Does language prevent policy implementation? Evidence from the Italian Start-up Act

Michele Cantarella*[1], Nicolò Fraccaroli[2], and Roberto Volpe[3]

[1]University of Helsinki, [2]Brown University, [3]LUISS Guido Carli

3rd November 2022

Abstract

Does ethnolinguistic diversity prevent policy implementation? The Italian Start-up Act of 2012 is an ideal policy to investigate this question. The Act sets up a scheme of benefits which young firms can access by registering as "innovative start-up" on a voluntary basis. Due to the ethnolinguistic divide in Trentino-Alto Adige, we find that firms with German-named CEOs are less likely to register as start-ups than firms with Italian-named CEOs, while performing similarly in the years preceding the introduction of the policy. These findings are robust to regional heterogeneity correlated with surname origin. We exploit these asymmetries as an exogenous source of access to the benefit scheme. Our findings suggest that registered start-ups do not perform any better than comparable unregistered firms.

Keywords: *Ethnolinguistic asymmetries, Start-ups, Innovation policy, Linguistic clusters*

JEL codes: M13, Z13, D83, O38, O32

*Corresponding author; email: michele.cantarella@helsinki.fi.

1 Introduction

Does ethnolinguistic diversity prevent policy implementation? Existing works highlighted that linguistic cleavages affect the supply of policies, by lowering the provision of public goods (Beach and Jones, 2017; Alesina et al., 2003; La Porta et al., 1999), hindering redistribution (Desmet et al., 2012; Dahlberg et al., 2012) and growth (Alesina et al., 2003), or worsening government performance (La Porta et al., 1999).[1] However, evidence that linguistic differences pose a barrier to policy implementation is scarce. Does access to a policy vary across ethnolinguistic groups based in the same country? Are ethnolinguistic minorities less likely to seek access to the benefits of a policy? And if so, can we exploit these exogenous information asymmetries to study the effectiveness of said policies?

In this paper we investigate these questions based on the case of the Italian *Start-up Act*, a policy introduced in 2012 with the aim of boosting investment in innovation among start-up firms. The policy, which is still in force, provides a set of economic and legal benefits to eligible firms, which are recorded as 'innovative start-ups'. These benefits include tax incentives, easier access to public credit guarantees and more flexible legal provisions (fewer restrictions on contracts, protection from bankruptcy, etc.). This specific policy suits our research question particularly well because it the most recent and widely advertised national industrial policy to which access remains voluntary: eligible firms that are aware of the policy must submit a request to be registered as innovative start-ups and enjoy its benefits. Similar policies, such as the ones targeted at innovative SMEs, feature a much smaller pool of recipients and much higher barriers to access.

We study the implementation of this policy on a sample of ∼20,000 firms based in the region of Trentino Alto-Adige/South Tyrol.[2] This Italian region located on the border with Austria provides a unique setting to study how linguistic cleavages affect policy access and estimate the impact of the Start-up Act on firms' performance. A historical linguistic divide splits the inhabitants of the region: while they are all Italian citizens, some of them speak Italian, whereas a considerable proportion (around 35%)[3] speaks German as their main language, and bilingualism is not very common. As documentation on the policy is available only in Italian, these linguistic differences, while presumably insufficient to make the regulations unintelligible to German-speaking entrepreneurs, appear to hamper the spread of information about the policy within their networks, giving rise to language-biased information asymmetries. In fact, eligible companies with a German-speaking management are less likely to apply for the status of innovative start-up than those with an Italian-speaking CEO that share similar characteristics even within the same municipalities.

[1] For a review, see Alesina and La Ferrara (2005).

[2] The official name of this region is "Trentino-Alto Adige/Südtirol", acknowledging the equal footing the German language has with Italian in the autonomous province of Bolzano (Bozen in German). This province is normally called, in English language sources, "South Tyrol", a translation of Südtirol. For clarity and concision only, we will use this name throughout the paper.

[3] Official statistics on linguistic groups are provided for South Tyrol, while Trentino is almost unilingually Italian. The German-speaking group accounts for 69.41% of the total population of the province of Alto Adige/South Tyrol. Source: ASTAT's population census of 2011. Available at: `https://redas.services.siag.it/redasArticlesAttachment?attachId=562911` (last accessed on 26 January 2022).

We find that this happens regardless of regional heterogeneity or policy shocks. Also, once these factors are controlled for, we find no traces of differences in performance or innovation propensity between the two linguistic groups, as we show that German-speaking firms performed just as good as their Italian-speaking counterparts in the years preceding the introduction of the policy. Showing that entrepreneurial activity is affected only by the geographical determinants of ethnolinguistic assignment, these findings suggest that differences in policy adoption cannot be explained by one group being more prone to innovation or more successful in business than the other one.

We identify the membership of a firm to an ethnolinguistic group by looking at the etymological origin of the surname of its CEO. To this end, we develop a simple text-based method to classify over 55,000 board members' surnames as of Germanic (German-speaking) or Romance (Italian-speaking) roots. While language can be acquired through an individual's lifespan, and is therefore endogenous, the surname is not. We hence exploit the etymological linguistic origins of the surname of the CEO as an exogenous source of variation to explain the probability that a firm accesses the policy. The fact that surnames are outside of the control of individuals means that the surname distribution is only geographically determined, so conditional independence is easier to obtain by controlling for regional factors alone.[4]

Our work contributes to several strands of the literature. First and foremost, we complement the broader literature of cultural economics (e.g., Desmet and Wacziarg, 2020; Alesina and Giuliano, 2015; Guiso et al., 2015; Guiso et al., 2008), and in particular those studies that focus on the economic costs that stem from linguistic differences. While prior research showed how language diversity shapes economic behaviour (Montalvo and Reynal-Querol, 2021; Wang and Steiner, 2015; Falk et al., 2018; Galor et al., 2020) and the supply of policies and institutions (Bazzi et al., 2019; Beach and Jones, 2017; Desmet et al., 2016; Spolaore and Wacziarg, 2016; Desmet et al., 2012; Dahlberg et al., 2012; Alesina et al., 2003; Alesina and La Ferrara, 2005; La Porta et al., 1999), evidence on how linguistic cleavages affect policy access is less widespread. Some studies, such as Chen (2013) and, on Trentino Alto Adige specifically, Sutter et al. (2018) have found that languages can affect intertemporal preferences, but there is little evidence on whether languages can actually influence entrepreneurship activity.

Our results enrich the knowledge on this matter by showing that membership to a linguistic minority is a source of asymmetric information affecting the likelihood of policy transfer, and that language-biased differences in entrepreneurial activity are attributable to regional factors determining the distribution of languages, and not to language itself. The Italian region is also an ideal setting in the context of language asymmetries because (1) national policy is exogenous to the region, and (2) one of the language groups retains a distinct informational advantage when it comes to bureaucratic procedures that are fully documented in national language only. This is not the case in other multilingual contexts, such as Switzerland or Belgium, where multiple languages are simultaneously official national languages. When policies and administrative procedures are fully

[4]The Italian Civil Code (article 143 bis) also requires all married women to keep their maiden name (which is usually not the case for German-speaking countries), making the Italian setting well suited for this identification strategy.

documented in all languages, there is no reason to assume that informational asymmetries of this kind might arise. Policies in these cases might also be endogenous in the sense that they might already take into account the presence of multiple language groups.

Second, our evidence contributes to the scholarship of innovation economics that studies the role of start-ups. This is particularly important in the context of language-biased asymmetries as small firms might be more susceptible to the costs arising from uneven information flows, which have been shown to affect innovation (Murray et al., 2016). Start-ups are considered engines of economic growth thanks to their high potential for innovation (Andrews et al., 2014; Baumol, 2004; Acs and Audretsch, 1987) and important sources of job creation (Buldyrev et al., 2020; 2007; Calvino et al., 2015; Haltiwanger et al., 2013; Criscuolo et al., 2014). As a response to the decline in firm dynamism (Decker et al., 2016) and trends in automation and offshoring (Autor, 2013), and especially during the recent Covid-19 crisis, governments of advanced economies increasingly engaged in the design of policies to promote the formation and growth of young innovative firms (Schot and Steinmueller, 2018; OECD, 2020c; Breschi et al., 2018). These actions led to the proliferation of start-ups, which today account for about 20% of employment across OECD countries (OECD, 2020c) and for almost one-half of productivity growth in the United States (Klenow and Li, 2020).

Assessing the impact of policies that support start-ups is generally challenging: several elements, such as market concentration (Acs and Audretsch, 1988), human capital (Protogerou et al., 2017) and different returns to innovation (Andrews and Criscuolo, 2013) affect the ability of small firms to be innovative or to create jobs, and empirical evidence has shown that start-up creation is endogenous (Cavallo et al., 2018; Colombelli, 2016). Firms that benefit from start-up policies may be more dynamic *ex ante*, and not thanks to the policies. We contribute to this literature by estimating whether a policy that supports start-ups is actually effective in boosting their performance in terms of innovation, access to finance, and employment growth net of this self-selection problem.

By doing so, our work is the first that identifies the causal effect of the Start-up Act, a flagship policy of the Italian government, on firms' performance. The value added of our research design can be further appreciated when compared with other analyses on the Start-up Act specifically. Overall, these studies found the policy to be effective in improving the conditions of treated firms in terms of investment rate (Finaldi Russo et al., 2016), job creation (Manaresi et al., 2021; Biancalani et al., 2021; Ferrucci et al., 2020), access to funding (Manaresi et al., 2021; Biancalani et al., 2021; Ferrucci et al., 2020; Giraudo et al., 2019; Menon et al., 2018; Finaldi Russo et al., 2016), and potential output (Fiorentino et al., 2021; Ferrucci et al., 2020), but these works do not always share the same findings. However, while these studies evaluate the policy beyond the scope of a single region, they all rely on methods that mitigate the self-selection bias only partially (e.g. propensity score matching and conditional difference-in-differences).

By contrast, we find the policy to have a much more moderate impact over the firms which benefited from start-up status. The differences between our results and the ones in these works further motivates the use of a quasi-experimental design for the evaluation of this policy. Addressing the issue of endogenous self-selection is not simply an econometric exercise, but it enables us to shed

a better light on the actual impact of the policy, at least in contexts of high income regions with historically varying innovation propensity.

The remainder of this paper is structured as follows. The next section provides a background on the Start-up Act policy framework and on the linguistic divide in the Trentino-Alto Adige region. Section 3 describes the data we use and their sources. Section 4 presents our research design and discusses its inherent assumptions. Results are outlined in Section 6. Section 7 concludes.

2 Background

2.1 Start-up Act

This paper studies the impact of a package of benefits created for young innovative companies based in Italy since early 2013, colloquially called the "Start-up Act".[5] As stated in its preamble, the goal of the Start-up Act is to promote "sustainable growth, technological development, innovative entrepreneurship and youth employment," and thereby contribute to social mobility and the attraction of innovative firms and capital in the country.

The policy consists of two main elements. First, it introduces a legal definition of "innovative start-up" company (*start-up innovativa*) and sets the eligibility requirements that firms must meet to qualify as such. Second, it provides a package of benefits to said firms aimed at improving their performance, as well as reducing market entry costs.

Innovative start-ups are assigned several facilitations that have the aim to benefit all stages of start-ups' life-cycle, from birth to maturity (MISE, 2019). Indeed, the policy gives firms access to a heterogenous set of measures. Most measures are of a regulatory nature, rather than financial. Their overall aim is to make start-up incorporation and management less costly, particularly in the early years of activity. Among them, there is a simplified company incorporation procedure, as well as exemptions from duties, fees, and corporate governance amendments such as a waiver from standard bankruptcy procedure.

On the financing side, the policy aims at increasing credit flows towards start-ups. The key tool is a public guarantee facility for access to credit for SMEs. The primary aim of this facility is to reduce risk on the issuing institution by introducing more favourable loan conditions and lower interest rates. While most SMEs can in principle access it, the facility is automatically applied to all loan applications from innovative start-ups, irrespective of their risk profile and credit rating. Unlike other SMEs, start-ups do not face any application fees.

The policy also supports equity financing via tax breaks for seed- and early-stage investors. Benefit rates have grown over the years: as of 2021, shareholders can deduct from their income tax up to 30% (50% in certain cases) of the capital invested in the company. However, the Start-up Act itself did not create any state-funded vehicle for direct equity investment: even though public

[5]The full text of the Act can be browsed on Italy's Official Gazette, at: `https://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=2012-12-18&atto.codiceRedazionale=12A13277` (Last accessed: 20 January 2021).

venture capital funds exist in Italy,[6] they are aimed at high-tech companies in general, irrespective of "innovative start-up" status.

To be eligible for the policy, start-ups need to be limited liability companies, less than five years old, not be publicly listed, and report an annual turnover lower than EUR 5 million. The incorporation of these companies should not be the result of a branch split or merger from a previous company, and they should not have distributed profits. In addition, firms must have an objects clause explicitly related to technological innovation, and should fulfil at least one of the following requirements: a R&D expenditure ratio higher than 15%; at least one third or two thirds of staff holding a PhD or a master's degree respectively; ownership of legal rights for a patent or a software.

