclusion that was also drawn in Knoben et al. (2016).

Pathway 2 also reflects one of the patterns found in Knoben et al. (2016), namely that face-to-face contacts contribute positively to productivity in urbanized (diverse) environments but not in localized (specialized) environments. In pathway 2 face-to-face contacts in conjunction with urbanization lead to high productivity. Moreover, there is no pathway in which face-to-face contacts together with localization lead to high productivity. This clearly indicates that whenever face-to-face contacts matter, they do so in diverse and not in specialized environments.

Pathway 3 echoes the finding of Knoben et al. (2016) that the strength of the internal knowledge base has a positive effect on firm productivity. However, the fsQCA shows this effect to be much more conditional as the positive effect appears only in conjunction with location in a localized, knowledge intensive but non-urbanized environment. In other words, the strength of the internal knowledge base mainly matters when surrounded by many highly capable (i.e. knowledge intensive) competitors (i.e. specialized). In those environments having a strong internal knowledge base might act as the absorptive capacity needed to benefit from the state-of-the-art knowledge available in the region and as a means to protect the intellectual property of the firm from 'leaking' to competitors.

A finding in the Knoben et al. (2016) paper that is reflected in the fsQCA reporting in Table 15.1, albeit weakly, concerns the econometric outcome that it is mainly medium-sized firms that benefit from being in an urbanized or localized environment. This finding is reflected in the fsQCA outcome for the pathways in which the region's level of urbanization or localization is a core condition (pathway 2 and 3) but being a large firm is not a core condition. This observation, combined with the dominance of firm size in explaining high productivity (and the absence of firm size in explaining the absence of high productivity; see Appendix Table A.15.2), warrants a further exploration into the effects for different firm sizes (see Tables 15.2 through 15.4).

**Table 15.2** High productivity in small firms

| | | 1 |
|---|---|---|
| *Firm-level* | Size | ● |
| | Internal knowledge base | ⊗ |
| | Face-to-face contacts | ⊗ |
| *Agglomeration-level* | Urbanization | ● |
| | Localization | ⊗ |
| | Knowledge intensity | ⊗ |
| Raw coverage | | 0.16 |
| Unique coverage | | 0.16 |
| Consistency | | 0.86 |
| **Overall solution coverage** | | 0.16 |
| **Overall solution consistency** | | 0.86 |

*Notes:* ● = core condition (present); ⊗ = core condition (absent); • = peripheral condition (present); ⊗ = peripheral condition (absent); blank space = the causal conditions may be present or absent.

**Table 15.3** High productivity in medium-sized firms

| **Causal conditions** | | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|---|
| *Firm-level* | Size | | | | ● | | ⊗ |
| | Internal knowledge base | ● | | | | ● | ● |
| | Face-to-face contacts | ⊗ | ⊗ | ⊗ | ⊗ | ● | ⊗ |
| *Agglomeration-level* | Urbanization | | ● | | ● | ⊗ | ⊗ |
| | Localization | | | ● | ⊗ | ⊗ | ● |
| | Knowledge intensity | ● | ● | ● | | ⊗ | |
| Raw coverage | | 0.10 | 0.13 | 0.18 | 0.14 | 0.03 | 0.08 |
| Unique coverage | | 0.03 | 0.01 | 0.06 | 0.06 | 0.02 | 0.03 |
| Consistency | | 0.73 | 0.75 | 0.67 | 0.86 | 0.85 | 0.83 |
| **Overall solution coverage** | | 0.35 | | | | | |
| **Overall solution consistency** | | 0.70 | | | | | |

*Notes:* ● = core condition (present); ⊗ = core condition (absent); • = peripheral condition (present); ⊗ = peripheral condition (absent); blank space = the causal conditions may be present or absent.

