



CONTRAFACTUAL-ML

AVALIAÇÃO CONTRAFACTUAL
DOS APOIOS ÀS EMPRESAS:
CRUZAMENTO ENTRE
ABORDAGENS DE INFERÊNCIA
CAUSAL E DE APRENDIZAGEM
AUTOMÁTICA

Lisboa, 10 de janeiro de 2023



Equipa:

Ricardo Paes Mamede (coordenador)

Maria da Conceição Figueiredo

Renato Pereira

Henrique Pereira

João Narciso

Marlon Santos

Matias Andrade

Tomás Martins

Índice

1. Introdução	6
2. Teorias da mudança das políticas sob análise.....	8
2.1. Teoria da mudança como base da avaliação de políticas	9
2.2. Teoria da mudança do SI Inovação Produtiva.....	11
3. Avaliação contrafactual de impactos.....	16
3.1. Experiências aleatórias	17
3.2. Variáveis instrumentais.....	20
3.3. Desenho de regressão descontínua.....	22
3.4. Diferença-em-diferenças	24
3.5. Equilíbrio de distribuições por emparelhamento (<i>matching</i>).....	25
3.6. Controlo Sintético	29
3.7. Possíveis contributos da aprendizagem automática	30
4. Causalidade, avaliação contrafactual e teorias da mudança.....	34
4.1. Diferentes abordagens à causalidade	34
4.2. Causalidade e avaliação contrafactual de impactos	37
4.3. Avaliação contrafactual de impactos e teorias da mudança	39
4.4. Conclusão	42
5. Dados utilizados e considerações sobre a sua análise	43
5.1. Fontes dos dados	44
5.2. Cruzamento dos dados	48
5.3. Construção de indicadores.....	49
5.4. Qualidade dos dados.....	49
5.5. Análise de dados	57
6. Análise contrafactual de impactos	64
6.1. Decisões críticas	64
6.2. Estimação de impactos	74
6.3. Análise de heterogeneidade	79
6.4. Análise de robustez.....	81
7. Referências	84
8. Anexos	88
8.1. Estatísticas Descritivas	88
8.2. Construção de indicadores.....	101

Índice de Tabelas

<i>Tabela 1: Percentagem de empresas no SCIE por primeiro e último ano com registo</i>	<i>45</i>
<i>Tabela 2: Percentagem de empresas no QP por primeiro e último ano com registo</i>	<i>46</i>
<i>Tabela 3: Dados em falta no QP-Trabalhadores.....</i>	<i>46</i>
<i>Tabela 4: Trabalhadores duplicados no QP-Trabalhadores.....</i>	<i>47</i>
<i>Tabela 5: Empresas com registo, cruzamento de bases de dados (SCIE e QP)</i>	<i>48</i>
<i>Tabela 6: Empresas com registos em falta em um ou mais anos entre o primeiro e o último registo</i>	<i>50</i>
<i>Tabela 7: Comparação entre amostras, segundo diferentes critérios de exclusão de observações.....</i>	<i>56</i>
<i>Tabela 8: Comparação de médias entre empresas não candidatas, candidatas não apoiadas e apoiadas.....</i>	<i>61</i>
<i>Tabela 9: Requisitos, vantagens e limitações de diferentes métodos de avaliação contrafactual.....</i>	<i>64</i>
<i>Tabela 10: Modelo da pontuação de propensão utilizando as variáveis consideradas importantes para explicar o tratamento</i>	<i>74</i>
<i>Tabela 11: Qualidade do emparelhamento (1:1) quando são utilizadas as variáveis importantes para explicar o tratamento e as variáveis importantes para explicar o resultado (quando o resultado é a média do log do IAF de t a t+2)</i>	<i>76</i>
<i>Tabela 12: Resultados obtidos segundo diferentes estratégias de estimação e indicadores.....</i>	<i>77</i>
<i>Tabela 13: Impacto estimado no investimento para subgrupos divididos segundo o endividamento.....</i>	<i>80</i>
<i>Tabela 14: Estimação do impacto do placebo para diferentes modelos de estimação.....</i>	<i>82</i>
<i>Tabela 15: Estimação de impacto na amostra sem outliers.....</i>	<i>83</i>
<i>Tabela 16: Resultados segundo a estratégia de emparelhamento</i>	<i>83</i>
<i>Tabela 17: Empresas candidatas aos SI do PT2020, por tipo de medida.....</i>	<i>88</i>
<i>Tabela 18: Empresas apoiadas em um ou mais projetos.....</i>	<i>89</i>
<i>Tabela 19: Número de projetos do PT2020 por início e fim.....</i>	<i>89</i>
<i>Tabela 20: Empresas candidatas aos SI do PT2020, por sector de atividade</i>	<i>90</i>
<i>Tabela 21: Empresas candidatas aos SI do PT2020, por classe de dimensão de empresas.....</i>	<i>91</i>
<i>Tabela 23: Contagem de empresas no SCIE e classificação por tamanho.....</i>	<i>91</i>
<i>Tabela 24: Percentagem de empresas no SCIE por primeiro e último ano com registo</i>	<i>91</i>
<i>Tabela 25: Descritivas gerais de variáveis selecionadas de todas as sociedades.....</i>	<i>92</i>
<i>Tabela 26: Descritivas gerais de variáveis selecionadas apenas de empresas pequenas e médias</i>	<i>92</i>
<i>Tabela 27: Percentagem de empresas no QP por primeiro e último ano com registo</i>	<i>93</i>
<i>Tabela 28: Distribuição de ganhos médios mensais por ano (euros).....</i>	<i>93</i>
<i>Tabela 29: Tipo de contrato dos trabalhadores por ano (%)</i>	<i>93</i>
<i>Tabela 30: Habilitações dos trabalhadores por ano (%).....</i>	<i>94</i>
<i>Tabela 31: Regime de duração do trabalho por ano (%)</i>	<i>94</i>
<i>Tabela 32: Habilitações dos trabalhadores e gestores, por dimensão de empresa (2019; %)</i>	<i>94</i>
<i>Tabela 33: Dados em falta no QP-Trabalhadores.....</i>	<i>95</i>
<i>Tabela 34: Empresas com registos em falta em um ou mais anos entre o primeiro e o último registo</i>	<i>95</i>
<i>Tabela 35: Trabalhadores duplicados no QP-Trabalhadores.....</i>	<i>95</i>
<i>Tabela 36: Empresas com outliers, segundo os intervalos de Tukey com $k = 3$, por classe de dimensão e apoio (PT2020).....</i>	<i>96</i>
<i>Tabela 37: Estatísticas descritivas gerais de variáveis selecionadas para empresas com volume de negócios = 0</i>	<i>96</i>
<i>Tabela 38: Estatísticas descritivas gerais de variáveis selecionadas para empresas com VAB < 0</i>	<i>97</i>
<i>Tabela 39: Empresas com registo, cruzamento de bases de dados (SCIE e QP)</i>	<i>97</i>
<i>Tabela 40: Empresas com registo, e apoiadas ou candidatas ao PT2020. Cruzamento de bases de dados (SCIE e QP).....</i>	<i>98</i>
<i>Tabela 41: Estatísticas descritivas gerais de indicadores selecionados de todas as empresas</i>	<i>99</i>
<i>Tabela 42: Descritivas gerais de indicadores selecionados só de empresas pequenas e médias</i>	<i>100</i>

Índice de Figuras

<i>Figura 1: Tratamento e impactos</i>	<i>9</i>
<i>Figura 2: Recursos, resultados e impactos.....</i>	<i>9</i>
<i>Figura 3: GAD de tratamento com variáveis de confundibilidade</i>	<i>34</i>
<i>Figura 4: Percentis de VAB para empresas com um trabalhador</i>	<i>52</i>
<i>Figura 5: Densidade de observações no espaço do VAB e Volume de Negócios</i>	<i>53</i>
<i>Figura 6: Distribuição do Volume de Negócios</i>	<i>54</i>
<i>Figura 7: Distribuição da Produtividade Aparente do Trabalho</i>	<i>54</i>
<i>Figura 8: Exemplo de árvore de decisão</i>	<i>59</i>
<i>Figura 9: Importância de variáveis para classificar empresas candidatas</i>	<i>59</i>
<i>Figura 10: Importância de variáveis para classificar cada empresa.....</i>	<i>60</i>
<i>Figura 11: Gráficos a pares, histogramas, caixas de bigodes e correlações.....</i>	<i>63</i>

1. Introdução

Este relatório sintetiza os resultados do projeto “Contrafactual-ML: Avaliação contrafactual dos apoios às empresas: cruzamento entre abordagens de inferência causal e de aprendizagem automática”. De forma mais geral, o “Contrafactual-ML” pretende contribuir para a melhoria dos processos de avaliação contrafactual de impactos dos sistemas de incentivos às empresas cofinanciados por fundos europeus em Portugal, sistematizando os principais desafios que se colocam a este tipo de avaliação e contribuindo para a capacitação de técnicos de organismos públicos e de jovens investigadores nestes domínios. O objetivo específico do projeto consiste em identificar, discutir e implementar a título ilustrativo abordagens recentes e inovadoras que cruzam métodos mais comuns de inferência causal com métodos de aprendizagem automática para efeitos de avaliação contrafactual de impacto.

Podemos definir vários objetivos com utilidade para uma avaliação contrafactual de impactos, todos eles passíveis de serem prosseguidos por um sistema de aprendizagem automática. Por exemplo, um objetivo pode ser saber qual o valor mais provável para entradas que tenham dados omissos. Este é um caso de aprendizagem supervisionada: para muitos outros exemplos sem dados omissos, sabemos qual o valor real que noutros casos pode estar omissos. Basta que o sistema aprenda como estes valores são gerados e consiga extrapolar para casos em que os dados são omissos para fazer imputação desses dados. Outro exemplo é a deteção de anomalias nos dados. Podia ser usada uma estratégia supervisionada, caso estejam identificados exemplos suficientes do que são anomalias. Contudo, pode também ser aplicada uma abordagem não supervisionada, em que o objetivo não é visível nos dados, e que não requer o trabalho moroso de identificar exemplos de anomalias previamente, mantendo desempenhos semelhantes. Outro objetivo ainda pode ser modelar diretamente variáveis observáveis e saber que variáveis são essenciais para o fazer, tarefa esta que é fundamental para a construção de um contrafactual (mais ou menos paramétrico). As abordagens de aprendizagem automática são mais flexíveis nessa modelação e mais dependentes dos dados (e menos do utilizador) apresentando vantagens relevantes. Além disso, como veremos, hoje é possível utilizar aprendizagem automática diretamente para inferência. É também possível identificar grupos em que o tratamento seja mais homogêneo que outros, revelando heterogeneidade, tudo isto de forma automatizada e deixando os dados falar por si. Por fim, é possível fazer previsões de quais seriam os impactos de diferentes apoios para diferentes públicos-alvos, abrindo caminho à otimização de políticas.

Neste relatório enfatizamos também a importância de articular a avaliação contrafactual com a elaboração de teorias da mudança das políticas sob análise. Como veremos, embora a abordagem contrafactual tenha subjacente uma visão peculiar da causalidade que a afasta das abordagens à avaliação baseadas na teoria, as lógicas configuracionais e generativas à causalidade (que partem da teoria da mudança das políticas para identificar os seus impactos) podem contribuir para um desenho mais adequado de avaliações contrafactuais. Isto reflecte-se em aspetos tão distintos como: a escolha da unidade de análise, a identificação das variáveis de resultado e de controlo relevantes, a definição dos momentos do tratamento e da aferição de impacto, ou o desenho das análises de heterogeneidade dos impactos.

O relatório está organizado da seguinte forma. No Capítulo 2 discutimos a teoria da mudança como base da avaliação de políticas e desenvolvemos uma teoria da mudança de uma política pública que

será usado como exemplo mais adiante (o SI Inovação Produtiva). No Capítulo 3 descrevemos as diferentes metodologias de avaliação contrafactual de impactos, apresentando exemplos de aplicação para cada uma delas. No final deste Capítulo, discutimos os possíveis contributos da aprendizagem automática para a avaliação contrafactual de impactos de políticas públicas. No capítulo seguinte discorremos sobre as diferentes visões da causalidade e o modo como estas se relacionam com as abordagens à avaliação baseadas na teoria e contrafactual. O Capítulo 5 apresenta os dados utilizados neste estudo, tecendo-se considerações sobre a sua qualidade e potencialidades de análise. Por fim, no Capítulo 6 concretizamos as reflexões desenvolvidas nos Capítulos 2 a 4 com base no exemplo do SI Inovação Produtiva, procurando mostrar na prática o contributo da teoria da mudança para o desenho da avaliação contrafactual e o papel que a aprendizagem automática pode ter na melhoria dos métodos de avaliação utilizados.

2. Teorias da mudança das políticas sob análise

A teoria da mudança (ou teoria do programa) consiste na explicitação dos diferentes componentes ativados por uma política que culminam num resultado antecipado. Todas as políticas contêm, mesmo que implicitamente, uma narrativa teórica que alimenta uma expectativa de sucesso (Weiss, 1972:84).

A teoria da mudança enquanto conceito central na avaliação de políticas advém dos esforços de avaliadores que manifestaram o seu descontentamento com o estado da disciplina que assentava (e ainda assenta, particularmente nas políticas consideradas “económicas”) excessivamente em métodos (quase-)experimentais. Estes métodos, que dependem de uma lógica causal contrafactual, são recomendados por serem ateóricos (Angrist & Krueger, 1999; Wolpin, 2013). A intervenção pode ser avaliada através da sua intermitência – isto é, a comparação de casos em que a política é acionada com aqueles em que não é.¹ Este tipo de método é apelidado de *black box* (‘caixa negra’), por omitir o mecanismo através do qual os recursos mobilizados pela política têm impacto. A política é avaliada pelos seus efeitos sem inquirir sobre como e porquê estes são atingidos.

Afastando-se de um modelo tipo *black box*, onde o foco está no mero estabelecimento de uma relação entre uma intervenção e os seus impactos finais, muitas abordagens à avaliação passaram a enfatizar a importância de explicar cada etapa das intervenções, discutindo os seus pressupostos e teorias subjacentes, os mecanismos de funcionamento e o contexto endógeno e exógeno das intervenções, de modo a obter uma compreensão mais credível acerca da atribuição causal dos efeitos às políticas.

Foi neste quadro que ganhou expressão a avaliação baseada na teoria, onde é investigada cada etapa da lógica da medida, implícita ou explícita, identificando relações causais e mecanismos de mudança em cada contexto de intervenção. Alguns dos principais contributos para a formulação de conceitos da disciplina estão associados a autores como Huey-Tsyh Chen, Carol Weiss, Ray Pawson e Peter Rossi. Conceptualmente, as avaliações baseadas na teoria tentam identificar a base teórica da política, posteriormente testando essa teoria com o objetivo de explicar como estas intervenções funcionam e porque causam os resultados pretendidos ou observados (Leeuw, 2012). Isto é, as políticas são teorias (Pawson & Tilley, 1997) e parte do objetivo de uma avaliação é testá-las.

A abertura desta ‘caixa negra’, que pressupõe a compreensão do funcionamento detalhado de uma intervenção, permite não só fortalecer a inferência causal de uma avaliação, mas também contribuir para um exercício cumulativo de aprendizagem sobre políticas públicas. Mais adiante, no Capítulo 4, discutiremos a relação entre teoria da mudança e as abordagens contrafactuais à avaliação de política (que constituem o foco deste estudo). O presente capítulo visa introduzir o debate sobre o papel da teoria da mudança com base de avaliação de políticas e apresentar um esboço de teoria da mudança de uma medida de apoio ao investimento empresarial que é utilizada como referência ao longo do estudo.

¹ Por exemplo, a identificação de um efeito parte da conceção de um grupo de empresas comparável, que simula o que aconteceria numa realidade alternativa, em que não existisse política.

2.1. Teoria da mudança como base da avaliação de políticas

Em geral, as políticas não afetam diretamente as condições que pretendem melhorar, antes agem sobre dimensões mais acessíveis, que por sua vez afetarão a condição que motiva a intervenção (Rossi et al., 2019). A pertinência da teoria da mudança decorre deste reconhecimento, exigindo que a sequência implícita nas ambições de sucesso da política seja explicitada.

Nas avaliações baseadas em métodos (quase-)experimentais, a política é vista como um tratamento que é ativado num grupo (o grupo de tratamento) e desativado no outro (grupo de controlo). Esta perspectiva, considerada “simplista” (Trochim, 1989b), é muitas vezes denominada por *black box*, aludindo ao facto de que os mecanismos que levam os recursos da política a surtirem o seu efeito não são refletidos na análise contrafactual, resumidos a uma mera dicotomia entre ausência/presença de tratamento. As Figura 1 e Figura 2 pretendem ilustrar a divergência ontológica das abordagens.

Figura 1: Tratamento e impactos

$$X > Y$$

Figura 2: Recursos, resultados e impactos

$$X_1 > X_2 > X_3 > Y$$

A Figura 1 aponta para a interpretação de que os impactos de uma intervenção decorrem e podem ser compreendidos pelo efeito de uma variável correspondente ao tratamento X numa variável de impacto. A Figura 2 apresenta mais granularidade. Como vimos, a política não afeta, diretamente, a dimensão de impacto pretendida. Esta começa por disponibilizar recursos (X_1) que levam a um resultado imediato (X_2). Por sua vez, os resultados intermédios serão atingidos (X_3), o que permite a existência de impactos (Y).

Por exemplo, no contexto dos apoios às empresas, uma política não exerce uma influência sobre os níveis de investimento, mas sobre as condições de financiamento. É expectável que a melhoria desta dimensão leve à decisão de investir. O racional da política assume, portanto, que as condições de financiamento limitam os níveis de investimento. A afetação dos recursos (financiamento) leva à mudança de comportamento (decisão de investir).

Mayne (2015) distingue entre simples modelos lógicos e teorias da mudança. O modelo lógico corresponde à sistematização da sequência de etapas causais da política, desde os *inputs* aos impactos. Os *inputs* são normalmente associados aos recursos alocados à intervenção. Nas políticas de apoios a empresas, por exemplo, estes são muitas vezes recursos financeiros. Um pressuposto que decorre imediatamente é a necessidade de as condições de financiamento oferecidas serem vantajosas, ou seja, abaixo das taxas de juro cobradas no mercado.

Uma teoria da mudança engloba um modelo lógico, mas não só. Além da identificação de uma sequência de relações causais, uma teoria da mudança robusta envolve uma descrição detalhada dos mecanismos que lhe estão subjacentes, dos seus pressupostos e dos riscos que se colocam à realização dos efeitos previstos. Os pressupostos são condições que têm de ser verificadas para que o elo causal decorra como planeado. Os riscos são entendidos como possíveis eventos, que ao acontecerem, condicionam o elo causal em questão.

A teoria da mudança deve ser, preferencialmente, construída antes da aplicação da política, não só servindo como guia para as ambições de quem a concebe, mas também facilitando o processo de

monitorização, detalhando os resultados intermédios esperados. A teoria da mudança, no entanto, deverá ser revista na altura da avaliação, dada uma série de evidências que se tornaram claras. Não só a interpretação dos *stakeholders* muda durante uma intervenção, mas já é possível observar os resultados intermédios, o que pode alterar a perceção sobre o modo como a política opera.

A informação que permite a construção das teorias da mudança deve ser retirada de diferentes fontes. Primeiro, a teoria (mesmo que implícita) de quem concebeu a política deve ser examinada. Para além das ambições refletidas nos resultados esperados, é de esperar que os responsáveis pela conceção da política tenham alguma ideia de como os recursos mobilizados irão originar a sequência de efeitos que levam aos resultados finais. Os beneficiários constituem outra fonte importante de informação para a teorização da política. Da sua ação depende o sucesso da política (Pawson & Tilley, 1997), pelo que a sua interpretação sobre os mecanismos relevantes é crucial.

Um aspeto muitas vezes realçado pelos proponentes de abordagem à avaliação baseadas na teoria, no entanto, é o facto de os decisores políticos não serem muitas vezes capazes de explicitar, de maneira satisfatória, a base teórica das intervenções. Mesmo que o fossem, a aceitação incondicional da narrativa elaborada por aqueles, ou até mesmo pelos demais *stakeholders*, não é um processo adequado. Weiss (1972), por exemplo, alerta para a necessidade de a elaboração de teorias do programa dialogarem com as ciências sociais, analisando de forma crítica as convicções declaradas por aqueles que estão diretamente envolvidos na política em causa.

2.2. Teoria da mudança do SI Inovação Produtiva

Nesta secção apresenta-se um esboço de teoria da mudança do Sistema de Incentivos (SI) Inovação Produtiva, uma medida de apoio às empresas apoiada no âmbito do Portugal 2020. Esta teoria da mudança servirá de referência em vários pontos do presente relatório.

2.2.1. A política

O SI Inovação Produtiva visa responder aos constrangimentos relacionados com a competitividade e a internacionalização do tecido empresarial português, decorrentes do perfil de especialização produtiva, das competências e estratégias das PME e das condições de contexto à atividade empresarial, nomeadamente as condições de financiamento.

O SI Inovação Produtiva comparticipa projetos de investimento empresarial de natureza inovadora. Segundo o *Aviso N.º 13/SI/2021 - Inovação Produtiva* (2021) a política visa promover a inovação empresarial nos seguintes domínios: a) Produção de novos bens e serviços ou melhorias significativas da produção atual através da transferência e aplicação de conhecimento; b) Adoção de novos ou significativamente melhorados processos ou métodos de fabrico, de logística e distribuição, bem como métodos organizacionais. Consideram-se enquadráveis os investimentos de natureza inovadora que se traduzam na produção de bens e serviços transacionáveis e internacionalizáveis e com elevado nível de incorporação nacional, exceto projetos de investimento de mera expansão ou de modernização.

Os projetos aprovados são financiados entre 15% a 75% do seu valor total. É utilizada uma forma híbrida de financiamento, que consiste na combinação de duas componentes, uma reembolsável e outra não reembolsável. A componente reembolsável é garantida através de um empréstimo bancário sem juros com uma duração de oito anos (pagamentos semestrais com dois anos de carência) e depende da aprovação do financiamento do projeto por parte da instituição bancária financiadora. Normalmente, cada componente representa metade do financiamento concedido a cada projeto, mas existem várias exceções onde pode ser garantido apenas um subsídio reembolsável e existe a possibilidade de a empresa financiar por capital próprio a componente que seria reembolsável, recebendo apenas o subsídio não reembolsável. A componente do incentivo não reembolsável é atribuída a título não definitivo até à avaliação dos resultados do projeto, em função do grau de cumprimento das metas contratualmente fixadas.

A incerteza em relação ao retorno do investimento e a instabilidade acerca da informação sobre a apropriação dos efeitos económicos produz o que se denominou como falhas de mercado (Arrow, 1972), reduzindo assim o interesse e o investimento privado em inovação. Devido à natureza pública do conhecimento e à presença de externalidades positivas, os retornos privados são inferiores aos retornos sociais (Bloom et al., 2013). Os apoios para promoção de investimento privado tem por base uma tentativa de correção desta suposta falha de mercado, pretendendo induzir os investimentos privados em inovação, com a garantia de que parte do investimento será reembolsado, o que reduz o esforço financeiro e o risco do investimento. A aplicação dessa medida funciona como uma partilha ou divisão dos riscos relativos às atividades de inovação e desenvolvimento das empresas (Hall & Lerner, 2010).

2.2.2. Resultados imediatos

O principal motivo para a utilização de subsídios ao investimento pelas empresas beneficiárias é o facto de permitirem reduzir os custos de investimento. Assim, a consequência esperada no imediato da implementação do SI Inovação Produtiva é um nível de investimento superior nas empresas apoiadas, com a realização de projetos de investimento de natureza inovadora que não seriam concretizados na ausência do subsídio.

Para que o resultado exposto no parágrafo anterior se verifique, é necessário que se verifiquem alguns pressupostos, seja para que os subsídios realmente reduzam os custos, seja para que essa redução de custos se traduza em mais investimento de natureza inovadora.

Para que a existência do SI Inovação Produtiva se traduza numa redução de custos de investimento, é necessário assumir que: 1) o programa disponibiliza financiamento às empresas apoiadas em condições mais atrativas que fontes de financiamento privadas; 2) os custos relacionados com o processo de candidatura são irrelevantes quando comparados com os benefícios esperados pelo acesso ao subsídio.

Para que a redução dos custos de investimento se traduza em mais investimento de natureza inovadora, é necessário assumir que: 1) os custos de financiamento são uma restrição relevante e suficiente para o não investimento das empresas apoiadas; 2) os projetos aprovados são executados até ao fim com os custos projetados e correspondem ao tipo de investimento desejado; 3) o financiamento é disponibilizado no tempo acordado com as empresas.

Os principais riscos que podem comprometer o aumento do nível de investimento de natureza inovadora nas empresas apoiadas prendem-se com a possibilidade de o nível e natureza dos investimentos realizados pelas empresas apoiadas não se distinguirem de forma relevante do que seriam realizados na ausência do apoio. Existem várias razões que podem contribuir para este cenário, nomeadamente: 1) os critérios de aprovação de empresas candidatas podem selecionar apenas as empresas com condições financeiras mais favoráveis, que são aquelas que à partida teriam menos restrições ao financiamento privado²; 2) os critérios de aprovação de projetos podem excluir projetos com elevado nível de incerteza, apenas aprovando projetos cuja rentabilidade é previsível e que à partida teriam menos restrições ao financiamento privado; 3) as instituições de crédito privadas (que têm de aprovar também os projetos) podem ter critérios demasiado restritivos e não aprovar projetos com risco elevado (que são aqueles que têm maior dificuldade de financiamento).

Podem existir outros riscos relacionados com a realização dos projetos, nomeadamente: 5) a empresa pode não ter as capacidades financeiras ou técnicas para realizar o projeto na prática; 6) os projetos podem não ter sido adequadamente planeados, ou depender de fatores externos que complicam a sua realização, implicando custos ou uma duração temporal adicional; e 7) existem atrasos na atribuição do apoio, que põe em risco a realização dos projetos.

² Esta possibilidade é estudada na literatura utilizando o tamanho da empresa. Visto que as maiores empresas são normalmente aquelas que têm condições de financiamento mais favoráveis, é de esperar que a adicionalidade promovida pelo subsídio seja maior nas empresas mais pequenas. Este pressuposto é testado e confirmado empiricamente para diferentes tipos de subsídios (Criscuolo et al., 2019; Srhoj et al., 2021; Szücs, 2020).

Para além dos efeitos mais diretos do apoio sobre a capacidade de investimento das empresas, é possível que a participação no programa provoque um efeito de certificação, sinalizando a investidores privados a qualidade das empresas apoiadas (Li et al., 2020; Stevenson et al., 2021; A. Wu, 2017; R. Wu et al., 2021), facilitando o seu financiamento por outras vias que não apenas o apoio público. De facto, uma vez que as empresas apoiadas tiveram de cumprir certos critérios de seleção, a aprovação da sua candidatura pode traduzir, de algum modo, a qualidade do seu projeto e da sua atividade face a outras empresas comparáveis. Para além disso, na medida em que na maior parte dos projetos em causa é requerida a participação de uma instituição financeira, o apoio pode contribuir para a criação de ligações de confiança e conhecimento entre as empresas apoiadas e as instituições de financiamento privadas. Sendo assim, é de esperar que as empresas apoiadas beneficiem de acesso a financiamento privado em melhores condições para outros investimentos que não o projeto subsidiado. Melhores condições de financiamento podem incentivar a um maior investimento geral pelas empresas apoiadas. Este efeito deve ser sobretudo sentido em jovens e pequenas empresas, as quais é mais provável terem poucas ligações com investidores privados e cuja informação é mais restrita para novos investidores.

Os pressupostos relacionados com o efeito de certificação atrás referido incluem os seguintes: 1) a falta de informação ou de ligações pré-estabelecidas é um fator importante que penaliza as condições de acesso a financiamento privado pelas empresas apoiadas; 2) o SI Inovação Produtiva e o processo de aprovação de candidaturas é percecionado por agentes privados como um indicador de qualidade das empresas apoiadas.

2.2.3. Resultados intermédios

Assumindo que os projetos subsidiados são efetivamente realizados, espera-se que a participação no SI Inovação Produtiva se traduza no desempenho das empresas apoiadas. Os projetos financiados pela medida têm o objetivo de dotar as empresas de maior capacidade produtiva relacionada com produtos transacionáveis e internacionalizáveis, diferenciadores e de qualidade, com elevado nível de incorporação nacional. É esperado que as empresas apoiadas aproveitem a capacidade produtiva adicional e a utilizem para aumentar a produção relacionada com estes produtos. De acordo com os objetivos do programa e com os critérios de seleção de projetos, no caso das não-PME apoiadas é também esperada a criação de emprego qualificado adicional e um efeito de arrastamento em PME.

Para que os efeitos referidos se verifiquem, é necessário que: 1) os projetos realizados estejam alinhados com os critérios do programa, permitindo às empresas expandir a sua capacidade produtiva relacionada com produtos e serviços inovadores transacionáveis e internacionalizáveis e/ou melhorar os processos produtivos; 2) as empresas tenham as capacidades técnicas e financeiras para aproveitar as oportunidades de expansão que o aumento da capacidade produtiva cria; 3) exista procura interna ou externa pelos novos produtos; 4) as empresas apoiadas tenham as competências necessárias para competir com outras empresas com produtos concorrenciais; 5) a nova produção (ou forma de produção) permita às empresa adicionar um elevado valor acrescentado ou que outras empresas nacionais adicionem um elevado valor acrescentado; 6) no caso das não-PME, o investimento realizado implique a criação de emprego qualificado e não substituto; 7) as empresas não-PME apoiadas tenham fortes ligações com empresas PME nacionais ou os projetos contribuam para a criação dessas ligações.

Alguns dos riscos que podem pôr em causa a obtenção dos resultados intermédios identificados são: 1) os efeitos esperados ao nível da melhoria do processo produtivo e/ou dos produtos são sobrestimados na fase de planeamento e aprovação dos projetos; 2) a empresa não tem as capacidades técnicas necessárias para fazer uso do investimento realizado, não sendo desenvolvidas atividades de formação complementares ao projeto ou sendo desenvolvidas atividades de formação insuficientes; 3) o contexto económico nacional ou internacional prejudica a atividade da empresa (por exemplo, quebra na procura, deterioração das condições de financiamento, encarecimento dos inputs, aumento da concorrência), impossibilitando-a de retirar benefícios da implementação do projeto.

Em referência à possível melhoria das condições de financiamento por via do efeito de certificação, é possível que as empresas apoiadas realizem mais projetos de investimento (para além do subsidiado) que permitam expandir a sua atividade, melhorar o seu processo produtivo ou beneficiar de menores custos pagos pelo financiamento que outras empresas não apoiadas. Os possíveis efeitos deste mecanismo são demasiado gerais, mas podem ser resumidos em: 1) as empresas apoiadas expandem a sua atividade produtiva ou melhoram os seus processos de produção, independentemente de essas atividades estarem relacionadas com os critérios do programa; e 2) as empresas apoiadas têm menores custos com o financiamento privado.

2.2.4. Resultados finais

Se os projetos financiados provocarem os resultados intermédios referidos anteriormente, é esperado que o SI Inovação Produtiva contribua para melhorias relevantes em termos da competitividade e internacionalização da economia portuguesa, nomeadamente: (i) aumento do investimento de natureza inovadora realizado pelas empresas à escala nacional; (ii) aumentando do peso das atividades de elevada intensidade tecnológica ou em conhecimento e de bens e serviços transacionáveis na estrutura produtiva nacional; (iii) aumento da competitividade internacional das empresas portuguesas e do valor acrescentado produzido em território nacional (refletido, por exemplo, na maior intensidade exportadora da economia e na melhoria do saldo da balança comercial; e (iv) aumento do emprego qualificado à escala nacional.

Podem ser identificados vários mecanismos que, pressupondo a realização dos resultados intermédios, levem à verificação dos resultados finais identificados. O mecanismo mais direto é o crescimento e aumento da produtividade nas empresas apoiadas como consequência do aumento e da melhoria da capacidade produtiva proporcionados pela realização dos projetos aprovados. Estes efeitos diretos também podem ser alcançados parcialmente pelo resultado do efeito de certificação. No entanto podem existir outros mecanismos importantes que acabam por alterar o comportamento de empresas não apoiadas para benefício da economia, nomeadamente: 1) efeito de arrastamento no desempenho em empresas não diretamente apoiadas pelo programa, mas que têm ligações importantes com as empresas apoiadas (especialmente relevante no caso das não PME apoiadas); 2) *spillovers* de conhecimento (por exemplo, as capacidades técnicas adicionais dos trabalhadores das empresas apoiadas são partilhadas com outras empresas, seja pela troca de conhecimentos entre trabalhadores seja pela circulação de trabalhadores entre empresas); 3) os resultados do investimento por parte das empresas apoiadas podem levar a um aumento de concorrência nos mercados

específicos dessas empresas e levar a que outras empresas não apoiadas que concorram nos mesmos mercados tenham a necessidade de investir e inovar para se manterem competitivas.