These benefits are not automatically applied: start-ups must first apply for inclusion in a "special section" of the Italian Business Registry ("Registro delle imprese"). The benefits are valid from the date of registration and for all the time during which they are registered in the "special section". Registration is revoked as soon as the company no longer fulfils legal requirements. This may be because the start-up no longer has a sufficient character of technological innovation, e.g. because its primary activity has changed, or because it "grows out" of innovative start-up status - registration automatically lapses 5 years after the date of incorporation of the company, or when its turnover exceeds EUR 5 million per year.

Self-selection is the distinctive feature of this regulatory framework: the benefits apply only to those eligible companies that voluntarily choose to register. This obviously implies that nascent and existing companies must be sufficiently informed about the policy to benefit from it, making its implementation vulnerable to knowledge gaps. Linkages with other players of the innovation ecosystem (e.g. start-up incubators and accelerators, investment funds and technology transfer institutions, as well as legal professionals and public bodies) may increase the likelihood to receive such information. As a consequence of this, it is widely assumed that a large population of potentially eligible companies exists outside the registry.

A second feature of the Start-up Act is its "non-discriminatory" approach. Besides the requirement of carrying out a "technologically innovative" business activity – which leaves room for arbitrary interpretations (Menon et al., 2018) – there are no economic sectors that are explicitly excluded. Moreover, there are no significant measures aimed at specific demographics, and most measures apply uniformly all over the country.

There are however several discernible patterns: start-ups are mostly concentrated in high-tech, low-capital sectors such as digital services. When compared to the average Italian SME, start-ups display a higher concentration of younger entrepreneurs (OECD, 2020b), and are more common in the north of the country and, in general, in high-income areas (MISE, 2021).

While free to pursue any kind of economic activity, in fact Italian start-ups tend to operate within specific economic sectors. In Appendix A, Table 3, we provide a list of the ten most common sectors of economic activity, which together cover more than half of the total number of start-ups.

[6] An example is the Fondo Nazionale Innovazione launched in 2019. More information: `https://www.cdpventurecapital.it/cdp-venture-capital/it/home.page` (last access: 26 December 2021)

We provide these figures for both Italy and the region of our analysis, Trentino-Alto Adige. The three most common sectors among those firms which achieved the start-up status are *Software production* (representing 22.9% of all start-ups in Italy, and 20.4% in Trentino-Alto Adige), *Experimental research and development in the fields of engineering and natural sciences* (10.1% and 16.9%), and *Web portals* (6.4%, and 4.1%).[7]

2.2 Ethnolinguistic divide in Trentino-Alto Adige

We propose that ethnolinguistic diversity generates information asymmetries which lowers the probability of policy adoption for one culturally distinctive group holding everything else equal. The region of Trentino-Alto Adige/South Tyrol is an ideal setting to study this question in the context of the Start-up Act due to its historical cultural divide.

The region is composed of two neighbouring provinces: the provinces of Trento ("Trentino") and Bolzano ("South Tyrol"). Trentino is the southern province, while the official name of the northern province, which shares its border with Austria, is Alto Adige/Südtirol, including both the Italian and German denomination.[8]

There is a clear divergence in terms of registration to the Start-up Act's benefits scheme between the two provinces. The southern half, Trentino, is the area in Italy where the Start-up Act is most popular: around 7.5% of all limited companies less than 5 years old registered as start-ups, by far the highest ratio in the country. Conversely, start-up density is much lower in South Tyrol: only 3.4% of all young limited companies registered as start-ups. While this statistic is in line with the national average, it is lower than most other affluent areas surrounding it. As a result, although the two provinces have a very similar firm population,[9] the number of registered start-ups in South Tyrol has been less than those that registered in the neighbouring province since the implementation of the policy.

In addition, the two provinces differ in terms of start-up geographical distribution. While start-ups are quite evenly distributed across the territory of Trentino,[10] in South Tyrol they are mainly concentrated in and around the provincial capital, Bolzano. While only 20% of the provincial population lives in the city, over two thirds of its innovative start-ups are based there. Such an intense concentration in the main population centre is not observed almost anywhere else in Italy.

Although some of these divergences could be explained by differences in political priorities and industrial structure between provinces, we propose that they derive from a historically-rooted cultural barrier which goes beyond aggregate differences between the two provinces. This barrier, which is attributable to the linguistic cleavage in the region, produces costs and information asymmetries

[7] The ATECO codes of these sectors are: 620100 for Software production, 721909 for Experimental research and development in the fields of engineering and natural sciences, and 631200 for Web portals.

[8] South Tyrol is part of the historical region of Tyrol, which includes the Austrian territories of North and East Tyrol.

[9] The average difference between the two provinces in terms of number of firms filing yearly accounts in each province was 433 between 2010 and 2019.

[10] Reflecting local population patterns, the ecosystem is polycentric, with start-up agglomerations also in semi-rural areas (OECD, 2020b).

between the two major ethnolinguistic groups.

While Italian is the main language in Trentino, South Tyrol is the only Italian province where Italian is not the majority language. Each South Tyrol resident is required to be registered into one of the three language communities. Almost 70% of South Tyrol's population declares German as their mother tongue, with Italian native speakers making up around 25%. The remainder speaks Ladin, a distinct Romance language typical of the eastern valleys of the province.

Putting language aside, Trentino and South Tyrol are similar in many respects. They have a similar size in terms of population (545,000 inhabitants in Trentino; 535,000 in South Tyrol) and are equally rural, with a large proportion of the population living outside the few mid-sized urban centres.[11] Their economic performance is comparable. They are among the richest provinces in the country, with South Tyrol ranking second and Trentino sixth in terms of gross domestic product across 100 Italian provinces,[12] and with a similar average household income[13] and economic structure (Banca d'Italia, 2021). Finally, from an administrative perspective, both provinces enjoy the status of *provincia autonoma*, which entrusts them with a large degree of legislative, fiscal and budgetary autonomy compared to other Italian local authorities.

Language is not only the main source of heterogeneity between the two provinces, but also within the South Tyrol province. Although education is bilingual (both Italian and German are compulsory subjects for members of both language groups from age six), effective bilingualism is not widespread. Several studies have pointed to a relatively low second language proficiency (L2) in the South Tyrolean population: only a minority of around 10-20% in either community is proficient (C1 level or above) in both languages, and the situation among high school pupils shows signs of deterioration in the last decade, with ever more students attaining only basic skills in their L2 (see Vettori et al., 2021; Abel et al., 2012; Vettori, 2016; Coia et al., 2012). There is evidence in support the idea that these language differences are strong enough to give rise to separate linguistic networks: German speakers strongly tend to interact more with other people sharing their same native tongue, and so do Italian speakers (Vettori et al., 2012; Vettori and Abel, 2017). Angerer et al. (2016) also find that cooperative behaviour towards children of the other language group tends to decrease with age.

It is reasonable to suspect that these language networks might have given rise to asymmetries that affect the probability of policy adoption. In line with this hypothesis, municipalities in South Tyrol with the highest density of Italian speakers are more likely to feature registered start-ups. Start-up density correlates with the proportion of Italian speakers at the municipal level even after controlling for population and urbanisation level (OECD, 2020a).

Following these considerations, we hypothesise that the language barrier has resulted in less

[11] Istat (2015): Principali dimensioni geostatistiche e grado di urbanizzazione del Paese. Available at: `https://www.istat.it/it/archivio/137001` (last accessed 27 January 2022).

[12] Data is from 2018. Source: Eurostat, Gross domestic product (GDP) at current market prices by NUTS 3 regions.

[13] In the period 2009-2018, the average household income in South Tyrol and Trentino was EUR 38,000 and EUR 34,000 respectively. Source: ISTAT elaboration based on EU-SILC, Household average income. Available at `http://dati.istat.it/Index.aspx?QueryId=22919#` (Last accessed: 21 January 2022).

effective policy transfer by giving rise to information asymmetries affecting the cost of registration. There are two factors that can contribute to the emergence of these asymmetries. First, official documentation in German[14] is extremely limited, and only refers back to the official documentation in Italian. Only the official norm is translated into German in the Official Regional Bulletin, in accordance with national law (*Decreto del Presidente della Repubblica n. 574, July 15th 1988*), but all the administrative formalities required for registration remain in Italian only. The online incorporation procedure was also entirely in Italian until the turn of the decade.

Second, linguistic networks shape and influence information flows. Cantarella et al. (2023) show that the linguistic networks in the two provinces affect the spread of online misinformation. In the same fashion, information on the Start-up Act, which is poorly documented in German, is also less likely to achieve the same degree of publicity that would have acquired elsewhere in the country, as it struggles to penetrate the German-speaking network of entrepreneurs alike. Accessing the Italian-speaking network might then come with non-negligible transaction costs to non-native speakers.

The lack of documentation in German on the fine details of the start-up legislation also makes it more difficult for professionals, such as business consultants, to offer high quality consulting services to entrepreneurs. To the best of our knowledge, there are no consultants in South Tyrol specialising on "innovative startups". We posit that, even if German-speaking consultants had full access to information about the policy, there would simply be no systemic incentives to specialise in this policy given the smaller pool of eligible recipients. This is an important disadvantage: available evidence on information channels for the spread of the Italian Start-up Act indicate that intermediaries play a crucial role in its diffusion, with up of two-thirds of entrepreneurs reporting the support of a *"commercialista"* (tax consultant) as a key factor for registration (Istat/Mise, 2018). Against this backdrop, it is likely more difficult for firms in municipalities with a German-speaking majority to access professional networks of consultants familiar with the policy.

These facts make the process of undertaking the effort to register much more costly for a firm operating within the German-speaking cluster. It should be noted that this hypothesis does not rest on the belief that German-speakers in South Tyrol do not understand Italian, but rather on the idea that, for entrepreneurs embedded in German-speaking networks, information on this policy found more barriers to its propagation than it did in mostly Italian-speaking ecosystems.

A potential counterargument is that the two groups might differ in the probability of policy take-up depending on different degree of trust in the government. For instance, German-speaking entrepreneurs might display a lower level of trust in the central government, which would discourage their policy access. If this was the case, then the mechanism would not necessarily lie on linguistic- and network-driven information asymmetries. However, we argue that since the benefits deriving from registration to the policy are *automatic*, and apply in areas which are beyond the scope of regional autonomy - e.g. corporate law, business taxation, and access to private finance - there

[14] Available at: `https://www.camcom.bz.it/de/dienstleistungen/handelsregister/dienste-des-handelsregisters/innovative-start-und-kmu` (Last accessed: 21 January 2022).

is no reason to believe that such benefits are less palatable to German speakers. If anything, the possibility that German speakers might distrust the policy because unaware of the extent of the benefits available supports the idea that information about the policy is not properly circulating within their linguistic network, and that, indeed, language-biased information asymmetries have arisen.

Another counterargument would imply that the take up rate is lower across the German-speaking simply because German-speaking firms do not innovate as much as Italian-speaking ones. A central objective of this paper is then to test whether differences in registration between the ethnolinguistic groups are robust to regional economic and policy heterogeneity correlated with language, or to differences in entrepreneurial activity and innovation propensity related to language.

3 Data

We construct our sample of firms using information from two main data sources. Our analysis primarily relies on business data from Bureau van Dijk's Aida database.[15] Aida is a subscription-based data resource on all Italian limited companies that have filed yearly accounts.[16] At the date of our extraction (November 2020) it covered information on 6.2 million companies, 22,944 of which are located in Trentino-Alto Adige. The available data cover years from 2010 to 2019, taking an unbalanced panel form, for a total of 115,000 observations. As we describe in the Appendix D, for the purpose of our analysis we reduce this sample to only companies that have been eligible for the status of innovative start-up at some point of their existence, leaving between 11,626 and 26,264 observations depending on the preferred strategy.

Data from Aida include company location, economic performance, and activity status, allowing us to control for a wide range of firm-specific characteristics. Moreover, the database provides profiling information on current and previous company directors including their full name, which is particularly important to define our instrumental variable.

We complement this information with open data from the Italian Business Registry. This dataset reports information on the list of companies registered in the "special section" for innovative start-ups and their date of registration.[17] This information allows us to accurately reconstruct the list of all companies registered as innovative start-ups for at least one week between 1 January 2013 and 31 December 2019, which serve as the reference dates for this work. After reconstructing the historical data on company directors,[18] we match the firm IDs of the two datasets, yielding the final

[15] Aida's page on Bureau van Dijk's website: `https://www.bvdinfo.com/en-gb/our-products/data/national/aida` (Last accessed: 22 January 2022).

[16] Aida does not include companies not limited by shares, such as partnerships, and individual entrepreneurs.

[17] The list can be downloaded free of charge, on registration, at the following address: `https://startup.registroimprese.it/isin/report?2&fileId=startup.zip`. The public list only offers a snapshot of all firms currently registered as innovative start-ups, meaning that firms that have been dissolved or have lost their start-up status will no longer appear. Since the introduction of the policy, the authors have collected and collated the information that has been published from this website, creating a single dataset comprising all firms that have ever obtained the start-up status. These data are available upon request.

[18] AIDA data only offer longitudinal information on firms' financials, while current and previous members of the

longitudinal dataset used in the estimation.