**Table 15.4** High productivity in large firms

| **Causal conditions** | | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|---|
| *Firm-level* | Size | ● | | | | |
| | Internal knowledge base | | ● | | | ● |
| | Face-to-face contacts | ⊗ | ⊗ | ⊗ | ⊗ | ⊗ |
| *Agglomeration-level* | Urbanization | ⊗ | | ● | | ⊗ |
| | Localization | | | | ● | ● |
| | Knowledge intensity | | ● | ● | ● | |
| Raw coverage | | 0.28 | 0.09 | 0.13 | 0.20 | 0.09 |
| Unique coverage | | 0.16 | 0.02 | 0.01 | 0.05 | 0.03 |
| Consistency | | 0.90 | 0.84 | 0.78 | 0.78 | 0.83 |
| **Overall solution coverage** | | 0.46 | | | | |
| **Overall solution consistency** | | 0.79 | | | | |

*Notes:* ● = core condition (present); ⊗ = core condition (absent); • = peripheral condition (present); ⊗ = peripheral condition (absent); blank space = the causal conditions may be present or absent.

Interestingly, Table 15.2 shows that there is only one pathway for small firms to be highly productive. This
p. 382 pathway requires the firms to be at the upper end of the size ↳ range for small firms and being located in an urbanized environment. This finding strongly echoes the findings of Knoben et al. (2016) that firms between the 5th and the 25th percentile of size benefit the most from being in an urbanized environment. Whereas
p. 383 Knoben et al. (2016) find a similar pattern for firms between the 5th and the ↳ 50th percentile of firm size to benefit from being located in a localized environment, the results in Table 15.2 show no such pathway. In fact, localization is an absent condition in the one pathway to high productivity for small firms.

Table 15.3 shows that for medium-sized firms there are many different paths towards high productivity. Pathways 2 and 3 show that medium-sized firms can be highly productive in urbanized and localized environments provided that those environments are also knowledge intensive. Pathway 5 on the other hand shows that in un-agglomerated environments medium-sized firms can be highly productive if they combine a strong internal knowledge base with high levels of face-to-face interaction. Pathways 1, 4, and 6 show interesting interactions between the firm and the regional level conditions. Being highly productive in a knowledge intensive (pathway 1) or localized environment (pathway 6) is conditional on medium-sized firms having a strong internal knowledge base. Being highly productive in an urbanized environment is conditional on firms being on the upper size limit of medium-sized firms (pathway 4). All these pathways are consistent with the general patterns found in Knoben et al. (2016) but provide more nuanced insights into how different firm and regional level characteristics interact to explain firm productivity.

The findings for large firms (see Table 15.4) closely match the findings for the complete sample. First, large firms can be highly productive just based on internal economies of scale (pathway 1). Second, they can combine a strong internal knowledge base with location in a knowledge intensive (pathway 2) or a localized environment (pathway 5). Third, and identical to the findings for medium-sized firms, large firms can be highly productive in a knowledge intensive and urbanized (pathway 3) or localized (pathway 4) environment.

p. 384 Overall, the findings of our fsQCA analyses are very much in line with those obtained by multilevel regression analyses of the same data. Based on the fsQCA analyses we confirm the dominant role that firm size plays in explaining high firm-level productivity. We also replicate the finding that medium-sized firms benefit the most from location in an urbanized environment. Finally, we show that face-to-face contacts matter only in urbanized environments and not in localized environments—a result that was found using interaction-terms in the multilevel analysis as well.

However, the fsQCA results provide richer and more nuanced findings that could not be obtained by multilevel regression analyses. For example, the fsQCA provides much more support for positive main effects of different kinds of agglomeration for all but the smallest firms. In the econometric models, only regional knowledge intensity had a positive effect on firm productivity and that effect was relatively small and marginally statistically significant. In contrast, the fsQCA shows that combinations of knowledge intensity and urbanization or localization lead to high productivity for both medium and large firms.

Another interesting nuance added by our analysis is the revelation of the more conditional role of the internal knowledge base of firms. In Knoben et al. (2016) it was hypothesized that the internal knowledge base would provide a firm with the absorptive capacity required to assess, access, and internalize externally available knowledge. As such, the internal knowledge base of a firm was expected to heavily influence (interact with) the effect of the external environment on a firm's productivity. Contrary to this hypothesis, almost no interaction effects were identified in the multilevel analysis. The fsQCA, however, does reveal such interaction effects of a firm's internal knowledge base. Several pathways show that the internal knowledge base is a requirement to benefit from particularly specialized environments.