Os principais pressupostos para a realização dos resultados finais identificados são: 1) a dotação orçamental do SI Inovação Produtiva tem uma dimensão suficiente para que o número e a dimensão dos projetos apoiados sejam relevantes relativamente ao universo empresarial português; 2) os critérios de seleção de empresas e projetos têm a qualidade necessária para financiar projetos onde os efeitos esperados sejam positivos e, de preferência, para selecionar aqueles em que os efeitos esperados são maiores; 3) o contexto económico (sobretudo o internacional) permite a transformação desejada da economia portuguesa; 4) não existem políticas que condicionem as transformações desejadas no perfil de especialização (e, de preferência, existem políticas complementares); 5) as cadeias de valor ao nível nacional permitem que as melhorias promovidas nas empresas apoiadas (especialmente nas não-PME) gerem externalidades positivas para o resto do tecido empresarial.

Alguns dos possíveis riscos que podem comprometer os resultados finais desejados são: 1) o programa de incentivos não tem a dotação orçamental suficiente para realizar mudanças relevantes na economia; 2) os constrangimentos ao investimento de natureza inovadora na maioria das empresas portuguesas não são colmatados pelo programa de incentivos (i.e., a maioria das empresas tem outros constrangimentos não financeiros, constrangimentos financeiros que não são respondidos pelo programa, ou existem poucas oportunidades de investimentos rentáveis, fazendo com que o número de empresas com interesse em serem apoiadas seja relativamente pequeno); 3) os critérios de seleção dos beneficiários acabam por selecionar, tendencialmente, empresas que fariam os mesmos investimentos na ausência do programa, não induzindo investimento adicional relevante ao nível agregado; 4) os critérios de seleção dos beneficiários acabam por selecionar, tendencialmente, projetos de pouca qualidade (relativamente a projetos que não são implementados mas teriam um maior retorno esperado – por exemplo por incapacidade técnica para prever o retorno dos programas ou por aversão ao risco); 5) as condições macroeconómicas (por exemplo a evolução cambial) são desfavoráveis aos setores apoiados; 6) os resultados podem não ser sustentáveis no tempo (i.e., a política pode só conseguir induzir investimento adicional sustentado se for prolongada eternamente, não alterando as decisões de investimento das empresas para além da redução dos custos pelo financiamento direto de parte do investimento); 7) podem não ser implementadas políticas complementares necessárias às transformações desejadas (ou existirem políticas que induzem transformações contrárias às tidas como desejadas pelo SI Inovação Produtiva).

3. Avaliação contrafactual de impactos

Um dos principais objetivos ao avaliar o impacto de uma política é saber qual foi o seu contributo para alterar a realidade que observamos. A hipótese basilar da avaliação contrafactual é que podemos identificar o impacto de uma política comparando aquilo que aconteceu com aquilo que teria acontecido caso a política não tivesse sido implementada. Observar estas duas realidades em simultâneo é logicamente impossível, pelo que os avaliadores têm de se contentar em tentar estimar a realidade que não é observada.

A estimação de um cenário contrafactual recorre a dados empíricos sobre unidades que não foram abrangidas pela política, mas que são semelhantes o suficiente àquelas que o foram, possibilitando uma comparação entre unidades tratadas e não tratadas. No caso das políticas de apoio às empresas, as estimações de contrafactual mais comuns utilizam dados de empresas não tratadas, permitindo inferir os efeitos das políticas ao nível da empresa média. Note-se, no entanto, que nem todas as empresas não tratadas são bons controlos. Um bom controlo deveria ser uma unidade exatamente igual à unidade tratada onde a única diferença seria a política em estudo. Como é impossível encontrar duas empresas exatamente iguais, a principal preocupação de qualquer análise contrafactual deve ser encontrar (ou construir) grupos de comparação onde a única diferença estatisticamente relevante, entre o grupo de empresas apoiadas e o grupo de controlo, seja a participação na política.

Mesmo quando se encontram ou constroem grupos de comparação semelhantes, existe uma hipótese essencial por detrás da avaliação contrafactual, que constitui um risco à sua validade: é necessário assumir que as unidades que servem como controlos não estão sujeitas a qualquer efeito direto ou indireto da política (i.e., não existam efeitos de *spillover*). Este requisito pode ser uma grande dificuldade para a avaliação de apoios a empresas porque, muitas vezes, os objetivos desses apoios são provocar efeitos de *spillover* positivos³. Mesmo quando não o são, é muito provável que qualquer apoio que provoque alterações nas empresas apoiadas acabe por provocar alguns efeitos indiretos nas restantes empresas, como consequência das interações que empresas não apoiadas têm com empresas apoiadas. A questão sobre se existem efeitos de *spillover* que podem enviesar significativamente a avaliação deve ser discutida antes da sua implementação. A validade de todos os métodos contrafactuais comuns (inclusive aqueles que serão apresentados neste capítulo), dependem da hipótese de ausência de efeitos de *spillover*.

Contrafactual: como seria a realidade se o tratamento (a política) não existisse.

Tratamento: o evento para o qual se quer estimar o impacto causal (neste relatório, os exemplos de tratamentos apresentados são geralmente medidas de política pública de apoio a empresas; no entanto, o termo tratamento e a abordagem contrafactual é generalizável a muitas outras circunstâncias e domínios).

Unidades tratadas: unidades que recebem os efeitos diretos do tratamento (neste caso, empresas apoiadas).

Unidades de controlo (ou controlos): unidades semelhantes na sua natureza às unidades tratadas (por exemplo, empresas ou regiões) e que poderiam ter recebido o tratamento, mas não o receberam. São utilizadas para estimar o contrafactual das unidades tratadas.

Grupo de tratamento: o conjunto de unidades tratadas consideradas na análise, para as quais vai ser estimado o efeito do tratamento (geralmente o efeito médio).

Grupo de controlo: o conjunto de unidades de controlo utilizadas na análise; é utilizado como o contrafactual do grupo de tratamento para estimar o efeito do tratamento.

³ Por exemplo, tais efeitos são uma das justificações principais para a existência de políticas de apoio à I&D (e.g., Arrow (1972)).

3.1.Experiências aleatórias

As experiências aleatórias são consideradas o ideal da análise contrafactual, porque permitem obter grupos de comparação semelhantes entre si, construídos diretamente a partir da seleção dos beneficiários da política.

Para compreender a particularidade deste método, consideremos primeiro uma das formas de seleção mais comuns de políticas de apoio a empresas. No caso de subsídios ao investimento/ inovação, o processo de seleção de empresas beneficiárias começa pela apresentação de candidaturas por parte das empresas interessadas. Numa segunda fase, as candidaturas são avaliadas e ordenadas por nível de adequabilidade aos objetivos da política. As empresas que acabam por ser as beneficiárias da política são as candidatas que obtêm as melhores avaliações.

Efeito de seleção: ocorre quando o grupo de tratamento tem características diferentes das do grupo de controlo antes de participar no tratamento. Se o efeito de seleção não for tido em conta, a estimativa do impacto do tratamento irá atribuir à política efeitos que decorrem de outros fatores, o que é indesejável.

Pré-tratamento: todo o período temporal antes do tratamento começar. Por definição, é impossível existirem efeitos causais do tratamento neste período e por isso os grupos de comparação devem ser semelhantes neste período de forma a garantir a não existência de efeito de seleção.

Pós-tratamento: todo o período temporal a partir do início do tratamento. É neste período que os efeitos causais são estimados e deve ser assumido que tudo aquilo que observamos neste período pode ter sido influenciado pelo tratamento.

Este método de seleção baseado no mérito coloca um grande obstáculo à validade da inferência causal: o efeito de seleção. O efeito de seleção ocorre quando o grupo de tratamento tem características diferentes das do grupo de controlo antes de participar no tratamento. Neste caso, é esperado que o grupo de empresas tratadas seja diferente tanto do grupo de empresas candidatas não apoiadas (porque as empresas são escolhidas com base nos critérios de avaliação), como do grupo de empresas não candidatas (porque existem diferenças que explicam o interesse de algumas empresas em apresentarem uma candidatura e outras não). A existência destas diferenças viola a legitimidade da inferência causal pela comparação entre qualquer destes grupos, visto que nunca poderemos saber com certeza se as diferenças observadas no pós-tratamento se devem à política ou às diferenças pré-existentes entre as empresas apoiadas e não apoiadas.

Teoricamente, existe uma forma simples de evitar o efeito de seleção: selecionar de forma aleatória os beneficiários da política. Ao selecionar desta forma um número elevado de unidades de um certo grupo garante-se que as características do subgrupo selecionado e do subgrupo não selecionado serão semelhantes. A semelhança entre subgrupos será tanto maior

quanto maior for o número de unidades total, visto que em pequenos números o processo de aleatorização terá uma elevada probabilidade de gerar diferenças significativas.

Se a seleção aleatória for bem-sucedida podemos comparar diretamente o resultado do grupo de tratamento com o grupo de controlo. Pode-se atribuir a totalidade da diferença registada ao tratamento, visto que (devido à aleatorização) o grupo de tratamento é estatisticamente equivalente ao grupo de controlo, sendo a única diferença a participação no tratamento. A diferença registada entre os grupos é então a estimativa do efeito causal da política.

Inferir causalidade a partir de uma seleção aleatória só é possível em casos em que a política é propositadamente desenhada para permitir essa avaliação. Na grande maioria das políticas de apoio a empresas, não é desejável nem possível que a seleção de beneficiários seja aleatória. Se a seleção

de beneficiários for aleatória, o alcance da política fica estrangido e podem ser levantados problemas éticos devido à exclusão de possíveis beneficiários. O cumprimento da seleção aleatória tem de ser garantido, o que é muito difícil em situações de apoios empresariais, porque implicaria forçar as empresas selecionadas a aceitarem os apoios públicos nas condições estabelecidas pelo Estado. A recolha de dados é também uma grande preocupação para a viabilidade da inferência causal, sendo necessário recolher dados de todas as empresas que participaram na seleção, de modo a garantir a fiabilidade da experiência (o que é extremamente difícil caso a equipa de avaliação tenha de realizar inquéritos e dependa da resposta voluntária das empresas).

EXEMPLO DE AVALIAÇÃO:

SUBSÍDIOS À COLABORAÇÃO EM I&D: O EFEITO CAUSAL DE VOUCHERS NA PERFORMANCE DE INOVAÇÃO DAS EMPRESAS (KLEINE ET AL., 2020)

O programa “Innovation Vouchers Programme” foi estabelecido pelo InnovateUK em 2012, com um orçamento de 4 milhões de libras. O programa é destinado a micro, pequenas e médias empresas inovadoras e oferece até 5 mil libras para a contratação de serviços de especialistas com os quais nunca trabalharam, com a finalidade de obter novos conhecimentos que possam ajudar os seus negócios a inovar e crescer. Nas 10 rondas que foram realizadas antes do estudo, mais de 6.600 empresas solicitaram um *voucher*, cerca de 3.100 foram concedidos e quase 2 mil *vouchers* foram resgatados. Para analisar a eficácia do programa foram utilizadas três rondas em 2015.

Existiram quatro etapas principais para a participação no programa: (1) inscrição, (2) lotaria e verificações de elegibilidade, (3) solicitação do *voucher* e (4) pagamento final. Na fase inicial de candidatura, as empresas indicaram o projeto de inovação específico que pretendiam desenvolver com ajuda externa. Os candidatos propuseram ainda um determinado parceiro externo com quem esperavam trabalhar e avaliaram o impacto potencial do projeto de inovação nos seus negócios. Além disso, as empresas responderam a um questionário que incluiu a recolha de informação sobre as características das empresas, atividades anteriores relacionadas com inovação, bem como planos para atividades futuras.

Na segunda etapa, realizou-se um sorteio. As lotarias selecionaram o número de beneficiários de acordo com o orçamento de cada ronda. As empresas selecionadas foram então analisadas por revisores independentes que verificavam o cumprimento dos critérios de elegibilidade. Para obter um grupo de controlo comparável às empresas do grupo de tratamento, as empresas que não foram selecionadas como beneficiárias pelo sorteio também foram submetidas à verificação de elegibilidade. Os revisores não sabiam se uma empresa tinha sido selecionada pela lotaria ou não. Após o processo de revisão, fez-se uma verificação de diligência prévia apenas para as empresas do grupo de tratamento. Um *voucher* de inovação foi oferecido se o solicitante passou na lotaria e em todos os processos de verificações.

Os candidatos tinham 10 dias para aceitar a oferta e até 6 meses para concluir o projeto proposto. Após a conclusão do projeto, o requerente enviava um pedido de pagamento. Por fim, o pedido era analisado por um funcionário do programa, havendo lugar ao pagamento do *voucher* em caso de aprovação.

Para assegurar a validade do estudo, os autores avaliaram a qualidade da seleção aleatória (confirmando se as características dos grupos de tratamento e de controlo eram estatisticamente equivalentes) e se existia algum enviesamento na distribuição de empresas que respondeu ao inquérito (os dados recolhidos dependem da resposta das empresas inquiridas).

Uma limitação importante à análise foi o facto de cerca de 335 empresas a quem foram oferecidos *vouchers* não os terem reclamado e 107 empresas falharem o teste de diligência prévia (que não pode ser efetuado às empresas do grupo de controlo por razões legais). Isto significa que uma grande parte das empresas selecionadas para o tratamento acabaram por não o receber. Para assegurar a validade da estimação, estas empresas foram incluídas no grupo de tratamento, pelo que aquilo que se estava, de facto, a estimar passou a ser o efeito médio da intenção de tratar (esta situação é abordada na secção seguinte sobre variáveis instrumentais).

Os resultados do estudo mostram que os *vouchers* foram bem-sucedidos, promovendo a execução de projetos de inovação com resultados positivos em termos de desenvolvimento de novos produtos e serviços, processos internos e propriedade intelectual.

Em síntese, as experiências aleatórias são a circunstância ideal para uma avaliação contrafactual. No entanto, é muito difícil garantir na prática as condições de uma experiência aleatória na atribuição de políticas de apoio a empresas. Avaliações que utilizam a seleção aleatória são por isso muito raras nesta área e, mesmo quando são possíveis, costumam deparar-se com dificuldades que limitam a validade da estimação.

3.2. Variáveis instrumentais

Quando não podemos recorrer a uma experiência aleatória, podemos tentar introduzir alguma aleatoriedade no nosso método de estimação recorrendo a variáveis instrumentais. O método das variáveis instrumentais permite controlar o efeito de seleção utilizando uma variável (a variável instrumental), obtendo estimativas não enviesadas do efeito causal.

Uma boa variável instrumental é uma variável que está fortemente correlacionada com a seleção para o tratamento, mas que não está correlacionada com as variáveis de resultado. O melhor exemplo disto é uma seleção aleatória com incumprimento (como no exemplo apresentado na secção anterior). Nos casos onde existe uma seleção aleatória, mas nem todas as unidades selecionadas para tratamento acabam por participar na política, não podemos inferir efeitos causais diretamente pela comparação entre as unidades participantes e não participantes na política. Neste caso, podemos identificar duas variáveis: uma variável Z que nos indica quem foi selecionado para tratamento, e uma variável P que nos indicia quem efetivamente participou no tratamento. Para obtermos uma estimativa não enviesada da participação na política (i.e., a variável P), teremos de recorrer à variável Z como variável instrumental. A variável Z é um bom instrumento porque a seleção inicial para tratamento influencia quem acaba por participar na política e, como a seleção foi aleatória, a variável Z não está correlacionada (é independente) com qualquer variável de resultado. A variável instrumental não necessita, no entanto, de ser uma variável de seleção aleatória, e o exemplo ilustrado anteriormente é válido para qualquer variável Z que cumpra os requisitos referidos.

Correlação: duas variáveis dizem-se correlacionadas se variam em conjunto (por exemplo, em qualquer grupo de empresas é provável que o número de trabalhadores esteja correlacionado com o montante de vendas, isto é, que as empresas com maior número de trabalhadores tendam a ser aquelas com o maior volume de vendas). Note-se que a existência de uma correlação estatística não implica uma relação de causalidade entre as variáveis em causa. Se existem diferenças significativas entre o grupo de tratamento e o grupo de controlo numa certa variável, então essa variável está correlacionada com a seleção para o tratamento

Independência: independência entre duas variáveis é essencialmente o contrário de correlação, significa que não existe qualquer relação estatística entre elas.

Variáveis de resultado: são as variáveis para as quais se pretende estimar o efeito do tratamento.

Na maioria dos casos de avaliação de apoios empresariais onde são utilizadas variáveis instrumentais, estas são variáveis regionais ou setoriais relacionadas com a seleção para a política e que se pressupõe que não influenciem de qualquer outra forma as variáveis de resultado estudadas.

A principal limitação ao uso de variáveis instrumentais é que nunca se pode garantir que a variável escolhida não influencie de alguma forma as variáveis de resultado, tendo esta escolha de ser argumentada teoricamente (à exceção de uma seleção aleatória com incumprimento). Visto que a maioria dos fatores que condicionam a participação

nos programas de apoio a empresas está quase sempre correlacionada com os resultados que se pretendem influenciar, as possibilidades de utilização deste método são limitadas.

ALGUNS EFEITOS CAUSAIS DE UMA POLÍTICA INDUSTRIAL (CRISCUOLO ET AL., 2019)

O “Regional Selective Assistance” começou em 1972 e desde o início dos anos 80 tem sido o principal programa de suporte empresarial no Reino Unido. O programa subsidia empresas em zonas desfavorecidas (níveis de PIB per capita relativamente baixos e elevado desemprego), tendo como principal objetivo a criação e manutenção de emprego no setor industrial. As empresas das áreas cobertas pelo programa podem apresentar candidaturas com projetos de investimento como a criação de novas fábricas ou a modernização de uma já existente. Se a candidatura for aprovada, o programa financia até 35% do custo do projeto. A percentagem financiada depende do “atraso económico” da região onde se localiza a empresa, e a fórmula é calculada ao nível da União Europeia, o que reduz preocupações com endogeneidade na determinação das percentagens de financiamento (o grau de atraso de cada região é determinado da mesma forma e relativamente a todas as regiões europeias a cada sete anos, o que impossibilita que os agentes nacionais decidam a intensidade do apoio de cada região em função da evolução da economia nacional).

Para obter estimativas do impacto causal do programa, os autores exploram as alterações na elegibilidade das regiões do Reino Unido ocorrida no ano 2000. São identificadas três razões que podem levar à alteração da elegibilidade: 1) alterações nos indicadores usados e no peso de cada indicador para a fórmula de cálculo do “atraso” (é exógeno visto ser determinado ao nível europeu); 2) evolução dos valores médios dos indicadores económicos (como o PIB per capita) ao nível europeu (as zonas são consideradas “desfavorecidas” em relação à média europeia; isto representa uma variação considerada exógena porque a variação desses indicadores ao nível do Reino Unido tem um peso pequeno para a variação da média europeia); 3) evolução económica de cada área (o que é claramente endógeno visto que a evolução económica de cada região depende dos resultados das suas empresas, e os resultados passados influenciam os resultados futuros que se pretende avaliar).

Visto que a variação da percentagem de comparticipação de cada região não é um indicador completamente exógeno (porque pode estar correlacionado com a evolução económica de cada região), os autores estimam uma variável exógena que representa o que seria a variação da percentagem de apoio em cada região caso não tivesse existido qualquer variação dos indicadores económicos regionais. Esta variável capta a variação das taxas de comparticipação causada pelos dois fatores exógenos referidos no parágrafo anterior e é uma boa variável instrumental (ou seja, está correlacionada com o tratamento, mas não com o resultado) por ser independente do desempenho das empresas do Reino Unido.

Os autores concluem que nas áreas elegíveis para maiores taxas de comparticipação se verificou um aumento significativo do emprego em manufatura e uma redução do desemprego. Quantitativamente, um aumento de 10 pontos percentuais na taxa de comparticipação máxima, estimula, em média, um aumento de 10% no emprego em manufatura da região. Numa análise de heterogeneidade, verifica-se que este impacto apenas é verificado nas pequenas empresas: as grandes empresas aceitam os subsídios sem aumentar a sua atividade.

3.3. Desenho de regressão descontínua

Existe um caso especial onde é possível aproximar a validade da avaliação à de uma experiência aleatória num subconjunto particular de unidades, utilizando o desenho de regressão descontínua (RDD). Para utilizar este método é necessário que exista uma única variável contínua que determine a seleção de unidades tratadas e que exista um ponto de corte identificável.

O exemplo mais comum na avaliação de políticas de apoio a empresas é a utilização de uma variável de classificação de mérito das candidaturas. Este tipo de variáveis existe quando, durante o processo de seleção, é feita uma avaliação sobre a qualidade das empresas ou dos projetos que se considera apoiar e essa informação é sintetizada numa única variável. Se essa variável for utilizada para selecionar as empresas que são apoiadas, excluindo do apoio aquelas que têm a classificação mais baixa, podemos utilizar a variável de classificação como ferramenta para a inferência causal.

A principal hipótese subjacente a este método é que ao selecionar beneficiários segundo a sua posição relativamente a um valor específico (o ponto de corte) numa única variável, a única diferença relevante entre os grupos de beneficiários e não beneficiários muito próximos do ponto de corte será a participação na política. Para garantir isto, é necessário que as unidades não consigam manipular a sua classificação de forma a posicionarem-se deliberadamente acima ou abaixo do ponto de corte. No caso dos apoios a empresas, esta é uma garantia que os processos de seleção geralmente têm de cumprir por razões éticas e de eficiência na distribuição de recursos.

Para utilizar a seleção por um ponto de corte como ferramenta de inferência, não é necessário que exista cumprimento total (i.e., que todos os indivíduos acima do ponto de corte recebam o tratamento e que todos os indivíduos abaixo do ponto de corte não recebam o tratamento). Desde que o ponto de corte influencie significativamente a seleção de beneficiários, podemos utilizar a seleção pelo ponto de corte como variável instrumental para estimar o impacto causal, na mesma lógica de uma seleção aleatória com incumprimento (esta estratégia denomina-se de *fuzzy RDD*).

Na interpretação das estimações de um RDD é necessário ter em conta a possibilidade de o impacto estimado não ser generalizável a todas as unidades beneficiárias. Em particular, as unidades cujo valor da variável de seleção (por exemplo, a classificação de mérito do projeto) é distante do ponto de corte poderão ser muito diferentes das unidades com valores mais próximos do ponto de corte (para as quais a estimação do impacto é válida). Em geral, o impacto estimado deve ser interpretado apenas como o efeito esperado da política para as unidades próximas do ponto de corte (o que pode ser uma limitação muito relevante para estudos que queiram determinar o impacto global de uma política, mas pode ser um bom indicador quando se quer determinar se o âmbito de uma política deve ser expandido ou reduzido).

As principais limitações ao uso de métodos de RDD para inferência de impactos de apoios empresariais são: a inexistência de informação sobre as empresas candidatas não selecionadas (o único grupo de controlo possível), o baixo número de empresas candidatas (o que torna o número de empresas próximas do ponto de corte bastante reduzido) ou até a inexistência de empresas candidatas e elegíveis para o apoio que não são apoiadas (devido, por exemplo, a uma elevada disponibilidade orçamental).

EXEMPLO DE AVALIAÇÃO:

PERFORMANCE EMPRESARIAL E FINANCIAMENTO PÚBLICO DA INOVAÇÃO: EVIDÊNCIA DO PROGRAMA INNOFUND CHINÊS (WANG ET AL., 2017)

O Innofund é um fundo de inovação para pequenas e médias empresas de base tecnológica, iniciado em 1999. É a maior fonte de apoio público à inovação para jovens empresas na China. O objetivo central do programa é financiar ideias de inovação em estado inicial que são promissoras, mas demasiado arriscadas para investidores privados.

Para que uma empresa possa beneficiar do apoio financeiro do Innofund é necessário que apresente uma candidatura com o projeto que pretende financiar. Numa segunda fase, as candidaturas são avaliadas por especialistas externos que atribuem uma pontuação técnica ao projeto e uma pontuação sobre a sustentabilidade financeira da empresa. As duas pontuações são agregadas numa única pontuação global da candidatura e essas pontuações são normalizadas a nível nacional. Os projetos são financiados de acordo com a classificação global da candidatura, sendo apoiados projetos em ordem decrescente até esgotar o orçamento nacional pré-estabelecido. Devido a este processo de seleção é criado organicamente um ponto de corte na classificação global dos projetos que permite utilizar a metodologia do RDD.

No entanto, existem exceções ao processo normal de avaliação: o Innofund reserva o direito de vetar a decisão de financiar uma empresa se for detetada manipulação de informação na candidatura, se a propriedade intelectual da sua tecnologia estiver sob disputa ou se houver suspeita de que o seu produto ou produção podem causar danos ambientais graves. Se a classificação financeira da empresa for menor que 60 (de 0 a 100), esta é excluída da possibilidade de receber apoio, independentemente da classificação técnica do seu projeto. Devido a estas exceções é necessário adotar um *fuzzy* RDD, visto que há projetos com classificação superior ao ponto de corte que não são apoiados.

Os autores testam três hipóteses essenciais para a validade do RDD: 1) a probabilidade de receber financiamento é diferente entre empresas acima e abaixo do ponto de corte; 2) as características das empresas próximas do ponto de corte são semelhantes; 3) as empresas não podem conseguir manipular o sucesso das suas candidaturas.

Com a aplicação do *fuzzy* RDD, não é encontrada evidência de que o Innofund melhorou os resultados das empresas apoiadas na amostra considerada. Apesar de as empresas apoiadas terem melhores resultados do que as não apoiadas, a diferença deve-se ao facto de o Innofund selecionar as melhores empresas. Quando se considera apenas as empresas próximas ao ponto de corte, não são encontradas diferenças significativas nos resultados pós-tratamento.

3.4. Diferença-em-diferenças

Em algumas circunstâncias pode ser adequado utilizar o método da diferença-em-diferenças (DID) para estimar o impacto causal. Neste caso é feita uma comparação entre o resultado antes e depois do tratamento no grupo de tratamento e no grupo de controlo (1^{as} diferenças). O impacto causal estimado será a diferença entre as 1^{as} diferenças do grupo de tratamento e controlo. Essencialmente, em vez de comparar os valores dos resultados no período pós-tratamento, compara-se a variação dos resultados verificada a partir do início do tratamento.

A comparação de variações é considerada um método em si mesmo porque apresenta duas grandes vantagens: 1) permite isolar o impacto do tratamento das diferenças pré-existentes entre o grupo de tratamento e de controlo; 2) permite isolar o impacto do tratamento de variações causadas por fatores temporais que afetem de forma semelhante o grupo de tratamento e de controlo.

O principal pressuposto inerente à utilização do método DID é que a variação da variável de resultado entre períodos seria semelhante no grupo de tratamento e no grupo de controlo na ausência do tratamento (a hipótese das tendências paralelas). Esta hipótese é demasiado exigente na maioria dos casos, mas pode ser aceitável em cenários onde haja fortes razões teóricas para a sustentar e quando existem dados sobre a variável de resultado no período pré-tratamento que corroborarem empiricamente aquele pressuposto.

Na avaliação de políticas de apoio a empresas, o método DID raramente é utilizado isoladamente, mas sendo frequente a sua utilização como complemento de outros métodos, sobretudo do emparelhamento (ver secção seguinte). A lógica subjacente ao DID serve também de base a um método recente que tem maiores possibilidades de aplicação: o método de controlos sintéticos (secção 3.6).

EXEMPLO DE AVALIAÇÃO:

EFEITOS DO APOIO PÚBLICO À I&D AO NÍVEL DA PRODUTIVIDADE EMPRESARIAL E EXISTÊNCIA DE *SPILLOVERS* (CRESPI ET AL., 2020)

Este artigo estima os efeitos diretos e indiretos de dois esquemas de subvenções destinados a promover o investimento em investigação e desenvolvimento (I&D) no Chile, ao nível da produtividade das empresas. Os dois programas são o Programa Nacional de Produtividade e Desenvolvimento Tecnológico (FONTEC), que subsidia I&D interna, e o Fundo de Desenvolvimento Científico e Tecnológico (FONDEF), que financia I&D externa realizada em colaboração com institutos de pesquisa. A análise de ambos os programas permite clarificar o processo de criação e difusão do conhecimento.

O conjunto de dados é obtido através da junção de várias rondas dos Inquéritos Nacionais de Indústria Chilena, que são recolhidos pelo Instituto Nacional de Estatística com dados que contêm informações sobre os beneficiários de ambos os programas. O conjunto de dados final é um painel de 17 anos cobrindo quase 9.000 empresas e 600 beneficiários.

Os autores utilizam o método da diferença-em-diferenças para estimar os efeitos da participação nos programas ao nível da produtividade da empresa (a variável de resultado é a variação logarítmica da produtividade). A metodologia permite controlar o efeito de seleção decorrente de efeitos

fixos ao nível da empresa e efeitos anuais globais e ao nível do setor e da região. A principal inovação deste artigo é que também são incorporados termos de *spillover* no modelo de estimação, permitindo controlar o enviesamento que a estimação do efeito direto teria caso fossem ignorados e, ao mesmo tempo, retirar conclusões sobre a existência ou não de *spillovers* positivos. São considerados possíveis *spillovers* regionais e setoriais.

A hipótese de tendências paralelas é testada empiricamente, concluindo-se que as diferenças entre o grupo de empresas tratadas e o grupo de empresas de controlo escolhido não é significativamente diferente no pré-tratamento. No entanto, a performance das empresas tratadas é significativamente melhor que a das empresas não tratadas.

Os resultados sugerem que, embora ambos os programas tenham tido um impacto positivo na produtividade dos participantes, apenas os projetos financiados pelo FONDEF geraram repercussões positivas na produtividade das empresas parceiras. A análise revela que os efeitos de *spillover* sobre a produtividade apresentam uma relação em U invertido com a intensidade do apoio público. Concluiu-se que os efeitos de *spillover* apenas ocorreram quando as empresas são geográfica e tecnologicamente próximas.

3.5. Equilíbrio de distribuições por emparelhamento (matching)

Os métodos apresentados até agora são aplicáveis em situações de seleção muito específicas (experiências aleatórias e RDD) ou em situações onde são verificadas certas particularidades (existência de uma boa variável instrumental ou hipótese de tendências paralelas). Os métodos que apresentaremos de seguida abrem a possibilidade de realizar inferências causais em situações onde nenhum dos métodos anteriores é aplicável.

Os métodos de equilíbrio de distribuições focam-se em construir, após a implementação da política, grupos de tratamento e controlo que eram o mais semelhantes possível entre si no período pré-tratamento. A forma mais simples de conseguir isto é selecionar, para cada unidade tratada, a unidade de controlo disponível com as características mais semelhantes – emparelhamento (ou *matching*, em inglês). No contexto de avaliação de políticas de apoio às empresas, isto implicaria selecionar, para cada empresa que participou no programa em estudo, a empresa não participante com níveis semelhantes (no pré-tratamento) de todas as características que se suspeite poderem influenciar os resultados para os quais se pretende estimar o impacto causal (por exemplo, volume de negócios, número de trabalhadores, produtividade, lucro, endividamento, setor de atividade, entre outras).

Características observáveis: informação sobre uma unidade que pode ser recolhida e adequadamente representada numa variável (por exemplo, o número de trabalhadores de uma empresa é uma característica observável, mas a motivação dos trabalhadores de uma empresa é uma característica não observável). Num contexto prático, as características observáveis são aquelas que são recolhidas e representadas sob a forma de variáveis.

Variáveis de controlo: são as variáveis usadas para controlar (i.e., eliminar) o efeito de seleção, permitindo obter uma boa estimativa do efeito causal do tratamento. A forma como estas variáveis são utilizadas depende da metodologia de cada avaliação.

É importante ressaltar que qualquer método de emparelhamento tem uma limitação adicional em relação aos métodos apresentados anteriormente: é necessário assumir que as diferenças relevantes entre os grupos de tratamento e de controlo são captadas por características observáveis. Isto implica também que é necessário recolher informação não só sobre a variável de resultado, mas também sobre todas as outras características que podem influenciar a participação no tratamento ou influenciar os resultados (usualmente denominadas variáveis de controlo). É por isso preferível explorar a possibilidade de utilizar algum dos métodos expostos anteriormente e apenas optar pelas técnicas de emparelhamento se nenhum outro for viável.

O emparelhamento permite selecionar uma amostra onde o grupo de tratamento e o grupo de controlo são estatisticamente equivalentes, aproximando-se do ideal das experiências aleatórias. No entanto, contrariamente a uma experiência aleatória, onde a semelhança entre grupos é inerente ao processo de seleção, o emparelhamento apenas permite assegurar que as características observáveis são estatisticamente equivalentes. Esta é uma grande desvantagem, porque podem existir características para as quais não são recolhidos dados (ou é impossível recolher dados) e que podem influenciar os resultados observados. Neste contexto, não é possível saber se os grupos criados pelo emparelhamento são semelhantes nas características não observadas, o que pode levar ao enviesamento da estimação do impacto causal.