In order to identify the ethnolinguistic cluster of a CEO we focus on the word endings of their surname. We define the language group derived by the etymological origins of a CEO's surname as the *assigned language*. The ending of a surname is a reliable indicator of whether the surname has West Germanic or Romance roots.[19] For instance, German surnames typically end in "-er", such as "Müller" or "Meyer", whereas Italian surnames do not. We use a simple text analysis approach to classify surnames to the two linguistic groups. We describe our methodology in detail in Appendix B. Based on this strategy, we create a binary indicator which equals one if the surname of the firm's CEO is German and zero otherwise.

4 Empirical Strategy

Our empirical strategy hinges on the reduced form connection between ethnolinguistic assignment and policy adoption. We exploit the etymological root of surnames as a proxy for assignment, which captures the intention to treat better than the language spoken by a CEO:[20] while language proficiency is endogenous as it can be acquired throughout life, surnames are assigned at birth and therefore do not depend on CEO's personal characteristics or choices. This accounts for the possibility of a CEO being bilingual or fluent in the non-native language, which might correlate with other factors such as skill, income or education.[21]

In terms of second stage, the linguistic instrument is particularly helpful to overcome a major limitation related to the evaluation of the Start-up Act: treated firms normally enjoy a high growth potential before benefiting from the policy, which makes it hard to disentangle the effect of the policy from the firm's ex ante growth potential. In other words, these companies might have performed just as well without benefiting from the scheme.

Nevertheless, a potential limitation related to the use of surnames as exogenous variables is that their geographical distribution can be correlated with local factors. The same factors are potentially linked to other determinants related to local economic dynamism and industrial structure, provincial policies or simply even geography (OECD, 2020b,a; Banca d'Italia, 2021). Differences in economic performance might also emerge across the two linguistic groups in specific situations, since ethnolinguistic diversity can generate economic externalities (Wang and Steiner, 2015, 2019).

We then want to ensure that we these factors are controlled for, as they would provide biased estimates of the connection between language assignment and policy adoption, and might also affect our second stage estimates if they affected the outcome through channels other than policy adoption. The advantage, here, is that these regional factors are easier to control for as regional factors are

board of directors are reported in cross-sectional form. We then reconstructed each member's years of activity by using information on their date of appointment and resignation.

[19] Some adjustments are needed to control for special cases, as we describe in detail in Section B of the Appendix.

[20] Also, AIDA does not contain information on the language of CEOs or board members.

[21] A CEO might belong to the German linguistic group and have an Italian surname or vice versa. This is not an issue for the estimation of the ATE, as long as the connection between surnames and policy adoption stays strong enough.

time-invariant and provincial policies feature higher-level variation. In this way, we can exploit variation in CEOs' language assignment net of the characteristics of the municipality their firm is based in.

To address these issues, we develop three estimation strategies, each testing the robustness of the connection between ethnolinguistic diversity and policy adoption under different specifications.

Section 5 discusses in detail the main conditions under which our identification strategy holds. In particular, the section focuses on how capturing the variation of surnames net of geographical variation requires a 'stricter' independence assumption; i.e. that surname assignment is not completely determined by location. Furthermore, we discuss and test the assumptions of validity and exogeneity which apply to the surname-assignment instrument. We do so by providing evidence for the connection between surnames and language in Trentino Alto Adige, and testing for parallel trends between the two language groups during the pre-policy years. Overall, our tests support the idea that, net of municipal and provincial variation, surnames predict ethnolinguistic assignment, yet fail explain differences in performance among firms.

4.1 Three-way fixed effects

In our first specification, we employ a three-way fixed effects strategy that holds municipality, year and municipality/year effects constant. This strategy ensures that all unobserved time-invariant and time-variant factors associated with each municipality are held fixed, and that we study heterogeneity across firms only.

We then estimate the following two-stages specification over the full sample of eligible firms i in the region between 2013 and 2019, for municipality m and year t:

$$Registered_{imt} = \alpha_{m1} + \beta_{t1} + \gamma_{mt1} + Surname_{imt}\delta_1 + Controls'_{imt}\zeta_1 + Financials'_{imt-1}\eta_1 + e_{1imt} \quad (4.1)$$

$$Y_{imt} = \alpha_{m2} + \beta_{t2} + \gamma_{mt2} + \widehat{Registered}_{imt}\delta_2 + Controls'_{imt}\zeta_2 + Financials'_{imt-1}\eta_2 + e_{2imt} \quad (4.2)$$

Equation 4.1 is the first stage. It yields the effect of the etymological origins of the CEO's surname on policy access. *Registered* is a binary indicator that takes value of one if a firm has registered as innovative start-up, and zero otherwise. The dummy variable *Surname* captures the effect of interest. This variable equals one if the CEO of firm i has a surname of German (Germanic) origins, and zero if it has Italian (Romance) origins.

Surname is excluded in the second stage equation 4.2, which measures the effect of being registered as innovative start-up on the dependent variable Y_{it}, for each firm i in year t. The treatment variable is now the fitted value for *Registered*. The coefficient of interest is δ, which describes the effect of the policy on Y_{it}.

We estimate the effect of the Start-up Act on three separate dependent variables contained in

Y_{it}: expenditure in R&D, bank loans, and number of employees (all in log terms). We select these variables to study the effect of the policy on innovation, credit access and job creation respectively, which as discussed were the main aims of the policy.

We include the vector $Controls$ to control for firm level characteristics. These include board size and CEO characteristics (age, gender, foreign nationality, foreign nationality × age). These controls are all interacted with dummies for the years of activity of a firm to account for heterogeneity related to firm age.

In the $Financials_{t-1}$ vector we include the logarithm of lagged assets at $t-1$, along with lagged outcomes (R&D, bank loans, and number of employees at $t-1$), along with dummies for each of these variables being zero.

The three-way fixed effects allow us to control for a number of unobserved factors. α_m will capture unobserved time-invariant municipality effects, while year effects are captured by β_t. We also control for unobserved time-variant heterogeneity across municipalities by interacting the two in γ_{mt}. Province-specific factors, along with time-variant policies which are collinear with these factors and are then fully captured as well. We intentionally do not include industry fixed effects as they can generate collider bias.[22]

4.2 Difference in differences: Municipality and province/year fixed effects

The previous model did not explicitly control for parallel trends among language groups. To offer an alternative model in which these differences are explicitly estimated, we develop a difference in differences strategy to hold differences among firms belonging to different linguistic groups as fixed.

Including available pre-treatment years, we estimate the following equation over all eligible firms between 2011 and 2019:

$$Registered_{impt} = \alpha_{m1} + \beta_{t1} + \gamma_{pt1} + Surname_{impt} * Post_t\delta_1 + Surname_{impt}\theta_1 + Controls'_{impt}\zeta_1 + Financials'_{impt-1}\eta_1 + Trends'_{impt}\lambda_1 + e_{1impt} \quad (4.3)$$

$$Y_{impt} = \alpha_{m2} + \beta_{t2} + \gamma_{pt2} + \widehat{Registered}_{impt}\delta_2 + Surname_{impt}\theta_2 + Controls'_{impt}\zeta_2 + Financials'_{impt-1}\eta_2 + Trends'_{impt}\lambda_2 + e_{2impt} \quad (4.4)$$

This specification is similar to the previous one, except that pre-policy differences between the two groups are held fixed as $Surname$ is now an included instrument. The language group effect on registration is now captured by the interaction between $Surname$ and the post-treatment dummy $Post$, denoting post-treatment years, in our first stage equation 4.4. As firms cannot register during their pre-treatment years, $Registered$ is still the variable of interest in our second stage 4.3.

This specification belongs to the family of staggered difference in differences (DiD) (such as

[22] While is tempting to add industry fixed effects, industry codes would be "bad controls". Firms select these codes, creating a potential source of selection bias: the choice of these codes might result from the same decision to register as a start-up, invalidating our estimates. This is an important issue especially when we consider the decision to apply for start-up status might predate the industry code chosen by the firm.

the models recently discussed by Wooldridge, 2021), but with a "fuzzy" relationship between group assignment and treatment. The only caveat is that the treatment is not "sharp", since language group membership influences the *likelihood* of treatment, hence the instrumental variable approach. The exposure effect will then be consistently estimated as long as this effect is stable and homogeneous (de Chaisemartin and D'Haultfœuille, 2017).

With none of the firms in our sample relocating their registered office, within municipality/year effects would absorb the time variation from the *Post* indicator, so we drop the municipality/year term. We still want to control for time-varying differences in, at least, provincial policies, which are captured specifically by the γ_{pt1} fixed effect. Province factors can become especially concerning if there is any time variance in provincial policies. Most significant interventions precede our sample timeframe by a few years, but lesser intervention might still affect our estimates. The higher-level province/year effect will ensure that this source of time-variant unobserved heterogeneity is held fixed.

We now add to the $Trends$ vector observable time-varying municipality controls, which were previously absorbed by the municipality/year interactions. These controls include firm density (measured as the natural logarithm of the number of firms in a given year and municipality), but also the logarithm of the average of all outcomes (employment, bank debt and R&D expenses) in year t and municipality m for all firms except i. In a difference in differences fashion, these average outcomes are also interacted with $Surname$ to account for different levels of investment in innovation across language groups.

All other controls remain identical to the ones previously listed in subsection 4.1.

4.3 Difference in differences: Firm fixed effects

None of the models discussed above can fully account for time-invariant firm heterogeneity. While not all startups were born as such, the registration status of innovative start-ups has very low variation (at the national level, 20% of all startups have registered in their first year). Nonetheless, the introduction of the policy in 2013 created a discontinuity that we can exploit, as we can expect to find a higher proportion of older firms applying in that year. With firm fixed effects, we can then study registration holding these firm fixed effects as fixed keeping the same instrumented difference in differences strategy from above, under the caveat that these results will be valid for the subset of firms which have changed their registration status at least once. We then develop the following model:

$$Registered_{it} = \alpha_{i1} + Surname_{it} * Post_t\delta_1 + Surname_{it}\theta_1 + Controls_{it}^{'}\zeta_1 + Financials_{it-1}^{'}\eta_1 + Trends_{it}^{'}\lambda_1 + e_{1it} \quad (4.5)$$

$$Y_{it} = \alpha_{i2} + \widehat{Registered}_{it}\delta_2 + Surname_{it}\theta_2 + Controls_{it}^{'}\zeta_2 + Financials_{it-1}^{'}\eta_2 + Trends_{it}^{'}\lambda_2 + e_{2it} \quad (4.6)$$

With this setting, we exploit the discontinuity in the introduction of policy to study whether firms which applied for startup status later in their life also had a Italian-assigned CEO (equation 4.5). We then exploit this source of exogeneity to estimate the policy effect in the second stage equation 4.6.

Year fixed effects would be collinear with firm age fixed effects, so we keep the latter. The initial year of registration is then absorbed by the firm fixed effects, while the average variation in outcomes for all other firms in municipality m and year t is controlled for again by our set of controls $Trends$. Lags are also dropped as they are now superflous with firm fixed effects.

It should be noted that this model does not exploit within-firm changes in the assigned language of the CEO as an instrument, but just the discontinuity in registration emerging from having a German-assigned CEO after the policy is introduced, holding pre-trends as fixed.

The reader might however be interested in looking at whether the variation in CEO assignment alone can affect policy adoption. This would require a slightly different model, but adding the $Post$ control to both stages is sufficient to produce these estimates.[23] We provide estimates for within-variation in CEO assignment for the first stage. For what concerns the second stage, as both startup registration and CEO assignment are nearly time invariant, estimates exploiting variation in CEO assignment alone will inevitably apply to a very small complier group of firms which have both changed the language group of their CEO and applied for startup status.[24]

5 Discussion of model assumptions

Our second stages rest on the typical instrumental variable assumptions. The first is independence: we assume that the surname of the CEO precedes policy assignment. The second is relevance, for which we assume that the surname of the CEO predicts the probability that the firm accesses the policy. The third is the exclusion restriction, for which we assume that there are no differences in performance among firms belonging to the two linguistic groups as defined by the CEOs' surnames.

Our focus on the first stage mechanism imposes a further assumption for the interpretation of the role of ethnolinguistic networks and policy adoption: that the origin of the surname is a good predictor of the ethnolinguistic group of a firm, which in turn leads to asymmetrical policy access for the minority group. This is linked to the relevance assumption, as it focuses on the mechanism underlying the validity of the first stage. Our second stage is indifferent to whether this specific connection between surname and ethnolinguistic grouping stands: our instrument is relevant as long as German-assigned CEOs are more likely not to register their firm as a start-up.

Concerning the first instrumental variable assumption, independence, we have discussed earlier how the origins of a CEO's surname are a broader indicator than the language spoken by the CEO, but have the advantage of being random at the individual level. Surnames by definition predate

[23] Equivalently, time fixed effects can be used.

[24] For this reason, these estimates are not particularly informative and feature bloated coefficients (their sign and significance, or lack of, is however still in line with the rest of our estimates), so we do not report them in this paper. They are, nonetheless, available upon request.

policy access and are exogenous to unobserved ability factors that could influence proficiency in Italian for a non-native speaker. However, to aid our interpretation of our first stage mechanism and strengthen our exclusion restriction, we would like to impose a stricter independence assumption further imposing that surnames are independent of the location of the firm and provincial policy shocks. We now discuss how our three estimation strategies can help us satisfy this assumption.