## 15.4 Discussion

Using a fsQCA toolkit, we analysed data previously investigated with a multilevel regression approach. After preparing the data for fsQCA (without losing the main characteristics of the observations), we were able to detail a large degree of consistency between the exploratory results of both methods. Both methods identify the same main patterns in the data. At the same time, the fsQCA approach is clearly able to identify more nuances and contingencies in how firm and agglomeration-level conditions like urbanization (diversity) and localization (specialization) interactively influence firm productivity. Moreover, with 'just' three firm-level and three agglomeration-level conditions, the multilevel regression approach was already reaching the limits of the amount of (cross-level) interactions it could deal with in terms of both collinearity problems and interpretability of the results. In principle, a (fs)QCA approach could deal with even larger numbers of both firm
p. 385 and regional level conditions. As such, we ↳ conclude that it is a useful and flexible method for various types of applications that want to identify potential factors influencing the firm-agglomeration relationship. This is important for several reasons.

First, given its focus on conjunction and equifinality, (fs)QCA is extremely useful to test inherently configurational theories and arguments (Fiss, 2011), especially when the number of observations is limited. The seminal argument by Porter (1998) that regional clustering of economic activity leads to competitiveness through a combination of agglomeration forces, inter-linkages between firms and complementary institutional arrangements is an example and can benefit from the nuanced insights generated by the fsQCA method (Speldekamp et al., 2020a, 2020b). Interestingly, such an analysis can be extended towards other performance and (spatial) context indicators, like innovation, network reliability, and the much-used concept of regional related variety highlighting local cognitive relatedness between actors as a main driver of productive relations (Frenken et al., 2007).

Second, (fs)QCA could serve as a method to inform researchers about which interactions to include in, for example, regression analyses. Regression analysis (including multilevel analysis) has, when appropriate conditions are met, advantages over QCA analysis in terms of causal identification possibilities of relations (Oshchepkov and Shirokanova, 2022). Especially in more exploratory analyses where the potential number of interactions to consider is very large, QCA can be a useful selection tool to help researchers identify interactions that are most useful to include. An example of exactly such an application can be found in the work of Torugsa and Arundel (2017).

Third, as shown by the application in this chapter, QCA is suitable to triangulate explorative results obtained by other research approaches. Examples are (multilevel) regression models, cluster analyses, and, given its ability to deal with small samples, qualitative studies. Given the presence of all of these approaches in fields such as economic geography and regional studies, QCA has the potential to bridge research and researchers using disparate research approaches.

Fourth and finally, QCA allows for asymmetry in results between the presence and the absence of the outcome (Fiss, 2011), and this makes it particularly suitable for theoretical arguments that predict such asymmetry. An example of such an argument can be found in the effect of agglomerations on start-ups where some agglomeration features stimulate start-up performance, but their absence may not be responsible for high failure rates (Stuart and Sorenson, 2003). Another example can be found in the research on firms' board structure, where board structure is not expected to have positive effects on firm performance in general but is expected to be able to prevent extremely poor performance of companies (Goranova et al., 2017).

Despite the many promising applications of QCA, we acknowledge that these are not without problems and challenges. As our re-analysis shows, this approach is rather limited in its ability to 'control for' within-sample heterogeneity. Controlling for productivity differences between industries is rather straightforward in various p. 386 types ↳ of regression analyses by including industry-level fixed effects. QCA, however, required standardizing the outcome condition across industries before calibrating the set membership. Similarly, including regional or time-level fixed effects that help with isolating impacts of interest requires a lot more effort and might not even be possible within the QCA approach. As such, QCA requires a sampling approach where, besides the conditions/sets of interest, homogeneity is an important sampling criterion.

A related challenge is that, even though it is ideally suited to study complexity (often called causal complexity in the QCA-literature), QCA actually offers fewer handholds to study causality (i.e. identification) compared to standard econometric approaches—and it is important to distinguish the two. Complexity refers to the notion that outcomes are often caused by several co-related factors being present (or absent) at the same time (Rutten, 2020). Following this notion, studying the contribution of single conditions to an outcome is not very fruitful. QCA was specifically developed to analyse situations in which complexity is expected to play a role. However, this complexity is easily confused with QCA being able to show causality (e.g. Rutten, 2022). Such an analysis, however, does not necessarily assure that the factors that it identifies to be complex are also causality related. For illustration, nothing in a QCA requires complex factors to temporally precede the outcome, whereas this is a key criterion for an effect to be causal. As such, it is important to separately consider the identification of causal effects within a QCA approach. Some identification strategies can potentially be incorporated, similar to regression methods (Cunningham, 2021). For example, discontinuity designs, such as spatial discontinuity designs (Lee and Lemieux, 2010), can be transformed into a crisp-set membership. Inclusion of this set in a QCA would automatically lead to the assessment of the interactions between the discontinuity and all other sets under scrutiny. However, compared to econometric approaches, the QCA approach is less equipped to include identifying instrumental variable approaches or matching procedures for difference-in-difference estimation. A way of dealing with unobserved factors (potentially also signalling variable complexity) may be applying panel data analysis (our analysis was cross-sectional). This has been applied more in multilevel than in QCA analysis (Hino, 2009; Audretsch et al., 2018; Furnari, 2018).