3.5.1. A pontuação de propensão (propensity score)

Uma questão central que se coloca no emparelhamento entre unidades tratadas e não tratadas é a definição da medida de semelhança entre elas. Como a quantidade de variáveis para as quais se quer assegurar que os grupos de comparação são estatisticamente equivalentes pode ser muito elevada, é necessário recorrer a técnicas de redução de dimensionalidade. A mais comum consiste em estimar uma pontuação de propensão (*propensity score*, ou PS, em inglês) que permite reduzir a informação sobre as diferenças relevantes para a probabilidade de tratamento numa única variável. O PS será um valor entre 0 e 1 que representa a probabilidade estimada de cada unidade participar no tratamento, baseada nas suas características pré-tratamento. Estimando o PS para todas unidades, podemos facilmente emparelhar as unidades tratadas e as unidades não tratadas com os valores de PS mais próximos para construir os grupos de tratamento e controlo. Desta forma, duas empresas – uma apoiada e outra não apoiada – serão emparelhadas se tiverem uma propensão para serem tratadas idêntica ou muito próxima.⁴

⁴ Intuitivamente, duas empresas com PS semelhantes têm a mesma probabilidade estimada de receber o tratamento. No entanto, isto não garante uma boa semelhança nas características entre qualquer par de empresas com PS semelhantes, porque os fatores que determinam a mesma probabilidade podem ser diferentes (por exemplo, se o investimento e produtividade estiverem positivamente correlacionadas com a atribuição do tratamento, uma empresa pode ter uma elevada probabilidade estimada de receber o tratamento porque realizou muito investimento, enquanto outra tem também uma probabilidade estimada semelhante, mas porque tem uma produtividade muito elevada). A vantagem decisiva do PS é que, para grupos construídos através do emparelhamento de empresas com a mesma probabilidade de tratamento, as distribuições das características dos grupos vão ser estatisticamente equilibradas, permitindo uma estimação válida dos efeitos médios.

Existem diversas formas de estimar o PS e de realizar o emparelhamento (incluindo também métodos de emparelhamento que não utilizam o PS). Na prática, a maioria dos estudos implementa várias dessas estratégias, como forma de aferir a robustez das conclusões. A decisão sobre qual das estratégias de emparelhamento implementar dependerá de cada caso particular. No entanto, o critério mais consensual é que a melhor forma será aquela que alcance um melhor equilíbrio entre as características do grupo de tratamento e do grupo de controlo (i.e., a validade interna). Outros critérios como a preservação do número de unidades e a semelhança dos grupos construídos com a amostra inicial também devem ser considerados, visto serem essenciais para aumentar a precisão estatística da estimação e a validade externa do estudo, respetivamente.

Outra estratégia comum de uso do PS é a estimação por ponderação. Neste caso, a estimativa do impacto do tratamento é a média ponderada das empresas tratadas, comparada com a média ponderada das empresas de controlo. A ponderação tem por base os valores de PS estimados. Ao contrário do emparelhamento, que acaba por ficar limitado a estimar o impacto local da amostra emparelhada (porque apenas as observações consideradas bons pares são mantidas, eliminando algumas unidades tratadas e unidades de controlo), a ponderação estima o impacto médio do tratamento para a totalidade da amostra. No entanto, é frequente que alguns PS estimados tomem valores extremos (muito próximos de 0 ou 1), o que leva à necessidade de os eliminar (fazer *trimming*, em inglês), voltando a um problema semelhante ao do emparelhamento.

Validade interna: depende da semelhança pré-tratamento entre o grupo de tratamento e o grupo de controlo – se os grupos forem semelhantes, então as estimações do efeito causal são válidas na amostra estudada.

Validade externa: depende da semelhança pré-tratamento entre o grupo onde se estimaram os efeitos (o conjunto do grupo de tratamento e controlo – a amostra estudada) e qualquer grupo para o qual se queira generalizar os efeitos. Por exemplo, se se realizar um estudo com validade interna sobre uma política que apoia pequenas e médias empresas industriais, é muito improvável que as conclusões sejam generalizáveis ao conjunto de todas as empresas portuguesas (porque o conjunto de empresas estudadas é muito diferente do conjunto de todas as empresas portuguesas e por isso não seria correto retirar qualquer conclusão sobre os efeitos que a política teria se fosse alargada a todas as empresas portuguesas).

Uma forma possível de melhorar a performance dos métodos de ajustamento que permitem o equilíbrio de distribuições é combiná-los com métodos de regressão. São chamados de métodos duplamente robustos e garantem consistência assintótica mesmo que um dos modelos esteja errado, sendo por isso sempre preferíveis. Ou seja, diferenças entre os grupos que não sejam ajustadas pelo ajustamento podem vir a ser controladas na regressão e vice-versa. Existem métodos que combinam as duas abordagens tanto em emparelhamento como em ponderação

3.5.2. Limitações gerais dos métodos de equilíbrio de distribuições

Como já foi referido, a principal limitação ao equilíbrio de distribuições (seja por emparelhamento ou ponderação) é que apenas é possível garantir a semelhança do grupo de tratamento e controlo nas características observáveis (*balance*). No entanto, é necessário também estabelecer quais as características para as quais é importante garantir a semelhança. Esta decisão deve estar dependente do método de seleção de cada política e das variáveis de resultado consideradas, mas acaba na maioria das vezes por ser arbitrária e pode influenciar decisivamente as conclusões do estudo. É também possível que em alguns casos

Balance: diz-se que existe balance quando as distribuições das características observáveis que interessa controlar são semelhantes no grupo de tratamento e controlo (i.e., quando as amostras estão balanceadas) – ou seja, quando existe validade interna. O objetivo dos métodos de emparelhamento e ponderação é gerar balance.

não existam unidades não tratadas suficientemente semelhantes a algumas unidades tratadas (nestes casos diz-se que o ‘suporte comum’ é diminuto), o que implica restringir a estimação apenas a uma parte das unidades. Se o número de unidades excluídas for elevado, as conclusões podem acabar por não ser generalizáveis e/ou a precisão da estimação pode ser reduzida.

EXEMPLO DE AVALIAÇÃO:

**IMPORTÂNCIA DO APOIO REGIONAL EUROPEU PARA A PERFORMANCE EMPRESARIAL
(BENKOVSKIS ET AL., 2019)**

Este artigo investiga o efeito da política regional da União Europeia na performance das empresas apoiadas na Letónia. São considerados os projetos apoiados pelo Fundo Europeu de Desenvolvimento Regional (FEDER) no âmbito do quadro financeiro plurianual de 2007 -2013. Este fundo procura alcançar três objetivos globais: 1) convergência de padrões de vida; 2) reforço da competitividade regional e do emprego; e 3) promoção da cooperação territorial europeia.

Os autores utilizam uma base de dados agregada com informação sobre mais de 88.000 empresas letãs. Depois de realizado um tratamento de dados (excluindo *outliers*, empresas com dados em falta e empresas que participaram noutras políticas europeias), a amostra fica reduzida a cerca de 10.000 a 15.000 empresas por ano, entre as quais 480 são empresas apoiadas pelo FEDER no período considerado.

Como as empresas apoiadas são muito diferentes da restante população, é utilizada a metodologia do PS para construir um grupo de controlo com empresas comparáveis. O PS foi estimado a partir de diversas variáveis sobre as características das empresas, como a produtividade, idade, emprego, volume de negócios, exportações, endividamento, entre outros.

Foram implementadas várias estratégias de emparelhamento após a estimação do PS, sendo selecionadas entre 435 e 1985 empresas para o grupo de controlo, dependendo da especificação do emparelhamento.

Na estimação do efeito foi utilizada uma especificação DID (ou seja, o efeito estimado corresponde à diferença das variações das variáveis de resultado). O período de análise inclui o ano de tratamento e os dois anos subsequentes.

Os resultados indicam que o FEDER teve um impacto positivo imediato no nível de emprego, volume de negócios e na intensidade de capital das empresas apoiadas. Na produtividade, o impacto positivo apenas é considerado estatisticamente significativo a partir do segundo ano. Uma análise de heterogeneidade mostra que as empresas mais beneficiadas pela participação no programa são aquelas com menores níveis iniciais de produtividade, mais endividadas e com menor intensidade de capital.

3.6. Controlo Sintético

O método do controlo sintético é uma derivação do método da diferença-em-diferenças, aplicado sobretudo em intervenções que afetam poucas unidades agregadas (como distritos, regiões ou países). A particularidade deste método é que, em vez de se comparar as variações em grupos de tratamento e controlo pré-existentes, é criada uma unidade de controlo sintética que serve de contrafactual para cada unidade de tratamento.

O controlo sintético é construído a partir de unidades de natureza semelhante à unidade tratada (i.e., se a unidade tratada é um distrito onde foi implementada uma certa política, o controlo sintético será construído a partir de outros distritos onde a política não foi implementada). O procedimento-base para a construção de um controlo sintético é a minimização das diferenças na variação da variável de resultado no período pré-tratamento entre a unidade tratada e a unidade de controlo sintética. Essencialmente, o controlo sintético é uma média ponderada de unidades não tratadas que, devido ao algoritmo de ponderação, segue uma tendência paralela à da unidade tratada. Ou seja, o método do controlo sintético é a construção artificial de uma unidade de controlo que, por cumprir o pressuposto das tendências paralelas, é comparável à unidade tratada. O impacto causal do tratamento é estimado a partir da diferença da variação da variável de resultado entre a unidade tratada e a unidade sintética, no período pós-tratamento. A principal vantagem do controlo sintético é a possibilidade de obter estimativas de impacto quantitativas em circunstâncias onde existem poucas unidades tratadas. Por ser um método relativamente recente, existem poucas aplicações em avaliações de políticas de apoio às empresas. No entanto, pode ser uma grande vantagem em situações onde se queira avaliar o impacto de uma política regional ou o impacto de projetos específicos para os quais resultados médios não são adequados.

EXEMPLO DE AVALIAÇÃO:

O EFEITO CAUSAL DE POLÍTICAS INDUSTRIAIS NO EMPREGO: ABORDAGEM DO CONTROLO SINTÉTICO (CASTILLO ET AL., 2017)

Neste artigo é utilizado o método do controlo sintético para analisar o impacto de longo prazo da Política de Desenvolvimento do Turismo (TDP) implementada pela província de Salta, na Argentina. TDP é o nome dado a um conjunto de políticas e intervenções públicas implementadas pelo governo regional com o objetivo de apoiar a expansão do turismo. Esta metodologia é escolhida por ser uma das únicas que permite estimar os efeitos de políticas regionais, quando existe apenas uma unidade (a região) tratada.

O método do controlo sintético é utilizado para estimar a situação contrafactual da região de Salta, construindo um controlo sintético que consiste na combinação ponderada de outras regiões da Argentina. Esta “Salta sintética” é construída por forma a que o nível de emprego mensal do setor do turismo seja o mais semelhante possível à do registado na Salta real no período pré-tratamento.

São utilizados dados mensais sobre as regiões argentinas no período de 1996-2013. Como a política começou em junho de 2003, existem 89 períodos (meses) de pré-tratamento. Para garantir a fiabilidade dos controlos como representação do contrafactual, são excluídas à partida regiões que implementaram também outras políticas regionais direcionadas ao turismo.

O estudo conclui que a política aumentou o emprego do setor de hotelaria na região apoiada de forma sustentada ao longo de 10 anos. O aumento equivale a um nível de emprego anual médio 11% superior ao que seria caso a política não tivesse existido, traduzindo-se em 1.376 empregos adicionais ao longo do período. É também comprovado que este aumento não aconteceu à custa de outros setores económicos na região e que a política gerou efeitos de *spillover* positivos noutras indústrias da região, traduzindo-se na criação de 2.750 empregos adicionais quando considerados estes efeitos indiretos.

3.7. Possíveis contributos da aprendizagem automática

Existe um grande potencial no cruzamento de métodos tradicionais de avaliação contrafactual com metodologias de aprendizagem automática (*machine learning*, ou ML, em inglês).

A aprendizagem automática dedica-se ao desenvolvimento e compreensão de sistemas computacionais que aprendem a desempenhar uma dada tarefa sozinhos. Um sistema de aprendizagem automática substitui o processo iterativo humano de encontrar a melhor solução para determinado problema.

Assim, os algoritmos de aprendizagem automática permitem uma maior independência face à discricionariedade do utilizador, nomeadamente em situações com muitas variáveis em que não é claro quais devam ser utilizadas nem como. Este é um problema frequente nas abordagens estatísticas à avaliação contrafactual de impactos. Por exemplo, nas abordagens baseadas no equilíbrio entre distribuições (por emparelhamento ou ponderação) é necessário decidir quais as variáveis de controlo a incluir nos modelos, não havendo em muito casos critérios robustos para orientar essa decisão. Como veremos, o recurso a algoritmos ML pode contribuir para apoiar essa e outras escolhas relevantes.

Nos anos recentes, foram publicados vários artigos científicos com muitas propostas teóricas, mas não encontramos nenhum artigo empírico que utilize metodologias de aprendizagem automática na avaliação contrafactual de políticas de apoio direto às empresas (esta observação é resultado de uma revisão de literatura exaustiva sobre este tipo de avaliações no período 2017-2021). Esta área de política é, no entanto, uma das que apresentam maior potencial para o cruzamento entre avaliação contrafactual e ML, devido ao elevado volume de dados.

Nos pontos que se seguem apresentamos alguns dos potenciais contributos que consideramos mais relevantes neste âmbito.

3.7.1. Na escolha de variáveis de controlo

Os métodos de emparelhamento visam identificar grupos de tratamento e de controlo que apresentem semelhança entre as distribuições de todas as variáveis que influenciam o resultado, de forma a maximizar a eficiência da estimação. Se tudo o que explica o resultado for semelhante entre os grupos, a variância da diferença de resultados corresponde apenas à variância do efeito de tratamento. Caso contrário, a estimação incorpora variância decorrente de diferenças de resultados que não se

devem ao efeito de tratamento, mas sim a diferenças nas características das amostras que ficaram por ser explicadas. Contudo, para obter estimativas consistentes basta apenas obter semelhança nas variáveis que simultaneamente afetam o resultado e a probabilidade de tratamento.

Parte da dificuldade em utilizar métodos que recorrem ao controlo de características observáveis passa pela escolha das variáveis de controlo a incluir e do modo de o fazer (i.e., a escolha das formas funcionais dos modelos). É comum essa escolha ser informada pela teoria e ser seguida de uma análise de sensibilidade estatística que demonstre que a estimação do impacto causal é pouco sensível a mudanças nas variáveis de controlo utilizadas ou nas formas funcionais.

A aprendizagem automática pode ajudar a guiar esta escolha e torná-la menos discricionária. Um dos métodos comuns de o fazer é *post selection*, em que um modelo é estimado recorrendo às variáveis e formas funcionais resultantes de uma regressão regularizada tipo LASSO. Numa regressão linear comum, o objetivo é minimizar a soma dos quadrados dos resíduos (erros de previsão do modelo). Numa regressão regularizada tipo LASSO, adicionalmente esse objetivo, procura-se o número ótimo de variáveis que o permite fazê-lo. Contudo, há estudos (e.g., Belloni et al., 2014) que mostram que tal abordagem pode introduzir enviesamentos substanciais.

A eliminação de certas variáveis imposta pela regularização tipo LASSO conduz a problemas de variáveis omitidas quando só se tenta ajustar o modelo de resultado. O problema com o ajustamento por LASSO resulta da existência de diferenças substanciais entre tratadas e controlos em variáveis cuja importância no modelo de resultado é reduzida e que, por isso, são excluídas pela penalização LASSO. Para lidar com este problema, utiliza-se a união de dois conjuntos de variáveis, um decorrente da seleção do LASSO para a variável de resultado, e o outro decorrente de uma regressão logística LASSO para a variável de tratamento.

3.7.2. Na modelação do propensity score

Como vimos, o PS é uma estratégia de emparelhamento, em que se procura emparelhar unidades tratadas e controlos com base na sua semelhança em termos de probabilidade de ser tratada. O PS é geralmente procurado de forma iterativa e avaliado pela sua capacidade de gerar equilíbrio entre as distribuições das variáveis relevantes dos grupos de unidades tratadas e não tratadas. Tal processo pode também beneficiar das vantagens da aprendizagem automática anteriormente elencadas. É aconselhada a utilização de algoritmos que façam seleção de variáveis, como o LASSO, para a escolha de formas funcionais para o PS (Imbens, 2014).

O emparelhamento também pode ser utilizado sem o uso de PS, diretamente nas variáveis. Contudo, existem sempre *trade offs* em termos de equilíbrio entre variáveis – isto é, ajustar melhor a semelhança numa variável implica piorar a semelhança noutra. Por vezes é aconselhável relaxar o emparelhamento em certas variáveis de forma a conseguir melhor equilíbrio noutras. No emparelhamento feito diretamente nas variáveis estas são todas consideradas como igualmente importantes, tendo o mesmo peso. A aprendizagem automática pode contribuir para melhorar os métodos de emparelhamento através da descoberta dos pesos relativos que cada variável deve ter para assegurar o melhor equilíbrio entre unidades tratadas e controlos.

A literatura também propõem o uso de aprendizagem automática para a estimação do PS e construção dos pesos (Lee et al., 2010), refletindo sobre as vantagens da aprendizagem automática já

discutidas anteriormente e dando conta de que estes pesos tendem a ser mais estáveis. Contudo, a abordagem continua a ser iterativa, devendo-se procurar a estratégia de estimação que conduza ao melhor equilíbrio das amostras.

Outros métodos focam-se apenas na geração de equilíbrio entre amostras, deixando de parte o PS. É o caso dos *stable balancing weights* (Zubizarreta, 2015) em que os pesos são encontrados diretamente através da resolução de um problema de otimização, sendo o objetivo equilibrar as médias das duas amostras. É possível utilizar tal método para equilibrar as distribuições marginais e conjuntas das variáveis entre as duas amostras através da utilização de uma matriz expandida que considere variáveis centradas, quadrados e quantis, permitindo equilibrar não só as médias, mas também as variâncias, as correlações e distribuição empírica inteira das duas amostras.

Quando o número de variáveis é muito elevado torna-se cada vez mais difícil obter bons equilíbrios em todas as variáveis, reduzindo a eficácia desta estratégia. Alguns autores (Athey et al., 2018) sugerem que o uso de aprendizagem automática em combinação com métodos de equilíbrio de amostras, mesmo sendo este equilíbrio apenas aproximado, é suficiente para eliminar os enviesamentos da estimação. Tal abordagem permite o uso de muitas variáveis e métodos flexíveis de estimação independentes do utilizador, mesmo quando as amostras são pequenas.

3.7.3. Como método de identificação de heterogeneidade

Uma grande limitação dos métodos de avaliação contrafactual tradicionais é o facto de se focarem nas condições para obter boas estimações do efeito médio do tratamento. No entanto, muitas vezes a estimação de efeitos médios pode ser pouco interessante. Mais interessante é saber para que unidades o tratamento é mais positivo e para quais o tratamento é pouco útil. Ou seja, estimar a heterogeneidade do tratamento.

Comumente, a heterogeneidade é identificada através da partição da amostra inicial em subamostras definidas *a priori* como hipótese de estas apresentarem efeitos de tratamento diferentes entre si. A tentativa de procurar subamostras iterativamente, sem as definir à partida, já no processo de estimação, leva à caça de fenómenos que podem ser espúrios, num processo apelidado de “*data snooping/P-hacking*”.

De forma a colmatar este problema, é proposto um método de aprendizagem automática baseado em árvores de regressão, apelidado de “árvores causais” (Athey & Imbens, 2016), para identificar subgrupos com efeitos de tratamento mais semelhantes entre si do que entre grupos, sendo os grupos divididos nas folhas terminais da árvore. As regras que esta árvore procura encontrar são aquelas que maximizam a variância de efeitos de tratamento entre folhas. A estratégia utilizada para evitar sobreajustamento e “*data snooping/P-hacking*” é nunca estimar os efeitos de tratamento na mesma amostra que é utilizada para definir a separação entre os grupos. Ou seja, a amostra inicial é dividida em duas de forma aleatória. Numa dessas amostras, são identificados os subgrupos. Noutra são estimados os efeitos de tratamento para esses subgrupos. A divisão de amostras assegura a capacidade de extrapolação dos subgrupos criados, sendo por isso “honesta”. O

Heterogeneidade: a heterogeneidade do tratamento significa que para unidades diferentes, o tratamento tem impactos diferentes (que dependem das características de cada unidade).

Por exemplo, uma política de apoio ao investimento por meio de empréstimos bonificados pode aumentar bastante o nível de investimento de uma pequena empresa constrangida financeiramente, mas não ter qualquer efeito se direcionada a uma grande empresa que tem acesso a uma grande variedade de fontes de investimento privado. rentável.

problema desta estratégia é que metade dos dados são desperdiçados devido à separação que tem de ser feita. Esse problema é resolvido na extensão desta abordagem a *random forests*, apelidada de “causal forests” (Athey et al., 2019), em que a amostra é dividida em várias iterações, utilizando *bootstrapping*, acabando todas as observações por ser utilizadas tanto na amostra que serve para a partição como na amostra utilizada na estimação. A vantagem da abordagem de árvore simples é ser facilmente interpretável, sendo fácil caracterizar os diferentes grupos e as regras que permitem dividi-los tal que o efeito de tratamento difira entre eles. A vantagem da abordagem baseada em *random forests* é ser mais flexível e não desperdiçar dados, permitindo previsões de efeitos de tratamento individuais mais granulares. Com recurso a estas previsões, é possível dizer à partida, dadas as características de certa unidade, qual o tratamento com melhores resultados esperados.

4. Causalidade, avaliação contrafactual e teorias da mudança

Na génese da adoção de um método de avaliação estão latentes diferentes visões da forma como o mundo opera e de como um investigador poderá interpretá-lo. Por outras palavras, diferentes abordagens de causalidade. Neste capítulo discutimos e contrastamos várias abordagens à causalidade, relacionando-as com a avaliação contrafactual de impactos e com a teoria de mudança.

4.1. Diferentes abordagens à causalidade

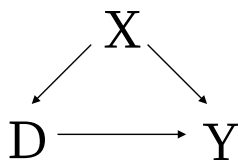
4.1.1. Contrafactual

A abordagem à causalidade que domina o paradigma de avaliação de políticas públicas é a *contrafactual*. Baseado no Método da Diferença de Mill, a abordagem contrafactual consiste em comparar duas instâncias que são idênticas em todas as dimensões exceto numa - a condição cujo efeito queremos aferir (Gates & Dyson, 2017; Stern et al., 2012). No contexto das políticas públicas, se uma unidade tratada é idêntica a outra em todas as dimensões, salvo na questão do tratamento, a discrepância no efeito observado é inteiramente atribuível à intervenção.

A aplicação desta metodologia pressupõe que as duas observações sejam idênticas numa infinidade de condições, menos na existência/ausência de política. Porém, os critérios de seleção usados pelo decisor político para definir quais as unidades passíveis de tratamento, bem como a condição de autoseleção para o tratamento por parte das unidades beneficiárias, perturbam este pressuposto. O facto de uma empresa ser (auto-)selecionada, indica que existem dimensões, porventura não-observáveis, em que esta tende a divergir das empresas não tratadas.

Na linguagem dos GAD (gráficos acíclicos diretos – ver Cunningham, 2021:118-129), os métodos (quase-)experimentais pretendem isolar as variáveis que influenciam, tanto a seleção ao tratamento como as variáveis de resultado. Observe-se a Figura 3.

Figura 3: GAD de tratamento com variáveis de confundibilidade



O efeito que uma avaliação pretende reportar é aquele encontrado entre D e Y , sendo que D é uma variável que representa o tratamento (e.g. que é igual a 1, se a empresa for abrangida pela política e 0, se o contrário se verificar) e Y uma variável de impacto, em que deverão ser capturadas as alterações provocadas pela política. X representa as variáveis que estão relacionadas, causalmente, com D e Y (ou seja, que influenciam tanto o tratamento como o resultado esperado do tratamento). Se estas variáveis não forem consideradas, o efeito que pretendemos identificar estará contaminado, visto que

ao invés de capturar, exclusivamente, a relação direta entre D e Y , também capturará a relação indireta e, neste caso, espúria ($D \leftarrow X \rightarrow Y$).

Posto isto, a preferência por experiências aleatórias (ver secção 3.1) torna-se evidente, dado que estas garantem que o processo de seleção é independente das características das unidades selecionadas (não existindo, portanto, variáveis X que influenciem tanto D como Y).⁵

Se o critério de seleção não é aleatório, é expectável que este esteja relacionado com a variável de resultado. Os métodos quase-experimentais (ou seja, abordagens contrafactuais que não se baseiam em experiências aleatórias) são julgados pela sua capacidade de se aproximarem à seleção aleatória. Ao considerarem as variáveis X , que influenciam tanto D como Y , estão a isolar o sistema de contaminação externa, restando apenas o efeito causal da política.

Aqueles que advogam o modelo causal de Pearl (e.g. Pearl, 2000) defendem que os métodos (quase-) experimentais necessitam do uso de teoria para conseguir identificar efeitos causais. Neste caso, o uso de teoria não pretende elucidar os mecanismos específicos através dos quais a política conduz a determinados resultados, apenas identificar quais as variáveis que estão a contaminar a relação estatística encontrada entre D e Y . No âmbito desta abordagem, o efeito da política é aferido com base na variação da variável de tratamento, mantendo tudo o resto contante (*ceteris paribus*).

4.1.2. Configuracional

A abordagem *configuracional*, como o nome indica, encontra causalidade em configurações de diferentes condições. Uma causa não é ativada num vácuo. São combinações de condições que produzem efeitos, ao invés de causas únicas (Ragin, 1987). É de notar que esta abordagem não defende que um efeito advém de múltiplas variáveis, como a noção de *ceteris paribus* indicaria, fulcral na perspectiva contrafactual. Uma variável, por si só, não produz um efeito marginal. São as combinações e interações de diferentes condições que produzem resultados.

Enquanto os métodos guiados pela abordagem contrafactual pretendem isolar os efeitos externos, esta abordagem defende que são estas condições ditas externas, quando combinadas com a política, que levam a efeitos. O trabalho do avaliador é, então, investigar que condições, quando combinadas com a política, permitem obter resultados desejados.

A análise de contribuição é uma das estratégias mais proeminentes no âmbito da avaliação baseada na teoria. Esta abordagem assume que as políticas não são condições suficientes, nem necessárias, a um determinado impacto, focando-se menos na atribuição de um efeito à intervenção do que na identificação de contribuições, onde os efeitos da política não ocorrem isoladamente, mas sim inseridos num pacote causal. Mayne (2012) faz referência às condições insuficientes, mas necessárias a um pacote causal, que em si é desnecessário, mas suficiente para causar impacto. Mackie (1974) definiu condições insuficientes como as componentes necessárias de um pacote causal, que por si só é

⁵ Além disso, é introduzida variação exógena na variável de tratamento. Numa situação não-aleatória, a variação da variável de tratamento pode existir devido a condições de (auto-)seleção que não são observáveis, mas que na mesma apontam para diferenças intrínsecas entre as unidades empíricas (ver Heckman et al., 1997).

suficiente para causar um efeito. A observação daquelas condições isoladas não garante a existência de uma consequência, esta só é assegurada quando é observada a totalidade de um pacote causal.

Um exemplo frequentemente apresentado (e.g., Mayne, 2012) é um incêndio que decorre de um evento, uma faísca. Note-se que a mera observação da faísca não garante que haverá um incêndio. Apenas na presença de material inflamável é que a faísca levará a uma chama. Deste modo, a conjunção de ambas as condições revela-se suficiente para garantir a existência de um incêndio. Porém, esta não é necessária, visto que outras combinações de eventos podem levar ao mesmo resultado.

Identificar as condições insuficientes e o pacote causal adjacente é o objetivo central da análise de contribuição. O pacote causal considera fatores que suportam o impacto da medida. Existe, assim, um contraste com as abordagens baseadas numa lógica contrafactual – um efeito não existe isolado do seu contexto. Por sua vez, um método que identifica um coeficiente *ceteris paribus* não parece ser consistente com uma análise de contribuição.

A perspectiva configuracional ocupa, assim, um espaço entre a contrafactual e a generativa (ver parágrafos subsequentes). O nível de detalhe da explicação não é tão profundo quanto as abordagens generativas, mas os pressupostos ontológicos são mais realistas do que o contrafactual.

4.1.3. Generativa

Outra abordagem distingue-se das antecedentes: o modelo *generativo* de causalidade. Enquanto nas abordagens anteriores a inferência causal resulta da identificação de padrões empíricos, sejam estes causas únicas ou constelações de fatores, a abordagem generativa implica a identificação e descrição de mecanismos causais e a contingência contextual da sua ativação (Pawson, 2013).

Na abordagem generativa, tal como nas anteriores, o investigador começa pela identificação de padrões e uniformidades. Assume-se que estes resultam da ação de um ou mais mecanismos que não são diretamente observáveis, dependendo da deliberação dos agentes. O objetivo da avaliação segundo uma abordagem generativa é, portanto, a produção de teorias que explicam “a presença e ausência de uniformidade” (Pawson, 2006). Os resultados de uma investigação empírica só são compreendidos se forem interpretados através de uma descrição dos mecanismos e os contextos que permitem a sua ativação. Vários níveis são considerados nas descrições contextuais, relativamente: ao indivíduo, às relações entre diferentes agentes e às condições institucionais e infraestruturais.

Segundo uma abordagem realista – que assume uma lógica de causalidade generativa – a investigação científica não é realizada através de experimentação laboratorial. Apesar de ser uma parte fundamental do processo científico, a experimentação procede, sempre, uma investigação teórica dos mecanismos causais (Sharpe & Bhaskar, 1976). Einstein não descobriu a força da gravidade como distorção do espaço-tempo apenas através do eclipse lunar, pelo que este foi apenas a “manipulação” que permitiu a validação empírica da sua construção teórica.

Um mecanismo não implica, necessariamente, um resultado empírico, mas sim uma “tendência causal”, devido à possível existência de outros processos que exercem pressão na direção oposta (Hedström, 2006). Os mecanismos apontam para regularidade, mas não uniformidade empírica (Pawson, 2008). O resultado antecipado por um mecanismo é contingente a um dado contexto.

Por exemplo, numa política de apoio às empresas, o facto de os incentivos públicos levarem a uma possível redução dos custos de financiamento não é condição suficiente para que a empresa invista. Para isso, é preciso que este mecanismo leve a uma mudança de comportamento, por via do impacto na decisão dos agentes. A avaliação realista propõe que a análise de políticas vá além de inquirir se uma política funciona, de forma a compreender para quem é que esta funciona e em que circunstâncias.

Esta abordagem, dita “mecanística”, entende que a mera associação estatística não é suficiente para poder inferir causalidade, dado que o foco é nas “causas dos efeitos” ao invés dos “efeitos das causas”. O realismo introduzido por uma visão generativa, ao identificar mecanismos e a contingência da sua ativação, tem a possibilidade de ser transfactual, e realizar-se noutras circunstâncias (Sharpe & Bhaskar, 1976).

A avaliação realista, cujos autores frisam preocupações nas definições de casualidade (e.g. Pawson & Tilley, 1997), afirmando-se no campo generativo, pretende identificar configurações CMR (contextos, mecanismos e resultados) que apontam para regularidades. Voltando ao exemplo de apoios a empresas, apenas naquelas empresas que partem de contextos com dificuldades de financiamento é que o apoio público pode ser compreendido como um mecanismo de redução de custos, levando a um aumento de investimento. Naquelas empresas cujos contextos que inibem investimento são outros (e.g. condições de procura), este mecanismo não é ativado.

4.2. Causalidade e avaliação contrafactual de impactos

A avaliação baseada na teoria e os métodos (quase-)experimentais partem de concepções completamente distintas da realidade e de como relações causais podem ser inferidas. Como tal, não é evidente que existam sinergias que resultem da combinação das diferentes abordagens.

Os autores associados à avaliação baseada na teoria partem do pressuposto de que os métodos baseados numa lógica contrafactual pretendem identificar efeitos causais da política. As lógicas configuracionais e generativas, por contraste, rejeitam a possibilidade de atribuir causalidade a condições singulares, apontando para a contingência introduzida pelas dimensões contextuais e as volições dos agentes, que constituem uma dimensão determinante do sucesso de uma política. Desta forma, é inadequado justificar uma metodologia pela sua capacidade de isolar um efeito num sistema social. É importante realçar que as críticas que estas abordagens tecem à lógica contrafactual vão além do argumento segundo o qual correlação não implica causalidade.

Reportar apenas o efeito de uma intervenção pouco acrescenta a uma disciplina de avaliação que necessita de um processo cumulativo de aprendizagem sobre políticas, sendo importante centrar uma

avaliação num objetivo de compreender o como e o porquê de uma política induzir sucesso, ou seja, que se debruce sobre as “causas dos efeitos”.