First, as we are using surnames as proxies for firm assignment, it could be argued that the surname of the CEO is endogenous to the firm, as the process leading to the selection of CEOs is not the result of a random process if the manager is hired by the firm. This could create a selection issue, à la Ham and Lalonde (1996). We argue that this is hardly the case, as arguably a firms's cultural milieu is also independent from registration as a startup. Previous studies have shown that cultural affinity can play an important role in hiring decisions for managerial roles (Westphal and Zajac, 1995; Damaraju and Makhija, 2018), so it is not a stretch to believe that firms would hire CEOs from their own language networks. This assumes that a hiring process has actually taken place as, for young firms such as the ones in the sample, the CEO is often also the founder, making the concerns for firm selection moot. For our intents and purposes, what matters is that the ethnolinguistic group of either the CEO or the firm is plausibly exogenous. The central assumption here is that, after controlling for regional factors, surnames become conditionally independent to the firm.

We test these arguments by providing balance tests for surname assignment in Appendix E, Table 5, in which we show that German-assignment is independent from other CEO and firm characteristics for all eligible firms during the pre-policy years. We show that age and gender are independent from language assignment suggesting that no particular linguistic group favour entrepreneurs based on age profiles or gender characteristics. Most importantly, board size and board language diversity are also independent from linguistic assignment, suggesting that neither language does provide a group with a significant advantage against the other. Foreign CEOs tend instead to be older and German-speaking, but this is easily explained by the contamination from neighbouring German-speaking areas.[25] Firm age is the only significant, and potentially endogenous, predictor of language assignment. The most important finding, however, is that municipality fixed effects absorb nearly all significance away from other predictors showing that the few differences found were to be attributed to geographic factors and not to differences in hiring practices between the two groups.

The Table also shows how the proportion of German-speakers on the board also strongly correlates with the assignment of the CEO, raising the model fit to more than 80%, which suggests that CEO assignment can instrument for language assignment of the firm. Finally, the proportion of German speakers in a municipality, added a control in the last two columns, is also shown to be independent from the assignment of the CEO after controlling for municipality fixed effects, confirming our conditional independence assumption.

Relevance of surname assignment to CEO language is assumed in the model, and relates to

[25] We then control for these factors in all our specifications.

the interpretation of our first stage mechanism. Abundant evidence shows that surnames are good predictors of language group assignment of a firm. Figure 1 plots the degree of diversity within all firms in the region by looking at the surnames of each member of the board of directors for each firm. The x-axis is the proportion of board directors with a German surname. A high density of mixed-surname firms would have indicated that either cultural affinity and language separation were not influential in terms of board compositions, or that the connection between surnames and language was simply not strong enough. Instead, the distribution has two fat tails, indicating that most firms employ a fully Italian- or German-assigned board. The figure also shows that and this result is impressively stable over the years, suggesting a high degree of cultural separation. In line with this evidence, Figure 2 shows that firms based in German-speaking areas are more likely to hire a CEO with a German surname, supporting our theoretical assumption that surnames are strong predictors of assignment to a linguistic group.

There is further evidence in support of the link between surnames and language. When comparing the frequency distribution of surnames (Astat, 2010) and the prevalence of linguistic groups (Astat, 2011) by municipality in Alto Adige, we notice that the most common surnames in municipalities with a German-speaking majority are almost entirely etymologically German. As an example, Table 4 in Appendix B shows the density of the ten most common surnames across the two province capital cities: Trento and Bolzano. In the Italian-speaking Trento, Italian surnames are much more common than in the majority German-speaking Bolzano. In addition, mixed marriages among partners of different linguistic groups are still largely uncommon: only 1 in 10 couples are mixed in terms of their native language, a statistic that has remained stable since 1981 (Buiting and Pokriefke, 2016).

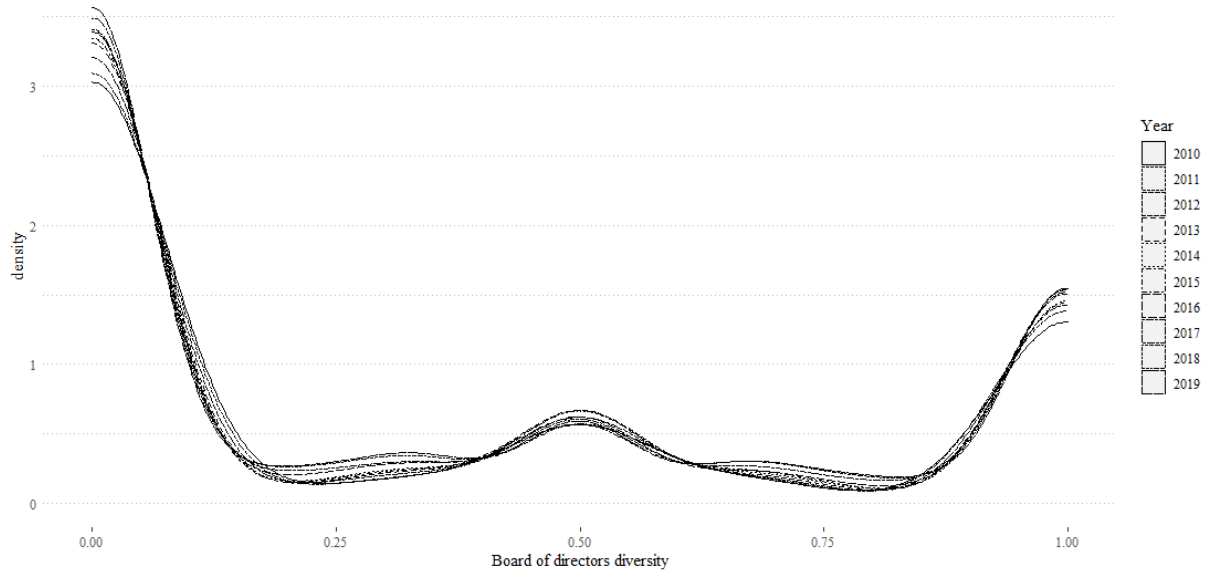

Figure 1: Density plots for diversity within board of directors surnames of firms in Trentino-Alto Adige. 0 = board composition made up of Italian-assigned members only; 1 = German-assigned members only. Firms with board size = 1 are excluded from the sample.

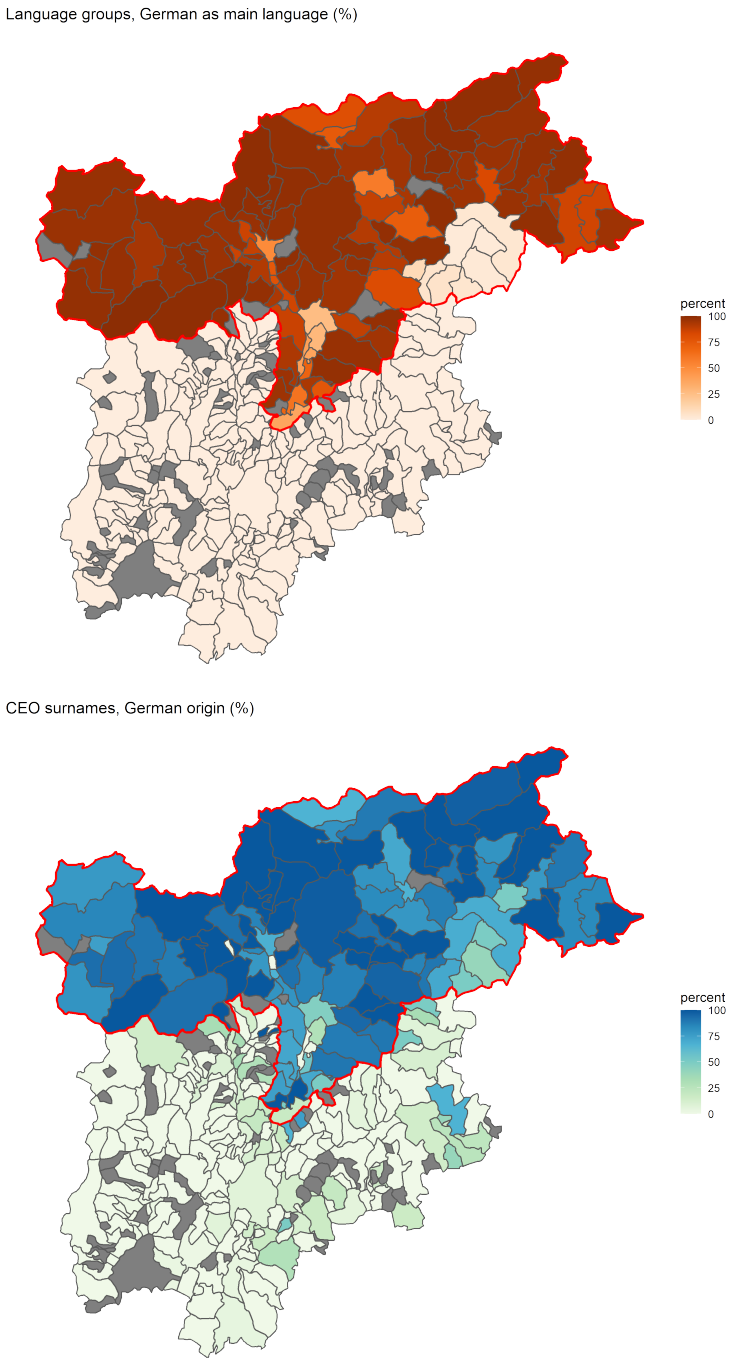

Figure 2: Language groups (2011) and etymological origin of CEO surnames (2010-2019 average) in Trentino-Alto Adige. Authors' calculations from census and AIDA data. Municipalities with no firm data are grey-shaded.

Since we have already discussed how strong ethnolinguistic networks are in the region (subsection 2.2), an unsettled question is whether these networks alone can justify heterogeneity in policy adoption. Disentangling whether the policy divide associated with language assignment is driven by language networks or rather other geographic, industrial and policy factors correlated with linguistic differences is a central part of our analysis. Here the advantage is that policies are time-dependent

and that municipalities do not move around, so our estimation strategies, focusing on holding time invariant, but also time variant, municipality and province factors fixed will allows us to study the relevance of the instrument net of these confounders.

These considerations about geographical and policy confounders also apply to the second stage only if these differences could lead to violations of the exclusion restriction. Also, some might wonder as to whether German-assigned firms perform better or worse than Italian-assigned ones regardless of where they are located.

There are, in fact, some industrial and policy differences between the provinces. Banca d'Italia (2021) shows that certain industries are subject to idiosyncratic shocks depending on whether these activities are based in the province of Trentino or South Tyrol. These shocks are specific to the alpine tourism and building sectors, and to the manufacturing industry focusing on exports to Austria and Germany. Since German-speaking CEOs are much more common in South Tyrol, the violation of the exclusion restriction might arise because firms in this area are more likely to operate in specific sectors with idiosyncratic performance, openness to credit and innovation propensity. At the same time, we need to investigate differences between the provinces in terms of investment in start-up creation (Banca d'Italia, 2021), which could provide a further source of heterogeneity between the provinces that might affect our estimates. In this regard, while we are not directly comparing the two provinces but rather CEOs conditional on their language, the correlation between geography and assigned language remains strong.

These concerns motivate our three different estimation strategies, as they focus specifically on keeping these policy and industrial factors as fixed. Furthermore, the firm fixed effect strategy also keeps between-firm variation as fixed, capturing all heterogeneity between firms that could relate to location factors. Additionally, our censoring approach based on sectors of economic activities described in Appendix D ensures that only firms which could realistically apply for startup status are considered, so the industries subject to provincial shocks are unlikely to relate to innovative firms anyways.[26]

After controlling for these regional and policy factors, we have no reason to believe that firms fared better or worse because of the assigned language of their administrator. Also, as discussed earlier, none of the benefits attached to the policy discriminate over linguistic assignment alone and as such are equally attractive to firms regardless of language, as any firm can equally benefit from tax breaks, credit guarantees and all the reliefs that come with the policy.

From this point of view, it is difficult to believe that, net of regional heterogeneity, one language group might be more prone to innovation or entrepreneurial activity without resorting to ethnic stereotyping. Even if outcomes still differed across groups, the difference in differences models discussed in subsections 4.2 and 4.3 would keep trends between group explicitly fixed, making these concerns moot.

Yet, we can test for this possibility by looking at the trends between firms managed by Italian

[26]Table 3 shows that none of these industries appear among the ten most common sectors of economic activity for start-ups in Trentino-Alto Adige.

and German-named CEOs during the pre-treatment years. We test this in Figure 3, in which we plot group-specific trends for all the outcome variables under analysis during the pre-treatment years (2010 to 2012) and after controlling for municipality fixed effects and all non lagged controls. This check is equivalent to a zero first stage test, as presented in Bound and Jaeger (2000), Altonji et al. (2005).

The figure provides striking evidence in support of our assumptions concerning the exclusion restriction. Not only did the two groups develop in parallel, but they also present no significant difference between outcomes in most instances: the mean value for one group is always well within the standard error of the mean of the other group, suggesting that any deviation from the latter is not different from random noise. Point deviations are also extremely contained, so it is difficult to argue that, after controlling for municipal characteristics alone, there was any difference in outcomes between the two groups. Repeating this same test with three-way fixed effects and firm fixed effect would only show closer trends, suggesting that the assigned ethnolinguistic group plays no effect at all over firm performance and that local factors could possibly be the only real source of endogeneity. This suggests that all three models outlined in section 4 are valid in terms of exclusion restriction.