Even though it is possible and defensible to make a tradeoff between identification strength and other important features of a research design, we do think researchers adopting QCA should be explicit about such tradeoffs and clearly acknowledge the difference between causal complexity and identifying causality. Yet, the analysis by Knoben et al. (2006) using multilevel analysis and our QCA analysis in this chapter reaching similar and complementary conclusions, while exploiting the same dataset, suggests that a much more fruitful marriage between the modelling approaches is possible. There is more than one way leading to consistent insight when exploring whether agglomeration is one firm's medicine while it may be another's poison. 'To make causal effects believable, prior knowledge is required in order to justify any claim of causal findings' (Cunningham 2021, p. 9). We show that the exploratory function of QCA is ideal for mapping such complex prior knowledge, suggesting which (combinations of) ways lead to Rome.

p. 387

## Notes

1. For a description of the survey, see Van Oort et al. (2006). Note that this data was also used by Van Oort et al. (2012).
2. Only regional knowledge intensity differs; see Table 1 in Knoben et al. (2016, p. 139).
3. That is, neither fully in nor out of the set. Please note that other variations of this technique exist, such as csQCA (crisp-set QCA), where only binary values exist, and mvQCA, which captures multinomial concepts (Ragin, 2012). The advantage of fsQCA compared with its counterpart csQCA is that less information is lost.

4. The SBI ’93 industry codes are as follows. Labour intensive: 17, 18, 19, 20, 28, 36. Knowledge intensive: 15, 16, 21, 25, 26. Capital intensive: 23, 24, 27, 29, 30, 31, 32, 33, 34, 35. Knowledge intensive services: 22, 72, 73, 74, 642 (CBS, 2004).

5. To illustrate, the complex solution prior to this reduction that is purely based on empirical observations may uncover that high productivity (Y) coheres with the presence of large size (A) and a strong internal knowledge base (B), and the absence of face-to-face contacts (c). Formally, this can be written as A•B•c→Y. There may not be empirical instances of A•B•C→Y (where face-to-face contacts are present and not absent). Yet, researchers may indicate that if such a combination were present in the data, it would likely lead to the outcome. This is an easy counterfactual, leading (in this example) to the simplification of A•B•c→Y and A•B•C→Y to A•B→Y. Please see Ragin and Sonnett (2005, pp. 9–11) for a detailed account.

6. The calibration strategy here was identical to the full sample. Small firms were in the first tercile, and medium and large firms in the second and third. All sets of conditions were similar to the full analysis, except for size, where the thresholds for non-membership, the crossover, and full membership were, as with the full analysis, the lowest, mean, and highest recorded values.

## References


Andersson, M., Larsson, J.P., Wernberg, J., 2019. The economic microgeography of diversity and specialization externalities – firm-level evidence from Swedish cities. *Research Policy* 48(6), 1385–1398.
Google Scholar WorldCat

Audretsch, D.B., Beletski, M., Desai, S., 2018. National business regulations and city entrepreneurship in Europe: A multilevel nested analysis. *Entrepreneurship Theory and Practice* 43, 1148–1165.
Google Scholar WorldCat

Baldwin, J.R., Brown, W.M., Rigby, D.L., 2010. Agglomeration economies: Microdata panel estimates from Canadian manufacturing. *Journal of Regional Science* 50, 915–934.
Google Scholar WorldCat

Beaudry, C., Schiffauerova, A., 2009. Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy* 38, 318–337.
Google Scholar WorldCat