4.2.1. Validade interna e externa

A metodologia de avaliação de políticas públicas tem sido dominada pelo paradigma contrafactual, cujo objetivo entende assegurar a validade interna das suas conclusões (ver subsecção 3.5.1). A preponderância deste objetivo, que remonta a Campbell et al. (1963), continua a ser afirmada pelas recentes contribuições nesta literatura (e.g. Imbens, 2013). O conceito de validade interna havia sido definido por Campbell com a seguinte questão: “...o estímulo experimental causou diferenças significativas nesta instância específica?” (Campbell, 1957). O foco é em apenas garantir que o efeito marginal do tratamento é devidamente identificado, portanto, a dimensão justificativa/explicativa deste tipo de análises cinge-se à ideia de que o efeito é devido ao tratamento em questão (e não a outras interferências). Nenhuma atenção é dada à compreensão de como os efeitos surgem e, como tal, a análise torna-se útil apenas para indicar quais as consequências diretamente atribuíveis à política.

No entanto, os “segredos” do desenvolvimento só podem ser descobertos através de análises que visem identificar mecanismos potencialmente generalizáveis e que sejam acompanhadas de explicações e descrições contextuais da sua ativação (Deaton, 2010). Consideremos, então a noção de validade externa, também defendida por Campbell (1957): “a que populações, contextos e variáveis pode este efeito ser generalizado?”.

Segundo Pawson (2008) uma análise que pertença a um nível de causalidade generativa deixa explícito aquelas dimensões mecanísticas que são deixadas implícitas pelos níveis configuracionais e contrafactuais. Uma descrição generativa acontece a níveis mais “profundos” do que aqueles abordados por visões mais simples. Num sentido próximo, uma abordagem generativa não é tanto uma alternativa ao contrafactual, como um acréscimo de requisitos.

Apesar dos métodos contrafactuais serem recomendados pelas suas características ateóricas (Wolpin, 2013), estes implicam sempre uma suposição teórica. No caso mais simples de selecção aleatória, é expectável que a exposição ao tratamento leve a resultados distintos aos do grupo de controlo. O problema é que uma hipótese teórica desta espécie convoca um número elevado de hipóteses concorrentes. Podem existir inúmeras razões para o grupo de tratamento exibir diferenças nas variáveis de impacto ditas significativas. O paradigma (quase-)experimental não é criticado pelos autores de avaliação baseada na teoria (o exemplo seminal sendo Chen & Rossi (1987) por ser ateórico, mas pelo facto de a construção teórica baseada numa intermitência de tratamento ser demasiado simples (Trochim, 1989a).

Se a construção teórica for complexificada, por exemplo, ao considerar resultados intermédios que apontam para os mecanismos por onde a política surte o seu efeito, ou compreender que tipos de contextos permitem o sucesso da política, a probabilidade de existirem construções teóricas concorrentes consistentes com os padrões de resultados diminui.

4.3. Avaliação contrafactual de impactos e teorias da mudança

A identificação de que resultados devem ser avaliados, sejam estes esperados ou inesperados pelo decisor político, é um passo essencial de uma avaliação contrafactual. Na ausência desta identificação, não é óbvio que variáveis devem ser utilizadas para medir os impactos, nem com que desfasamento temporal face à introdução da política tais impactos devem ser medidos. Uma aceitação incondicional dos efeitos esperados pelos decisores políticos pode levar a que uma avaliação não considere outros efeitos, inesperados, que representam impactos relevantes da política. Para que estes sejam descobertos é necessário deliberar acerca das alterações imediatas e intermédias que resultam da política, e determinar os seus possíveis impactos.

Quando uma avaliação assente num método contrafactual de uma política indica que esta é infrutífera, o avaliador não consegue compreender a que se deve este fracasso. Será que a teoria que motiva este tipo de política pública é inconsistente, sendo o seu efeito insignificante mesmo em contextos mais promissores? A literatura chama a esta situação uma falha da teoria da política. Uma das etapas cruciais de uma avaliação baseada na teoria é a reflexão da viabilidade da teoria de um programa. Se esta for logicamente coerente, assente no conhecimento científico existente, compreendida e aceite pelos *stakeholders*, existe uma probabilidade razoável de sucesso.

Em alternativa, poderão existir falhas de implementação. Mesmo que a política seja, do ponto de vista teórico, coerente e viável, a sua implementação pode não o ser. A teoria da mudança – e a sequência causal que sugere – fornece um ângulo adequado para compreender que etapa condicionou a política a não surtir o efeito desejado. Uma das vantagens introduzidas pelas avaliações baseadas na teoria é que para além da possibilidade de reportarem se a política teve sucesso ou não, esforçam-se para compreender que fatores levaram a certo resultado.

Em termos gerais, uma avaliação baseada na teoria pressupõe a construção de uma teoria do programa que aponta para os mecanismos desencadeados pela política que se espera causarem os resultados pretendidos. A dimensão empírica deste exercício envolve o levantamento de dados sobre os construtos definidos acima, de maneira a poder validar se estes ocorreram (Cook, 2000).

4.3.1. Mediadores

O interesse por mediadores representa uma tendência recente na literatura contrafactual de não só reportar o efeito de uma política, mas também inquirir sobre os mecanismos que permitem a uma política causar certos efeitos. Este objetivo – que partilha com a literatura de avaliação baseada na teoria, ainda que aliado a uma estratégia de identificação que dá primazia à validade interna – oferece um método que inquire não só sobre os “efeitos das causas”, mas também sobre “as causas dos efeitos”.

A análise de mediação subentende, tal como a teoria da mudança, que a política não exerce diretamente a sua influência sobre as dimensões que pretende, em última instância, impactar, mas antes sobre dimensões mais acessíveis. Estas dimensões, consideradas resultados intermédios, são aquelas que quando afetadas, são capazes de produzir as alterações de impacto desejadas. A teoria da

mudança como explicitação mecanística da política, informa sobre a presença de diferentes vias por onde os efeitos se fazem sentir – ou seja, aponta para as variáveis mediadoras a considerar.

A já referida literatura dos GAD (gráficos acíclicos diretos) realça o carácter benevolente dos mediadores (ao invés de variáveis de confundibilidade ou *colliders*). Se uma variável (B) aparece na sequência entre o tratamento (A) e o efeito a medir (C) – ou seja, $A \rightarrow B \rightarrow C$ – aquela pode ser eliminada da análise (ou seja, no que respeita à identificação do efeito de A em C, $A \rightarrow B \rightarrow C$ é equivalente a $A \rightarrow C$). Por outras palavras, em primeira instância, o desenvolvimento de uma teoria da mudança nada parece oferecer ao método contrafactual. Os processos descobertos, envolvidos no caminho da mudança, serão ignorados pelo método estatístico, dado que em nada facilitam na identificação pretendida.

Voltando ao exemplo dos apoios ao investimento empresarial, a identificação do efeito de financiamento público na expansão da capacidade produtiva não depende da observação da variável *custo de capital*. No entanto, uma análise de mediação pode ajudar a compreender alguma da heterogeneidade observada: por exemplo, os apoios terão menos impacto na expansão da capacidade produtiva para as empresas cujo custo de capital não é reduzido pela política, devido ao facto de a taxa de juro oferecida pelo Estado não ser atrativa face às alternativas existentes.

Mas o que significa se, ao tentar identificar o efeito de tratamento, mesmo após controlar o mediador, este for estatisticamente significativo? Neste caso podemos assumir que existem outros mecanismos, desencadeados pela política, que interferem na expansão produtiva. Imagine-se que o apoio público tem um efeito de certificação das empresas apoiadas (ver secção 2.2). Tal significa que existe uma outra via que medeia o efeito de tratamento. Uma teoria da mudança, baseada no conhecimento dos *stakeholders* e na literatura relevante, pode apontar para diferentes canais causais, antes do método estatístico ser aplicado.

Imagine-se agora que os efeitos de certificação não são estatisticamente significativos. Neste caso, o pressuposto de que as condições de financiamento devem ser suficientemente atrativas para reduzir o custo de capital é fulcral – e o canal relativo à certificação pode ser descartado. Este tipo de abordagem parte de uma posição teórica (uma teoria da mudança) cuja validação é realizada pela análise de mediadores num contexto quase-experimental.

Uma relação de mediação estatística é dada como “completa” (Judd & Kenny, 1981) quando, ao considerar os efeitos dos mediadores relevantes na variável de resultado, a variável de tratamento deixa de ser estatisticamente significativa. Se ainda existir um efeito da variável de tratamento na variável dependente, mesmo após controlar o efeito de mediação, esta não é prova de que os mecanismos alternativos considerados na teoria da mudança estejam corretos, mas é evidência da existência de outros efeitos causados pelo tratamento, para além do mecanismo principal. Mesmo que destes mecanismos alternativos não decorram variáveis mensuráveis, a análise de mediadores consegue testar a sua presença.

A análise de mediadores é facilmente compreendida através de uma representação gráfica. MacKinnon (2008) operacionaliza o modelo de Chen (1990) para obter dois efeitos resultantes de uma análise de mediação. O primeiro trata-se do impacto que a política tem no mediador, ou seja, traduz

as etapas da teoria da mudança relativamente aos mecanismos que transformam as atividades e recursos disponibilizados pela política (tipicamente denominadas por teoria de ação). Importa ao avaliador validar a capacidade da política em interferir com a dimensão, considerada relativamente mais acessível, que inicia o encadeamento causal. O segundo elo, que representa a teoria conceptual, é a medição do efeito que o moderador tem sobre o resultado. Embora a política pretenda afetar as condições mais próximas, se estas não despoletarem diferenças nas dimensões de impacto que se ambiciona alterar, a intervenção terá sido inconsequente. A análise de mediadores permite aferir se determinada política falhou devido a problemas de implementação ou, alternativamente, porque a sustentação teórica da política era frágil e que, por isso, as alterações levadas a cabo não produziram o impacto desejado, embora a dimensão próxima (mediadores) tenham sido impactados pelos recursos da política.

4.3.2. Variáveis interativas

O avaliador que, apesar de recorrer a métodos (quase-)experimentais, partir de uma abordagem generativa para descrever os mecanismos ativos de uma política, irá aprimorar a sua capacidade de inferência, do ponto de vista tanto da validade interna como externa.

Leamer (2010) questiona a noção de efeito de tratamento médio local (Local Average Treatment Effect, ou LATE – ver Imbens & Angrist, 1994), recorrente nas abordagens contrafactuais. Um LATE, se devidamente especificado, informa o decisor político acerca do impacto que certa política teve, mas pouco dirá o que deve fazer em futuras rondas da política, ou prever que sucesso terão outros decisores políticos em replicar esta política noutros contextos (temporais ou geográficos).

Muitas vezes, o efeito de tratamento na variável de resultado não é um coeficiente (como é assumido na literatura contrafactual), mas sim uma variável, visto que esta interage com variáveis contextuais, quase nunca consideradas. Falar de um efeito de tratamento médio é, por isso, inadequado, dado que este varia com as características da população. Ainda para mais, a localidade deste efeito exclui qualquer tipo de generalização a outros contextos.

A metodologia proposta por Leamer ocorre no sentido de reconhecer heterogeneidade ao longo de diferentes configurações. Implicitamente, o autor tenta estender a análise econométrica para além de uma abordagem à causalidade contrafactual, aproximando-a de uma lógica configuracional e generativa (típica das abordagens de avaliação baseada na teoria). Invoca, assim, Heckman que várias vezes apelou à necessidade de estudar “mecanismos causais” (e.g., Heckman, 1991), bem no sentido generativo. Tal implica que o estudo atento dos mecanismos causais deve procurar revelar as variáveis que representam os efeitos contextuais que interagem com a política. Na mesma linha, Deaton (2010) havia manifestado a importância dos objetivos de uma avaliação realista (não só medir um efeito, mas explicar como este aconteceu e em que casos foi mais pronunciado) para um processo cumulativo de aprendizagem sobre políticas, defendendo que a relevância do método contrafactual depende da capacidade de responder a estes objetivos.

Uma das componentes centrais da teoria da mudança, particularmente enfatizada por Mayne (e.g., 2012) trata-se das influências externas e fatores que suportam os mecanismos idealizados da política.

Shadish et al. (2002) interpretam o problema de heterogeneidade como um problema de validade interna (ou mais especificamente, um problema de validade de conclusão estatística). A heterogeneidade vai “obscurer” a relação tratamento-resultado, visto que esta envolve mais desvio padrão na variável de tratamento. Além disso, se o investigador não considerar as dimensões que interagem com a relação causa-efeito, esta interação será contabilizada no resíduo, diminuindo a capacidade de inferir causalidade. O problema das variáveis interativas não se cinge à capacidade de assegurar validade externa, mas dificulta também a dimensão correspondente à validade interna. Por esta razão é difícil assegurar que os métodos (quase-)experimentais serão capazes de identificar parâmetros relevantes se não partirem de uma base teórica robusta.

4.4. Conclusão

Os métodos contrafactuais são muitas vezes associados a avaliações *black box*, algo que decorre de uma fraca conceptualização teórica da política e que limita a compreensão da avaliação. Assente num foco restrito na validade interna das suas conclusões, o contrafactual limita-se a reportar se a política esteve associada a um efeito significativo. Se este for o caso, esta abordagem é limitada na identificação dos processos fundamentais para o sucesso da política e a quem este sucesso foi mais pronunciado. De igual modo, a conclusão de que a política não surtiu o efeito esperado não é acompanhada de explicação. Idealmente, gostaríamos de compreender se a política no seu todo é inadequada e dificilmente teria sucesso noutros contextos, se o nível de dosagem foi demasiado baixo para causar uma resposta observável ou se a política foi mal implementada.

O avaliador que, apesar de recorrer a métodos (quase-)experimentais, partir de uma abordagem generativa para descrever os mecanismos ativos de uma política, irá aprimorar a sua capacidade de inferência, de um ponto de vista tanto da validade interna como externa. A análise de mediadores força a interação entre a metodologia empírica e a teoria causal da política. As abordagens aqui descritas partem de modelos teóricos conceptualizados ao nível generativo, em busca de conhecimento generalizável, que é testado a níveis de causalidade mais simples, como é aquele do contrafactual. A análise de mediadores e a consideração de variáveis interativas permite aprimorar a validade interna das conclusões de um contrafactual, visto que explica heterogeneidade, mas também a validade externa, ao compreender os mecanismos e as interações que advêm de contingências contextuais.

5. Dados utilizados e considerações sobre a sua análise

Uma avaliação contrafactual de impactos tem uma natureza quantitativa, dependendo por isso da disponibilidade de informação que permita não só quantificar as dimensões onde os impactos são esperados, como também as dimensões que possam introduzir confundibilidade na análise. Como vimos anteriormente, confundibilidade é tudo aquilo que pode impedir a atribuição do efeito a uma certa causa, e é o grande desafio das análises contrafactuais.

No Capítulo 2 foi apresentado parte do manancial de metodologias estatísticas que permite lidar com este desafio. Um ponto comum a todas elas é que estas dependem da quantidade, variedade e qualidade da informação quantitativa disponível: Quantidade, no sentido em que os métodos estatísticos gozam de boas propriedades assintóticas, ou seja, têm melhor performance à medida que a quantidade de observações aumenta; Variedade, no sentido em que é necessário observar uma quantidade tão diversa de dimensões quão mais sofisticado for o mecanismo de seleção das empresas apoiadas; E qualidade, uma vez que todos os métodos são falíveis a erros existentes nos dados, conduzindo a conclusões erróneas (por exemplo, as empresas falsearem os resultados que reportam).

A avaliação contrafactual de impactos utiliza, por norma, dados estruturados (e.g. tabelas com informação financeira). Contudo, não é de descurar as possibilidades de utilização de informação não estruturada (e.g. texto das candidaturas) e semi-estruturada (e.g. redes de relações de empresas) em análises deste género. Por exemplo, quando se querem comparar empresas apoiadas com empresas candidatas não apoiadas, é importante estabelecer uma base de comparabilidade (e.g. empresas de apoiadas de tamanho médio com empresas não apoiadas com tamanho médio também). Uma possibilidade de aumentar a base de comparabilidade de empresas é utilizar informação sobre a sua candidatura (e.g., comparar empresas apoiadas que se propõem a comprar fornos com outras não apoiadas que também se propõem a comprar fornos). Existem vários ramos da aprendizagem automática focados na análise deste tipo de informação que podem ter contributos úteis não só em avaliações contrafactuais como também em outras abordagens mais qualitativas como as discutidas nos Capítulos 2 e 4. Contudo, esses tópicos saem fora do âmbito do presente relatório.

Neste capítulo, pretende-se descrever sucintamente as fontes de dados utilizadas nos exemplos práticos apresentados no Capítulo 6 e que correspondem a parte da informação disponível e indispensável à avaliação de políticas de apoio a empresas em Portugal. São discutidas as limitações dos mesmos, possíveis estratégias para lidar com estas (preenchimento de valores omissos, deteção de anomalias e *outliers*) e implicações destas para a análise. Por fim, são dados exemplos de como a aprendizagem automática pode ser útil tanto no tratamento dos dados como na análise dos mesmos, nomeadamente demonstrando a forma como esta pode ser útil para informar análises contractuais.

5.1. Fontes dos dados

Neste estudo, foram utilizadas três fontes de dados:

- Sistema de Contas Integradas das Empresas (SCIE), com informação financeira sobre empresas individuais e sociedades.
- Quadros de Pessoal (QP), com duas tabelas distintas: QP-Empresas, que inclui informação por empresa, e QP-Trabalhadores, que inclui informação por trabalhador, sendo possível cruzar uma com a outra.
- PT2020, com informação sobre os instrumentos de política agregada por empresa candidata.

Os dados foram disponibilizados pelo INE para este fim específico, ao abrigo do Protocolo entre este instituto e o Ministério da Educação e Ciência. Os dados foram previamente anonimizados, consoante cada uma das tabelas um número identificativo fictício de cada empresa, que é o que permite fazer o cruzamento da informação.

Seria uma mais valia a utilização de fontes adicionais de dados, nomeadamente as que dizem respeito a outras políticas que tenham ocorrido anterior e simultaneamente à política alvo de estudo (e.g. SIFIDE, QREN, Linhas de Crédito), sendo aspetos importantes de controlar por serem importantes fatores de confundibilidade. Adicionalmente, poder-se-ia complementar os dados utilizados com mais informação disponível em Portugal, nomeadamente a que diz respeito a certificações, a propriedade industrial, a atividades inovadoras e de investigação (e.g. IPCTN), entre outras. Este tipo de informação permitiria robustecer a validade dos impactos identificados. Contudo, tal não faz parte do propósito do presente estudo. Certo é que tamanha quantidade de informação acarreta dificuldades computacionais e mesmo estatísticas. As estratégias baseadas em aprendizagem automática referidas anteriormente ganham mais relevância neste cenário dado a sua capacidade de contornar problemas de maldição da dimensionalidade. Um exemplo de como a maldição da dimensionalidade surge em análises contrafactuais é o método de emparelhamento discutido anteriormente. Acontece que encontrar observações semelhantes torna-se difícil em espaços de alta dimensionalidade (muitas variáveis), por natureza esparsos, pois todas as observações se vão revelar muito diferentes umas das outras.

5.1.1. PT2020

As tabelas do PT2020 correspondem aos dados da política em estudo. É a partir daqui que é retirada informação sobre que empresas são candidatas não apoiadas e apoiadas, em que instrumento de política, com que montantes de apoio e em que períodos.

Para fins de anonimização, as variáveis quantitativas relativas ao PT2020 foram agregadas, sendo apenas observável o valor do percentil dessa variável a que dada observação corresponde. Por este motivo, parte importante da informação para fins de avaliação contrafactual não foi utilizada (e.g. o montante de incentivo, que possibilitaria abordagens com tratamentos contínuos). Outra limitação da transformação de variáveis contínuas em percentis é que deixa de ser fidedigno considerar como apoiadas as empresas que tiveram montantes de execução superiores a zero (podem coexistir no primeiro percentil de valor não nulo tanto empresas apoiadas como não apoiadas). Empresas que tenham sido candidatas a mais do que um projeto sofreram também agregações, de forma a existir uma entrada única por empresa, sendo por isso omitidas as múltiplas tipologias de apoio a que se candidataram.

Estas agregações inviabilizam contagens realistas do número de candidaturas e apoios por cada instrumento de política, sendo por isso pouco pertinente descrevê-las aqui. Tal descrição está disponível em anexo (Tabela 17). Uma vez que o propósito do estudo não é realizar avaliações contrafactuais válidas, estas limitações não se apresentam como problemas de maior.

5.1.2. Sistema de Contas Integrado das Empresas (SCIE)

As tabelas do SCIE são compostas por sociedades e empresas individuais, sendo que a maior parte da informação só está disponível para as primeiras. Neste relatório só consideramos os dados relativos a sociedades.

Uma particularidade da avaliação de apoios públicos a empresas é que é necessário acompanhar estas ao longo de vários anos (antes e depois do apoio). Acontece que a percentagem de empresas que é possível ser acompanhada sem quebras na série de dados corresponde a um terço do total de empresas que surge alguma vez nos anos disponibilizados. A Tabela 1 pretende exemplificar isso. Tal limita a amostra de potenciais controlos, bem como de potenciais empresas apoiadas a analisar.

Tabela 1: Percentagem de empresas no SCIE por primeiro e último ano com registo

Primeiro ano	Último ano									
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2010	4,2	4	3,4	2,7	2,3	2	1,6	1,4	1,4	31,5
2011		0,5	0,6	0,4	0,3	0,3	0,2	0,2	0,2	2,3
2012			0,3	0,4	0,4	0,3	0,2	0,2	0,2	2,2
2013				0,4	0,4	0,4	0,3	0,2	0,2	3
2014					0,3	0,4	0,4	0,3	0,2	3,1
2015						0,3	0,5	0,4	0,3	3,6
2016							0,3	0,4	0,4	3,8
2017								0,3	0,5	4,5
2018									0,4	5,5
2019										6,5

Importa notar que, considerando-se apenas sociedades, não existem dados em falta no SCIE para além dos que correspondem a anos em que certas empresas deixam de estar na tabela. Não existem também entradas duplicadas.

5.1.3. Quadros de Pessoal (QP)

Os Quadros de Pessoal contêm duas tabelas distintas. A QP-Empresas, que inclui informação por empresa; e a QP-Trabalhadores, que inclui informação por trabalhador. Da primeira, só foi utilizada uma variável: a percentagem de capital social estrangeiro. Tais variáveis são indispensáveis para caracterizar a composição da força de trabalho das diferentes empresas e a sua gestão, sendo a informação ao nível do trabalhador agregada para o nível da empresa.

Tal como acontece no SCIE, também nos QP existem empresas que deixam de reportar informação ao longo da série, como exemplifica a tabela 17.

Tabela 2: Percentagem de empresas no QP por primeiro e último ano com registo

Primeiro ano	Último ano									
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2010	4,8	4,6	3,6	2,6	2,4	1,9	1,7	1,7	2,7	27,5
2011		1	0,7	0,5	0,4	0,3	0,3	0,3	0,6	2,7
2012			0,7	0,5	0,4	0,3	0,2	0,2	0,3	2,1
2013				0,6	0,6	0,4	0,3	0,3	0,4	2,7
2014					0,7	0,5	0,4	0,4	0,5	2,8
2015						0,7	0,5	0,5	0,5	3,1
2016							0,7	0,6	0,6	3,1
2017								0,8	0,8	3,4
2018									1,0	3,8
2019										4,9

Nos QP, ao contrário do SCIE, existem diversos valores omissos conforme reportados na Tabela 3, bem como registos duplicados, reportados na Tabela 4. Estes são duplicados em todas as variáveis, existindo também casos em que existe mais do que uma entrada para o mesmo trabalhador, seja na mesma empresa (não sendo claro se correspondem a cargos diferentes ou erro de introdução da identificação do trabalhador) ou em empresas diferentes (que aconteceria naturalmente caso os trabalhadores trabalhem em mais do que uma empresa). Ao longo da série de um mesmo trabalhador, é possível encontrar discrepâncias em termos de idade e nível de educação do mesmo (e.g. diminuir o nível de educação).

Tabela 3: Dados omissos no QP-Trabalhadores

	QP	
	Total	%
Nacionalidade	11	0,00
Antiguidade / Data de admissão	2 316	0,07
Idade	25 009	0,77
Tipo de Contrato e Regime de Duração	186 495	5,77
Rendimento e Número de Horas	187 079	5,79

Tabela 4: Trabalhadores duplicados no QP-Trabalhadores

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Total	280	714	396	434	574	793	1 350	1 603	1 673	1 894
%	0,01	0,03	0,02	0,02	0,02	0,03	0,05	0,05	0,05	0,06

5.2. Cruzamento dos dados

Tal como dito anteriormente, as diferentes tabelas contêm um número de identificação fictício que permite o cruzamento das mesmas. Contudo, existem empresas que não constam em todas as tabelas em períodos simultâneos. Estas situações são descritas pela Tabela 5.

Tabela 5: Empresas com registo, cruzamento de bases de dados (SCIE e QP)

Ano	QP	SCIE			
		Não	Sim	Não (%)	Sim (%)
2010	Não	0	145 562	0,0	40,2
2010	Sim	1 168	215 673	0,3	59,5
2011	Não	0	146 378	0,0	40,3
2011	Sim	1 577	215 473	0,4	59,3
2012	Não	0	149 481	0,0	41,6
2012	Sim	3 489	206 288	1,0	57,4
2013	Não	0	149 534	0,0	41,5
2013	Sim	3 409	207 043	1,0	57,5
2014	Não	0	150 684	0,0	41,2
2014	Sim	2 219	212 672	0,6	58,2
2015	Não	0	156 933	0,0	41,9
2015	Sim	2 178	215 268	0,6	57,5
2016	Não	0	161 660	0,0	42,2
2016	Sim	2 159	219 275	0,6	57,2
2017	Não	0	171 703	0,0	43,3
2017	Sim	1 870	223 264	0,5	56,3
2018	Não	0	186 253	0,0	44,8
2018	Sim	1 714	227 514	0,4	54,8
2019	Não	0	212 833	0,0	48,4
2019	Sim	1 020	226 126	0,2	51,4

Este fenómeno é bastante mais comum no cruzamento entre o SCIE e os QP para a generalidade das empresas, e menos comum entre as empresas que foram candidatas ao Portugal 2020 (ver Tabela 39 em anexos). Estes casos são considerados como valores omissos na tabela que resulta do cruzamentos entre QP, SCIE e PT2020.

5.3. Construção de indicadores

A construção de indicadores foi estruturada em 12 dimensões, correspondendo a 47 variáveis de desempenho das empresas (cujas fórmulas subjacentes se encontram na secção 8.2 em anexo). As dimensões de desempenho consideradas foram as seguintes:

- Caracterização geral das empresas
- Investimento
- Situação financeira
- Inovação
- Internacionalização
- Competitividade
- Ecoeficiência
- Qualificações
- Qualidade do emprego
- Remunerações
- Igualdade de Género

De modo a permitir a comparação entre empresas e diluir o efeito de dimensão das mesmas, a maior parte dos indicadores construídos são rácios. Evitou-se usar denominadores que pudessem ter um valor negativo (uma vez que estes podem sugerir realidades pouco coerentes com a dimensão em análise) ou zero (que resulta em indeterminações e, por isso, casos de valores omissos). Na Tabela 40 em anexo são apresentadas estatísticas descritivas gerais sobre alguns destes indicadores, sem serem sujeitos a tratamento prévio, para todas as empresas. Note-se que, mesmo com preocupações acrescentadas sobre a formulação dos rácios, existem ainda proporções consideráveis de dados em falta nos rácios construídos.

5.4. Qualidade dos dados

A análise da qualidade dos dados aqui exemplificada é focada no propósito do exercício em mãos, não sendo tentadas abordagens exaustivas seja na identificação de problemas, seja na proposta de soluções. Tal exercício seria útil, acarretando grandes mais valias para posteriores trabalhos, sobretudo para trabalhos mais exploratórios e abrangentes quanto à amostra alvo do estudo.

Existindo, neste caso, uma amostra alvo bem delimitada (empresas apoiadas pelo SI Inovação no PT2020), as principais preocupações de qualidade são pensadas de acordo com os cuidados que estas inspiram para uma avaliação contrafactual. Contudo, são aqui descritos alguns dos problemas identificados, estratégias para os identificar e hipóteses de solução para estes.

5.4.1. Duplicados e discrepâncias em séries do QP-Trabalhadores

Os trabalhadores duplicados em todas as variáveis do QP-Trabalhadores são reduzidos a uma observação. Trabalhadores que surjam em mais do que uma empresa são considerados independentemente disso, uma vez que se trata de uma situação enquadrável na realidade dos trabalhadores portugueses. Em casos de discrepâncias na informação do mesmo trabalhador entre empresas, seria de considerar a eliminação de observações devido às dúvidas que isso deixa sobre a credibilidade da informação. Uma vez que o objetivo é ter informação agregada ao nível da empresa, erros na introdução dos dados nos trabalhadores individuais terão pouco impacto ao nível agregado (e.g. casos em que os

trabalhadores têm incongruências ao nível das suas qualificações, quando corrigidos, surtiriam pouco impacto em termos de percentagem de trabalhadores com determinada qualificação a nível da empresa). A opção deste estudo foi não ser exaustivo na correção destas incongruências, mas tal deve ser reconsiderado para análises em que a amostra alvo seja de trabalhadores e não de empresas.

5.4.2. Correção dos indicadores

Devido aos casos de indeterminações ou de denominadores negativos, parte das informações sobre indicadores resulta em valores omissos. Esses casos, caso surjam precedidos por anos em que a informação é completa, são imputados os valores omissos através de médias móveis a 3 anos. Outra solução possível, aplicável a casos em que os valores omissos são intervalados por casos de informação completa, seria fazer-se uma interpolação entre as entradas com observações. O impacto que este tipo de correção pode ter nos resultados da análise não é de descurar, pelo que deve ser procedida de análise de robustez em que se compare as amostras resultantes de cada tipo de correção com a amostra original. Não houve necessidade de aplicar mais correções deste tipo à amostra sobre estudo, uma vez que as unidades de tratamento não eram afetadas por estes erros e existiam suficientes unidades de controlo para a análise ser possível, sendo por isso razoável a exclusão de observações problemáticas.

5.4.3. Valores omissos resultantes do cruzamento de dados

Para além das observações que existem em determinada fonte de dados e não noutra, existem quebras em séries que resultam da ligação de tabelas de diferentes anos da mesma fonte. A tabela Tabela 6 descreve estas situações. Alguns destes casos representam empresas que suspenderam atividade durante esses anos, outros casos poderão sinalizar dados em falta.

Casos com apenas um ano em falta foram corrigidos por interpolação entre o valor anterior e o valor posterior ao correspondente ao ano em falta. Mesmo que a empresa tenha suspenso atividade durante um ano, tal não compromete a sua utilização como controlo numa análise contrafactual, no sentido em que a sua atividade teve continuidade e a própria suspensão de atividade e as suas consequências, caso aconteça em momentos após o período de tratamento, pode ser uma boa representação da realidade contrafactual. Outras situações foram excluídas da análise por serem menos comuns e requererem maior extrapolação.

Tabela 6: Empresas com registos em falta em um ou mais anos entre o primeiro e o último registo

	SCIE		QP	
	Nº	%	Nº	%
1 ano	15 728	2,4	24 482	6,0
2 anos consecutivos	6 804	1,0	8 216	2,0
3 anos consecutivos	3 351	0,5	3 992	1,0

4 anos consecutivos	1 687	0,6	2 119	0,5
5 anos consecutivos	941	0,1	1 205	0,3
6 anos consecutivos	460	0,1	711	0,2
7 anos consecutivos	218	0,03	354	0,1

5.4.4. Anomalias, outliers e valores extremos

Parte da pertinência do estudo das anomalias, *outliers* e valores extremos nos dados em causa prende-se com a identificação de empresas em que o processo de geração de dados é duvidoso, provavelmente associado a erros de reporte. Outra preocupação prende-se com a inclusão de valores de tal maneira extremos que influenciam em demasia as estimações a serem feitas.