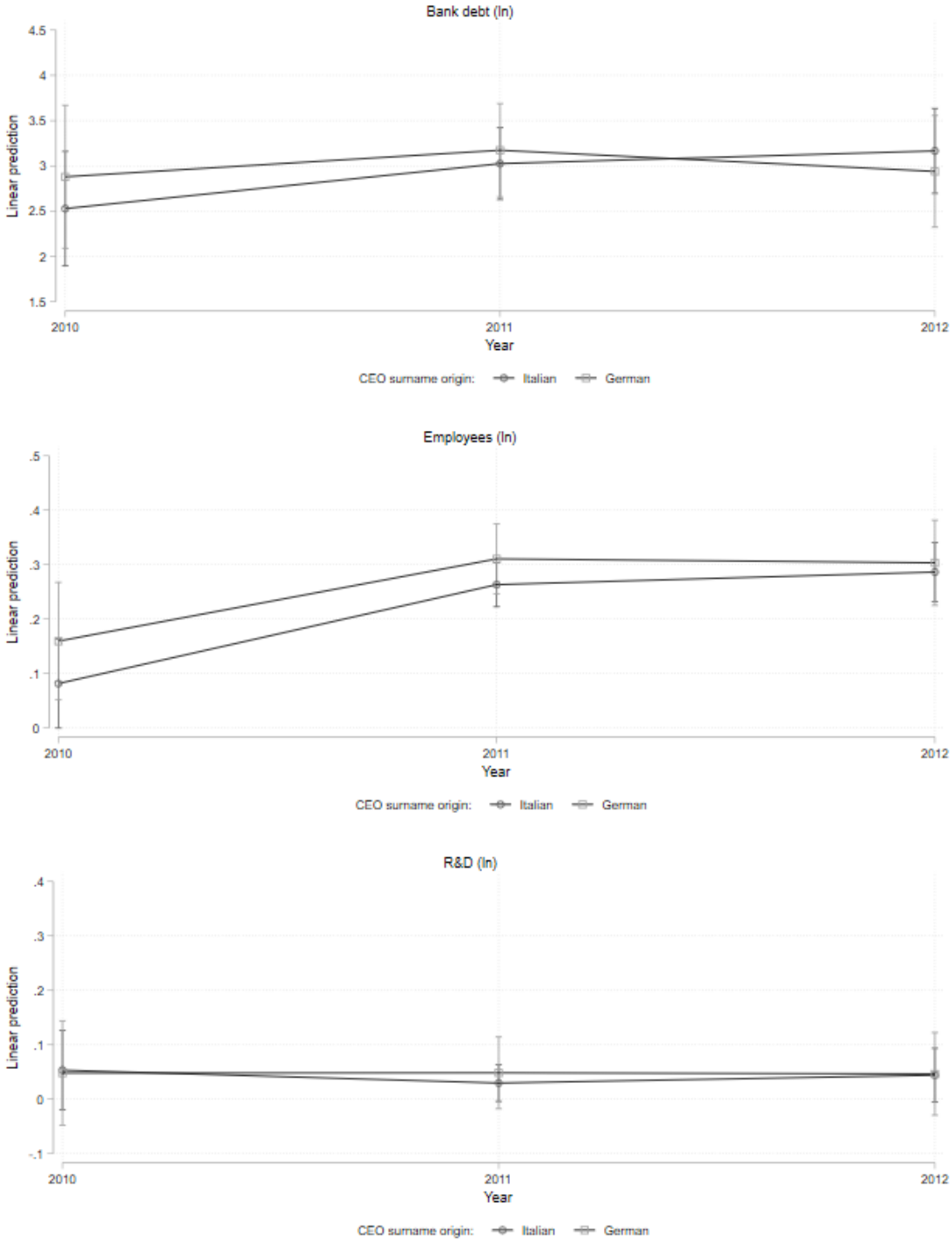

Figure 3: Etymological origin of CEO surnames and firm performance in Trentino-Alto Adige prior to the enactment of the Start-up Act. Linear prediction, with controls for year, years of activity, firm density, CEOs characteristics and municipality fixed effects. Standard errors clustered by firm.

The zero first stage tests also serve as further balance tests for linguistic assignment. Against the backdrop of our evidence on the independence of surnames, it is difficult to argue that language assignment, net of geographical factors, is sufficient to generate significant differences in performance that could affect our results other than through the asymmetries in access to information they generate.

6 Results

Table 1: First stages: Ethnolinguistic assignment and policy adoption

	(1)	(2)	(3)	(4)
	Registered as startup			
VARIABLES	OLS	OLS	OLS	OLS
Surname	-0.020***	-0.125	-0.018	-0.002
	(0.005)	(0.114)	(0.072)	(0.074)
Surname × Post		-0.053***	-0.027***	-0.033***
		(0.012)	(0.004)	(0.007)
Post				0.009
				(0.008)
Firm density (ln)		-0.075	-0.069***	-0.073***
		(0.049)	(0.015)	(0.014)
Observations	26,264	12,260	11,626	11,626
Adjusted R-squared	0.044	0.187	0.882	0.882
F-Test	17.90	20.49	53.72	22.79
Year fixed effects	Yes	Yes	No	No
Years of activity fixed effects	Yes	Yes	Yes	Yes
Municipality fixed effects	Yes	Yes	No	No
Municipality/Year fixed effects	Yes	No	No	No
Province/Year fixed effects	No	Yes	No	No
Firm fixed effects	No	No	Yes	Yes

Notes: SE clustered by firm in parentheses. Other controls: for column (1), see equation 4.1; for column (2), see equation 4.3, for columns (3) and (4), see equation 4.5.

*p<.05; **p<.01; ***p<.001

We first discuss the link between language assignment and policy access. Table 1 illustrates this connection under each of the specifications discussed in section 4.

Column (1) features results from the three-way fixed effects model (subsection 4.1), holding municipality, year and municipality/years as fixed. The estimated effect of assignment to the German-

speaking is negative and statistically significant at the 0.001 level. In other words, firms with CEOs with an etymologically German surname are found 2% less likely to apply for startup status, regardless of municipality, year, and time-variant shocks exclusive to each municipality.

Second, we examine the difference in differences models. With the introduction of pre-policy years, the coefficent of interest is now the one of the interaction between surname and the post-treatment years $Suname \times Post$.

In column (2) we display the results from the diff-in-diff model with municipality fixed effects and province-year shocks (subsection 4.2). We find a slightly stronger effect of language assignment, which drops to -5.3%. These results constitute a lower bound for our estimates, as they allow for time variant municipality-specific shocks to operate within language assignment net of higher-level, time-variant province-specific shocks. With only province/year fixed effects, the effect is equally significant and estimated at -6.8%.[27]

Column (3) replaces all fixed effects with firm fixed effects (subsection 4.2), exploiting the discontinuity in policy introduction to show results which are valid for all firms which were not born as startups. Our estimates indicate that among this group of firms those with German-assigned administrators were 2.7% less likely to apply for the policy. Column (4) further investigates the issue adding back time variation (through the *Post* control) to study whether firms which have changed a CEO for another one from the opposite language group were more or less likely to apply for the policy. Our results are particularly important as they indicate that switching to an Italian administration increased the probability of applying for startup status, as firms which switched to a German one are estimated 3.3% less likely to apply.[28]

All results presented are robust to the inclusion of controls for firm characteristics and municipality trends, when the latter are allowed. Apart from firm density, all controls for trends in outcomes (omitted from the Table for brevity) have a near zero effect on registration, including interactions between language groups and these trends. Firm density alone seems to play a role in registration in the firm, but only in the firm fixed effects models.

In all cases, the coefficients of interests are statistically significant at the 0.001 level, with the language instruments passing conventional weak-instrument F-Tests. Standard errors are always clustered at the level of the firm.

All these results point to a strong connection between language assignment and policy adoption. Are then surnames good predictors of this policy divide? As discussed, the results presented so far are reduced form estimates for the effect of language group assignment on policy adoption.

It follows that these results do not necessarily suggest that language alone is the only source of asymmetries in policy adoption, even if most of the evidence seems to highlight a very strong connection between surname and language (see discussion in the previous section). The robustness of these results to time-variant and invariant municipality and province shocks, and firm fixed effects, suggests that these differences cannot be explained by idiosyncratic municipal and firm

[27] Not shown in the table for brevity. Estimates are available upon request.

[28] Equivalently, employing two-way firm-year fixed effects leads to the same result.

Table 2: Second stages: The effects of startup registration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Has R&D		R&D (ln)		Has Bank Debt		Bank Debt (ln)		Has Employees		Employees (ln)	
VARIABLES	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
						Model 1: Three-way f.e.						
Registered	0.038***	0.083	0.437***	0.607	-0.003	0.037	-0.101	0.037	-0.012	0.336	-0.071***	0.458
	(0.007)	(0.077)	(0.081)	(0.731)	(0.009)	(2.767)	(0.087)	(2.767)	(0.014)	(0.334)	(0.018)	(0.555)
Observations	26,264	26,264	26,264	26,264	26,264	26,264	26,264	26,264	26,264	26,264	26,264	26,264
Adjusted R-squared	0.248	0.200	0.277	0.239	0.380	0.363	0.428	0.363	0.512	0.453	0.646	0.605
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years of activity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality/Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
						Model 2: DiD, municipality-province/year f.e.						
Registered	0.043***	0.034	0.485***	0.282	0.002	0.043	-0.051	0.958	0.044***	-0.099	0.024	0.295
	(0.008)	(0.074)	(0.084)	(0.717)	(0.011)	(0.348)	(0.100)	(3.972)	(0.015)	(0.302)	(0.018)	(0.487)
Surname	-0.030	-0.032	-0.217	-0.271	-0.037	-0.026	0.238	0.504	0.203	0.166	0.206	0.278
	(0.030)	(0.031)	(0.281)	(0.286)	(0.137)	(0.168)	(1.445)	(1.805)	(0.143)	(0.162)	(0.211)	(0.260)
Observations	12,260	12,260	12,260	12,260	12,260	12,260	12,260	12,260	12,260	12,260	12,260	12,260
Adjusted R-squared	0.239	0.203	0.270	0.236	0.339	0.264	0.386	0.313	0.508	0.428	0.606	0.545
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years of activity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province/Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
						Model 3: DiD, firm f.e.						
Registered	0.064***	-0.124	0.760***	-0.152	0.014	0.264	0.206	4.834	0.200***	-0.880	0.256***	0.091
	(0.024)	(0.170)	(0.285)	(1.503)	(0.038)	(1.000)	(0.398)	(11.514)	(0.047)	(0.787)	(0.063)	(1.088)
Surname	-0.187**	-0.197**	-1.723**	-1.770**	-0.223	-0.210	-1.221	-0.981	0.242	0.186	0.122	0.113
	(0.080)	(0.080)	(0.757)	(0.749)	(0.294)	(0.301)	(3.090)	(3.179)	(0.203)	(0.216)	(0.347)	(0.346)
Observations	11,626	11,626	11,626	11,626	11,626	11,626	11,626	11,626	11,626	11,626	11,626	11,626
Adjusted R-squared	0.347	-0.027	0.363	0.001	0.504	0.003	0.558	-0.002	0.747	-0.103	0.822	0.092
Years of activity fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: SE clustered by firm in parentheses. ther controls: for Model 1, see equation 4.1; for model 2, see equation 4.3, for model 3, see equation 4.5.

*p<.05; **p<.01; ***p<.001

factors, provincial policy shocks or simply by proximity to Italy or German-speaking countries. In other words, even in German-speaking municipalities, CEOs of innovative startups tend to belong to the Italian surname group.

These results then suggest that ethnolinguistic networks do affect policy adoption. Differences in language proficiency are only one of the possible causes, what is important is that an exogenous mechanism is in place making policy adoption more costly, either in terms of access to information or in terms of registration itself. This does not mean that German-speaking CEOs do not understand Italian at all, or that they are completely unaware of the policy. More simply, this indicates that the lack of documentation in German makes accessing the policy more costly for firms operating in German-speaking networks. Young innovative startups might be particularly exposed to these network-related costs due to their relative financial fragility compared to larger and older establishments.

We then move to the second stage estimates in Table 2. We show results for the policy effect for both OLS and IV specifications, so that we can benchmark our estimates against the existing literature on the Startup Act. For each of our outcomes, we provide two specifications. In the first set of specifications, the dependent variable is a dummy that denotes whether the firm has accessed a certain market - i.e. whether the firm has any investment in R&D, any bank debt, or any employee. This first set allows us to study whether the policy activated the participation of treated firms to these markets (innovation, credit and labour market). The second set of variables provides the continuous indicators. With this type of specification we can study if the policy access made firms more likely to have higher investment in innovation, access to credit and employment than comparable untreated firms.