Beaudry, C., Swann, G.M.P., 2009. Firm growth in industrial clusters of the United Kingdom. *Small Business Economics* 32, 409–424.
Google Scholar WorldCat

Cassiman, B., Veugelers, R., 2006. In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition. *Management Science* 52, 68–82.
Google Scholar WorldCat

CBS, 2004. SBI '93 versie 2004. https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/activiteiten/sbi-2008-standaard-bedrijfsindeling-2008/sbi-93-versie-2004 (accessed 4.8.21).
WorldCat

p. 388 CBS, 2021. Stedelijkheid. https://www.cbs.nl/nl-nl/nieuws/2019/44/meeste-afval-per-inwoner-in-minst-stedelijke-gemeenten/stedelijkheid (accessed 4.9.21).
WorldCat

Chappin, M.M.H., Cambré, B., Vermeulen, P.A.M., Lozano, R., 2015. Internalizing sustainable practices: A configurational approach on sustainable forest management of the Dutch wood trade and timber industry. *Journal of Cleaner Production* 107, 760–774.
Google Scholar WorldCat

Ciccone, A., 2002. Agglomeration-effects in Europe. *European Economic Review* 46, 213–227.
Google Scholar WorldCat

Cohen, W.M., Klepper, S., 1996. Firm size and the nature of innovation within industries: The case of process and product R&D. *The Review of Economics and Statistics* 78, 232–243.
Google Scholar WorldCat

Cunningham, S., 2021. *Causal Inference.* Yale University Press, New Haven.
Google Scholar Google Preview WorldCat COPAC

De Groot, H.F.L., Poot, J., Smit, M., 2016. Which agglomeration externalities matter most and why? *Journal of Economic Surveys* 30, 756–782.
Google Scholar WorldCat

Faggio, G., Silva, O., Strange, W.C., 2017. Heterogeneous agglomeration. *Review of Economics and Statistics* 99, 80–94.
Google Scholar WorldCat

Fainshmidt, S., Witt, M.A., Aguilera, R.V., Verbeke, A., 2020. The contributions of Qualitative Comparative Analysis (QCA) to international business research. *Journal of International Business Studies* 51, 455–466.
Google Scholar WorldCat

Fiss, P.C., 2007. A set-theoretical approach to organizational configurations. *Academy of Management Review* 32, 1180–1198.
Google Scholar WorldCat

Fiss, P.C., 2011. Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal* 54, 393–420.
Google Scholar WorldCat

Fiss, P.C., Sharapov, D., Cronqvist, L., 2013. Opposites attract? Opportunities and challenges for integrating large-N QCA and econometric analysis. *Political Research Quarterly* 66, 191–198.
Google Scholar WorldCat

Frenken, K., Cefis, E., Stam, E., 2015. Industrial dynamics and clusters: A survey. *Regional Studies* 49, 10–27.
Google Scholar WorldCat

Frenken, K., Van Oort, F.G., Verburg, T., 2007. Related variety, unrelated variety and regional economic growth. *Regional Studies* 41, 685–697.
Google Scholar WorldCat

Furnari, S., 2018. *Four approaches to longitudinal QCA: Opportunities and challenges.* Paper presented at the Academy of Management, Chicago (2018).
Google Scholar Google Preview WorldCat COPAC

Goranova, M.L., Priem, R.L., Ndofor, H.A., Trahms, C.A., 2017. Is there a 'dark side' to monitoring? Board and shareholder monitoring effects on M&A performance extremeness. *Strategic Management Journal* 38, 2285–2297.
Google Scholar WorldCat

Greckhamer, T., Furnari, S., Fiss, P.C., Aguilera, R.V., 2018. Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. *Strategic Organization* 16, 482–495.
Google Scholar WorldCat

Greckhamer, T., Misangyi, V.F., Fiss, P.C., 2013. The two QCAs: From a small-N to a large-N set theoretic approach, in: Fiss, P.C., Cambré, B., Marx, A. (Eds.), *Configurational Theory and Methods in Organizational Research.* Emerald Group Publishing Limited, Bingley, 49–75.
Google Scholar Google Preview WorldCat COPAC

Grillitsch, M., Nilsson, M., 2017. Firm performance in the periphery: On the relation between firm-internal knowledge and local knowledge spillovers. *Regional Studies* 51, 1219–1231.
Google Scholar WorldCat