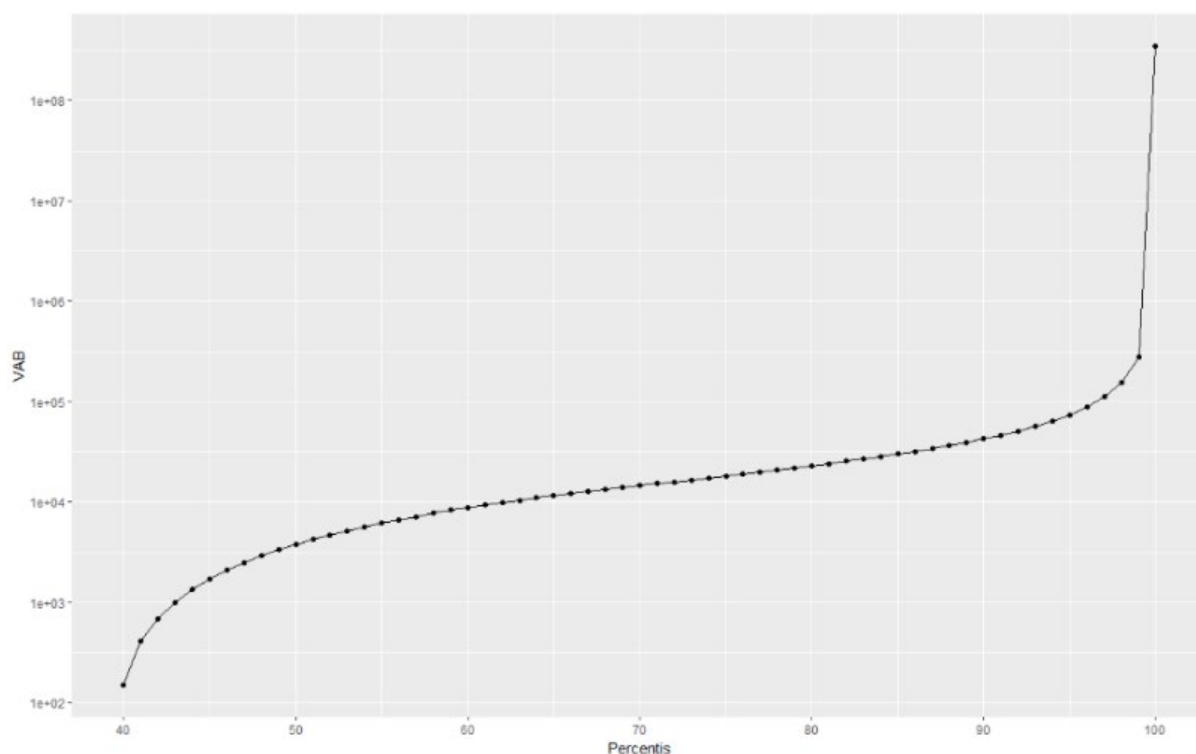
Por exemplo, tome-se a distribuição de valores de VAB para empresas com apenas 1 trabalhador presente na Figura 4. Note-se que 40 % dos valores são negativos ou nulos, não aparecendo no gráfico devido a este estar em escala logarítmica. Entre o percentil 50 e o 90, o VAB cresce de forma exponencial (apresentando-se como uma linha parcialmente reta). Nos últimos 10 percentis, o VAB cresce mais que exponencialmente. Não é de esperar tamanha dispersão para empresas de dimensão equivalente.

Quando o objetivo passa por fazer estimação e inferência, um *outlier* pode ser um indivíduo para o qual nunca faria sentido generalizar as conclusões do estudo que está a ser conduzido e, ao serem incluídos na amostra, estes podem ser responsáveis por aumentos das medidas de erro que nunca vão permitir rejeitar hipóteses que seriam perfeitamente rejeitáveis para os indivíduos que são alvo do estudo em causa.

Quando se pretende realizar uma análise contrafactual de subsídios ao investimento em empresas, em que se procuram efeitos de tratamento médios, incluir empresas que são radicalmente diferentes do grosso da amostra de empresas tratadas faz com que um efeito de tratamento médio faça pouco sentido.

Em termos de inferência, tal tem como implicação que o erro associado a essa média seja de tal maneira grande que não é possível tirar conclusões sobre as hipóteses a serem testadas. O mesmo se aplica à estimação de médias, mesmo que o propósito não seja inferência.

Figura 4: Percentis de VAB para empresas com um trabalhador

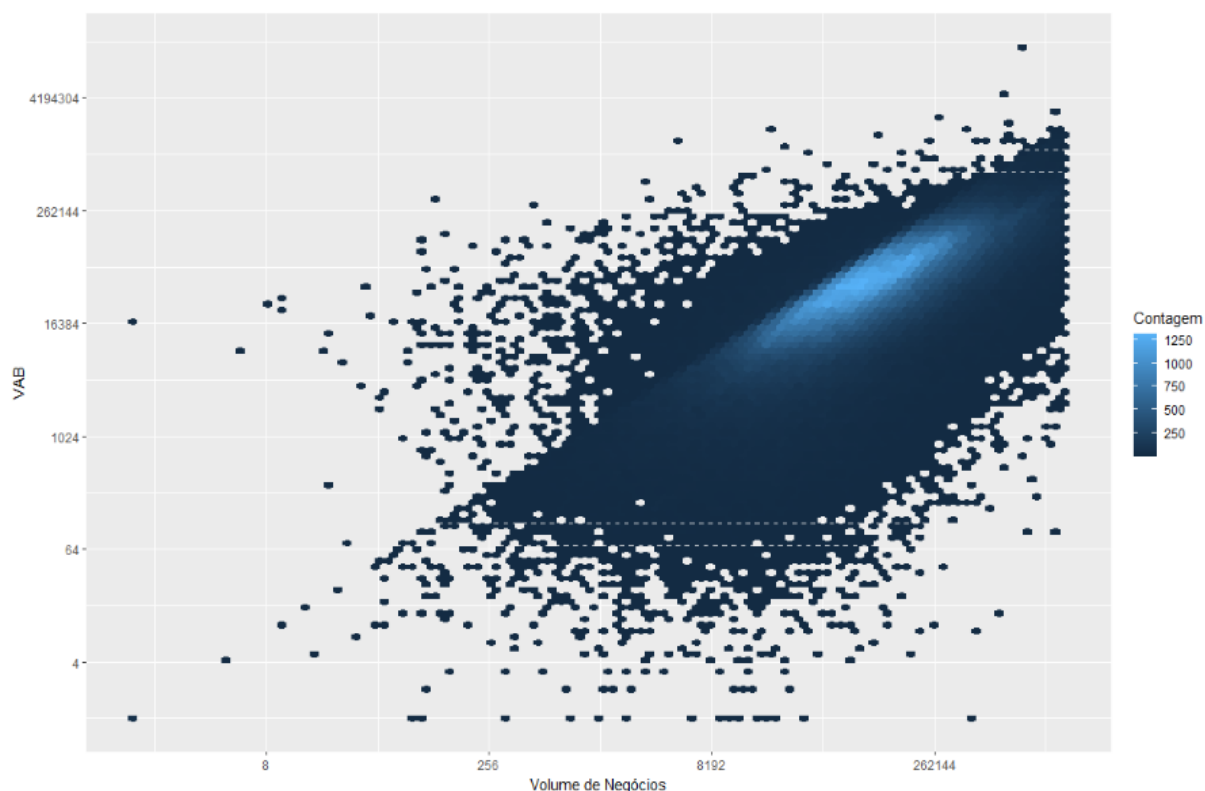


Por exemplo, quando se pretende comparar empresas apoiadas pelos fundos com empresas não apoiadas, algumas médias podem estar a ser influenciadas por casos extremos que não faz sentido incluir na amostra. A sua presença faz com que a média não ilustre bem as diferenças que existem entre os grupos, perdendo-se o sentido de utilizar esta para comparação de amostras. Nestes casos, a opção de exclusão é justificável. Para análises mais exploratórias e menos focadas, a pertinência do estudo de *outliers* passa maioritariamente pela sua identificação e posterior análise mais fina, na medida em que estes podem revelar realidades relevantes e inclusive mais interessantes para caracterizar, por exemplo, a economia nacional.

Como vimos no caso do VAB, existem valores suspeitos que requerem maiores cuidados de análise. Um critério possível para identificar este tipo de situações é o de densidade. Os valores mais prováveis são aqueles em que existe maior densidade de observações. Por oposição, os erros seriam valores onde existe baixa densidade de observações (assumindo-se que estes são uma raridade, e não a norma).

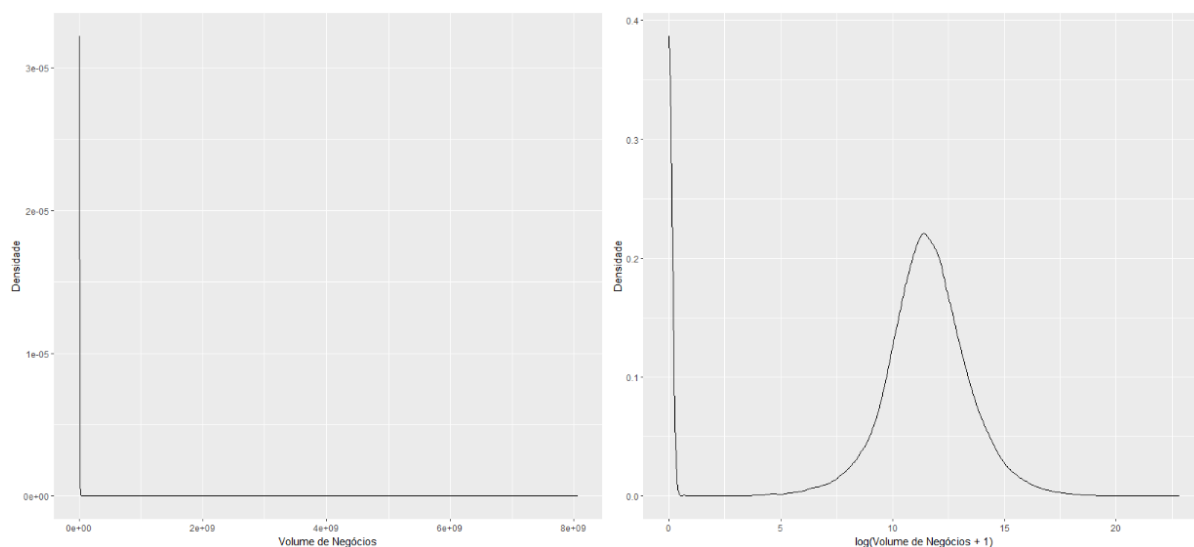
Veja-se, como exemplo, a Figura 5. As zonas mais claras são regiões em que existe maior densidade de observações. Pode-se notar que, mesmo nas zonas de maior densidade, existe um corte abrupto desta sempre que o VAB é superior ao Volume de Negócios. Tal faz todo o sentido (sendo $VAB = VN - \text{Custos Intermédios}$). Não é descabido utilizar este critério para definir observações anómalas. A título ilustrativo, só considerando micro empresas, se se utilizar este critério excluir-se-iam 2.3% das observações.

Figura 5: Densidade de observações no espaço do VAB e Volume de Negócios



Quanto a valores extremos, a definição de regras do gênero da anterior pode ser mais difícil. Uma hipótese alternativa é considerar as características da distribuição dos dados e analisar o que são valores pouco prováveis dado essa distribuição. Existem métodos não paramétricos que vão nesse sentido. Por exemplo, o método dos intervalos de *Tuckey* decorre do estudo de distribuições normais, aplicando-se por isso exclusivamente a distribuições desse tipo. Segundo este método, *outlier* é qualquer valor que se encontre fora do intervalo $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ de uma distribuição, sendo Q_1 e Q_3 o primeiro e terceiro quartil, e k uma constante arbitrária. Acontece que este tipo de distribuição é rara nos dados utilizados neste estudo. Apesar disto, para certas variáveis é possível aproximá-la. Veja-se, a título de exemplo, a distribuição do Volume de Negócios apresentada na Figura 6. Quando a variável é logaritmizada e os zeros são excluídos, a distribuição é aparentemente normal.

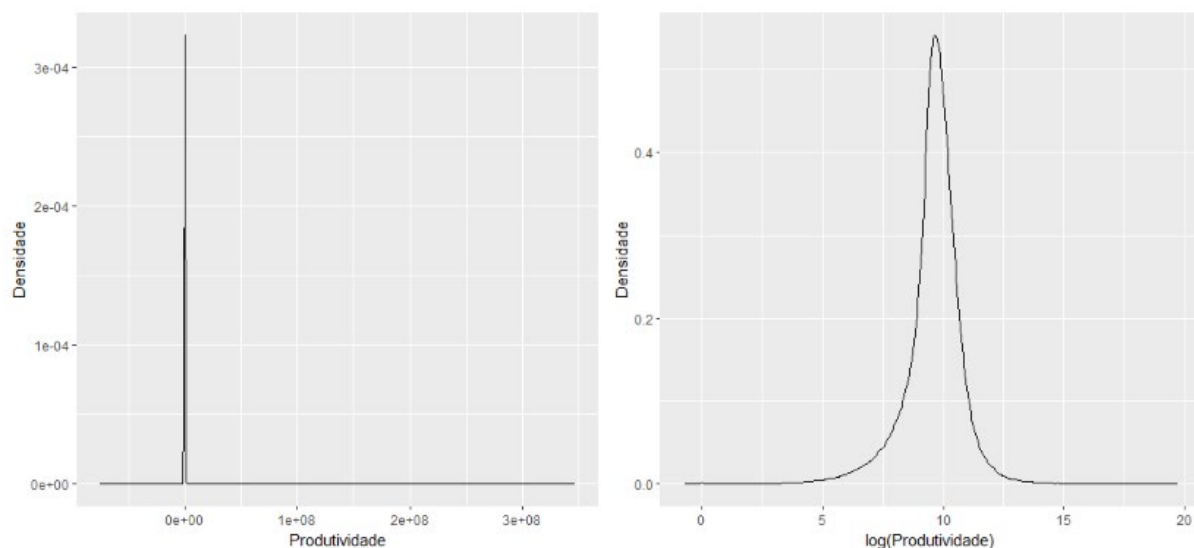
Figura 6: Distribuição do Volume de Negócios



Para essa distribuição, o intervalo de *Tuckey* com $k=1.5$ é [1 373.57, 8 172 414]. Com base nestes valores, considerar-se-iam como *outliers* 16% das empresas. Caso só se considerassem valores extremos à direita da distribuição apenas se excluiria 1% das empresas. Contudo, empresas com volume de negócios desta magnitude são interessantes de se estudar, nomeadamente pelo peso que têm na economia nacional, e excluí-las pode ser erróneo. Devemo-nos perguntar se tal exclusão faz sentido para o tipo de estudo em causa.

O caso da Produtividade Aparente do Trabalho é semelhante ao do Volume de Negócios, no sentido em que a distribuição desta é aproximável a uma normal por transformação. Neste caso ter-se-ia como limites de Tuckey com $k=1.5$ o intervalo [1 936.7, 132 366]. No total, 27% das empresas têm valores de Produtividade fora deste intervalo, mas apenas 1.5% têm valores superiores ao que seria de esperar para esta distribuição. Estas são maioritariamente microempresas (98%), nomeadamente empresas com apenas 1 trabalhador (70%).

Figura 7: Distribuição da Produtividade Aparente do Trabalho



Caso se utilizasse o critério de *outliers* exemplificado com o Volume de Negócios, estaríamos a eliminar empresas com base na sua dimensão. No caso da Produtividade já estamos a cruzar dois aspetos das empresas, dimensão e VAB. Se assumirmos que o VAB varia mais ou menos em torno de numa proporção do número de trabalhadores, consideraríamos que o rácio que dá origem aos valores de Produtividade (VAB/número de trabalhadores) apresentaria valores com menos dispersão e o seu valor esperado seria mais previsível que o do volume de negócios. Seguindo este raciocínio, valores extremos de produtividade são mais anómalos que valores extremos de volume de negócios.

Dependendo do tipo de estudo em causa, poder-se-ia estabelecer diferentes conjuntos de regras para a identificação de anomalias e *outliers*, seja para exclusão ou seja para análise posterior. Contudo, tal processo será sempre moroso, dependente das escolhas do utilizador e, por vezes, tal como exemplificado, difícil de justificar.

Existem alternativas baseadas em aprendizagem automática desenhadas para esta tarefa. Por exemplo, o método de deteção de anomalias *Isolation Forests* é um método de aprendizagem automática não supervisionado, baseado em árvores de decisão, que permite identificar situações anómalas em dados tabulares. A intuição base é a de que, numa árvore estruturada de acordo com valores de variáveis escolhidos aleatoriamente, as empresas que se situam em pontos mais profundos da árvore são mais difíceis de separar ou isolar e, portanto, é menos provável que se trate de anomalias. Neste sentido, observações fáceis de isolar com base em regras aleatórias apresentam-se como sendo mais estranhas que outras.

Tendo como vantagem ser um processo automatizado que tem em conta a natureza multivariada dos dados (por exemplo, determinado valor de volume negócios pode apenas ser estranho se se considerar o número de trabalhadores que a ele está associado), tem também como desvantagem, comum a diferentes métodos baseados em aprendizagem automática, de se ter de definir a proporção esperada de anomalias nos dados. Ou seja, o algoritmo produz uma pontuação para cada observação, baseada no quão fácil foi isolar essa observação do resto das observações. Contudo, continua a depender do utilizador a definição de qual é o valor desta pontuação em que se começa a considerar que as observações são estranhas ou não. Além disso, devido à natureza aleatória das regras, este algoritmo carece de explicabilidade, o que o pode tornar menos atrativo.

A Tabela 7 ilustra as diferentes ideias que se podem ter quando se comparam as amostras de empresas não candidatas, candidatas e apoiadas com execução de fundos europeus do Portugal 2020, excluindo-se *outliers* com base em diferentes critérios. Regra geral, as apoiadas são mais produtivas que as restantes e as amostras são sempre estatisticamente diferentes (segundo o teste *Kruskal-Wallis*). Contudo, quando se começam a eliminar *outliers* estas diferenças ficam mais nítidas, nomeadamente quando se eliminam os *outliers* à direita de Produtividade.

Note-se que quando se eliminam o 5% de empresas mais isoladas, a média que sofre maior alteração é a das empresas apoiadas, fazendo-nos refletir sobre a excecionalidade destas (que advém da seletividade da política) e se o “isolamento” das observações é um bom critério quando se lida com empresas, nomeadamente quando a análise se prende com a atribuição de apoios públicos.

No caso específico deste estudo, a estratégia que foi seguida foi a exclusão de *outliers* relativos às distribuições de empresas apoiadas nas várias variáveis selecionadas e utilizar essa amostra como teste de robustez dos resultados obtidos com a amostra completa.

Tabela 7: Comparação de médias entre amostras, segundo diferentes critérios de exclusão de observações

		Não Candi- datas	Candidatas	Apoiadas com Execu- ção
Produtividade Aparente do Trabalho em 2019	Todas	25 229.78 (671 369.8)	26 709.33 (357 902)	28 104.73 (126 055.9)
	Sem VAB>VN	24 769.34 (676 137.6)	26 632.82 (361 410.8)	28 080.25 (127 716.6)
	Sem VAB>VN e outliers à direita de VN	18 485.21 (160 963.9)	21 762.46 (55 771.88)	25 287.19 (43 628.28)
	Sem VAB>VN e outliers à direita de Produtividade	12 847.44 (13 2341.2)	20 573.22 (27 118.78)	25 330.53 (114 526.4)
	Sem VAB>VN e outliers à direita de VN e Produtividade	12 641.93 (132 434.9)	19 414.79 (26 668.73)	23 739.72 (32 950.55)
	Sem 1% mais isoladas	19 275.56 (97 701.76)	23 640.64 (39 312.91)	26 593.76 (36 726.64)
	Sem 3% mais isoladas	17 349.57 (54 740.23)	21 602.94 (32 491.22)	23 455.52 (33 175.43)
	Sem 5% mais isoladas	16 493.11 (42 816.16)	20 039.03 (26 005.84)	21 421.05 (23 928.11)

Existem variadas estratégias possíveis de serem seguidas que aqui não foram exploradas. Por exemplo, pode ser utilizado o algoritmo *local outlier factor* para a detecção de anomalias; o erro de reconstrução de espaços de dimensionalidade reduzida (e.g. por via de PCA ou *autoencoders*); modelos preditivos de fraude em reporte de informação financeira previamente estimados, como o *Beneish M-score*; ou leis probabilísticas que descrevem como certas sequências de números são improváveis, como a lei de *Benford*.

O que aqui foi apresentado pretende ser apenas exemplificativo das dificuldades e, simultaneamente, da necessidade de se preceder uma avaliação contrafactual de uma análise anomalias e valores extremos. Salienta-se a existência de observações deste tipo nos dados em causa, que podem introduzir erro ou ruído nas estimações. Apontam-se possíveis direções para a solução desta limitação e discutem-se as suas consequências para a análise, reforçando a natureza contextual das decisões a serem tomadas bem como a dificuldade que pode existir em as justificar.

5.5. Análise de dados

Segue-se um exemplo de análise aplicada aos exemplos presentes no Capítulo 6. Este tipo de análise é genericamente aplicável a vários problemas de identificação de impactos, e é especialmente informativa para avaliações contrafactuais mais complexas, como as que se aplicam no contexto de apoios públicos a empresas. O tipo de informação útil para este propósito prende-se maioritariamente com o que permite explicitar os critérios de seleção e autosseleção utilizados, os determinantes das variáveis de resultado e a definição de grupos de empresas interessantes para o estudo de heterogeneidade.

Relativamente à análise sobre preditores, o objetivo é encontrar variáveis influentes na determinação do apoio das empresas e na realização da variável de resultado onde se pretende identificar impactos. Para tal análise, recorre-se a algoritmos de aprendizagem supervisionada destinados a tarefas de regressão (no caso da variável de resultado) e de classificação (no caso da predição do apoio). Embora tais algoritmos sejam uma *black box*, ou seja, o seu funcionamento é difícil de esmiuçar, existem hoje estratégias que permitem perceber a influência de cada variável na determinação dos valores previstos para cada observação. Neste exemplo, serão utilizados para esse fim os *shapley values*. O exemplo a ser apresentado é o de análise de preditores da decisão de candidatura ao SI Inovação Produtiva por parte das empresas.

Relativamente à análise que permite informar sobre possíveis fontes de heterogeneidade, pretende-se encontrar grupos de empresas que sejam homogéneas e significativamente diferentes entre si. A intuição que esta estratégia segue é que empresas semelhantes terão efeitos semelhantes, e que grupos de empresas diferentes darão origem a impactos substancialmente diferentes. A estratégia a seguir recorrerá a um tipo de aprendizagem não supervisionada também conhecida como *clustering*. Contudo, tal análise não permitirá identificar as fontes de heterogeneidade, como será discutido na secção 6.3

De seguida, apresenta-se um exemplo da análise de preditores. Relativamente à análise de *clusters*, os exercícios conduzidos indicam que o grupo de empresas em causa (empresas apoiadas pelo SI Inovação em 2016) não é facilmente aglomerável em grupos distintos. Assim, uma análise *clusters* para este grupo de empresas não é informativa sobre possíveis fontes de heterogeneidade e, como tal, a descrição desse exercício é omitida neste relatório.

5.5.1. Preditores

Como introduzido anteriormente, o objetivo desta análise é o de tentar separar as empresas que se candidataram aos fundos das empresas que nunca se candidataram. Para tal, foi utilizado o ano de candidatura de 2016 (referido como *t* adiante) e considerou-se apenas o instrumento de política SI Inovação Produtiva.

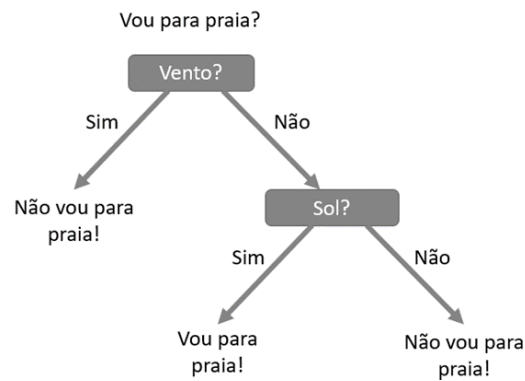
Utilizaram-se as seguintes variáveis (resultado da transformação de alguns dos indicadores descritos na secção 8.2):

- Log da idade (t-1)
- Log do volume de negócios (t-1)
- Variação do log do vol. negócios (t-1)
- Log de ativos (t-1)
- Variação do log de ativos (t-1)
- Log do número de trabalhadores (t-1)
- Variação do log do número de trabalhadores (t-1)
- Percentagem de trabalhadores com formação superior (t-1)
- Variação da percentagem de trabalhadores com formação superior (t-1)
- Log da produtividade do trabalho (t-1)
- Variação do log da produtividade do trabalho (t-1)
- Log dos gastos médios com pessoal (t-1)
- Variação do log dos gastos médios com pessoal (t-1)
- Peso das remunerações no VAB (t-1)
- Rentabilidade do volume de negócios (t-1)
- Variação da rentabilidade do volume de negócios (t-1)
- Intensidade exportadora (t-1)
- Variação da intensidade exportadora (t-1)
- Endividamento (t-1)
- Efeito dos juros suportados (t-1)
- Autonomia financeira (t-1)
- Solvabilidade (t-1)
- Liquidez geral (t-1)
- Rentabilidade dos ativos (t-1)
- Rácio do ativo intangível (t-1)
- Intensidade do uso de combustíveis (t-1)
- Intensidade do uso de eletricidade (t-1)
- Log da média do Investimento (t-3 a t-1)

A amostra foi dividida em duas subamostras, 50% para treino e 50% para teste, de forma aleatória. O objetivo de tal separação prende-se com a avaliação dos modelos. É importante garantir que o que modelo está a aprender não seja espúrio, sendo a aprendizagem tão fidedigna quanto seja possível generalizar os seus resultados para observações que nunca foram utilizadas no processo de aprendizagem/treino (ou seja, exemplos completamente novos). A divisão em 50% (em vez dos tradicionais 70/30 ou 80/20) prende-se com o reduzido número de observações.

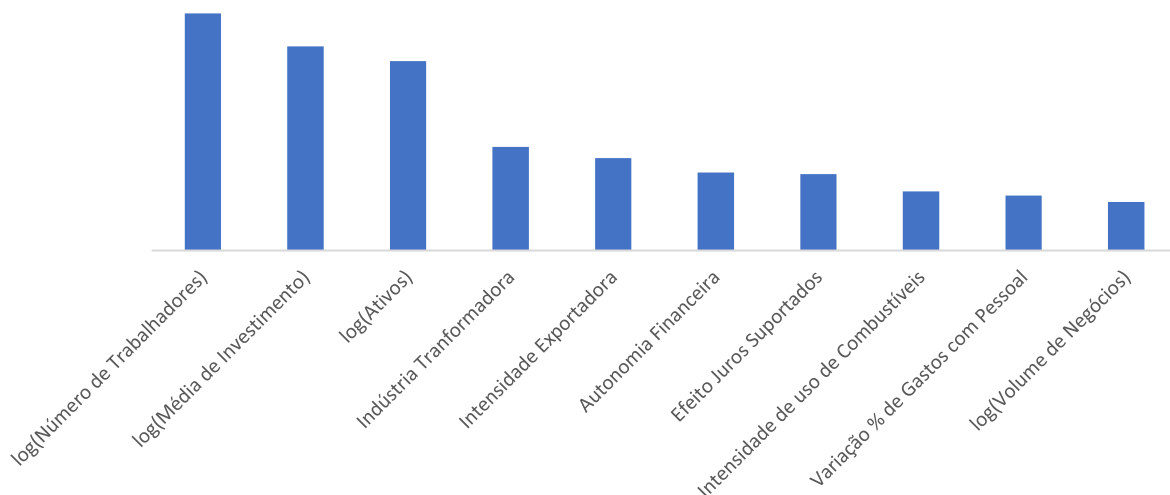
Utilizou-se o algoritmo extreme gradient *boosting* para a tarefa de classificação, tido como sendo um dos modelos com melhores resultados em dados tabulares. Este algoritmo recorre a várias árvores de decisão para separar as observações, utilizando-as de forma sucessiva e limitando por turno a amostra às observações mais difíceis de classificar. Uma árvore de decisão procura as regras que melhor separam os dados de acordo com os resultados esperados, como mostra o exemplo da Figura 8 (e.g. se uma empresa for da indústria transformadora, é mais provável candidatar-se ao SI Inovação Produtiva). No fundo, a intuição é que certas observações (nomeadamente as mais difíceis) precisam de um modelo de separação mais complexo (com mais árvores) do que outras. No final, as várias árvores de decisão são conjugadas (chamado *essemble*) e dão origem à classificação final.

Figura 8: Exemplo de árvore de decisão



Uma vantagem deste algoritmo é que seleciona primeiro as variáveis que melhor separam os dados, dando pouco relevância, ou mesmo ignorando, variáveis com pouco poder preditivo, pelo que se pode dizer que faz seleção de variáveis internamente (chamado de *embedded feature selection*). Outra vantagem é que o contributo de cada variável para a separação de observações é mensurável e agregável, permitindo ter uma ideia nítida de quais as variáveis mais relevantes. A Figura 9 sumariza a importância relativa das dez variáveis mais importantes na classificação do estatuto de empresa candidata ao SI Inovação em 2016. Estas importâncias decorrem da média da melhoria que resulta na medida de separação empregue para julgar se as amostras estão bem separadas (índice de Gini) depois de se aplicar a regra respetiva à variável considerada.

Figura 9: Importância de variáveis para classificar empresas candidatas



O melhor modelo identificado consegue distinguir adequadamente as candidatas das não candidatas em 84% das observações (taxa de acerto, também chamada de *accuracy*), sendo que de todas as candidatas deteta corretamente 73% destas (cobertura do modelo, também chamada de *sensitivity*) e de todas as empresas que prevê como sendo candidata, 79% são de facto candidatas (precisão do modelo).

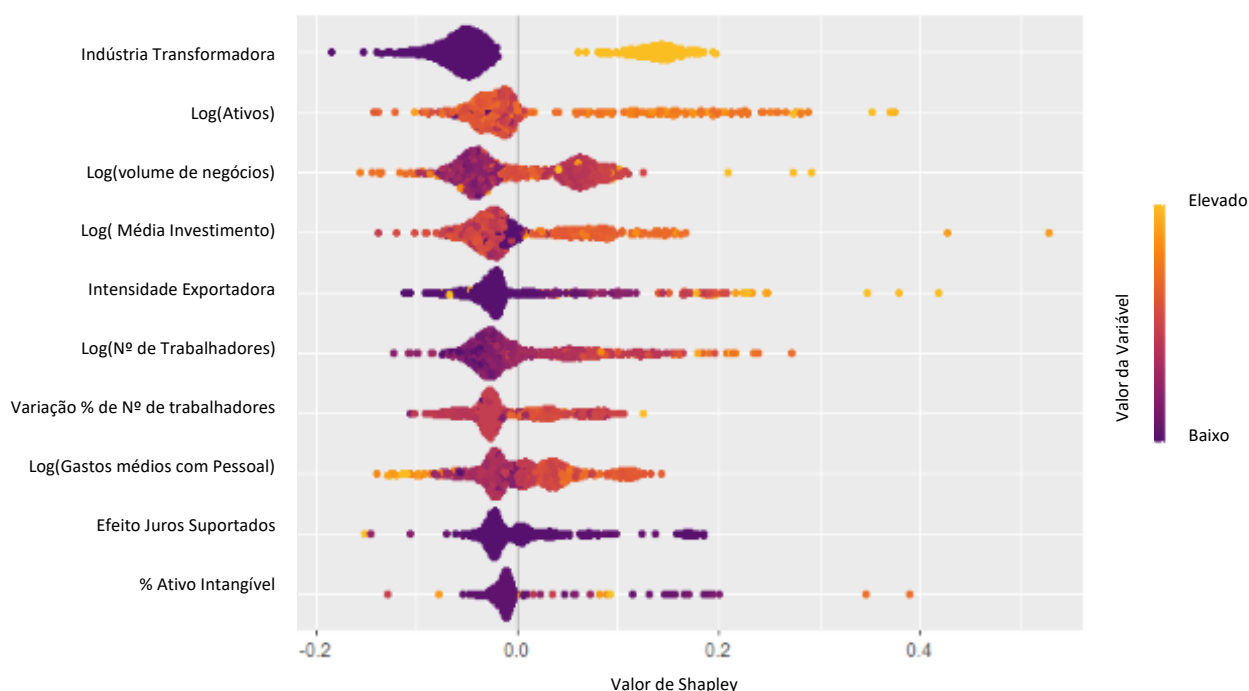
Outra estratégia orientada à explicabilidade de modelos de aprendizagem automática recorre aos valores de *shapley*. Esta estratégia, oriunda da teoria de jogos cooperativos, analisa o contributo marginal esperado de cada cooperante (neste caso, as variáveis) depois de todas as combinações de

possíveis de contributos (cruzamento entre variáveis, em que diferentes variáveis são utilizadas à vez) ter sido considerada.

Por motivos computacionais, reduziu-se a amostra de empresas não candidatas para 1 000 observações de forma aleatória.

A Figura 10 apresenta, para cada variável, um ponto por empresa. O eixo horizontal deve ser interpretado como o contributo que determinada variável está a dar para que o modelo atribua o estatuto de candidata a cada observação. A cor de cada ponto indica se, para dada observação, a variável possui um valor elevado ou baixo (claro ou escuro).