Estimates are presented for each of the three models presented in section 4, and include all controls for firm and CEO characteristics, lagged financials and municipality trends, when applicable.[29] The coefficients of control variables are omitted for brevity.[30]

In general, controlling for endogenous registration reveals how some of the correlations estimated in the OLS regressions might be spurious. For the IV results, the coefficients of interest will estimate a local treatment effect in the sense that they apply to the subset of observations which were induced to take up the treatment, and not to all firms in the sample.

Columns (1) to (4) present the results on the effect of registration on research and development expenditure. We find a strong and statistically significant effect of the policy over both access into R&D and R&D expenditure in the OLS regressions: in the three-way fixed effects model the probability of doing any investment in R&D increases by 3.8% with the three-way fixed effects model, and to 4.3% (municipality, province/year fixed effects) and 6.4% (firm fixed effects) for the difference in differences models. On average, startups are expected to spend 54.8% more than

[29] We also tested different specifications for model 4.3, using province/year fixed effects only, finding stronger first stages (F = 32.41) but identical results. Similarly, the estimates from way firm fixed effects model discussed at the end of subsection 4.3 feature the same statistical significance (or lack thereof) for all variables. These results are available on request.

[30] Full specifications are available upon request.

comparable firms in R&D with the three-way fixed effects specification, a figure that rises to 63.4% for the municipality and province/year fixed effects model, and to 113.8% with the firm fixed effects model.

These estimates, however, do not hold when we introduce the etymological origin instrument. Under all specifications, the coefficients are not statistically different from zero. This result suggests that self-selection leads to biased results in terms of R&D. In other words, the positive effect of the policy on investment might result from firms increasing their R&D expenses in order to satisfy the access criteria of the policy. The difference in differences procedure seems to be justified for the firm fixed effects model only, the only one for which surname assignment pre-trends appear to have a statistically significance influence in outcomes. This suggests that provincial policy shocks might have affected firms' propensity to innovate, further motivating all three of our estimation models.

We show the results for credit access in Columns (5) to (8), where the dependent variable is bank debt. We find little statistical evidence that start-up registration facilitates credit. The OLS regressions do not display statistically significant coefficients for credit access in most year. The IV estimates likely suggest non-significant effect too. This finding is in contrast with previous works on the Start-up Act. [31]

Our evidence suggests that credit institutions are not influenced by the start-up status when granting credit application. Start-ups might already have had access to credit or, in turn, have problems accessing credit before registering. A straightforward interpretation of this result is that the public loan guarantee provided by the policy is not sufficient for granting credit requests, as lenders usually request the value of the collateral to be as high as the loan to be secured. Another possible reason is that banks do not consider the start-up status as a determinant factor when screening the financing conditions of a firm.

In Columns (9) to (12) we estimate the impact of the policy on job creation. For the two diff-in-diff models, OLS results indicate that policy access has increased the probability of employing workers by 4.4% and 20% approximately. The estimates show that firms with start-up status have employed between -6.85% (three-way fixed effects) and 29.17% more workers than other comparable firms. Other coefficients are not statistically significant. These results are generally in line with the literature on the Start-up Act, but underline how control strategies alone can suffer for inconsistency depending on the preferred model. The IV estimates are much more unambiguous, but reveal again that OLS results were driven by self-selection since we find no evidence for a statistically significant effect of startup status in terms of job creation under any instances.

These results strongly contrast with the positive employment effects in the literature on the Start-up Act and on start-up policies in general. This finding opens interesting questions about how human capital is remunerated across young and small firms. It is likely that employing work from no persons other than the owners of the firm is a common pattern for firms of this nature, and this might explain the non-significant effect across our IV specifications. Another possibility is that

[31] Previous studies found the Start-up Act to be effective in easing access to bank debt (Manaresi et al., 2021; Biancalani et al., 2021; Ferrucci et al., 2020).

the policy benefits are pushing sole proprietors into filing as limited liability companies, which is a requirement of the policy. This would likely not occur if the firm were not eligible for innovative start-up status.

Our three strategies allow us to rule out that differences between the two provinces might have affected our results. Our approach also suggest that our results are valid net of idiosyncratic municipality characteristics, including the proportion of each linguistic group, and general differences in economic performance and industrial structure. Time variance for any of these factors is also accounted for by the three-way fixed effects model. When applicable, language-specific difference in outcomes within municipalities are also found to be statistically not significant, suggesting that our strategy can account for most of these factors. This is not shown in the table only for brevity. As trends between linguistic groups are also kept fixed in the differences in difference model, showing also that pre-trends are in most cases superfluous, we find difficult to believe that linguistic assignment could have played any role in firms' outcomes, other than through policy adoption.

7 Conclusions

This paper has two main findings. The first is that ethnolinguistic diversity has the potential to jeopardise the implementation of a policy, penalising minority groups merely on the basis of their language. We have shown that the linguistic divide, proxied by the surnames of firms' CEOs, is crucial to determine the voluntary participation to a policy even among firms based in the same country. In particular, young firms with a CEO belonging to a German-speaking group are significantly less likely to access the package of benefits they are eligible for. We provide solid evidence that this likelihood is driven by the exogenous allocation to a linguistic group alone.

The fact that these asymmetries persist within the same municipalities, and even after accounting for policy shocks, suggests that residual differences in policy adoption can only be attributed to surname assignment and, by extension, linguistic networks. The persistence of asymmetries even after controlling for firm fixed effects further suggests that linguistic assignment influences the likelihood of registering as startup throughout the life of the firm. Our focus on specific knowledge-intensive industries, along with our difference in differences strategy, further allow us to rule out that differences in policy adoption could also be explained by differences in economic performance between the two groups. In fact, we find no evidence of differences in economic performance between the two linguistic groups during the pre-policy years.

The only viable explanation, then, is that these differences are cultural in origin. Net of municipality, province and industrial characteristics, the only difference between the two communities remains the main language they speak. Due to the lack of official documentation about the policy in German, we hypothesise that the outcome we observe is the result of language-biased cost asymmetries. Such asymmetries affect the information flow within German-speaking entrepreneurship networks and make registration more difficult, as German-speaking entrepreneurs have less access to information about the policy, unless they seek help from consultants embedded in Italian-speaking

networks. These costs could be not indifferent for small and notoriously fragile firms such as start-ups. As the policy offers immediate benefits to firms which cannot possibly generate heterogeneous effects across language groups, it is difficult to believe that German-assigned firms would not apply only because of distrust in the policy or because they fear they would not benefit as much as other firms. We simply believe that, to German-speaking firms, the presence of these language-biased cost asymmetries makes the policy adoption process simply not worth the effort.

Whether these legal benefits are actually beneficial relates to our second finding. Our results indicate that the Italian Start-up Act has had no major impact over treated firms in terms of investment in R&D, financing and job creation, contradicting the evidence from other related studies in the field. Some might argue that these results could detract from our findings on policy adoption and ethno-linguistic diversity: would language barriers have been overcome if the policy was more effective? We do not believe this to be that case as what matters is that the policy has been, until this moment, widely signalled as strongly effective.

We do not find any significant correlation between language and economic performance, which suggests that our results are not biased by exclusion restriction violations. Before the policy was introduced, (eligible and ineligible) firms with German-assigned CEOs performed just as good as firms with Italian-assigned CEOs.

Based on our findings, we do not rule out the possibility that the policy might have had beneficial effects on a subset of participants only, nor we imply that the policy failed to select innovative firms in general. Our results suggest instead that the Italian Start-up Act did not fulfil its stated ambitions: unleashing systemic effects for a wide population of innovative enterprises.

As a final word of caution, it should be underlined that the external validity of our results might be limited to contexts similar to the one of this paper. Trentino-Alto Adige is an ideal case study because of its economic and demographic homogeneity, which makes language the most meaningful divide between the two groups. Importantly, the fact that Italian speakers in South Tyrol have facilitated access to information and professional networks from the broader Italian-speaking domestic entrepreneurship community makes this setting preferable to other instances of bilingualism, such as Switzerland or Belgium. These are multilingual countries, while German is a minority language in Italy, providing Italian-speakers in South Tyrol with an inherent linguistic advantage.

Future works could exploit the surname-based instrument that we propose in this paper, or the presence of linguistic minorities in general, to study similar contexts where language poses a barrier to linguistic minorities in accessing a policy. More precisely, this research design is applicable to those cases where (i) the groups are close enough in terms of economic performance and entrepreneurship opportunities, at least after controlling for observable differences, (ii) language differences are sufficiently strong to justify the emergence of these asymmetries, and finally, if used for assignment, (iii) surnames correlate with language, and the stochastic component in the distribution of surnames is not entirely determined by geography. While more work is certainly needed to capture the etymological origins of surnames, future research could exploit recent advancements (such as

Desmet et al., 2012 and Galor et al., 2020) to construct new surname-based measures of linguistic diversity.

Further studies could also focus on disentangling the mechanism behind the language biased-asymmetries we have observed. Importantly, we do not suggest that policymakers should differentiate the provision of their policy across linguistic groups: this would heighten language-based inequalities in the long-run (Desmet et al., 2012). On the contrary, the results of this paper encourage the creation of policy documentation accessible to minority languages and the support of multilingual professional networks. Understanding whether these asymmetries are caused by the speed and quality of information flows within non-majoritarian groups, or simply by the opportunity costs of seeking professional expertise beyond one's own language network, can prove valuable in this direction.

References

Abel, A., Vettori, C., and Forer, D. (2012). Learning the neighbour's language: The many challenges in achieving a real multilingual society—the case of second language acquisition in the minority–majority context of south tyrol. *European Yearbook of Minority Issues Online*, 9(1):271–303.

Acs, Z. J. and Audretsch, D. B. (1987). Innovation, market structure, and firm size. *The Review of Economics and Statistics*, 69(4):567.

Acs, Z. J. and Audretsch, D. B. (1988). Innovation in large and small firms: An empirical analysis. *The American Economic Review*, 78(4):678–690.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2):155–194.

Alesina, A. and Giuliano, P. (2015). Culture and institutions. *Journal of Economic Literature*, 53(4):898–944.

Alesina, A. and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *The Journal of Human Resources*, 40(4):791–821.

Andrews, D. and Criscuolo, C. (2013). Knowledge-based capital, innovation and resource allocation. OECD Economics Department Working Papers No. 1046, Organisation for Economic Co-Operation and Development (OECD).

Andrews, D., Criscuolo, C., and Menon, C. (2014). Do resources flow to patenting firms? cross-country evidence from firm level data. OECD Economics Department Working Papers No. 1127, Organisation for Economic Co-Operation and Development (OECD).

Angerer, S., Glätzle-Rützler, D., Lergetporer, P., and Sutter, M. (2016). Cooperation and discrimination within and across language borders: Evidence from children in a bilingual city. *European Economic Review*, 90:254–264. Social identity and discrimination.

Astat (2010). Nachnamen in südtirol - cognomi in provincia di bolzano. Astat schriftenreihe 172, Landesinstitut für Statistik - Istituto provinciale di statistica - ASTAT.

Astat (2011). Censimento della popolazione 2011: Determinazione della consistenza dei tre gruppi linguistici della Provincia Autonoma di Bolzano-Alto Adige. Astat Informazioni 38, Landesinstitut für Statistik - Istituto provinciale di statistica - ASTAT.

Autor, D. H. (2013). The "task approach" to labor markets: an overview. *Journal for Labour Market Research*, 46(3):185–199.

Banca d'Italia (2021). L'economia delle province autonome di Trento e di Bolzano. Economie Regionali - Rapporto annuale N. 4, Banca d'Italia.

Baumol, W. J. (2004). Entrepreneurial enterprises, large established firms and other components of the free-market growth machine. *Small Business Economics*, 23(1):9–21.

Bazzi, S., Gaduh, A., Rothenberg, A. D., and Wong, M. (2019). Unity in diversity? how intergroup contact can foster nation building. *American Economic Review*, 109(11):3978–4025.

Beach, B. and Jones, D. B. (2017). Gridlock: Ethnic diversity in government and the provision of public goods. *American Economic Journal: Economic Policy*, 9(1):112–36.

Biancalani, F., Czarnitzki, D., and Riccaboni, M. (2021). The italian start up act: a microeconometric program evaluation. *Small Business Economics*.

Bound, J. and Jaeger, D. A. (2000). *Do compulsory school attendance laws alone explain the association between quarter of birth and earnings?*, volume 19, pages 83–108. Emerald Group Publishing Limited.

Breschi, S., Lassébie, J., and Menon, C. (2018). A portrait of innovative start- ups across countries. OECD Science, Technology and Industry Working Papers No. 2018/02, Organisation for Economic Co-Operation and Development (OECD).

Buiting, S. and Pokriefke, E. (2016). 12 familienstrukturen und heiratsmuster. In Atz, H., Haller, M., and Pallaver, G., editors, *Ethnische Di erenzierung und soziale Schichtung in der Südtiroler Gesellschaft*, pages 245–264. Nomos Verlagsgesellschaft mbH & Co. KG.

Buldyrev, S. V., Growiec, J., Pammolli, F., Riccaboni, M., and Stanley, H. E. (2007). The growth of business firms: Facts and theory. *Journal of the European Economic Association*, 5(2-3):574–584.

Buldyrev, S. V., Growiec, J., Pammolli, F., Riccaboni, M., and Stanley, H. E. (2020). *The Rise and Fall of Business Firms: A Stochastic Framework on Innovation, Creative Destruction and Growth*. Cambridge University Press.