Hagedoorn, J., Cloodt, M., 2003. Measuring innovative performance: Is there an advantage in using multiple indicators? *Research Policy* 32, 1365–1379.
Google Scholar WorldCat

Henderson, J.V., 2003. Marshall's scale economies. *Journal of Urban Economics* 53, 1–28.
Google Scholar WorldCat

Hervas-Oliver, J.L., Sempere-Ripoll, F., Rojas Alvarado, R., Estelles-Miguel, S., 2018. Agglomerations and firm performance: Who benefits and how much? *Regional Studies* 52, 338–349.
Google Scholar WorldCat

p. 389 Hino, A., 2009. Time-series QCA: Studying temporal change through boolean analysis. *Sociological Theory and Methods* 24, 247–

265.
Google Scholar WorldCat

Hotho, J.J., 2014. From typology to taxonomy: A configurational analysis of National Business Systems and their explanatory power. *Organization Studies* 35, 671–702.
Google Scholar WorldCat

Illeris, S., 1996. *The Service Economy: A Geographical Approach.* John Wiley, Chichester.
Google Scholar Google Preview WorldCat COPAC

Jacobs, J., 1969. *The Economy of Cities.* Random House, New York.
Google Scholar Google Preview WorldCat COPAC

Knoben, J., Arikan, A.T., Van Oort, F., Raspe, O., 2016. Agglomeration and firm performance: One firm's medicine is another firm's poison. *Environment and Planning A* 48, 132–153.
Google Scholar WorldCat

Knoben, J., Oerlemans, L.A.G., 2006. Proximity and inter-organizational collaboration: A literature review. *International Journal of Management Reviews* 8, 71–89.
Google Scholar WorldCat

Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48, 281–355.
Google Scholar WorldCat

Martin, R., 2021. Putting the case for a pluralistic economic geography. *Journal of Economic Geography* 21, 1–28.
Google Scholar WorldCat

McCann, B.T., Folta, T.B., 2011. Performance differentials within geographic clusters. *Journal of Business Venturing* 26, 104–123.
Google Scholar WorldCat

McCann, P., Van Oort, F.G., 2019. Theories of agglomeration and regional economic growth: A historical review, in: Capello, R., Nijkamp, P. (Eds.), *Handbook of Regional Growth and Development Theories.* Revised and extended second edition. Edward Elgar Publishing, Cheltenham, 6–23.
Google Scholar Google Preview WorldCat COPAC

Misangyi, V.F., Greckhamer, T., Furnari, S., Fiss, P.C., Crilly, D., Aguilera, R., 2017. Embracing causal complexity: The emergence of a neo-configurational perspective. *Journal of Management* 43, 255–282.
Google Scholar WorldCat

Oshchepkov, A.Y., Shirokanova, A., 2022. Bridging the gap between multilevel modeling and Economic Methods. Social Science Research, 102689.

Park, Y., Fiss, P.C., Sawy, O.A. El, 2020. Theorizing the multiplicity of digital phenomena: The ecology of configurations, causal recipes, and guidelines for applying QCA. *MIS Quarterly: Management Information Systems* 44, 1493–1520.
Google Scholar WorldCat

Porter, M.E., 1998. *On Competition.* Harvard Business School Press, Boston.
Google Scholar Google Preview WorldCat COPAC

Ragin, C.C., 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond.* Chicago University Press, Chicago.
Google Scholar Google Preview WorldCat COPAC

Ragin, C.C., 2012. Qualitative comparative analysis using fuzzy sets (fsQCA), in: Rihoux, B., Ragin, C.C. (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques.* SAGE Publications, Inc., Thousand Oaks, CA, 87–122.