Figura 10: Contributo de cada variável para classificar cada empresa como candidata



Daqui é possível interpretar o seguinte:

- Pertencer à indústria transformadora dá, invariavelmente, um contributo positivo para a probabilidade de uma empresa ser candidata. Note-se que existe maior densidade de observações escuras (valor igual 0, uma vez que se trata de uma variável binária) e que estas se encontram sempre à esquerda do referencial de valor de *shapley* zero.
- Valores intermédios de volume de negócios também contribuem para a probabilidade de uma empresa ser candidata, enquanto valores muito reduzidos reduzem essa probabilidade para a maioria das observações. Já valores muito elevados, em certas ocasiões estão associados a probabilidades altas e noutras a probabilidades baixas. Tal deverá depender da combinação desta com outras variáveis (e.g. ser uma que não exporta pode fazer com que o contributo da dimensão da empresa para a probabilidade de ser candidata seja irrelevante).
- Valores mais elevados de investimento nos períodos anteriores à candidatura estão em larga medida associados a maiores probabilidades de a empresa se candidatar, mas tal conclusão não se aplica a todo o tipo de empresas. Contudo, pode-se dizer com alguma certeza que valores baixos nunca contribuem para aumentar a probabilidade de candidatura.

- Valores mais elevados de intensidade exportadora estão associados a probabilidades de candidatura muito superiores, embora haja casos em que tal característica parece contribuir negativamente (pode ser o caso de empresas muito exportadoras em setores que não estão abrangidos pelos apoios).

Um estudo deste género pode ser feito caso a caso, examinando o que contribuiu, para empresas específicas, para que esta fosse considerada candidata.

Note-se que o exercício aqui feito é meramente ilustrativo. A validade das conclusões é limitada pelas várias decisões tomadas, nomeadamente a de se considerar apenas um ano, a de se ter limitado a amostra, a de se ter reduzido o conjunto de variáveis a utilizar e a escolha do modelo; por motivos de simplificação, foram omitidos pormenores técnicos do processo de treino dos modelos que não são de descurar para fins de interpretação e replicabilidade.

O intuito deste exercício é exemplificar como a aprendizagem automática pode contribuir para a avaliação contrafactual. Ficou claro que a classificação de atividade económica, a dimensão, a atividade exportadora e os investimentos passados são fatores de autosseleção relevantes que devem ser especialmente controlados aquando da análise empírica.

5.5.2. Estatísticas descritivas

A Tabela 8 sumariza as médias e desvios para parte das variáveis utilizadas no estudo apresentado no Capítulo 6, comparando-se as empresas apoiadas com as não candidatas e as candidatas não apoiadas. A pertinência deste exercício prende-se com a ilustração da seletividade da política sobre estudo. Note-se que as empresas apoiadas são, à partida, bastante diferentes das outras, nomeadamente apresentando melhores performances económicas. São estas diferenças à partida que os métodos a serem empregues devem conseguir corrigir.

Tabela 8: Comparação de médias entre empresas não candidatas, candidatas não apoiadas e apoiadas.

Variáveis	Não Candidatas	Candidatas não Apoiadas	Apoiadas
Autonomia Financeira	0,41 (0,29)	0,38 (0,19)	0,42 (0,19)
Efeito Juros Suportados	0,01 (0,02)	0,01 (0,02)	0,03 (0,14)
% de Endividamento	0,38 (0,39)	0,42 (0,36)	0,43 (0,28)
Intensidade Exportadora	0,06 (0,18)	0,19 (0,29)	0,36 (0,36)
Intensidade uso Combustíveis	0,02 (0,02)	0,02 (0,03)	0,01 (0,02)
Ativos	317 038 (3 212 442)	641 097 (1 501 062)	3 817 866 (8 844 387)
Gastos Médios com Pessoal	14 340 (7 194)	14 257 (5 197)	18 425 (5 301)
Idade	18,76	18,16	23,90

	(12,66)	(9,39)	(18,09)
Média do Investimento	41 705 (541 359)	79 884 (231 080)	401 997 (732 557)
Produtividade Aparente do Trabalho	22 712 (45 355)	23 670 (12 886)	31 585 (19 090)
Nº de Trabalhadores	13,20 (85,14)	26,09 (58,72)	72,46 (104,60)
Volume de Negócios	1 426 432 (8 989 325)	2 169 522 (7 148 103)	10 294 857 (19 079 562)
Rentabilidade de Volume de Negócios	0,03 (0,13)	0,05 (0,07)	0,00 (0,36)
Trabalhadores com Formação Superior	0,16 (0,27)	0,12 (0,17)	0,13 (0,15)

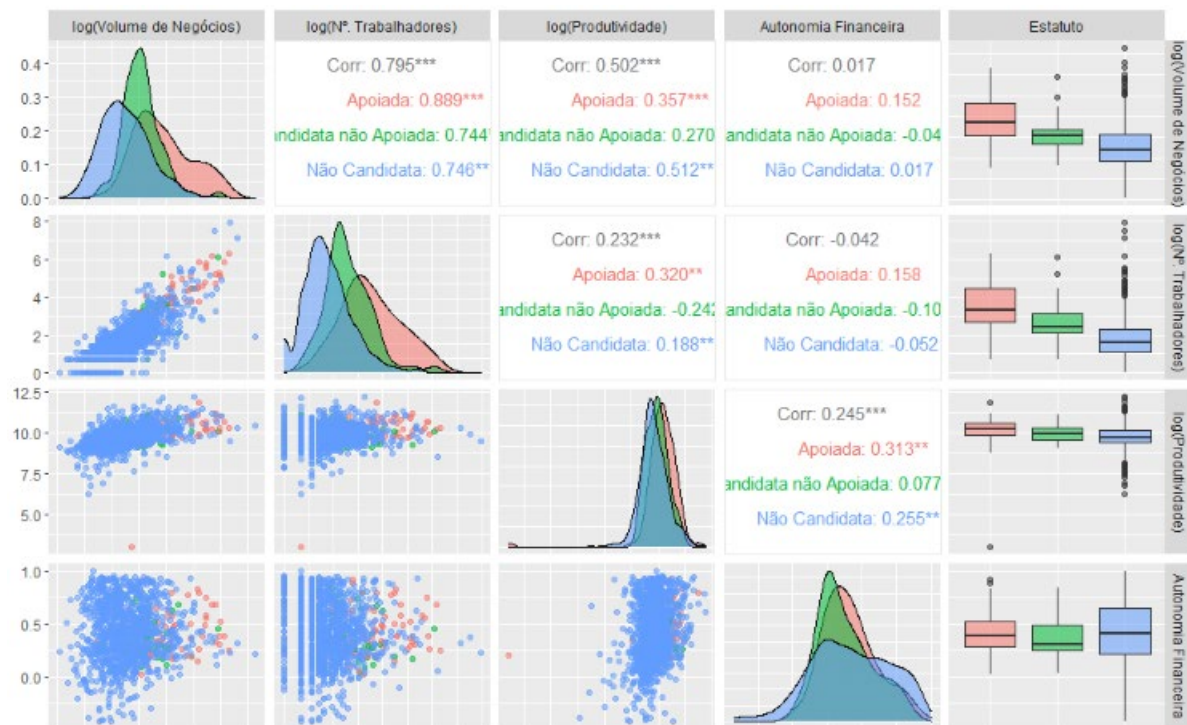
A Figura 11 ilustra parte das diferenças discutidas anteriormente, com a vantagem de ser possível comparar a distribuição inteira de cada amostra para as variáveis selecionadas (através, por exemplo, dos histogramas na diagonal principal ou das caixas de bigodes na coluna mais à direita).

As empresas apoiadas apresentam sempre distribuições com maiores densidades em zonas onde as outras têm pouca densidade. Tal demonstra a excepcionalidade desta amostra e o grau de seletividade da política. A variável em que tal é menos patente é a Autonomia Financeira.

Este exercício permite averiguar a exequibilidade de uma avaliação contrafactual, na medida em que é possível verificar se a amostra de empresas apoiadas tem suporte (tem correspondência em termos de valores) nas outras amostras para cada variável, de forma a que é possível escolher controlos adequados.

Esta visualização levanta algumas preocupações relativamente à exequibilidade de uma análise contrafactual, nomeadamente quando se tem em conta o cruzamento de diferentes variáveis (gráficos de pontos no triângulo inferior da matriz). Acontece que, deste cruzamento de características, resultam pontos referentes a empresas apoiadas bastante isolados dos restantes. Para as empresas correspondentes a estes será difícil encontrar um bom controlo e a estimação de um contrafactual recorrerá a maior extrapolação. Provavelmente parte destas empresas será colocada de fora da análise por motivos de validade interna, embora comprometendo a validade externa.

Figura 11: Gráficos a pares, histogramas, caixas de bigodes e correlações



6. Análise contrafactual de impactos

Neste capítulo pretende-se ilustrar o processo de uma avaliação contrafactual de impactos através de um estudo de caso sobre o SI Inovação Produtiva. Na secção 6.1 são discutidas e exemplificadas decisões fulcrais necessárias a qualquer avaliação contrafactual de apoios a empresas. A secção 6.2 apresenta e discute os resultados obtidos da estimação do impacto do SI Inovação Produtiva, tendo por base as decisões discutidas em 6.1. A secção 6.3 foca-se no problema da heterogeneidade (muito relevante nos apoios empresariais). Por fim, são exemplificados alguns testes de robustez na secção 6.4.

No âmbito deste relatório, escolheu-se utilizar o SI Inovação Produtiva para realizar o exemplo de avaliação. A decisão é motivada pela boa adequação da avaliação contrafactual para estimar impactos sobre esta política e pela existência de um número de empresas apoiadas suficiente para realizar a avaliação. Nesta secção, partimos da teoria da mudança apresentada na secção 2.2 sobre o SI Inovação Produtiva e dos dados apresentados no Capítulo 5, focando-nos apenas nas questões diretamente relacionadas com a avaliação contrafactual. A nossa avaliação é limitada às empresas apoiadas em 2016, porque é o ano que nos permite ter um período de pré-tratamento e pós-tratamento adequado.

6.1. Decisões críticas

6.1.1. Escolha do método

A questão da escolha do método depende essencialmente do processo de seleção de cada política, como vimos no Capítulo 3. A Tabela 9 sintetiza os requisitos, vantagens e limitações de cada um dos métodos apresentados anteriormente, esclarecendo em que situações é possível e desejável utilizá-los.

Tabela 9: Requisitos, vantagens e limitações de diferentes métodos de avaliação contrafactual

Método	Requisitos	Vantagens	Limitações
Experiências aleatórias	Os grupos de tratamento e controlo têm de ser criados de forma aleatória. No limite, apenas é necessário recolher dados sobre a variável de resultado.	A seleção aleatória garante que ambos os grupos são (em geral) idênticos em todas as características relevantes. Se for aplicada corretamente, as estimativas do impacto são garantidamente válidas.	O grupo tratado e o grupo de controlo podem não cumprir as regras de seleção iniciais. Problemas éticos por excluir possíveis beneficiários de forma aleatória. É necessário que a avaliação seja planeada antes da implementação da política, o que pressupõe custos financeiros e temporais adicionais.

Método	Requisitos	Vantagens	Limitações
Variáveis instrumentais	<p>Pode ser utilizada quando existe uma variável independente das variáveis de resultado, mas correlacionada com a atribuição do tratamento (denominada de variável instrumental).</p> <p>No limite, apenas é necessário recolher dados sobre as variáveis de resultado e instrumental.</p>	<p>É um método com bastante validade, dependendo da qualidade da variável instrumental.</p> <p>Pode ser aplicada sem ser necessário planejar a avaliação, desde que se encontre uma variável instrumental apropriada.</p>	<p>É muito difícil encontrar uma variável instrumental apropriada na maioria dos casos.</p> <p>É impossível ter a certeza de que a variável escolhida é independente das variáveis de resultado (porque não é possível testar empiricamente).</p>
Desenho de regressão descontínua (RDD)	<p>Aplicável quando a atribuição de tratamento é feita segundo um ponto de corte numa única variável.</p> <p>No limite, apenas é necessário recolher dados sobre a variável de resultado e a variável segundo a qual é feita a seleção.</p>	<p>É um método com boa validade para estimar o efeito do tratamento nas unidades próximas ao ponto de corte.</p> <p>O método de atribuição do tratamento pode ter menos problemas éticos do que a seleção aleatória, tornando o RDD aplicável a um maior número de situações.</p>	<p>É necessário que exista um número elevado de unidades próximas ao ponto de corte.</p> <p>A escolha do ponto de corte deve ser independente dos valores da classificação das unidades estudadas (i.e., o valor do ponto de corte e o valor atribuído a cada unidade não pode ter sido manipulado de forma a beneficiar algumas unidades).</p> <p>Os resultados apenas são válidos para unidades próximas ao ponto de corte, impossibilitando generalizações sobre os efeitos da política.</p>
Diferença-em-diferenças	<p>É aplicável quando haja razões para acreditar que a variação da variável de resultado seria semelhante no grupo de tratamento e controlo, se não existisse tratamento.</p> <p>No limite, é necessário recolher dados sobre a variável de resultado no pré e pós-tratamento, de preferência para muitos períodos pré-tratamento de forma a testar a assunção de tendências paralelas.</p>	<p>Caso a existência de tendências paralelas seja verdadeira, o método permite obter estimativas válidas apenas recorrendo à variável de resultado.</p> <p>Noutras circunstâncias, é um método que pode ser utilizado como complemento dos métodos de emparelhamento.</p>	<p>A existência de tendências paralelas no período-chave é sempre questionável.</p> <p>É muito improvável encontrar ou ser possível construir uma seleção que garanta grupos de comparação onde existam tendências paralelas.</p> <p>É necessário recolher dados para múltiplos períodos antes do tratamento.</p>

Método	Requisitos	Vantagens	Limitações
Emparelhamento (e outros métodos de controlo por observáveis)	<p>Não tem requisitos quanto à forma de seleção do tratamento.</p> <p>É necessária uma base de dados bastante extensa, com informação sobre a variável de resultado no pré e pós-tratamento, mas também com todas as variáveis sobre as características das unidades que se pressuponha que possam influenciar a variável de resultado ou que tenham influenciado a seleção para o tratamento.</p>	<p>Permite obter estimativas de impacto independentemente da forma como a seleção é feita (desde que se saiba como a seleção é feita).</p> <p>Pode ser aplicada em situações onde a política não foi planeada de forma a ser avaliada, desde que exista uma base de dados muito completa.</p> <p>É possível ir para além dos efeitos médios e obter outras estimativas de impacto locais.</p>	<p>Exige uma grande quantidade de dados disponíveis.</p> <p>A validade da estimação depende da assunção de que o tratamento apenas é determinado por características observáveis.</p> <p>Os resultados podem ser muito sensíveis à abordagem de controlo por observáveis escolhida (exigindo uma análise de sensibilidade de forma a garantir que os resultados são robustos e não manipuláveis).</p> <p>As estimações do impacto obtidas apenas são válidas para a subamostra onde existe suporte comum (que na maioria dos casos é menor que a amostra inicial), reduzindo as possibilidades e qualidade de generalização.</p> <p>A proficiência técnica necessária para realizar a avaliação é muito mais elevada do que para outros métodos.</p>
Controlo sintético	<p>No mínimo, requer dados extensivos temporalmente para a variável de resultado e um número suficiente de possíveis controlos.</p> <p>É aconselhável ter um grande número de períodos pré-tratamento observados e dados sobre outras características observáveis.</p>	<p>É aplicável a situações onde existe apenas uma única unidade tratada.</p> <p>Apesar de recorrer a uma construção artificial de um controlo, a interpretação e aplicação é bastante transparente.</p>	<p>É necessário que exista um número elevado de possíveis controlos e de períodos pré-tratamento observados de forma a conseguir construir um bom controlo sintético.</p> <p>É impossível obter bons indicadores de validade estatística dos impactos estimados se apenas se tiver uma ou poucas unidades tratadas.</p>

Caso prático

No caso do SI Inovação Produtiva, podemos excluir à partida a metodologia da seleção aleatória, porque o processo de seleção baseia-se num concurso, onde as empresas interessadas apresentam as suas candidaturas que são depois avaliadas e apoiadas com base no seu mérito. Excluimos também a metodologia das variáveis instrumentais porque é um caso bastante particular, e não temos indício sobre qualquer variável apropriada. A metodologia do controlo sintético é desnecessária neste caso por termos um elevado número de empresas apoiadas e não seria adequada por não termos um longo período de pré-tratamento (apenas de 2011 a 2015).

Face ao processo de seleção de empresas apoiadas, o desenho de regressão descontínua (RDD) seria o mais apropriado. No entanto, o número de empresas candidatas é pequeno e não temos informação completa sobre a nota de candidatura dos projetos. Desta forma, acabamos por não ter as condições necessárias para implementar um RDD.

Optamos por utilizar a estratégia de balanceamento das amostras nas características observáveis, em conjunto com diferença-em-diferenças, para realizar a nossa avaliação. Esta metodologia é adequada ao nosso caso porque temos uma base de dados bastante extensa sobre as características das empresas apoiadas e de um número extremamente elevado de empresas não apoiadas. Sendo que este é um dos métodos mais comumente utilizados na avaliação contrafactual de apoios a empresas e existe a possibilidade de introduzir contributos de aprendizagem automática, é o método ideal para explorar como exemplo neste relatório.

Os métodos que recorrem a balanceamento das amostras e que não dependem apenas de modelos paramétricos dão maiores garantias de consistência em análises de realidades complexas como a que é alvo deste estudo. Isto deve-se essencialmente às dificuldades de modelação correta da variável de resultado. A par disso, o balanceamento de amostras pode ser observado e testado empiricamente, validando mais ou menos a estratégia prosseguida. Desta forma, restam duas estratégias para balanceamento das amostras: emparelhamento ou ponderação. Dado o número reduzido de empresas tratadas, as estratégias baseadas em ponderação tornam-se menos atrativas por apresentarem pior performance em amostras reduzidas.

Escolhendo a estratégia de emparelhamento, importa definir em que espaço o emparelhamento é feito. A escolha tradicional recai sobre a pontuação de propensão (PS), mas uma alternativa é emparelhar diretamente em todas as variáveis consideradas relevantes. Esta alternativa requer que o praticante escolha quais as variáveis em que é importante fazer emparelhamento e qual a importância relativa de cada variável. Essa escolha é tradicionalmente guiada pela teoria, devendo no mínimo escolher-se todas as variáveis que afetem simultaneamente o tratamento e o resultado, mas preferencialmente escolher todas as variáveis que influenciem o resultado, por motivos de eficiência na estimação (secção 6.1.5). Tal estratégia revela-se altamente dependente das escolhas do utilizador e propensa a comportamentos menos honestos como práticas de *p-hacking*. Além disso, a utilização de muitas variáveis para fazer emparelhamento reduz a performance do método por via da maldição da dimensionalidade, enquanto aumenta o *trade-off* entre bons pares numa variável por piores pares noutra, podendo acontecer não chegar a ser bom par em nenhuma variável. Para evitar isto, é necessário dar mais importância a certas variáveis do que a outras quando existe pouca informação sobre

como e porque o fazer – mais uma vez, tornando os resultados muito dependentes das escolhas do utilizador. Por outro lado, fazer emparelhamento no PS resolve parte destes problemas: a dimensionalidade fica reduzida a uma variável, não se incorrendo na maldição da dimensionalidade; a importância relativa de cada variável é dada pela importância que esta tem para aferir a probabilidade de tratamento, não sendo manipulável pelo utilizador. Contudo, a escolha da forma funcional do PS levanta parte dos mesmos problemas. A bondade de diferentes formas funcionais pode ser averiguada pela capacidade de gerar *balance* (um bom equilíbrio estatístico entre as características do grupo de tratamento e do grupo de controlo) depois do emparelhamento. É comum procurar-se a melhor forma funcional iterativamente julgando-as pelo *balance* conseguido, num processo moroso e, novamente, altamente dependente do utilizador. Uma alternativa prática a esse processo iterativo e mais independente do utilizador é a utilização de aprendizagem automática, nomeadamente estimando o PS através de uma regressão regularizada tipo LASSO que faz seleção de variáveis internamente.

O estado da arte sugere a utilização de métodos duplamente robustos, devendo existir um ajustamento de enviesamento após o emparelhamento com recurso a regressão. Tal regressão pode também ser feita com recurso a aprendizagem automática, nomeadamente com post-LASSO, em que as variáveis que devem entrar na regressão são escolhidas pelo algoritmo de forma a minimizar o erro deste em validação cruzada e de seguida utilizadas numa regressão OLS.

6.1.2. Variáveis de resultado

Uma das razões para a atratividade das avaliações baseadas no contrafactual é a possibilidade de estimar impactos quantitativos. É tentador interpretar números como evidência objetiva sobre o sucesso de uma política, no entanto, deve ser tida muita cautela nas conclusões que se retiram.

Se a escolha de variáveis de resultado (i.e. as variáveis numéricas para as quais são estimados impactos quantitativos) não for informada por teoria que esclareça o porquê da sua adequabilidade para extrair conclusões relevantes sobre a beneficidade da política, a avaliação contrafactual não tem a utilidade desejada.

Se o objetivo for avaliar o sucesso de uma política, as variáveis de resultado selecionadas devem ser as que melhor permitam retratar as alterações pretendidas pela política. Ter uma boa teoria da mudança sobre a política é a melhor forma de identificar quais as variáveis onde é crucial estimar o impacto.

Em quase todos os casos, será necessário recorrer a um conjunto de variáveis que captem diferentes objetivos que a política pretende alcançar. Em alguns, a variável de resultado ótima não existirá, seja porque não há dados para algumas variáveis comuns, seja porque alguns objetivos não são apropriadamente retratados por variáveis numéricas.

A forma como a variável é utilizada também pode influenciar as conclusões ou a própria estimação quantitativa (por exemplo, pensando no “investimento” com uma variável de resultado, as conclusões poderão ser diferentes se a variável for o investimento absoluto, o investimento relativo aos ativos, a variação absoluta, a variação relativa...). Se não houver uma razão forte para que uma variável seja

utilizada de uma certa forma específica, é boa prática utilizar várias especificações para confirmar os efeitos estimados.

No mínimo, qualquer avaliação deve justificar a adequabilidade das variáveis escolhidas para retratar os resultados esperados antes de implementar a avaliação.

Caso prático

As variáveis de resultado escolhidas devem ser variáveis capazes de retratar os resultados pretendidos pela implementação da política. No nosso caso, podemos basear a escolha de variáveis de resultado na teoria da mudança elaborada na secção 2.2.

O SI Inovação Produtiva tem como resultado imediato identificado o aumento do investimento em capacidade produtiva das empresas apoiadas, por meio da comparticipação financeira de projetos de investimento. Como resultados intermédios principais, pretende-se aumentar a produção de bens transacionáveis e inovadores nas empresas apoiadas, bem como a respetiva produtividade. Por fim, pretende-se provocar alterações no perfil de especialização produtiva português, melhorias na competitividade das empresas e um aumento do investimento e do emprego à escala nacional.

Devido à natureza da avaliação contrafactual, não é possível avaliar os resultados finais da política, pois são objetivos a uma escala agregada (nacional) que não tem unidade de comparação adequada.⁶ As opções para avaliar os resultados imediatos e intermédios são constrangidas pelos dados existentes, tendo o avaliador de escolher variáveis existentes que melhor retenham esses resultados. Por brevidade de exposição e exemplificação, vamos limitar a discussão apenas à avaliação dos resultados imediatos.

Para avaliar o “Aumento do investimento de natureza inovadora”, a variável ideal seria uma variável com informação sobre o valor líquido de aquisições de capital fixo (i.e., o investimento) que tenha como finalidade o aumento ou a melhoria da capacidade produtiva de bens de natureza inovadora (i.e., que promova o tipo de produção pretendida pelos critérios de seleção da política). Variáveis personalizadas a este nível raramente existem e por isso é necessário optar por variáveis que são comumente recolhidas e que, apesar de não serem perfeitas, são bons substitutos. Neste caso, uma variável de investimento das empresas será o mais adequado e é um indicador satisfatório.

Na base de dados de que dispomos, existem diversas variáveis de investimento, no entanto as categorias de divisão são contabilísticas e não são úteis para segregar o tipo de investimento que gostaríamos. As variáveis que identificámos como mais adequadas foram a “Formação bruta de capital fixo (FBCF)” e o “Investimento em ativos fixos tangíveis” (IAFT). As duas variáveis estão fortemente correlacionadas e como a FBCF toma valores negativos (inclui cessões e depreciações), optámos pelo IAFT. Os problemas de variáveis com valores negativos que pretendemos evitar é que é impossível

⁶ Sendo correto, a estimação contrafactual dos efeitos macroeconómicos de políticas públicas é possível mas muito difícil e com validade bastante questionável, especialmente no caso de políticas de apoio a empresas; no entanto, já foram feitos estudos contrafactuais com comparações ao nível nacional para políticas e eventos cujos efeitos se pressupõe serem suficientemente grandes, como o Brexit ou a entrada na zona euro.

avaliar variações relativas (que é a medida mais interessante para comparar efeitos ao nível da empresa) ou logaritmizar essas variáveis.

6.1.3. Momento de tratamento

A determinação do momento de tratamento (i.e., o momento temporal específico em que se assume que o tratamento começou) é uma decisão fundamental para a validade da avaliação contrafactual. A principal preocupação é que podem existir efeitos que se façam sentir antes do momento considerado e que enviesem a estimação de efeitos (denominados efeitos antecipatórios).

Por exemplo, podemos considerar que a data em que os beneficiários da política recebem a aprovação da sua candidatura é o momento de início do tratamento. No entanto, podem já ter existido efeitos assim que a política é anunciada: se a política é atrativa e tem como critério que apenas as empresas com lucros são apoiadas, algumas empresas poderão alterar a sua atividade de forma a gerar lucros contabilísticos no ano $t-1$ de forma a poder beneficiar da política no ano t . Neste caso, a política estaria já a provocar efeitos no ano anterior ao identificado como ano de início do tratamento e a estimação de efeitos seria enviesada.

Apesar disto, o momento de início de tratamento escolhido também não deve ser antecipado de forma irrealista pois a precisão da estimação pode ser posta em causa. O importante é ter em conta que, em alguns casos, os efeitos antecipatórios são uma forte possibilidade e a escolha do momento de tratamento deve tentar minimizá-los.

Caso prático

No caso do SI Inovação Produtiva, poderíamos considerar várias datas como potencialmente relevantes: 1) a data de anúncio da política; 2) data das candidaturas das empresas; 3) data de aprovação da candidatura; 4) data de início do projeto; 5) data de recebimento do subsídio; 6) data do final do projeto.

Recorrendo ao exemplo apresentado anteriormente, poderíamos colocar a hipótese de que a política altera de alguma forma o comportamento das empresas simplesmente ao ser anunciada, pelas necessidades de cumprir critérios de candidatura. No entanto, as alterações não deverão ser muito significativas e deverão impactar tanto empresas apoiadas como empresas não apoiadas, visto que nessa data ainda não é possível saber quem efetivamente serão os beneficiados.

A data que consideramos mais adequada é a data de aprovação da candidatura – quando as empresas apoiadas passam a saber que vão receber o apoio. Apesar de ainda não terem recebido nenhum financiamento nem dado início ao projeto associado, é muito provável que a atividade das empresas comece a diferenciar-se ao receberem a confirmação de que vão receber financiamento público para realizar o projeto de investimento planeado.

Empiricamente, apenas temos acesso aos anos das datas 2, 4 e 6. Escolhemos estabelecer o início do tratamento através da data 2 (o ano em que a empresa se candidata) pelas razões do parágrafo

anterior. De qualquer forma, as datas 2 e 4 são semelhantes para a maioria das empresas, apenas sendo diferentes quando a empresa começa o projeto no ano seguinte ao da ronda de candidaturas.

6.1.4. Momento de estimação do impacto

A decisão sobre o momento de estimação, assim como a decisão sobre que variáveis de resultado utilizar, é fulcral no momento de retirar conclusões. A resposta simples para a pergunta “Em que momento estimar impactos?” é: em todos os períodos a seguir ao momento identificado como início de tratamento. No entanto, a resposta à pergunta “Em que momento retirar conclusões?” é mais complexa e depende da política em si.

Nesta questão, a principal preocupação é que o período de pós-tratamento disponível possa ser demasiado reduzido para que alguns objetivos possam ser avaliados. Por exemplo, se o objetivo da política é aumentar a produtividade das empresas apoiadas por base na comparticipação de projetos de investimento, só é expectável que o aumento da produtividade seja perceptível alguns anos após o início do tratamento (após a conclusão dos projetos). Neste caso, se apenas existirem dados para dois anos depois do início do tratamento e os projetos demorarem três anos até à sua conclusão, não se estimará qualquer impacto sobre a produtividade das empresas. Um avaliador que queira concluir se a política teve impactos sobre a produtividade, poderá responder que não foram encontrados quaisquer efeitos significativos nas empresas apoiadas. Um avaliador informado pela teoria da política e que tenha em consideração o momento de estimação de impacto adequado, responderá que não é possível retirar quaisquer conclusões no horizonte temporal disponível.

Caso prático

Empiricamente, apenas temos acesso a dados sobre a política de 2015 a 2019. Escolhemos avaliar o efeito do SI Inovação Produtiva nas empresas apoiadas no ano de 2016, de forma a ter simultaneamente um bom período de pré-tratamento e um período de pós-tratamento suficientemente longo para averiguar os impactos. Como o foco da nossa avaliação recai sobre os resultados imediatos (mais concretamente, sobre de que forma o investimento das empresas apoiadas é alterado), consideramos que a estimação de impactos até dois anos após o início do tratamento é adequada. Visto que o investimento pode ser uma variável bastante volátil, utilizamos a média do investimento nos primeiros três anos após o tratamento (incluindo o ano de tratamento) como principal variável de resultado – mas averiguamos também os efeitos em cada ano.

6.1.5. Variáveis de controlo

A necessidade de utilização de variáveis de controlo depende do método de avaliação contrafactual aplicado. Quando se utilizam métodos cujo foco é alcançar o equilíbrio das distribuições dos grupos de comparação após a política ser implementada (como o emparelhamento), as variáveis de

controle são a base da validade da metodologia. Por essa razão, o foco desta seção será a decisão sobre que variáveis de controle incluir neste tipo de métodos. Noutros métodos onde o equilíbrio das distribuições é garantido pela seleção (como a seleção aleatória ou o RDD), a utilização de variáveis de controle não é necessária, mas pode aumentar a precisão da estimação e demonstrar a sua robustez.

As variáveis de controle são utilizadas para controlar o efeito de seleção decorrente da atribuição de tratamento, de forma a isolar o impacto do tratamento nos resultados estudados. As variáveis indispensáveis são, por isso, variáveis que simultaneamente influenciam a atribuição do tratamento e os resultados.

Se uma variável influenciar a seleção para o tratamento, mas não influenciar os resultados por outra via que não o próprio tratamento, seria desnecessário (teoricamente) incluí-la como variável de controle. No entanto, como é impossível garantir a inexistência empírica dessa relação, é recomendável incluir todas as variáveis que influenciam o tratamento como variáveis de controle. No caso dos apoios empresariais esta recomendação é ainda mais forte porque os métodos de seleção dessas políticas focam-se em selecionar empresas segundo fatores que estão relacionados com a sua performance.

Se uma variável influenciar os resultados medidos, mas não influenciar a seleção para o tratamento, a sua inclusão não serve para controlar o efeito de seleção porque ele não existe para essa variável. No entanto, a sua inclusão aumenta a precisão estatística da estimação do impacto, sendo por isso recomendável incluir este tipo de variáveis em qualquer método de avaliação.

A escolha das variáveis de controle deve ser informada pela teoria: uma teoria da mudança é a melhor forma de compreender quais os fatores que influenciaram a seleção para a política e, em conjunto com literatura relevante, fornecerá as ferramentas para identificar quais as variáveis que à partida são fulcrais na determinação dos resultados. No entanto, é também possível utilizar dados de forma a investigar empiricamente que variáveis que influenciam o tratamento e as que influenciam os resultados (na ausência de tratamento). A aprendizagem automática é uma boa ferramenta para lidar com situações em que existe um elevado número de variáveis e de unidades.

Caso não sejam incluídas todas as variáveis necessárias para controlar o efeito de seleção (e de uma forma que permita efetivamente controlar o efeito de seleção), a estimação do impacto será incorreta e poderá levar a conclusões erradas sobre o efeito da política. Devem por isso ser considerado um número alargado de possíveis variáveis, mesmo que à partida a probabilidade de algumas delas serem relevantes seja reduzida.

O maior problema é que, em muitos cenários, existem fatores que influenciam o resultado e o tratamento, mas que simplesmente não são observados. Caso haja oportunidade, é importante determinar quais os fatores sobre os quais é necessário obter dados (utilizando a teoria da mudança) para controlar o efeito de seleção (como referido anteriormente, se existirem fatores não observados relevantes, a estimação do impacto será incorreta, sendo esta a principal limitação de métodos de avaliação que recorrem a variáveis de controle).