Calvino, F., Criscuolo, C., and Menon, C. (2015). Cross-country evidence on start-up dynamics. OECD Science, Technology and Industry Working Papers No. 2015/06, Organisation for Economic Co-Operation and Development (OECD).

Cantarella, M., Fraccaroli, N., and Volpe, R. (2023). Does fake news affect voting behaviour? *Research Policy*, 52(1):104628.

Cavallo, A., Ghezzi, A., Colombelli, A., and Casali, G. L. (2018). Agglomeration dynamics of innovative start-ups in Italy beyond the industrial district era. *International Entrepreneurship and Management Journal*, 16(1):239–262.

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, 103(2):690–731.

Coia, V., Boschi, I., Trombetta, F., Cavulli, F., Montinaro, F., Destro-Bisol, G., Grimaldi, S., and Pedrotti, A. (2012). Evidence of high genetic variation among linguistically diverse populations on a micro-geographic scale: a case study of the italian alps. *Journal of human genetics*, 57(4):254–260.

Colombelli, A. (2016). The impact of local knowledge bases on the creation of innovative start-ups in italy. *Small Business Economics*, 47(2):383–396.

Criscuolo, C., Gal, P. N., and Menon, C. (2014). The dynamics of employment growth: New evidence from 18 countries. OECD Science, Technology and Industry Policy Papers No. 14, Organisation for Economic Co-Operation and Development (OECD).

Dahlberg, M., Edmark, K., and Lundqvist, H. (2012). Ethnic diversity and preferences for redistribution. *Journal of Political Economy*, 120(1):41–76.

Damaraju, N. L. and Makhija, A. K. (2018). The role of social proximity in professional CEO appointments: Evidence from caste/religion-based hiring of CEOs in india. *Strategic Management Journal*, 39(7):2051–2074.

de Chaisemartin, C. and D'Haultfœuille, X. (2017). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, 85(2):999–1028.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2016). Where has all the skewness gone? The decline in high-growth (young) firms in the U.S. *European Economic Review*, 86:4–23.

Desmet, K., Ortuño-Ortín, I., and Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of Development Economics*, 97(2):322–338.

Desmet, K., Ortuño-Ortín, I., and Wacziarg, R. (2016). Linguistic cleavages and economic development. In Ginsburgh, V. and Weber, S., editors, *The Palgrave handbook of economics and language*, pages 425–446. Springer.

Desmet, K. and Wacziarg, R. (2020). The cultural divide. *The Economic Journal*, 131:2058–2088.

Dosch, J. and Lakatos, M. (2020). South tyrol and Åland: Collective identity in the interplay of old and new minorities. *Studies in Ethnicity and Nationalism*, 20(2):188–207.

Dräger, K. and Schmuck, M. (2009). The German Surname Atlas Project—Computer-based surname geography. In Ahrens, W., Embleton, S., and Lapierre, A., editors, *Names in multi-lingual, multi-cultural and multi-ethnic contact. Proceedings of the 23rd International Congress of Onomastic Sciences*, volume 1722, page 2008.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.

Ferrucci, E., Guida, R., and Meliciani, V. (2020). Financial constraints and the growth and survival of innovative start-ups: An analysis of italian firms. *European Financial Management*, 27(2):364–386.

Finaldi Russo, P., Magri, S., and Rampazzi, C. (2016). Innovative start-ups in italy: their special features and the effects of the 2012 law. Questioni di Economia e Finanza (Occasional Papers) 330, Banca d'Italia.

Fiorentino, R., Longobardi, S., and Scaletti, A. (2021). The early growth of start-ups: innovation matters. evidence from italy. *European Journal of Innovation Management*, ahead-of-print(ahead-of-print).

Galor, O., Özak, Ö., and Sarid, A. (2020). Linguistic traits and human capital formation. *AEA Papers and Proceedings*, 110:309–13.

Giraudo, E., Giudici, G., and Grilli, L. (2019). Entrepreneurship policy and the financing of young innovative companies: Evidence from the italian startup act. *Research Policy*, 48(9):103801.

Guiso, L., Sapienza, P., and Zingales, L. (2008). Social capital as good culture. *Journal of the European economic Association*, 6(2-3):295–320.

Guiso, L., Sapienza, P., and Zingales, L. (2015). Corporate culture, societal culture, and institutions. *American Economic Review*, 105(5):336–39.

Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2013). Who creates jobs? small versus large versus young. *Review of Economics and Statistics*, 95(2):347–361.

Ham, J. C. and Lalonde, R. J. (1996). The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. *Econometrica*, 64(1):175–205.

Istat/Mise (2018). *Startup Survey 2016: The first survey on innovative startups in Italy*. Istituto nazionale di statistica.

Klenow, P. J. and Li, H. (2020). Innovative growth accounting. NBER Working Paper 27015, National Bureau of Economic Research (NBER).

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. (1999). The quality of government. *The Journal of Law, Economics, and Organization*, 15(1):222–279.

Manaresi, F., Menon, C., and Santoleri, P. (2021). Supporting innovative entrepreneurship: an evaluation of the italian "start-up act". *Industrial and Corporate Change*.

Menon, C., DeStefano, T., Manaresi, F., Soggia, G., and Santoleri, P. (2018). The evaluation of the italian "start-up act". OECD Science, Technology and Industry Policy Papers 54, Organisation for Economic Co-Operation and Development (OECD).

MISE (2019). The Italian Startup Act. Italy's policy framework to support innovative startups: Executive summary. policy summary 07/2019, Italian Ministry of Economic Development.

MISE (2021). Relazione annuale al parlamento sullo stato di attuazione e l'impatto delle policy a sostegno di startup e pmi innovative. report, Italian Ministry of Economic Development.

Montalvo, J. G. and Reynal-Querol, M. (2021). Ethnic Diversity and Growth: Revisiting the Evidence. *The Review of Economics and Statistics*, 103(3):521–532.

Murray, F., Aghion, P., Dewatripont, M., Kolev, J., and Stern, S. (2016). Of mice and academics: Examining the effect of openness on innovation. *American Economic Journal: Economic Policy*, 8(1):212–52.

OECD (2020a). An insight into the innovative start-up landscape of south tyrol. OECD Local Economic and Employment Development (LEED) Papers No. 2020/07, Organisation for Economic Co-Operation and Development (OECD).

OECD (2020b). An insight into the innovative start-up landscape of trentino. OECD Local Economic and Employment Development (LEED) Papers No. 2020/09, Organisation for Economic Co-Operation and Development (OECD).

OECD (2020c). Start-ups in the time of covid-19: Facing the challenges, seizing the opportunities. OECD Policy Responses to Coronavirus (Covid-19) 13 May 2020, Organisation for Economic Co-Operation and Development (OECD).

Protogerou, A., Caloghirou, Y., and Vonortas, N. S. (2017). Determinants of young firms' innovative performance: Empirical evidence from europe. *Research Policy*, 46(7):1312–1326.

Schot, J. and Steinmueller, W. E. (2018). Three frames for innovation policy: R&d, systems of innovation and transformative change. *Research Policy*, 47(9):1554–1567.

Spolaore, E. and Wacziarg, R. (2016). Ancestry, language and culture. In Ginsburgh, V. and Weber, S., editors, *The Palgrave Handbook of Economics and Language*, pages 174–211. Palgrave Macmillan.

Sutter, M., Angerer, S., Glätzle-Rützler, D., and Lergetporer, P. (2018). Language group differences in time preferences: Evidence from primary school children in a bilingual city. *European Economic Review*, 106:21–34.

Vettori, C. (2016). The long way to bilingualism: the peculiar case of multilingual South Tyrol. *International Journal for 21st Century Education*, 3(2):25.

Vettori, C. and Abel, A., editors (2017). *KOLIPSI II. Gli studenti altoatesini e la seconda lingua: indagine linguistica e psicosociale. / Die Südtiroler SchülerInnen und die Zweitsprache: eine linguistische und sozialpsychologische Untersuchung.* Eurac Research, Bolzano/Bozen.

Vettori, C., Abel, A., and Wisniewski, K., editors (2012). *KOLIPSI. Gli studenti altoatesini e la seconda lingua: indagine linguistica e psicosociale. / Die Südtiroler SchülerInnen und die Zweitsprache: eine linguistische und sozialpsychologische Untersuchung.* Eurac Research, Bolzano/Bozen.

Vettori, C., Colobo, S., and Abel, A. (2021). Multilingualism in south tyrol: between old fears and new challenges. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 1(26):223–249.

Wang, C. and Steiner, B. (2015). Can Ethno-Linguistic Diversity Explain Cross-Country Differences in Social Capital?: A Global Perspective. *Economic Record*, 91(294):338–366.

Wang, C. and Steiner, B. (2019). Can social capital explain business performance in Denmark? *Empirical Economics*, 59(4):1699–1722.

Westphal, J. D. and Zajac, E. J. (1995). Who Shall Govern? CEO/Board Power, Demographic Similarity, and New Director Selection. *Administrative Science Quarterly*, 40(1):60–83.

Wooldridge, J. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Mimeo*.

Does language prevent policy implementation? Evidence from the Italian Start-up Act - Appendix

A Distribution of start-ups

Table 3: Distribution of the 10 most common areas of economic activity among startups

Activity description	Ateco Code	freq.	percent.
Italy			
Software production	620100	3,808	0.230
R&D in engineering and natural sciences	721909	1,674	0.101
Web portals	631200	1,078	0.065
Other IT services	620909	909	0.055
Consultancy in IT	620200	796	0.048
R&D in biotechnology	721100	457	0.028
Other consultancy in business & administration	702209	422	0.025
Retail trade via internet	479110	248	0.015
Other technical consultancy	749093	243	0.015
Integrated engineering design services	711220	204	0.012
Other sectors of economic activity	–	6,753	0.407
Trentino Alto-Adige/Südtirol			
Software production	620100	105	0.246
R&D in engineering and natural sciences	721909	68	0.160
Web portals	631200	18	0.042
Non-ferrous metals & semi-finished products manufacturing	244500	11	0.026
Other IT services	620909	11	0.026
Consultancy in IT	620200	10	0.023
Manufacturing of other measuring and regulating equipment	265129	9	0.021
R&D in biotechnology	721100	8	0.019
Development of software packages	582900	7	0.016
Other consultancy in business & administration	702209	6	0.014
Other sectors of economic activity	–	173	0.406

*N*otes: Authors' calculations based on data from Camere di Commercio. Figures attain to the universe of startups registered between 2013 and 2010. Percentages are shown in reference to the relative startup population, with firms active in multiple years being counted once.

B CEO language identification

To find the linguistic group of the firm, we exploit information on the CEO name, included in the Aida dataset, to determine whether the name of the CEO is of German origin. We do so by using a text analysis algorithm.

Our intuition is that in most cases surnames of Italian and German origin are distinguishable through their morphology with little ambiguity. We focus specifically on word endings, where the

different grammar rules between West Germanic and Romance languages are most evident.[32]

Many German family names derive either from the occupation or the place of origin of a forebearer (Dräger and Schmuck, 2009). In both cases, these surnames typically end in "-er", following grammar rules common in West Germanic languages (including English), where this suffix indicates an agent noun or an adjective related to origin. This also applies to closely related spelling variants (e.g., "Mair" or "Mayr" for "Meyer"). This kind of family name is especially common in South Tyrol, where all the 30 most common surnames province-wide end in "-er" (most cases), "-ir", or "-yr" (Astat, 2010).

By contrast, "-er" endings are uncommon in Italian family names. As most Italian words, surnames tend to end with a vowel, and exceptions are rare and mostly owing to the influence of regional languages such as Venetian or Friulian (e.g. "Zanier"), or are direct borrowings from Latin (e.g. "Paternoster").

Even beyond occupational and place-based surnames, German family names often diverge from those of Italian origins in ways that are obvious to proficient speakers: some German suffixes such as "-sch", "tsch", "kl", "-gg", "-gl", "-hl", "th" or "-aas" are never found in Italian. Umlauts, such as in "Thöni", and other letters that are not proper of standard Italian (such as "K" and "W") are useful giveaways as well.

Certain ambiguous cases still exist: the German diminutive suffix "-le", found e.g. in "Hammerle", also appears in Italian surnames. In that case, however, we can refer to the relative distribution of that surname in South Tyrol compared to the rest of Italy, where there are no other sizeable German-speaking cultural areas. A surname in "-erle" which is mostly found in South Tyrol and surroundings will be identified as German. This logic also works the opposite way: an "-er" surname which is mostly found in Italian regions outside South Tyrol will most likely be of non-German origin.

We develop a system of classification based on word endings distinctive in German, which are then manually compiled into a dictionary. As German surname structure is more "predictable" than Italian, since most native family names can be traced back to the categories touched upon here (Dräger and Schmuck, 2009), we expect to be able to capture the vast majority of the residents' family names. We compiled our dictionary after observing all surnames born by at least 10 residents of the province of Bolzano on 31 December 2010 (N: 6414), as resulting from local population registers and compiled by Astat (2010). The dictionary is made up of a list of 304 "German" word endings, or complete surnames in a few cases (e.g. some monosyllabical names like "Fink").[33]

In the first step, the dictionary is then applied to the family names of all directors [34] of the firms in our sample, in South Tyrol as well as in Trentino, whose nationality is recorded by Aida as "Italian" (N: 55,947). Each individual is classified in either "surname language" group ("German-speaking" and "non-German-speaking") via a binary variable.