Google Scholar Google Preview WorldCat COPAC

Ragin, C.C., 2017. *User's Guide to Fuzzy-Set / Qualitative Comparative Analysis,* fsQCA 3.0. ed. University of California, Irvine, Irvine, CA.
Google Scholar Google Preview WorldCat COPAC

Ragin, C.C., Drass, K.A., Davey, S., 2019. *Fuzzy-set/Qualitative Comparative Analysis 3.1b.* University of California, Irvine, Irvine, CA.
Google Scholar Google Preview WorldCat COPAC

Ragin, C.C., Sonnett, J., 2005. Between complexity and parsimony: limited diversity, counterfactual cases, and comparative analysis, in: Kropp, S., Minkenberg., M. (Eds.), *Vergleichen in Der Politikwissenschaft.* VS Verlag für Sozialwissenschaften, Wiesbaden, 180–197.
Google Scholar Google Preview WorldCat COPAC

Rauch, J.E., 1993. Does history matter only when it matters little? The case of city-industry location. *The Quarterly Journal of Economics* 20, 843–867.
Google Scholar WorldCat

Rigby, D.L., Brown, W.M., 2015. Who benefits from agglomeration? *Regional Studies* 49, 28–43.
Google Scholar WorldCat

Rutten, R., 2020. Comparing causal logics: A configurational analysis of proximities using simulated data. *Zeitschrift fur Wirtschaftsgeographie* 64, 134–148.
Google Scholar WorldCat

p. 390 Rutten, R., 2022. Applying and assessing large-N QCA: Causality and robustness from a critical realist perspective. Sociological Methods & Research 51, 1211–1243.

Saka-Helmhout, A., Chappin, M., Vermeulen, P., 2020. Multiple paths to firm innovation in Sub-Saharan Africa: How informal institutions matter. *Organization Studies* 41, 1551–1575.
Google Scholar WorldCat

Schneider, C., Wagemann, C., 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis.* Cambridge University Press, Cambridge.
Google Scholar Google Preview WorldCat COPAC

Speldekamp, D., 2021. *Reconfiguring Clusters: From Contradictions to Pathways.* PhD-thesis, Radboud University Nijmegen.
Google Scholar Google Preview WorldCat COPAC

Speldekamp, D., Knoben, J., Saka-Helmhout, A., 2020b. Clusters and firm-level innovation: A configurational analysis of agglomeration, network and institutional advantages in European aerospace. *Research Policy* 49, 103921.
Google Scholar WorldCat

Speldekamp, D., Saka-Helmhout, A., Knoben, J., 2020a. Reconciling perspectives on clusters: An integrative review and research agenda. *International Journal of Management Reviews* 22, 75–98.
Google Scholar WorldCat

Stavropoulos, S., Van Oort, F.G., Burger, M., 2020. Heterogeneous relatedness and firm productivity. *The Annals of Regional Science* 65, 403–437.
Google Scholar WorldCat

Steijn, M., Koster, H., Van Oort, F.G., 2022. The dynamics of industry agglomeration: Evidencefrom 44 Years of coagglomeration patterns. Journal of Urban Economics 130, 103456.

Stuart, T., Sorenson, O., 2003. The geography of opportunity: Spatial heterogeneity in founding rates and the performance of

biotechnology firms. *Research Policy* 32, 229–253.
Google Scholar WorldCat

Torugsa, N., Arundel, A., 2017. Rethinking the effect of risk aversion on the benefits of service innovations in public administration agencies. *Research Policy* 46, 900–910.
Google Scholar WorldCat

Van Oort, F., 2015. Unity in variety? Agglomeration economics beyond the specialization-diversity controversy, in: Karlsson, C., Andersson, M., Norman, T. (Eds.), *Handbook of Research Methods and Applications in Economic Geography.* Edward Elgar, Cheltenham, 259–271.
Google Scholar Google Preview WorldCat COPAC

Van Oort, F., Van Brussel, J., Raspe, O., Burger, M., Van Dinteren, J., Van der Knaap, B., 2006. *Economische netwerken in de regio.* RPB, Den Haag.
Google Scholar Google Preview WorldCat COPAC

Van Oort, F., Burger, M.J., Knoben, J., Raspe, O., 2012. Multilevel approaches and the firm-agglomeration ambiguity in economic growth studies. *Journal of Economic Surveys* 26, 468–491.
Google Scholar WorldCat

Van Soest, D., Gerking S., Van Oort, F., 2006. Spatial impacts of agglomeration externalities. *Journal of Regional Science* 46, 881–89.
Google Scholar WorldCat