Caso prático

No nosso caso, devido ao facto de o método escolhido se focar na obtenção de distribuições semelhantes entre os grupos de comparação para as variáveis de controlo relevantes, a decisão sobre as variáveis a incluir é fulcral para a validade da estimação.

A primeira restrição consiste no leque de variáveis incluídas nas bases de dados disponíveis. Em alguns casos, pode fazer sentido proceder a uma recolha de dados, se existirem razões para acreditar que há fatores importantes sobre os quais não existe informação disponível. Dados os objectivos ilustrativos do presente estudo, tal possibilidade não foi considerada. Assim, assumimos que as variáveis da nossa base de dados são suficientes para controlar o efeito de seleção, ou seja, que não há características não observáveis que influenciem simultaneamente a seleção para o tratamento e o resultado.

Como referido na secção 5.4, foram criadas algumas variáveis com o objetivo de permitir retirar informação económica relevante sobre a performance e atividade das empresas. A partir delas, e em conjunto com as variáveis originais, selecionámos algumas variáveis que consideramos possivelmente relevantes para diferenciar o nível de investimento das empresas (a nossa variável de resultado).

A principal preocupação ao criar a lista de variáveis inicial é incluir informação sobre o maior número possível de características das empresas antes do tratamento. Logaritmizamos as variáveis que não são rácios de forma a obter distribuições normais e informação sobre a variação percentual pré-tratamento. Para todas as variáveis, é utilizado o valor do ano imediatamente anterior ao tratamento (2015) ou da variação entre 2014 e 2015 (para as variáveis em variação). A única exceção é a variável “log.media.IAFT_pre”, que é o logaritmo da média do Investimento em Ativos Fixos Tangíveis nos três anos (2013 a 2015) pré-tratamento. Trata-se da variável de resultado no pré-tratamento, que é utilizada desta forma pelas razões apresentadas na secção sobre o momento de estimação de impactos.⁷

O processo de seleção a partir desta lista de variáveis iniciais e de construção dos modelos de estimação é apresentada na próxima secção.

⁷ O investimento é uma variável muito volátil para uma empresa, e não é surpreendente que, ao escolher um único ano específico, se verifiquem valores muito anormais em relação ao investimento normal dessa empresa.

6.2. Estimação de impactos

6.2.1. Estimação da pontuação de propensão

Devido ao método selecionado, antes de proceder à estimação do impacto é necessário estimar a pontuação de propensão (PS) das empresas, que representa a probabilidade de participarem no SI Inovação Produtiva. O modelo de PS é estimado em duas fases.

Em primeiro lugar, estimamos um modelo LASSO em que o tratamento é função de todas as variáveis em X, as variáveis em X ao quadrado e todas as interações entre as variáveis em X. O objetivo é selecionar as variáveis, ou interações entre variáveis, mais importantes para prever o tratamento. O mesmo processo é aplicado utilizando a variável de resultado (o logaritmo da média do investimento de 2016 a 2018), de forma a selecionar as variáveis importantes para explicar o resultado. Esta segunda regressão é estimada apenas nas empresas que não participaram na política porque o objetivo é selecionar os fatores importantes para explicar o resultado na ausência do tratamento.

Após obtermos a lista de variáveis importantes para explicar a atribuição do tratamento e a lista de variáveis importantes para explicar o resultado, estimamos o PS como uma regressão logística utilizando as variáveis selecionadas. Numa versão utilizamos apenas as variáveis selecionadas como importantes para explicar o tratamento e noutra utilizamos o conjunto das duas listas.

A Tabela 10 apresenta o modelo estimado utilizando as variáveis consideradas importantes para explicar o tratamento (identificadas segundo o método LASSO) como regressores. É importante ressaltar que a finalidade deste exercício não é retirar conclusões sobre os fatores que causam a seleção para o tratamento, mas sim criar uma ferramenta (o PS) que permita construir um grupo de tratamento e controlo equilibrados estatisticamente nas características pré-tratamento. Os coeficientes do modelo de PS não devem por isso ser interpretados como indicadores de causalidade e não têm utilidade para a nossa avaliação. O modelo apresentado na Tabela 10 serve apenas como ilustração.

Tabela 10: Modelo da pontuação de propensão utilizando as variáveis consideradas importantes para explicar o tratamento

Variáveis	Coeficientes	Erros Padrão
Log de ativos (t-1)	0.315 ***	0.11
Log de trabalhadores (t-1)	0.359 ***	0.14
Log da produtividade do trabalho (t-1)	0.390	0.27
Log dos gastos médios com pessoal (t-1)	0.444	0.40
Peso das remunerações no VAB (t-1)	0.028 ***	0.01
Variação da rentabilidade do volume de negócios (t-1)	-0.647	0.58
Intensidade exportadora (t-1)	1.040 ***	0.33
Variação da intensidade exportadora (t-1)	1.324	0.89
Efeito dos juros suportados (t-1)	3.103 ***	1.19
Rácio do ativo intangível (t-1)	7.184 *	3.75

Intensidade do uso de eletricidade (t-1)	5.027	4.73
Log da média do IAFT (nos 3 anos pré-tratamento)	0.216 **	0.09
Efeitos fixos de setor		Sim
Número de observações		53 465
Pseudo R ² de McFadden		0.32

Notas: *(**)[***] denota significância ao nível 0.1(0.05)[0.01].

6.2.2. Emparelhamento na pontuação de propensão

Depois de calcularmos os PS das empresas da amostra, utilizamos os valores de PS para realizar o emparelhamento de empresas apoiadas com as empresas não apoiadas com probabilidades de tratamento o mais semelhante possível. O objetivo deste exercício é selecionar um grupo de empresas não tratadas que seja o mais semelhante possível ao grupo de empresas tratadas (em termos da distribuição das características observadas) de forma a poder estimar o impacto do tratamento pela comparação dos dois grupos.

A estratégia de emparelhamento usada é selecionar uma empresa de controlo por cada empresa tratada. A empresa de controlo é aquela com um valor de PS estimado o mais semelhante ao de cada empresa tratada. Adicionalmente, é implementada uma restrição de distância de 0.1 (um *caliper* de 0.1)⁸.

Para aferir a qualidade do método de emparelhamento, é importante confirmar empiricamente a semelhança do grupo de tratamento e controlo selecionados. A Tabela 11 apresenta estatísticas sobre as características dos grupos emparelhados quando são consideradas as variáveis importantes para explicar o tratamento e as variáveis importantes para explicar o resultado, tendo com variável de resultado a média do logaritmo do IAFT de t a t+2. O emparelhamento melhora claramente o equilíbrio estatístico entre os grupos, mas não consegue equilibrar completamente as distribuições.

⁸ Ou seja, se para uma certa empresa tratada o PS da empresa de controlo mais próxima for diferente em mais de 0.1, essa empresa tratada é eliminada do grupo de tratamento final. É importante incluir este tipo de restrições de forma a garantir que os pares emparelhados são suficientemente semelhantes em termos da probabilidade estimada de participar no tratamento.

Tabela 11: Qualidade do emparelhamento (1:1) quando são utilizadas as variáveis importantes para explicar o tratamento e as variáveis importantes para explicar o resultado (quando o resultado é a média do log do IAF de t a t+2)

Variáveis	Diferença das médias estandardizada		Rácio de variâncias	
	Antes do emparelhamento	Após emparelhamento	Antes do emparelhamento	Após emparelhamento
Log da idade (t-1)	0.29 *	-0.07	1.18	1.13
Log do volume de negócios (t-1)	1.29 *	0.09	1.34	0.94
Variação % do vol. negócios (t-1)	0.01	-0.13	1.21	0.61
Log de ativos (t-1)	2.02 *	0.04	0.37 *	0.85
Variação % de ativos (t-1)	0.11	-0.28 *	0.36 *	0.49 *
Log do número de trabalhadores (t-1)	1.48 *	0.06	1.45	0.94
Variação % do número de trabalhadores (t-1)	0.10	-0.26 *	0.96	0.53
Percentagem de trabalhadores com formação superior (t-1)	-0.19	-0.01	0.32 *	0.82
Variação da percentagem de trabalhadores com formação superior (t-1)	-0.05	-0.04	0.81	1.45
Log da produtividade do trabalho (t-1)	0.39 *	0.06	1.93 *	0.74
Variação % da produtividade do trabalho (t-1)	-0.12	0.05	2.92 *	0.44 *
Log dos gastos médios com pessoal (t-1)	0.97 *	0.01	0.54	0.57
Variação do log dos gastos médios com pessoal (t-1)	0.07	0.03	0.42 *	0.64
Peso das remunerações no VAB (t-1)	0.11	0.00	190.09 *	0.14 *
Rentabilidade do volume de negócios (t-1)	-0.06	0.08	8.03 *	0.32 *
Variação da rentabilidade do volume de negócios (t-1)	-0.13	0.09	5.51 *	0.38 *
Intensidade exportadora (t-1)	0.84 *	0.15	3.85 *	0.90
Variação da intensidade exportadora (t-1)	0.23	0.06	2.66 *	0.63
Endividamento (t-1)	0.18	0.02	0.54	0.54
Efeito dos juros suportados (t-1)	0.15	-0.05	37.71 *	0.14 *
Autonomia financeira (t-1)	0.04	-0.12	0.45 *	0.61
Solvabilidade (t-1)	-0.25 *	-0.21	0.46 *	0.70
Liquidez geral (t-1)	-0.34 *	0.00	0.68	2.03 *
Rentabilidade dos ativos (t-1)	0.01	0.07	0.84	0.88
Rácio do ativo intangível (t-1)	0.12	-0.01	1.45	0.85
Intensidade do uso de combustíveis (t-1)	-0.22	-0.02	0.68	1.08
Intensidade do uso de eletricidade (t-1)	0.41 *	0.03	1.18	0.73
Log da média do IAF (t-3 a t-1)	1.73 *	0.04	0.44 *	0.65
Número de empresas tratadas	87	85	87	85
Número de empresas de controlo	53 378	2 008	53 378	2 008

Notas: * denota uma diferença na média estandardizada superior a 0.25 ou um coeficiente de variação fora do intervalo [0.5,2]. Estes são critérios comuns para considerar se as distribuições são significativamente diferentes. O número de empresas de controlo na amostra emparelhada é ponderado de forma a ser igual ao de empresas tratadas.

6.2.3. Estimação do impacto médio

O impacto médio pode ser estimado pela comparação direta da variável de resultado do grupo de tratamento e do grupo de controlo obtidos através do emparelhamento, mas isto pressupõe que as diferenças pré-tratamento entre os dois grupos foram perfeitamente eliminadas. Como esta é uma hipótese demasiado forte (e confirmámos na secção anterior que podem existir algumas diferenças significativas), utilizamos estimadores de diferença-em-diferenças (ou seja, estamos a comparar as variações dos dois grupos) de forma a eliminar diferenças causadas por efeitos fixos. Na nossa especificação preferencial, recorreremos também a uma regressão com as variáveis seleccionadas como importantes aquando da estimação do PS para corrigir o enviesamento decorrente de desequilíbrios estatísticos não corrigidos pelo emparelhamento (como discutido em 6.1.1).

A Tabela 12 apresenta as estimações de impacto obtidas segundo o processo de emparelhamento (com três formas diferentes de calcular o PS) e regressão de OLS. Para além da variável de resultado preferencial (a primeira linha), são apresentadas as estimações de impacto obtidas seguindo diferentes decisões críticas. O processo de estimação para estas diferentes decisões segue exactamente a mesma metodologia, incluindo a estimação de um modelo de PS para cada caso.

Tabela 12: Resultados obtidos segundo diferentes estratégias de estimação e indicadores

Indicadores	OLS	Emparelhamento (variáveis importantes para explicar o tratamento)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado, com correção de enviesamento)
Média do log do IAFI (t a t+2)	1.039 *** (0.336)	1.363 *** (0.238)	1.672 *** (0.308)	1.407 *** (0.257)
Log do IAFI (t)	0.733 * (0.414)	1.378 *** (0.479)	1.009 *** (0.361)	0.986 ** (0.406)
Log do IAFI (t+1)	1.369 *** (0.426)	1.952 *** (0.493)	1.371 *** (0.386)	1.345 *** (0.397)
Log do IAFI (t+2)	1.503 *** (0.436)	1.577 *** (0.387)	1.505 *** (0.398)	1.462 *** (0.398)
Média do log do IAFI usando apenas candidatas não apoiadas (t a t+2)	0.453 (0.333)	0.403 (0.376)	0.320 * (0.188)	-0.669 (0.873)

Média do log do IAFI usando o início de tratamento alternativo (t a t+2)	1.115 *** (0.354)	1.119 *** (0.265)	1.245 *** (0.267)	1.249 *** (0.268)
Log da produtividade do trabalho (t)	0.059 (0.038)	0.067 *** (0.026)	0.052 ** (0.027)	0.064 ** (0.026)
Log da produtividade do trabalho (t+1)	0.065 (0.044)	0.058 (0.058)	0.067 (0.053)	0.061 ** (0.031)
Log da produtividade do trabalho (t+2)	0.076 (0.049)	0.073 (0.063)	0.067 (0.061)	0.058 (0.036)

Notas: *(**)[***] denota significância ao nível 0.1(0.05)[0.01]. Na estimação por OLS são incluídas todas as variáveis consideradas importantes para explicar o tratamento e resultado como controlos. O emparelhamento é sempre 1 para 1. Entre parênteses constam as medidas de erro.

Como podemos observar na Tabela 12, estima-se que o SI Inovação Produtiva teve um efeito positivo sobre o investimento das empresas apoiadas (segundo a nossa estimação preferencial). O efeito representa um aumento bastante elevado – as empresas apoiadas tiveram um investimento em cerca de 140% superior ao que teriam se não tivessem sido apoiadas, nos primeiros três anos.

Quando avaliamos o aumento do investimento por ano, chegamos a conclusões semelhantes em termos de dimensão e significância das estimações, sendo que os efeitos positivos parecem ser maiores no 2º e 3º anos (t+1 e t+2). Esta conclusão era esperada porque o primeiro ano pode incluir projetos que começaram no final desse ano ou apenas no ano seguinte.

Quando consideramos um critério diferente para a determinação do início do tratamento – o ano de início do projeto (discutido em 6.1.3), obtemos impactos ligeiramente inferiores. Esta diferença não é surpreendente porque a escolha do momento de tratamento foi motivada pela preocupação de que podem existir efeitos assim que as empresas sabem que vão receber o financiamento e não apenas quando iniciam os projetos.

Adicionalmente, restringimos a nossa amostra de controlo apenas às empresas que se candidataram à política, mas acabaram por não ser selecionadas para financiamento. Esta questão é importante porque a diferença na decisão de apresentar uma candidatura pode sinalizar diferenças não observáveis entre empresas candidatas e não candidatas, que não são controladas pelas variáveis que temos à nossa disposição. No nosso caso em particular, a decisão de considerar apenas empresas candidatas acaba por prejudicar bastante a qualidade do emparelhamento, porque o número de empresas candidatas não apoiadas é bastante reduzido (tornando impossível encontrar bons pares). Optamos por não dar importância às estimações obtidas segundo esta estratégia porque consideramos que não tem validade (apesar de não apresentarmos as estatísticas de balance, apresentamos um exercício de robustez – na seção 6.4– que ilustra bem a falta de qualidade do emparelhamento com empresas candidatas).

Por fim, apresentamos as estimações obtidas para o impacto do SI Inovação Produtiva sobre a produtividade. Já discutimos anteriormente o porquê de optarmos pelo investimento como variável preferencial para a nossa avaliação, mas pretendemos ilustrar que ao utilizar diferentes variáveis de resultado, podemos chegar a conclusões diferentes sobre o impacto da política. A estimação do impacto sobre o investimento acaba por ilustrar também um efeito positivo, mas menos forte do que sobre a produtividade. Não é surpreendente que os efeitos sobre o investimento sejam mais fortes do que sobre a produtividade, porque segundo a teoria da mudança, o impacto sobre o investimento é esperado como um resultado imediato, enquanto o aumento da produtividade é esperado apenas como um resultado de médio-prazo. As estimações obtidas acabam por surpreender um pouco porque estimamos efeitos significativamente positivos no primeiro ano e não no segundo, quando postulávamos que, se obtivéssemos estimações significativamente positivas, elas deveriam aparecer no último ano (no entanto, os coeficientes são semelhantes nos três anos e não é surpreendente que a significância estatística diminua devido ao aumento da variância que ocorre em estimações com horizontes temporais mais longos).

6.3. Análise de heterogeneidade

Todos os métodos de avaliação contrafactual têm limitações no que toca à generalização de conclusões sobre as estimações obtidas. É muito importante saber que conclusões não devemos retirar em função da metodologia utilizada, e as avaliações devem expor essas limitações. No nosso caso, utilizámos um método de emparelhamento que se foca em selecionar empresas de controlo o mais semelhante possível a cada empresa tratada, estimando o efeito de tratamento médio para as empresas tratadas (ATT). Devido a isto, as estimações obtidas na secção anterior são estimações sobre o efeito médio do tratamento para as empresas tratadas. Ou seja, o efeito de apoiar um grupo de empresas diferentes daquelas que realmente receberam o apoio pode ser completamente diferente. Para além disto, o efeito médio de tratamento acaba por oferecer informação limitada sobre o impacto da política porque diferentes empresas poderão ser impactadas de forma muito assimétrica.

É importante realizar uma análise de heterogeneidade quando o contexto permite. Uma análise de heterogeneidade permite retirar conclusões importantes sobre quais os grupos que mais benefícios retiraram da política e oferecer pistas sobre a forma como a política transforma as empresas apoiadas. Estas questões são fundamentais para aperfeiçoar as políticas, tornando-as mais benéficas ao seu público-alvo.

Caso prático

No nosso caso, a análise de heterogeneidade é muito difícil devido ao pequeno número de empresas apoiadas. O principal problema é que, querendo investigar se a política impacta empresas com diferentes características de forma assimétrica, é necessário termos um elevado número de empresas onde diferentes combinações dessas características possam ser verificadas e que permitam realizar estimações com validade estatística. Apesar disto, apresentamos um exemplo onde testamos a existência de heterogeneidade em função do endividamento das empresas. Utilizamos o endividamento existente como indicador de constrangimentos ao financiamento privado das empresas apoiadas. Como vimos na teoria da mudança discutida na secção 2.2, um dos pressupostos do aumento do investimento por parte das empresas apoiadas é a existência de constrangimentos financeiros ao

investimento privado. Investigamos, por isso, se a política tem maior impacto em empresas com maiores níveis de endividamento (ou seja, que estão à partida mais constrangidas financeiramente).

Existem várias estratégias para investigar a existência de heterogeneidade. A mais comum e simples é dividir a amostra de empresas tratadas segundo a variável onde queremos verificar a existência de heterogeneidade (no nosso caso, o endividamento). Depois, realizamos a estimação do impacto da política em cada subgrupo (e.g., um subgrupo é a metade das empresas tratadas com menor endividamento e o outro é a metade das empresas tratadas com maior endividamento).

A Tabela 13 ilustra os resultados deste exercício. Como podemos verificar, o nosso pressuposto inicial sai confirmado, visto que as empresas com maiores níveis de endividamento são aquelas onde o investimento aumentou mais (um aumento 50% superior) por receber o apoio. Esta estratégia tem, no entanto, uma grande limitação: é impossível saber se esta diferença se deve ao endividamento em si ou a alguma outra característica que seja diferente nos dois grupos e que esteja correlacionada com o endividamento. Ou seja, ao utilizar esta estratégia, não devemos retirar conclusões sobre as causas da heterogeneidade.

Tabela 13: Impacto estimado no investimento para subgrupos divididos segundo o endividamento

Grupo	OLS	Emparelhamento (variáveis impor- tantes para explicar o tratamento)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado, com correção de enviesamento)
A metade das em- presas tratadas com menor endi- vidamento	0.879 * (0.470)	1.131 *** (0.252)	1.281 *** (0.305)	1.222 *** (0.407)
A metade das em- presas tratadas com maior endivi- damento	1.510 *** (0.485)	1.408 *** (0.218)	1.589 *** (0.195)	1.873 *** (0.315)

Notas: A variável de resultado é a média do log do IAFI (de t a $t+2$). *(**)[***] denota significância ao nível 0.1(0.05)[0.01]. Na estimação por OLS são incluídas todas as variáveis consideradas importantes para explicar o tratamento e resultado como controlo e efeitos fixos de setor.

Outra estratégia mais complexa passa por tentar modelar os efeitos individuais estimados para cada empresa em função das características de cada empresa. Desta forma, podemos investigar simultaneamente como diferentes características das empresas impactam o efeito da política. No entanto, isto pressupõe que o método de estimação utilizado consiga produzir estimações individuais robustas, enquanto o foco da maioria dos métodos de estimação é em produzir estimações válidas do impacto médio. Isto implicaria utilizar outra metodologia, como por exemplo o controlo sintético ou aprendizagem automática, mas, sobretudo por falta de número de empresas apoiadas, não conseguimos explorar esta estratégia neste contexto.

6.4. Análise de robustez

Os testes de robustez são essenciais a qualquer avaliação contrafactual (sobretudo às que utilizam métodos de controlo por observáveis) para testar a validade dos resultados obtidos. Os testes de robustez servem para aferir se a estimação de impactos se modifica significativamente ao alterar algumas especificações da metodologia, nomeadamente ligeiras alterações ao método ou aos dados utilizados. Também se pode voltar a estimar o impacto utilizando um método completamente diferente se não houver fortes razões para a preferência do método principal. Podem ser também realizados testes de placebo (ver secção seguinte).

Na secção 6.2 apresentámos algumas estimações que podem ser consideradas como testes de robustez (e.g., ao estimar o impacto utilizando a data de início do projeto como início de tratamento e utilizando apenas as candidatas não apoiadas como possíveis controlos). Nesta secção, apresentamos mais alguns testes de robustez que não estão relacionados com as decisões críticas discutidas no início do Capítulo .

6.4.1. Teste de placebo

Os testes de placebo pretendem aferir se o método escolhido consegue controlar o efeito de seleção em situações onde se sabe que não existe efeito de tratamento. A melhor forma de testar isto é recorrer à variável de resultado num período pré-tratamento que não tenha sido usado como variável de controlo. No nosso caso, este exercício serve também para testar a existência de tendências paralelas entre os grupos de empresas emparelhadas.

A Tabela 14 utiliza os grupos criados por emparelhamento para estimar o efeito do tratamento no

IAFT das empresas em $t-4$ e $t-5$ (i.e., quatro e cinco anos antes de receberem o apoio, respetivamente). Obviamente, sabemos à partida que os apoios oferecidos em 2016 não tiveram qualquer efeito no investimento de 2011 ou 2012. Este exercício serve para comprovar se conseguimos chegar a essa conclusão através do nosso método de estimação (ou se, pelo contrário, os grupos de tratamento e de controlo apresentavam diferenças relevantes antes da política produzir qualquer efeito).

Os resultados apresentados na Tabela 14 dão-nos indícios claros sobre o enviesamento da estimação apenas no caso onde são utilizadas exclusivamente empresas candidatas como possíveis controlos. Este resultado deve-se ao número de empresas candidatas ser muito pequeno, tornando impossível construir grupos de tratamento e controlo com características semelhantes.

Tabela 14: Estimação do impacto do placebo para diferentes modelos de estimação

Modelo testado	OLS	Emparelhamento (variáveis importantes para explicar o tratamento)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado, com correção de enviesamento)
Média do log do IAFI (t a t+2)	-0.511 (0.340)	-0.117 (0.318)	0.145 (0.381)	-0.111 (0.291)
Log do IAFI (t)	-0.206 (0.340)	-0.117 (0.318)	-0.088 (0.323)	-0.268 (0.292)
Log do IAFI (t+1)	-0.217 (0.340)	-0.117 (0.318)	-0.088 (0.323)	-0.107 (0.305)
Log do IAFI (t+2)	-0.262 (0.339)	-0.117 (0.318)	-0.218 (0.300)	-0.240 (0.301)
Média do log do IAFI usando apenas candidatas não apoiadas (t a t+2)	-1.076 ** (0.481)	-0.364 (0.645)	-0.892 * (0.495)	-2.159 (1.794)
Média do log do IAFI usando o início de tratamento alternativo (t a t+2)	-0.223 (0.356)	-0.332 (0.366)	-0.297 (0.368)	-0.256 (0.310)
Log da produtividade do trabalho (t)	0.030 (0.05)	0.034 (0.05)	0.049 (0.05)	0.052 (0.05)
Log da produtividade do trabalho (t+1)	0.030 (0.05)	0.035 (0.05)	0.050 (0.05)	0.046 (0.05)
Log da produtividade do trabalho (t+2)	0.027 (0.05)	0.035 (0.05)	0.046 (0.05)	0.031 (0.05)

Notas: *(**)[***] denota significância ao nível 0.1(0.05)[0.01]. Na estimação por OLS são incluídas todas as variáveis consideradas importantes para explicar o tratamento e resultado como controlos.

A variável de resultado onde o placebo é estimado é a “média do log do IAFI (t-5 a t-4)” quando a variável de resultado original é o IAFI, e a “média do log da produtividade (t-5 a t-4)” quando a variável de resultado original é a produtividade.

Os testes de placebo não nos permitem comprovar indubitavelmente que o nosso método de estimação é adequado. No entanto, permitem testar se ele é inadequado e descartar essas metodologias (no nosso caso, descartamos a utilização de apenas empresas candidatas como controlos).

6.4.2. Estimação sem outliers

Uma questão que pode ser relevante explorar é a existência de *outliers* e o seu possível impacto na estimação de efeitos médios. A principal preocupação é que um pequeno número de empresas muito diferentes da empresa média estejam a dominar a estimação do impacto médio.

A Tabela 15 mostra o resultado da estimação quando retiramos *outliers* da nossa amostra de dados. Os resultados acabam por ser muito semelhantes aos obtidos originalmente, não havendo razão para acreditar que seja um problema no nosso caso.

Tabela 15: Estimação de impacto na amostra sem outliers

Indicadores	OLS	Emparelhamento (variáveis importantes para explicar o tratamento)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado, com correção de enviesamento)
Média do log do IAFI (t a t+2)	1.501 *** (0.325)	1.806 *** (0.256)	1.447 *** (0.229)	1.446 *** (0.23)

Notas: *(**)[***] denota significância ao nível 0.1(0.05)[0.01]. Na estimação por OLS são incluídas todas as variáveis consideradas importantes para explicar o tratamento e resultado como controlos. Considera-se que uma empresa é um *outlier* se o valor de qualquer variável estiver acima do 3º quartil, ou abaixo do 1º quartil, numa dimensão superior a 1.5 vezes o intervalo interquartil da distribuição de empresas tratadas.

6.4.3. Emparelhamento com múltiplas empresas de controlo

Por fim, exploramos a hipótese de emparelhar cada empresa tratada com mais do que uma empresa de controlo. Esta estratégia permite testar se a estimação é dependente do grupo de empresas de controlo composto unicamente pelas empresas mais próximas a cada empresa tratada. À partida, se o número de empresas tratadas for estatisticamente grande, não haverá razões para que isto aconteça. A Tabela 16 mostra que expandir o emparelhamento às 2 ou 5 empresas não tratadas mais próximas, não muda os resultados obtidos.

Tabela 16: Resultados segundo a estratégia de emparelhamento

Tipo de emparelhamento	Emparelhamento (variáveis importantes para explicar o tratamento)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado)	Emparelhamento (variáveis importantes para explicar o tratamento e resultado, com correção de enviesamento)
O vizinho mais próximo	1.363 *** (0.238)	1.672 *** (0.308)	1.407 *** (0.257)
Os 2 vizinhos mais próximos	1.418 *** (0.23)	1.553 *** (0.247)	1.421 *** (0.247)
Os 5 vizinhos mais próximos	1.460 *** (0.225)	1.503 *** (0.297)	1.420 *** (0.304)

Notas: *(**)[***] denota significância ao nível 0.1(0.05)[0.01]. Na estimação por OLS são incluídas todas as variáveis consideradas importantes para explicar o tratamento e resultado como controlos.

7. Referências

- Angrist, J. D., & Krueger, A. B. (1999). *Empirical Strategies in Labor Economics* (pp. 1277–1366). [https://doi.org/10.1016/S1573-4463\(99\)03004-7](https://doi.org/10.1016/S1573-4463(99)03004-7)
- Arrow, K. J. (1972). Economic Welfare and the Allocation of Resources for Invention. In *Readings in Industrial Economics* (pp. 219–236). Macmillan Education UK. https://doi.org/10.1007/978-1-349-15486-9_13
- Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360. <https://doi.org/10.1073/pnas.1510489113>
- Athey, S., Imbens, G. W., & Wager, S. (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4), 597–623. <https://doi.org/10.1111/rssb.12268>
- Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2). <https://doi.org/10.1214/18-AOS1709>
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81(2), 608–650. <https://doi.org/10.1093/restud/rdt044>
- Beňkovskis, K., Tkačevs, O., & Yashiro, N. (2019). Importance of EU regional support programmes for firm performance*. *Economic Policy*, 34(98), 267–313. <https://doi.org/10.1093/epolic/eiz003>
- Bloom, N., Schankerman, M., & Reenen, J. Van. (2013). Identifying Technology Spillovers and Product Market Rivalry. *Econometrica*, 81(4), 1347–1393. <https://doi.org/10.3982/ECTA9466>
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297–312. <https://doi.org/10.1037/h0040950>
- Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.
- Castillo, V., Figal Garone, L., Maffioli, A., & Salazar, L. (2017). The causal effects of regional industrial policies on employment: A synthetic control approach. *Regional Science and Urban Economics*, 67, 25–41. <https://doi.org/10.1016/j.regsciurbeco.2017.08.003>
- Chen, H.-T. (1990). *Theory-driven evaluations*. Sage .
- Chen, H.-T., & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning*, 10(1), 95–103. [https://doi.org/10.1016/0149-7189\(87\)90025-5](https://doi.org/10.1016/0149-7189(87)90025-5)
- Aviso N.º 13/SI/2021 - Inovação Produtiva, (2021) (testimony of COMPETE2020).
- Cook, T. D. (2000). The false choice between theory-based evaluation and experimentation. *New Directions for Evaluation*, 2000(87), 27–34. <https://doi.org/10.1002/ev.1179>
- Crespi, G., Figal Garone, L., Maffioli, A., & Stein, E. (2020). Public support to R&D, productivity, and spillover effects: Firm-level evidence from Chile. *World Development*, 130, 104948.

<https://doi.org/10.1016/j.worlddev.2020.104948>

Criscuolo, C., Martin, R., Overman, H. G., & Van Reenen, J. (2019). Some Causal Effects of an Industrial Policy. *American Economic Review*, 109(1), 48–85. <https://doi.org/10.1257/aer.20160034>

Cunningham, S. (2021). *Causal Inference*. Yale University Press. <https://doi.org/10.12987/9780300255881>

Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48(2), 424–455. <https://doi.org/10.1257/jel.48.2.424>

Gates, E., & Dyson, L. (2017). Implications of the Changing Conversation About Causality for Evaluators. *American Journal of Evaluation*, 38(1), 29–46. <https://doi.org/10.1177/1098214016644068>

Hall, B. H., & Lerner, J. (2010). *The Financing of R&D and Innovation* (pp. 609–639). [https://doi.org/10.1016/S0169-7218\(10\)01014-2](https://doi.org/10.1016/S0169-7218(10)01014-2)

Heckman, J. J. (1991). *Randomization and Social Policy Evaluation: Revisited*.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, 64(4), 605–654. <https://doi.org/10.2307/2971733>

Hedström, P. (2006). *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge University Press.

Imbens, G. (2013). Book Review Feature: Public Policy in an Uncertain World. *The Economic Journal*, 123(570), F401–F411. <https://doi.org/10.1111/ecoj.12065>

Imbens, G. (2014). *Matching Methods in Practice: Three Examples*. <https://doi.org/10.3386/w19959>

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467. <https://doi.org/10.2307/2951620>

Judd, C. M., & Kenny, D. A. (1981). Process Analysis. *Evaluation Review*, 5(5), 602–619. <https://doi.org/10.1177/0193841X8100500502>

Kleine, M., Heite, J., & Rosendahl Huber, L. (2020). Subsidized R&D Collaboration: The Causal Effect of Innovation Vouchers on Innovation Performance. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3648126>

Leamer, E. E. (2010). Tantalus on the Road to Asymptopia. *Journal of Economic Perspectives*, 24(2), 31–46. <https://doi.org/10.1257/jep.24.2.31>

Lee, B. K., Lessler, J., & Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29(3), 337–346. <https://doi.org/10.1002/sim.3782>

Leeuw, F. L. (2012). Linking theory-based evaluation and contribution analysis: Three problems and a few solutions. *Evaluation*, 18(3), 348–363. <https://doi.org/10.1177/1356389012452051>

Li, J., Lee, R. P., & Wan, J. (2020). Indirect effects of direct subsidies: an examination of signaling effects. *Industry and Innovation*, 27(9), 1040–1061.