[32] Conversely, this is not true for first names, where there is much more overlap between German and Italian, and particularly so for women.

[33] In the case of double-barrelled surnames, only the last element is considered.

[34] Legal persons only. We use all directors and not the CEO only in order to have a larger, more representative population for surname classification.

We then review the most common surnames in the sample, seeking both false negatives (German surnames mislabelled as non-German) and false positives (non-German surnames classified as German). To capture more German surnames, we added more suffixes to the dictionary. To limit the number of false positives, we reviewed the most common surnames in the sample for individuals born in Italian-majority municipalities, and tested the reliability of this classification via external sources (primarily the `forebears.io` website). If a family name classified as "German" is significantly more common in areas where German is not natively spoken, the name is classified as a "false positive" and the binary variable value manually corrected to 0 (i.e., "non-German"). Given the small size of South Tyrol compared to the the rest of Italy, these surnames are normally most common elsewhere by a factor of 10 to 1 or more. Thanks to this method we could forego the introduction of custom exceptions and of other frequency-based systems.[35]

We end up with 2,808 valid observations in the German-assigned group. Overall, 32.7% of highest executives are classified in the German surname group. This includes 28.7% of Italian citizens in the sample, and 67.5% of all individuals born in South Tyrol, which is close to the share of German speakers in the province as reported in official statistics (69.4%).

It is important to stress that our classification is not intended to detect which language the CEO speaks, but rather which language group the surname of a CEO belongs to. In fact, the surname origin will not necessarily indicate which language is spoken by a person, as it is not really possible to draw a line between "German" and "Italian" cultural milieu in a mixed language context, when there has been cultural exchange as well as intermarriage over many centuries, and where fully bilingual families exist – although not as many as it may be assumed (Vettori et al., 2012; Vettori and Abel, 2017).

With multicultural families in mind, we will also assume that, for married women, the family name in Aida is the maiden name. In Italy women do not normally change their surname when getting married, which leads us to believe that the assumption will be correct in the overwhelming majority of cases.

A significant proportion (14.1%) of CEOs in our sample are expatriates from other countries. We work under a conservative assumption that immigration of CEOs is endogenous to firm performance, and flag all individuals whose country of nationality is not Italy as non-nationals. Non-nationals from German-speaking countries[36] are however still grouped in the German etymological group, while not nationals from other countries will be always flagged as "non-German". In a few cases, nationality data may not available, and we then impute the nationality from the place of birth.[37]

Similarly, individuals born in Italy that may well be German speakers but that have a family name of foreign origin will not be captured. South Tyrol has experienced some immigration in the last few decades (Dosch and Lakatos, 2020), and new residents have the option to register in either

[35] Twenty-three surnames were removed with this method. The most common are "Piffer", "Merler" and "Coser", all very common in Trentino and much less so in South Tyrol.

[36] In our classification, Austria, Germany, Luxembourg, Switzerland, and the German-speaking community of Belgium.

[37] It should be noted that in Aida individuals with missing nationality data are always non-Italian nationals. If neither the nationality nor place of birth are available, the observation is dropped from the sample.

the Italian or German community according to their preference. In this work, however, they will be all be considered "non-German". While being a simplification, there is empirical evidence (Vettori and Abel, 2017) that children with an immigration background are much more likely to attend Italian schools and identify as Italian-speaking.

Finally, one minor limitation of our approach is that the dictionary will not classify Ladin surnames appropriately: some will be flagged as German, others will not. Ladin is a Romance language with substantial Germanic influence, meaning that surnames typical of the Ladinia are a patchwork of Latinate ("Clara", "Irsara") and West German ("Ploner", "Demetz") features. There are further differences across Ladin valleys, with some having more German-sounding family names than others. While it is improper to associate Ladins to either the Italian or the German language group, we will treat their surnames as a random source of noise. In any case, start-ups are very rarely found in Ladin-majority municipalities (where most Ladins live), and this noise is easily absorbed by municipality fixed effects in a our robustness specifications.

As discussed earlier, these caveats are entirely expected as we are not interested in which language the CEO actually speaks but only in detecting the linguistic origin of the surname as a marker for the intention-to-treat. As long as some connection between surname origin and native language remains, the former will stand as a superior instrument choice as the language spoken is not exogenous, while the surname origin is.

We expect our approach to be valid – i.e., surnames are a valid proxy of linguistic diversity and, by extension, cultural origin – if this connection still exists. As discussed earlier, the proportion of board members classified in the German language group strongly correlates with the proportion of German speakers in each municipality, as resulting from census data. This is shown in Figure 2.Ignoring the few Ladin majority municipalities, regressing the share of the estimated German speaking CEOs over the share of German speaking citizens (according to census data) yields a 0.92 correlation coefficient with a 0.93 r-squared.

C Ethymology of surnames, an example

Table 4: Distribution of the 10 most common surnames in the municipalities of Trento and Bolzano

Municipality: Trento				Municipality: Bolzano			
Surname	freq.	percent.	origin	Surname	freq.	percent.	origin
TOMASI	893	0.76	IT	PICHLER	299	0.29	DE
DEGASPERI	584	0.49	IT	FERRARI	263	0.25	IT
FERRARI	550	0.47	IT	ROSSI	259	0.25	IT
PEDROTTI	543	0.46	IT	GASSER	233	0.22	DE
BORTOLOTTI	519	0.44	IT	MAIR	224	0.22	DE
NARDELLI	491	0.42	IT	KOFLER	178	0.17	DE
FRANCESCHINI	456	0.39	IT	GRUBER	170	0.16	DE
AGOSTINI	373	0.32	IT	EGGER	160	0.15	DE
GIOVANNINI	371	0.31	IT	HOFER	156	0.15	DE
TAMANINI	360	0.31	IT	LARCHER	156	0.15	DE

Sources: Astat (2011) and Servizio sviluppo economico, studi e statistica, Provincia di Trento (2017)

D Adjusting for access criteria

It is important to ensure that either only firms who have been in the position to access the policy are included in the estimation sample, or that all factors that can influence access to the policy independently of language groups are controlled for. Using the full sample of 22,944 unique firms we run the risk of weakening our first stages by including firms which, no matter the language of the CEO, would have never been eligible to access the policy or are operating in industries which have experienced little innovation. This means that we are only using the information we really need for our estimation purposes.

To overcome this issue, we adjust the sample and the methodology to ensure that only firms which satisfy the criteria set by regulators are included. First, in order to apply for start-up status, firms should not be older than five years. We then remove all firms aged more than four. Furthermore, a firm cannot apply if its yearly revenues are larger than EUR 5 million. We then keep in the sample all firms which, in the first four years of life, have not exceeded EUR 5 million in revenues at least once.[38] Also, as firms were only offered the opportunity to register as start-ups in late 2012, we drop from the sample all observations related to years before 2013.

Innovative start-ups should also be unlisted and legally registered as a limited liability company (where specific limitations on profit distributions are in place). As a result, we remove the few (five in total) listed firms based in Trentino-Alto Adige, and remove firms whose legal form was not a limited liability company. Firms benefiting from the Start-up Act should also not be distributing profits, and therefore we remove all firms with non-zero dividends.

Innovative start-ups should not be formed as the result of a company merger, split-up or selling-off. Legislators introduced this criterion to prevent existing firms from exploiting the benefit scheme

[38]The highest revenue for a startup in the region was around EUR 3 millions. This strategy is only adopted to ensure that firms that have already been born as "large" firms are filtered out of the sample.

and creating new firms by splitting up or selling existing companies. While the available data does not allow us to track mergers, split-ups or sell-offs, the inclusion of firms borne from corporate restructuring does not pose a serious issue for our estimates, as it makes little sense for firms not registered as innovative start-ups to commit to mergers, split-ups or sell-offs to benefit from a policy they have no interest in applying for.

The final compulsory criterion set by the Start-up Act is that firms must have the development, production and marketing of a product or service with a high technological value as their exclusive or predominant corporate purpose. This is a subjective criterion, as firms are free to provide any description for their corporate purpose, and it is up to the relevant government authorities to assess its technological value. To overcome this complication, we offer an alternative approach based on the distributional structure of the economic activities of Italian innovative start-ups.

While free to pursue any kind of economic activity, in fact Italian start-ups tend to operate within specific economic sectors. Out of all firms which ever achieved the start-up status, the three most common sectors are *Software production* (representing 22.9% of all start-ups in Italy, and 20.4% in Trentino-Alto Adige), *Experimental research and development in the fields of engineering and natural sciences* (10.1% and 16.9%), and *Web portals* (6.4%, and 4.1%).[39] In Table 3 we provide a list of the ten most common sectors of economic activity, which together cover more than half of the total number of start-ups.

We obtain the frequency distribution of all sectors of economic activity within the universe of start-ups from Trentino-Alto Adige using their ATECO codes. We create a dummy variable that captures whether the industry a firm is operating in has seen any start-up activity since the introduction of the policy. In this way, firms operating in industries that have seen no start-up activity at the macro level are removed from the analysis. Under the reasonable assumption that the start-up distribution by economic activity at the regional level is representative of a natural level of concentration of start-ups within sectors of economic activity, this approach ensures that firms for which no first stage mechanism is in place are removed from the analysis. This approach provides a proxy for corporate purpose. It should be noted that this approach is adopted to filter out firms operating in industries where innovative activities are not possible, including the industries where large regional differences have been identified (Banca d'Italia, 2021). After removing observations belonging to these groups, we do not include ATECO fixed effects in the final specifications, allowing firms to freely choose any area of economic activity within industries where start-up activity can emerge. To better allow for capturing economic and industrial shocks, this filtering on the ATECO codes is not employed for the three-way f.e. model described in subsection 4.1.

We are then left with the three *à la carte* innovation criteria. As firms only needed to satisfy one of these criteria, we add proxies for them as controls when the model allows. In particular, we control for R&D expenses, average salaries and patent costs at year $t-1$ (as outlined in Subsections 4.1 and 4.2).

[39]The ATECO codes of these sectors are: 620100 for Software production, 721909 for Experimental research and development in the fields of engineering and natural sciences, and 631200 for Web portals.

E Robustness checks

Table 5: Assigned language and ceo characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	German-assigned CEO							
Gender of CEO	0.044	0.012	0.042	0.012	0.022	0.020	0.018	0.020
	(0.036)	(0.032)	(0.036)	(0.032)	(0.014)	(0.015)	(0.014)	(0.015)
Age of CEO	0.000	0.001	0.000	0.001	0.000	-0.000	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Foreign CEO	0.986***	0.944***	0.985***	0.944***	0.528***	0.541***	0.542***	0.542***
	(0.115)	(0.111)	(0.116)	(0.111)	(0.088)	(0.089)	(0.086)	(0.089)
Foreign CEO × Age of CEO	-0.010***	-0.012***	-0.010***	-0.012***	-0.008***	-0.009***	-0.009***	-0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Firm age: 1	-0.025*	-0.021*	-0.025*	-0.021*	-0.009*	-0.010*	-0.009*	-0.010*
	(0.013)	(0.011)	(0.013)	(0.011)	(0.005)	(0.006)	(0.005)	(0.006)
Firm age: 2	-0.034*	-0.024	-0.034*	-0.025	0.001	0.000	0.000	0.000
	(0.018)	(0.016)	(0.018)	(0.016)	(0.008)	(0.008)	(0.008)	(0.008)
Firm age: 3	-0.063***	-0.035	-0.063***	-0.035	-0.005	0.003	-0.003	0.003
	(0.024)	(0.022)	(0.024)	(0.022)	(0.012)	(0.012)	(0.012)	(0.012)
Firm age: 4	-0.025	0.016	-0.025	0.016	0.016	0.032**	0.023*	0.032**
	(0.027)	(0.023)	(0.027)	(0.023)	(0.013)	(0.013)	(0.013)	(0.013)
Firm density (ln)	-0.069***	0.077	-0.069***	0.076	-0.010***	-0.016	-0.001	-0.018
	(0.007)	(0.082)	(0.007)	(0.081)	(0.003)	(0.043)	(0.003)	(0.042)
Board size	-0.006	-0.007	-0.006	-0.007	-0.001	-0.002	-0.001	-0.002
	(0.004)	(0.006)	(0.004)	(0.006)	(0.001)	(0.002)	(0.001)	(0.002)
Board diversity			-0.113	-0.046	0.093	0.116	0.110	0.116
			(0.107)	(0.116)	(0.100)	(0.102)	(0.100)	(0.102)
% German-assigned in board					0.870***	0.808***	0.812***	0.808***
					(0.016)	(0.022)	(0.021)	(0.022)
% German-assigned in municipality							0.229***	0.132
							(0.032)	(0.092)
Observations	3,743	3,729	3,743	3,729	3,743	3,729	3,743	3,729
Adjusted R-squared	0.280	0.489	0.281	0.489	0.802	0.817	0.809	0.817
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality fixed effects	No	Yes	No	Yes	No	Yes	No	Yes

Notes: SE clustered by firm in parentheses.