# Appendix

**Table A.15.1** Set operationalization and calibration (full sample)

| | | | | Fuzzy-set calibration | | |
|---|---|---|---|---|---|---|
| **Sets** | | **Measure** | | **Fully out** | **Crossover** | **Fully in** |
| *Outcome* | Productivity | Establishments' added value per full-time employee in 2004, calculated as the yearly gross turnover minus purchases (including all intermediate goods and services). The added value includes taxes, subsidies, wages, and profits. Converted to Z-scores per industry type (labour intensive (AI), capital intensive (CI), and knowledge intensive (KI) industries, and firms in knowledge intensive services (KIS)). | | -1 | 0 | 1 |
| *Firm-level* | Size | Establishments' gross sales in 2004. Converted to Z-scores per industry type for firms in the third tercile. | AI | -0.2096 | 0.6936 | 8.7961 |
| | | | CI | -0.2027 | 0.7314 | 5.3194 |
| | | | KI | -0.1537 | 0.3755 | 13.8917 |
| | | | KIS | -0.1097 | 0.2571 | 20.1025 |
| | Internal knowledge base | Establishments' share of knowledge intensive jobs in 2004. For manufacturing establishments knowledge intensive jobs are occupations in R&D, and for business services it is in marketing and design-related consulting. Converted to Z-scores per industry type for firms in the third tercile. | AI | 0.0005 | 0.5343 | 9.5072 |
| | | | CI | 0.1921 | 1.0756 | 6.1495 |
| | | | KI | -0.0223 | 0.6100 | 5.3150 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | KIS | 0.0060 | 0.9454 | 2.4367 |
| | Face-to-face contacts | Establishments' level of local connectedness as the share of inter-organizational contacts that take place face-to-face in 2004. Converted to Z-scores per industry type for firms in the third tercile. | AI | 0.2113 | 1.1222 | 2.6666 |
| | | | CI | 0.5142 | 1.1221 | 2.5817 |
| | | | KI | 0.4526 | 1.0725 | 3.5150 |
| | | | KIS | 0.3213 | 1.4866 | 2.8861 |
| *Agglomeration-level* | Urbanization | The total number of jobs per square kilometre in 2004. | | 500[a] | 1000[a] | 1500[a] |
| | Localization | Employment location quotient for the region and industry in which an establishment operates in 2004. | | 1 | 1.5 | 2 |
| | Knowledge intensity | Percentage of all establishments in a region that have generated technological innovations in 2000–2002 (new or improved products, services, or processes). | | 57[b] | 61[b] | 65[b] |

a Based on the CBS (2021) urbanization classification (highly urbanized, moderately urbanized, little urbanization).

b These values respectively correspond to the 90th (fully in), 70th (crossover), and 50th (fully out) percentiles (rounded to nearest whole).

p. 392 **Table A.15.2** Full sample QCA: Absence of outcome

| **Causal conditions** | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
|---|---|---|---|---|---|---|---|---|---|---|
| *Firm-level* | Size | ⊗ | ⊗ | ⊗ | ⊗ | ⊗ | ⊗ | ⊗ | ⊗ | ⊗ |
| | Internal knowledge base | ● | | ● | | | ⊗ | | ● | ● |
| | Face-to-face contacts | | ● | ⊗ | | ● | ● | ⊗ | ● | |
| *Agglomeration-level* | Urbanization | ● | ● | | ● | ⊗ | | ⊗ | | ⊗ |
| | Localization | ⊗ | ⊗ | | ⊗ | ● | ⊗ | ● | ⊗ | ● |
| | Knowledge intensity | | • | ● | ● | ⊗ | ● | ● | ⊗ | ⊗ |
| Raw coverage | | 0.05 | 0.03 | 0.09 | 0.08 | 0.05 | 0.04 | 0.13 | 0.03 | 0.07 |
| Unique coverage | | 0.01 | 0.00 | 0.03 | 0.02 | 0.02 | 0.01 | 0.06 | 0.01 | 0.03 |
| Consistency | | 0.86 | 0.86 | 0.82 | 0.83 | 0.86 | 0.88 | 0.87 | 0.89 | 0.88 |
| **Overall solution coverage** | | 0.32 | | | | | | | | |
| **Overall solution consistency** | | 0.82 | | | | | | | | |

*Notes:* ● = core condition (present); ⊗ = core condition (absent); • = peripheral condition (present); ⊗ = peripheral condition (absent); blank space = the causal conditions may be present or absent.

p. 393