<https://doi.org/10.1080/13662716.2019.1685375>

- Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*. Oxford, Clarendon Press.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Taylor & Francis Group/Lawrence Erlbaum Associates.
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18(3), 270–280. <https://doi.org/10.1177/1356389012451663>
- Mayne, J. (2015). Useful theory of change models. *Canadian Journal of Program Evaluation*, 30(2), 119–142.
- Pawson, R. (2006). *Evidence-Based Policy*. Sage.
- Pawson, R. (2008). Invisible Mechanisms. *Evaluation Journal of Australasia*, 8(2), 3–13. <https://doi.org/10.1177/1035719X0800800202>
- Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. SAGE Publications Ltd. <https://doi.org/10.4135/9781473913820>
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.
- Ragin, C. C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. In University of California Press. University of California Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.
- Sharpe, R. A., & Bhaskar, R. (1976). A Realist Theory of Science. *The Philosophical Quarterly*, 26(104), 284. <https://doi.org/10.2307/2219031>
- Srhoj, S., Lapinski, M., & Walde, J. (2021). Impact evaluation of business development grants on SME performance. *Small Business Economics*, 57(3), 1285–1301. <https://doi.org/10.1007/s11187-020-00348-6>
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations* (No. 38).
- Stevenson, R., Kier, A. S., & Taylor, S. G. (2021). Do policy makers take grants for granted? The efficacy of public sponsorship for innovative entrepreneurship. *Strategic Entrepreneurship Journal*, 15(2), 231–253. <https://doi.org/10.1002/sej.1376>
- Szücs, F. (2020). Do research subsidies crowd out private R&D of large firms? Evidence from European Framework Programmes. *Research Policy*, 49(3), 103923. <https://doi.org/10.1016/j.respol.2020.103923>
- Trochim, W. M. K. (1989a). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, 12(1), 1–16. [https://doi.org/10.1016/0149-7189\(89\)90016-5](https://doi.org/10.1016/0149-7189(89)90016-5)
- Trochim, W. M. K. (1989b). Outcome pattern matching and program theory. *Evaluation and Program*

Planning, 12(4), 355–366. [https://doi.org/10.1016/0149-7189\(89\)90052-9](https://doi.org/10.1016/0149-7189(89)90052-9)

Wang, Y., Li, J., & Furman, J. L. (2017). Firm performance and state innovation funding: Evidence from China's Innofund program. *Research Policy*, 46(6), 1142–1161. <https://doi.org/10.1016/j.respol.2017.05.001>

Weiss, C. H. (1972). *Evaluation research: Methods for assessing program effectiveness*. Prentice-Hall.

Wolpin, K. I. (2013). *The limits of inference without theory*. MIT Press.

Wu, A. (2017). The signal effect of Government R&D Subsidies in China: Does ownership matter? *Technological Forecasting and Social Change*, 117, 339–345. <https://doi.org/10.1016/j.techfore.2016.08.033>

Wu, R., Liu, Z., Chen, X., & Liao, S. (2021). Certification effect of R&D subsidies on debt financing: do institutional forces matter? *R&D Management*, 51(5), 538–550. <https://doi.org/10.1111/radm.12465>

Zubizarreta, J. R. (2015). Stable Weights that Balance Covariates for Estimation With Incomplete Outcome Data. *Journal of the American Statistical Association*, 110(511), 910–922. <https://doi.org/10.1080/01621459.2015.1023805>

8. Anexos

8.1. Estatísticas Descritivas

Tabela 17: Empresas candidatas aos SI do PT2020, por tipo de medida

Medida	Nº de candidatas	Nº de apoiadas com execução
Mais do que uma medida	9 623	5 148
ADAPTAR PME	2 445	20
Formação Autónoma	104	9
Formação-Ação para PME	121	48
I&D - Copromoção - COVID-19	12	0
I&D - Individuais - COVID-19	27	0
I&D - Infraest. Ensaio e Otimização - COVID-19	2	0
I&DT - Copromoção	159	18
I&DT - Copromoção - RCI	1	0
I&DT - Demonstradores Copromoção	8	1
I&DT - Demonstradores Individuais	17	1
I&DT - Individuais	175	38
I&DT - Individuais - RCI	6	0
I&DT - Internacionalização	3	0
I&DT - Núcleos Copromoção	4	0
I&DT - Núcleos Individuais	5	0
I&DT - Programas Mobilizadores	1	1
I&DT - Vales	328	88
Inovação - Produtiva - COVID-19	484	76
Inovação - Empreendedorismo	417	34
Inovação - Produtiva	2 148	360
Inovação - RCI	33	7
Inovação - Vales	1 551	350
QI PME - Conjuntos	27	6
QI PME - Individuais	2 380	645
QI PME - Vales	3 022	704
Todas	23 103	7 554

Notas: Consideram-se empresas apoiadas com execução aquelas cujo percentil de valor de execução seja superior a zero. As contagens correspondentes a cada medida correspondem a empresas com apenas um projeto.

Tabela 18: Empresas apoiadas em um ou mais projetos

Número de projetos	Número de empresas apoiadas
1	2 406
2	1 749
3	1 132
4	813
5 ou +	1 454

Tabela 19: Número de projetos do PT2020 por início e fim

Ano de início de projeto	Ano de fim de projeto									
	2015	2016	2017	2018	2019	2020	2021	2022	2023	Sem ano de fim
2014	1	0	0	0	1	0	0	1	0	0
2015	33	1 263	939	553	304	408	750	293	58	0
2016		72	674	1 238	681	348	494	176	37	0
2017			53	343	870	450	328	70	14	0
2018				19	445	652	391	57	4	0
2019					25	383	1 260	120	12	0
2020						1 534	1 990	373	25	0
2021							4	18	13	0
Sem ano de início									0	5 326

Notas: Empresas com mais do que um projeto são consideradas como tendo um único projeto, com início no início do seu primeiro projeto e fim no fim do seu último projeto.

Tabela 20: Empresas candidatas aos SI do PT2020, por sector de atividade

	Nº de candidatas	Nº de candidatas com execução
Agricultura, produção animal, caça, floresta e pesca	388	127
Indústrias extrativas	92	41
Indústrias transformadoras	6 502	3 186
Eletricidade, gás, vapor, água quente e fria e ar frio	24	2
Captação, tratamento e distribuição de água; saneamento, gestão de resíduos e despoluição	130	41
Construção	940	308
Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos	4 541	1 335
Transportes e armazenagem	399	109
Alojamento, restauração e similares	1 879	310
Atividades de informação e de comunicação	1 558	565
Atividades financeiras e de seguros	0	0
Atividades imobiliárias	292	25
Atividades de consultoria, científicas, técnicas e similares	2 525	785
Atividades administrativas e dos serviços de apoio	791	211
Administração Pública e Defesa; Segurança Social Obrigatória	0	0
Educação	326	38
Atividades de saúde humana e apoio social	601	53
Atividades artísticas, de espetáculos, desportivas e recreativas	441	91
Outras atividades de serviços	435	216
Atividades das famílias empregadoras de pessoal doméstico e atividades de produção das famílias para uso próprio	0	0
Atividades dos organismos internacionais e outras instituições extraterritoriais	0	0
Total	21 864	7 443

Notas: A classificação de atividade económico é referente à empresa, tal como classificada no SCIE, e não à classificação de atividade económica do projeto.

Tabela 21: Empresas candidatas aos SI do PT2020, por classe de dimensão de empresas

	Nº de candidatas	Nº de apoiadas com execução
Micro	11 403	2 906
Pequena	7 356	2 792
Média	2 542	1 424
Grande	563	321
Total	21 864	7 443

Notas: A classe de dimensão da empresa é construída segunda a informação contida SCIE..

Tabela 22: Contagem de empresas no SCIE e classificação por tamanho

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Sociedades	361.235	361.851	355.769	356.577	363.356	372.201	380.935	394.967	413.767	438.959
Micro %	87,3	88,0	88,8	89,2	89,2	89,0	88,9	88,6	88,6	88,8
Pequenas %	10,5	9,9	9,2	8,8	8,8	9,0	9,1	9,3	9,3	9,1
Médias %	1,8	1,7	1,6	1,6	1,6	1,6	1,6	1,7	1,7	1,7
Grandes %	0,4	0,4	0,4	0,5	0,4	0,4	0,4	0,4	0,4	0,4

Tabela 23: Percentagem de empresas no SCIE por primeiro e último ano com registo

1º ano	Último ano									
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2010	4,2	4	3,4	2,7	2,3	2	1,6	1,4	1,4	31,5
2011		0,5	0,6	0,4	0,3	0,3	0,2	0,2	0,2	2,3
2012			0,3	0,4	0,4	0,3	0,2	0,2	0,2	2,2
2013				0,4	0,4	0,4	0,3	0,2	0,2	3
2014					0,3	0,4	0,4	0,3	0,2	3,1
2015						0,3	0,5	0,4	0,3	3,6
2016							0,3	0,4	0,4	3,8
2017								0,3	0,5	4,5
2018									0,4	5,5
2019										6,5

Tabela 24: Descritivas gerais de variáveis selecionadas de todas as sociedades

	Média	Mediana	Desvio-Padrão	Mínimo	Máximo	% de zeros	% de negativos
Pessoal ao serviço	7,4	2	89	1	26 559	0	0
Pessoal ao serviço afeto à I&D	0	0	1	0	374	99,6	0
Gastos com pessoal	138	19	2 157	-1	694.485	21,1	0
Formação bruta de capital fixo (FBCF)	51	0	1 735	-305 968	384 854	55,4	4,5
Juros e gastos similares suportados	16	0	1 320	0	567 070	56,5	0
Vendas no mercado comunitário	130	0	7 052	0	3 321 273	87,6	0
Vendas no mercado extracomunitário	50	0	3 690	0	1 842 884	93,4	0
Volume de negócios	904	76	22 438	0	8 050 298	13,5	0
Valor acrescentado bruto (VABcf)	221	26	3 898	-112 713	1 108 653	0,5	21,1
Resultado líquido	44	1	4 358	-2 184 797	754 395	0,1	39,8
Capital próprio	495	21	17 022	-1 116 994	6 724 581	0,6	20,3

Notas: os valores monetários estão expressos em milhares de euros

Tabela 25: Descritivas gerais de variáveis selecionadas apenas de empresas pequenas e médias

	Média	Mediana	Desvio-Padrão	Mínimo	Máximo	% de zeros	% de negativos
Pessoal ao serviço	27	18	28	1	200	0	0
Pessoal ao serviço afeto à I&D	0	0	1	0	129	98,5	0
Gastos com pessoal	545	299	808	0	51 580	1,5	0
Formação bruta de capital fixo (FBCF)	143	15	2.214	-305 968	244 012	20,2	6,9
Juros e gastos similares suportados	51	3	798	0	69 646	22	0
Vendas no mercado comunitário	391	0	1.702	0	49 488	63	0
Vendas no mercado extracomunitário	163	0	1.113	0	45 632	77,5	0
Volume de negócios	3 159	1 410	5 244	0	49 938	0,4	0
Valor acrescentado bruto (VABcf)	853	427	1 634	-112 713	39 753	0,01	1,6
Resultado líquido	226	29	5 284	-75 349	754 395	0,04	17,2
Capital próprio	1 840	308	26 499	-1 116 994	3 634 591	0,1	7,7

Nota: os valores monetários estão expressos em milhares de euros

Tabela 26: Percentagem de empresas no QP por primeiro e último ano com registo

1º ano	Último ano									
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2010	4,8	4,6	3,6	2,6	2,4	1,9	1,7	1,7	2,7	27,5
2011		1	0,7	0,5	0,4	0,3	0,3	0,3	0,6	2,7
2012			0,7	0,5	0,4	0,3	0,2	0,2	0,3	2,1
2013				0,6	0,6	0,4	0,3	0,3	0,4	2,7
2014					0,7	0,5	0,4	0,4	0,5	2,8
2015						0,7	0,5	0,5	0,5	3,1
2016							0,7	0,6	0,6	3,1
2017								0,8	0,8	3,4
2018									1,0	3,8
2019										4,9

Tabela 27: Distribuição de ganhos médios mensais por ano (euros)

	Média	P10	P25	P50	P75	P90	P99	Máximo
2010	883	119	510	668	1.006	1.680	4.315	333.334
2011	951	374	548	703	1.058	1.739	4.454	333.334
2012	962	373	554	713	1.074	1.763	4.487	333.334
2013	954	350	551	706	1.067	1.752	4.443	285.715
2014	948	340	561	705	1.053	1.734	4.398	285.715
2015	949	334	562	708	1.056	1.731	4.400	416.667
2016	956	334	580	717	1.059	1.729	4.418	654.166
2017	973	334	606	740	1.077	1.739	4.426	654.166
2018	1.001	342	633	769	1.110	1.780	4.488	325.000
2019	1.032	351	655	800	1.151	1.831	4.545	361.288

Tabela 28: Tipo de contrato dos trabalhadores por ano (%)

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Contrato sem termo	70,0	68,7	68,9	67,2	65,1	63,9	62,6	61,1	59,9	60,3
Contrato a termo	18,5	19,3	19,3	20,2	21,6	22,8	23,9	25,4	26,5	26,0
Contrato por tempo indeterminado	3,9	4,6	4,4	5,3	5,9	6,1	6,5	6,9	7,1	7,3
Não enquadrável	1,1	0,8	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,6

Tabela 29: Habilitações dos trabalhadores por ano (%)

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Ensino Básico	60,5	58,8	57,6	56,2	54,9	52,6	52,4	51,1	49,4	47,6
Ensino Secundário	22,3	23,3	23,4	24,3	25,1	25,9	26,9	27,8	28,9	30,1
Ensino Superior	16,3	17,1	18,3	18,9	19,4	20,0	20,3	20,6	21,2	21,9

Tabela 30: Regime de duração do trabalho por ano (%)

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Tempo Inteiro	87,5	87,2	86,7	86,7	86,4	86,5	86,7	86,9	87,2	87,2
Tempo Parcial	6,1	6,2	6,6	6,7	6,9	7,0	7,0	7,0	6,8	7,1

Nota: a soma por coluna não totaliza 100 devido aos valores omissos anteriormente reportados.

Tabela 31: Habilitações dos trabalhadores e gestores, por dimensão de empresa (2019; %)

	Nível máximo de qualificações		
	Básico	Secundário	Superior
Trabalhadores			
Micro empresas	31,6	34,9	33,5
Pequenas empresas	10,1	24,9	65,0
Médias empresas	1,4	5,3	93,3
Grandes empresas	0,4	1,5	98,1
Total	32,9	31,9	35,3
Gestores			
Micro empresas	41,6	28,6	29,8
Pequenas empresas	42,7	24,9	32,4
Médias empresas	28,9	23,4	47,7
Grandes empresas	16,4	19,4	64,1
Total	42,1	28,0	29,8

Nota: As percentagens totais podem ser superiores às percentagens discriminadas por dimensão de empresa, uma vez que no cálculo das mesmas são incluídas as empresas que, por não estarem presentes na base de dados do SCIE, não têm informação de dimensão.

Tabela 32: Dados em falta no QP-Trabalhadores

	QP	
	Nº	%
Nacionalidade	11	0,00
Antiguidade / Data de admissão	2.316	0,07
Idade	25.009	0,77
Tipo de Contrato e Regime de Duração	186.495	5,77
Rendimento e Número de Horas	187.079	5,79

Tabela 33: Empresas com registos em falta em um ou mais anos entre o primeiro e o último registo

	SCIE		QP	
	Nº	%	Nº	%
1 ano	15.728	2,4	24.482	6,0
2 anos consecutivos	6.804	1,0	8.216	2,0
3 anos consecutivos	3.351	0,5	3.992	1,0
4 anos consecutivos	1.687	0,6	2.119	0,5
5 anos consecutivos	941	0,1	1.205	0,3
6 anos consecutivos	460	0,1	711	0,2
7 anos consecutivos	218	0,03	354	0,1
8 anos consecutivos	95	0,01	109	0,03

Tabela 34: Trabalhadores duplicados no QP-Trabalhadores

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
N	280	714	396	434	574	793	1.350	1.603	1.673	1.894
%	0,01	0,03	0,02	0,02	0,02	0,03	0,05	0,05	0,05	0,06

Tabela 35: Empresas com outliers, segundo os intervalos de Tukey com $k = 3$, por classe de dimensão e apoio (PT2020)

	Empresas apoiadas		Candidatas sem execução		Não candidatas		Todas	
	% do total	% das empresas com outliers	% total	% das empresas com outliers	% total	% das empresas com outliers	% total	% das empresas com outliers
Micro	37,7	28,1	62,7	44,6	90,7	73,8	88,8	69,9
Pequenas	40,6	46,7	29,3	43,2	7,8	21,8	9,1	24,2
Médias	17,6	20,5	6,7	10,2	1,2	3,5	1,7	4,8
Grandes	4,1	4,8	1,3	2,0	0,3	0,9	0,4	1,2
Total	100	100	100	100	100	100	100	100

Tabela 36: Estatísticas descritivas gerais de variáveis selecionadas para empresas com volume de negócios = 0

	Média	Desvio-Padrão	Mínimo	Máximo	10%	25%	75%	90%
Pessoal ao serviço	1	4	1	622	1	1	1	1
Gastos com Pessoal	5	74	0	11 955	0	0	0	9
Resultado Operacional	9	2 216	-324 266	315 887	-22	-7	0	3
Juros Suportados	14	1 440	0	346 233	0	0	0	0
Capital Próprio	350	9 675	-523 092	1 373 581	-83	-11	52	341
Financiamentos Obtidos	396	21 007	0	5 016 138	0	0	12	257
Total Passivo	829	28 077	-1 573	6 400 136	0	3	245	912
VAB	-6	181	-26 414	14 338	-15	-4	0	0
Excedente Bruto de Exploração	-11	174	-26 417	5 498	-22	-7	0	0

Nota: os valores monetários estão expressos em milhares de euros

Tabela 37: Estatísticas descritivas gerais de variáveis selecionadas para empresas com VAB < 0

	Média	Desvio-Pa- drão	Mínimo	Máximo	10%	25%	75%	90%
Pessoal ao serviço	2	8	1	1 748	1	1	1	2
Volume de Negócios	80	3 446	0	778 260	0	0	10	51
Gastos com Pessoal	12	178	-1	23 086	0	0	7	18
Resultado Operacional	-11	3 343	-741 138	422 552	-44	-17	-1	0
Juros Suportados	16	1 627	0	346 233	0	0	0	1
Capital Próprio	382	16 806	-999 229	3 823 074	-101	-24	31	251
Financiamentos Obtidos	408	24 082	0	5 016 138	0	0	15	194
Total Passivo	863	35 246	-1.573	8 003 357	0	7	211	792
VAB	-23	540	-112 713	0	-27	-9	-1	0
Excedente Bruto de Ex- ploração	-36	617	-113 833	0	-45	-17	-1	0

Nota: os valores monetários estão expressos em milhares de euros

Tabela 38: Empresas com registo, cruzamento de bases de dados (SCIE e QP)

Ano	QP	SCIE			
		Não	Sim	Não (%)	Sim (%)
2010	Não	0	145.562	0,0	40,2
2010	Sim	1.168	215.673	0,3	59,5
2011	Não	0	146.378	0,0	40,3
2011	Sim	1.577	215.473	0,4	59,3
2012	Não	0	149.481	0,0	41,6
2012	Sim	3.489	206.288	1,0	57,4
2013	Não	0	149.534	0,0	41,5
2013	Sim	3.409	207.043	1,0	57,5
2014	Não	0	150.684	0,0	41,2
2014	Sim	2.219	212.672	0,6	58,2
2015	Não	0	156.933	0,0	41,9
2015	Sim	2.178	215.268	0,6	57,5
2016	Não	0	161.660	0,0	42,2
2016	Sim	2.159	219.275	0,6	57,2
2017	Não	0	171.703	0,0	43,3
2017	Sim	1.870	223.264	0,5	56,3
2018	Não	0	186.253	0,0	44,8
2018	Sim	1.714	227.514	0,4	54,8
2019	Não	0	212.833	0,0	48,4
2019	Sim	1.020	226.126	0,2	51,4

Tabela 39: Empresas com registo, e apoiadas ou candidatas ao PT2020. Cruzamento de bases de dados (SCIE e QP)

Ano	QP	SCIE			
		Não	Sim	Não (%)	Sim (%)
2010	Não	0	1.559	0,0	12,5
2010	Sim	207	10.733	1,7	85,9
2011	Não	0	1.697	0,0	12,8
2011	Sim	228	11.306	1,7	85,5
2012	Não	0	1.961	0,0	14,0
2012	Sim	253	11.762	1,8	84,2
2013	Não	0	2.177	0,0	14,4
2013	Sim	276	12.672	1,8	83,8
2014	Não	0	2.400	0,0	14,6
2014	Sim	263	13.798	1,6	83,8
2015	Não	0	3.109	0,0	16,8
2015	Sim	318	15.057	1,7	81,5
2016	Não	0	3.456	0,0	17,6
2016	Sim	321	15.848	1,6	80,8
2017	Não	0	3.717	0,0	18,3
2017	Sim	318	16.288	1,6	80,2
2018	Não	0	3.918	0,0	18,9
2018	Sim	322	16.448	1,6	79,5
2019	Não	0	4.399	0,0	21,1
2019	Sim	301	16.171	1,4	77,5

Tabela 40: Estatísticas descritivas gerais de indicadores selecionados de todas as empresas

	Média	Mediana	Desvio-Padrão	Mínimo	Máximo	% de zeros	% de negativos	% NAs
Taxa de investimento em formação (M3A)	20,7	0,0	333	0,0	681.33	93,1	0,0	0,0
Produtividade aparente do trabalho (M3A)	25	12	659	-75.000	346 000	0	0	0
Peso das remunerações no VAB (M3A)	0,9	0,31	47	0	15 306	27,68	0	13,74
Pessoal em I&D (% do total)	0	0	0	0	24	99,62	0	0
Diversificação geográfica das exportações	0,02	0	0	0	1	83,93	0	9,29
Intensidade exportadora	0,06	0	0	0	1	75,38	0	9,29
Intensidade do uso de combustíveis	0,05	0,01	1	0	524	30,28	0	9,29
Intensidade do uso de água	0,01	0	1	0	235	45,4	0	9,29
Intensidade do uso de eletricidade	0,05	0	3	0	973	35,17	0	9,29
Ativo intangível (% do ativo total)	0,01	0	0	-0,39	2,76	86,72	0,01	0,32
Rentabilidade do ativo	-1,53	0,01	113	-37 857	10 903	0,1	38,99	0,76
Rentabilidade do volume de negócios	-0,03	0,02	281	-42 291	115 534	0,07	28,07	11,13
Rácio de dívida financeira	1,81	0	243	0	137 633	55,75	0	0,32
Autonomia financeira	-7,52	0,31	850	-354 336	31	0,02	24,54	0,69
Solvabilidade	46,94	0,42	11411	-1	7 457 337	0,02	25,23	0,89
Efeito dos juros suportados	0,39	0	95	0	55 614	47,83	0	9,29
Rácio administradores com formação superior	0,38	0,33	0	0	1	0	0	90,99
Rácio de trabalhadores com formação superior	0,48	0,4	0	0	1	0	0	77,35
Rácio trabalhadores com contrato sem termo	0,66	0,67	0	0	1	0	0	59,63
Incidência de contratos atípicos	0,49	0,5	0	0	1	0	0	73,78
Incidência de salários mínimos	0,63	0,64	0	0	1	0	0	75,3
Rácio mulheres	0,57	0,5	0	0,01	1	0	0	62,26
% capital social estrangeiro	2,55	0	15	0	100	50,02	0	48,49
Média dos ganhos	826,85	727,13	475	0	24 286	0,3	0	53,3
Idade média dos trabalhadores	42	42	9	18	67	0	0	53,36
Idade média dos administradores	48	48	10	18	67	0	0	70,97
<i>Gender pay gap</i>	121,2	58,54	483	-11 073	17 645	1,37	7,71	77,16

Tabela 41: Descritivas gerais de indicadores selecionados só de empresas pequenas e médias

	Média	Mediana	Desvio-Pa- drão	Mínimo	Máximo	% de zeros	% de ne- gativos	% de NA
Taxa de investimento em formação (M3A)	33,1	0,0	301,2	0	47.800	78	0,0	0,0
Produtividade aparente do trabalho (M3A) (euros)	82.283,2	22.082,5	740.320,4	-9.187.239	39.753.111	0	1,6	0,0
Peso das remunerações no VAB (M3A)	0,7	0,5	7,5	0	1.449	2	0,0	0,6
Pessoal em I&D (% do total)	0,0	0,0	0,0	0	1	98	0,0	0,0
Diversificação geográfica das exportações	0,0	0,0	0,1	0	1	77	0,0	0,3
Intensidade exportadora	0,1	0,0	0,3	0	1	59	0,0	0,3
Intensidade do uso de combustíveis	0,0	0,0	2,5	0	524	9	0,0	0,3
Intensidade do uso de água	0,0	0,0	1,0	0	209	23	0,0	0,3
Intensidade do uso de eletricidade	0,0	0,0	1,1	0	211	13	0,0	0,3
Ativo intangível (% do ativo total)	0,0	0,0	0,1	0	1	67	0,0	0,0
Rentabilidade do ativo	-0,1	0,0	18,0	-3.706,08	31	0	17,2	0,0
Rentabilidade do volume de negócios	0,1	0,0	43,1	-2.785,67	8.366	0	16,9	0,3
Rácio de dívida financeira	0,2	0,1	0,7	0	90	37	0,0	0,0
Autonomia financeira	0,1	0,4	13,1	-2.039,42	31	0	8,5	0,0
Solvabilidade	1,5	0,6	7,7	-1	1.051	0	8,5	0,0
Efeito dos juros suportados	0,0	0,0	0,7	0	74	22	0,0	0,3
Rácio de administradores com formação superior	0,1	0,1	0,1	0	1	0	0,0	82,8
Rácio de trabalhadores com formação superior	0,3	0,2	0,3	0	1	0	0,0	35,8
Rácio trabalhadores com contrato sem termo	0,6	0,6	0,3	0	1	0	0,0	15,4
Média dos ganhos mensais (euros)	989,7	849,7	558,5	0	20.097	0	0,0	11,4
Incidência de salários mínimos	0,3	0,3	0,3	0	1	0	0,0	48,7
Incidência de contratos atípicos	0,4	0,4	0,3	0,01	1	0	0,0	24,3
Rácio mulheres	0,4	0,4	0,3	0,01	1	0	0,0	16,3
% capital social estrangeiro	6,4	0,0	23,8	0	100	82	0,0	11,2
Idade média dos trabalhadores	40,9	41,2	5,8	18	67	0	0,0	11,4
Idade média dos administradores	49,0	49,0	9,1	18	67	0	0,0	52,5
Gender pay gap	142,3	90,2	473,1	-8.994,79	14.700	0	25,9	21,5

8.2. Construção de indicadores

Indicador	Fórmula de cálculo	Base de dados	Dimensão
Incidência trabalhadores com formação superior	$\frac{ Trabalhadores(habil \geq 4) }{ Trabalhadores }$	QP	Qualificação
Incidência administradores com formação superior	$\frac{ Trabalhadores(habil \geq 4 \cap sitpro = 1) }{ Trabalhadores(sitpro = 1) }$	QP	Qualificação
Incidência trabalhadores com formação básica	$\frac{ Trabalhadores(habil > 1 \cap habil < 3) }{ Trabalhadores }$	QP	Qualificação
Incidência trabalhadores CT	$\frac{ Trabalhadores(prof = 211 \cup prof = 212 \cup prof = 213 \cup prof = 213 \cup prof = 214 \cup prof = 221 \cup prof = 222) }{ Trabalhadores }$	QP	Qualidade do Emprego
Incidência trabalhadores qualificados	$\frac{ Trabalhadores(nqual \geq 1 \cap nqual \leq 3) }{ Trabalhadores }$	QP	Qualificação
Incidência trabalhadores sem termo	$\frac{ Trabalhadores(tipo_contr \leq 14) }{ Trabalhadores }$	QP	Qualidade do emprego
Incidência contratos atípicos	$\frac{ Trabalhadores(tipo_contr \neq 1) }{ Trabalhadores }$	QP	Qualidade do emprego
Incidência da duração de contratos atípicos	$\frac{ Trabalhadores(sitpro = 3 \cap (pnt < 32 \cup pnt > 48)) }{ Trabalhadores }$	QP	Qualidade do emprego
Incidência do salário mínimo	$\frac{ Trabalhadores(tipo_{contr} \neq 1) }{ Trabalhadores }$	QP	Remunerações
Proporção de mulheres na administração	$\frac{ Trabalhadores(sitpro = 1 \cap sexo = 2) }{ Trabalhadores(sitpro = 1) }$	QP	Género
Proporção de mulheres empregadas	$\frac{ Trabalhadores(sexo = 2) }{ Trabalhadores }$	QP	Género
Média de ganhos	\overline{Rganho}	QP	Remunerações
Salários dos homens	$\overline{Rganho(sexo = 1)}$	QP	Remunerações
Salários das mulheres	$\overline{Rganho(sexo = 2)}$	QP	Remunerações
Gender gap	$\overline{Rganho(sexo = 2)} - \overline{Rganho(sexo = 1)}$	QP	Género
Idade média dos administradores	$\overline{idade(sitpro = 1)}$	QP	Outros
Idade média dos trabalhadores	$\overline{idade(sitpro \neq 1)}$	QP	Outros
Dispersão salarial	$\frac{\sigma Rganho}{\overline{Rganho}}$	QP	Remunerações
Rácio de capital estrangeiro	Direto	QP	Caracterização
Idade	Direto	QP	Caracterização
Autonomia financeira	Total capital próprio / Total ativo líquido	SCIE	Situação financeira

Indicador	Fórmula de cálculo	Base de dados	Dimensão
Efeito dos juros suportados	Resultado antes impostos / Resultado operacional	SCIE	Situação financeira
Liquidez geral	Total ativo corrente / Total passivo corrente	SCIE	Situação financeira
Solvabilidade	Total capital próprio / Total passivo	SCIE	Situação financeira
Rácio de dívida financeira	Financiamentos obtidos (passivo não corrente) / Total ativo líquido	SCIE	Situação financeira
Taxa de investimento com formação	Gastos com pessoal (formação) / Pessoal ao serviço	SCIE e QP	Investimento
Proporção pessoal I&D	Pessoal I&D / Pessoal ao serviço	SCIE e QP	Inovação
Diversificação geográfica das exportações	(Total vendas mercado extracomunitário + Prestações de serviços mercado extracomunitário) / Volume de negócios (vendas e serviços prestados)	SCIE	Internacionalização
Exportações para a UE	Total vendas mercado comunitário + Serviços prestados mercado comunitário	SCIE	Internacionalização
Exportações extra-UE	Total vendas mercado comunitário + Serviços prestados mercado comunitário	SCIE	Internacionalização
Exportações	Exportações UE + Exportações extra-UE	SCIE	Internacionalização
Importações	Compras mercado comunitário + Compras mercado extracomunitário	SCIE	Internacionalização
Intensidade exportadora	Exportações / Volume de negócios (vendas e serviços prestados)	SCIE	Internacionalização
Inserção no comércio internacional	(Exportações + Importações) / VABcf	SCIE	Internacionalização
Grau de transformação de produção	VABcf / Produção	SCIE	Competitividade
Produtividade aparente do trabalho	VABcf / Pessoal ao serviço	SCIE	Competitividade
Rentabilidade do volume de negócios	Resultado líquido do período / Volume de negócios	SCIE	Competitividade
Rentabilidade do ativo	Resultado líquido do período / Total ativo líquido	SCIE	Competitividade
Rentabilidade do volume de negócios (2)	Resultado operacional / Volume de negócios	SCIE	Competitividade
Rentabilidade do ativo (2)	Resultado operacional / Total ativo líquido	SCIE	Competitividade
Intensidade de uso de água	Fornecimento de serviços externos (energia e fluídos - água) / VABcf	SCIE	Ecoeficiência
Intensidade de uso de combustíveis	Fornecimento de serviços externos (energia e fluídos - combustíveis) / VABcf	SCIE	Ecoeficiência
Intensidade de uso de eletricidade	Fornecimento de serviços externos (energia e fluídos - combustíveis) / VABcf	SCIE	Ecoeficiência
Gastos médios com pessoal	Gastos com pessoal / Pessoal ao serviço	SCIE	Remunerações

Indicador	Fórmula de cálculo	Base de dados	Dimensão
Peso das remunerações no VAB	Gastos com pessoal (Remunerações pessoal) /VABcf	SCIE	Remunerações
Ativo fixo sobre pessoal	Ativos fixos tangíveis / Pessoal ao serviço	SCIE	Competitividade
Rácio de ativo intangível	Ativos intangíveis / Total ativo líquido	SCIE	Competitividade