

MESTRADO
ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

O IMPACTO DAS CONSULTAS À CENTRAL DE
RESPONSABILIDADE DE CRÉDITO SOBRE A TAXA DE
JURO (2017)

MAURICIO SILVA DE CARVALHO

OUTUBRO - 2018

MESTRADO EM
ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

O IMPACTO DAS CONSULTAS À CENTRAL DE
RESPONSABILIDADE DE CRÉDITO SOBRE A TAXA DE
JURO (2017)

MAURICIO SILVA DE CARVALHO

ORIENTAÇÃO:

PROF. DOUTORA ISABEL MARIA DIAS PROENÇA

DR. HOMERO ALEXANDRE MARTINS GONÇALVES

OUTUBRO - 2018

Como não poderia ser diferente, dedico, por uma vez mais, este trabalho aos meus pais, Lícia Nara e Alberto Azemiro, que, além de base e fundamento do meu caminho, são os meus maiores incentivadores por propiciarem o privilégio, as condições e a oportunidade de aprofundar minha formação como acadêmico e acima de tudo, como pessoa.

AGRADECIMENTOS

À minha família – especialmente, meus pais, minha irmã e minha namorada – pela paciência e incentivo até aqui despendido. O carinho e atenção são indissociáveis dos resultados alcançados. A distância durante esses anos não foi fácil, por vezes inquietante, e é por isso também que agradeço pelo conforto que tanto precisei ao longo dessa nova caminhada. Sem eles, nada disso seria possível.

Aos meus amigos de longa data que, sem embargo, muito contribuíram e, que também apesar da distância, não deixaram jamais de participar da minha constante formação, acima de tudo, como pessoa. Destaco, entre eles, Alejandro Padrón no qual me ajudou, de forma direta, na resolução dos problemas surgidos durante o trabalho. Ressalto ainda as minhas novas amigas – maiores presentes por morar fora do meu país – conquistadas nesses dois anos de ISEG (Lisbon School of Economics and Management). A todos estes, meu agradecimento.

A todos os meus amigos, construídos durante os 6 meses enriquecedores do Estágio Curricular, do Departamento de Estatística (DDE) e demais departamentos do Banco de Portugal por todo o acolhimento, atenção e apreço despendido. A estes, também, meu profundo agradecimento.

Para não cometer nenhuma injustiça, ademais, agradeço a todos aqueles que contribuíram, mesmo que indiretamente, para a construção do meu caráter, da minha pessoa e da formação dos meus ideais. A todos, muito obrigado.

Por fim, mas não menos importante, agradeço aos meus orientadores, Prof. Doutora Isabel Maria Dias Proença e Dr. Homero Alexandre Martins Gonçalves, pela oportunidade da orientação e do estágio, engrandecedor, no Banco de Portugal. Certamente, o presente trabalho seria inalcançável sem suas intervenções e orientações.

RESUMO

Esta dissertação tem como objetivo estudar o impacto da solicitação por parte das instituições de crédito, nomeadamente, os bancos, da consulta à Central de Responsabilidade de Crédito (CRC) sobre a taxa de juro média dos contratos de crédito. O estudo divide-se em cinco partes. No primeiro capítulo apresenta-se a crescente literatura sobre a partilha de informação (*information sharing*). O Capítulo 2 introduz com maiores detalhes a descrição dos dados utilizados e a estratégia necessária para um melhor entendimento sobre a relação (ligação) entre banco e empresa, fulcrais para a análise aplicada nesta investigação. No terceiro capítulo expomos a metodologia adotada, pautada no conceito de resultado potencial (*potential outcome*) e estimação por procedimento de *matching*. É feito uso tanto do procedimento de *matching* nas próprias covariáveis do modelo – através do *nearest neighbor estimator* – quanto o *matching* baseado no *propensity score*. Ademais, segundo sugere a literatura, as estimativas OLS também são incluídas como base de comparação. No Capítulo 4, para além da exposição dos resultados alcançados, traçamos as discussões quanto à aplicabilidade das hipóteses adotadas, nomeadamente a hipótese de *overlap*, como também da adequação dos próprios resultados, através da análise do balanceamento (*balance*) das covariáveis. Por fim, o Capítulo 5 tece considerações finais, apontando as suas limitações e evidenciando as principais conclusões do estudo.

Palavras-chave: Acesso ao crédito, partilha de informação, efeitos de tratamento, estimação por *matching*.

ABSTRACT

This dissertation aims to analyze the impact of the inquiries' request by the credit institutions, namely the banks, to the *Central de Responsabilidade de Crédito* (CRC) – Portuguese Public Registry – on the average interest rate of credit contracts. The study is divided into five parts. The first chapter presents the growing literature on information sharing. Chapter 2 introduces the data used and the necessary strategy for a better understanding of the relationship between the bank and the company, central to the analysis applied in this research. In Chapter 3 we present the methodology adopted, based on the concept of potential outcome and estimation by matching procedure. Both the matching procedure in the covariates of the model – through the nearest neighbor estimator – and the matching under the propensity score are used. Furthermore, as suggested in the literature, OLS estimates are also included for comparing purposes. In Chapter 4, in addition to presenting the results achieved, we discuss the applicability of the hypotheses adopted, namely the overlap hypothesis, as well as the adequacy of the results themselves, by analyzing the balance of the covariates. Finally, Chapter 5 brings the final considerations of this research, pointing out its limitations and highlighting the main conclusions.

Keywords: Access to credit, information sharing, treatment effects, matching estimation.

SUMÁRIO

INTRODUÇÃO	9
CAPÍTULO 1 - Revisão da Literatura.....	10
CAPÍTULO 2 - Dados e Estatística Descritiva.....	14
CAPÍTULO 3 - Metodologia e Métodos de Estimação	21
3.1. Efeitos de tratamento e a sua problemática para a Inferência	21
3.2. Hipóteses	24
3.3. Métodos de Estimação.....	26
3.3.1. <i>Matching on covariates – Nearest Neighbor Matching (NNM)</i>	26
3.3.2. <i>Matching on Propensity Score (PSM)</i>	28
3.3.3. <i>OLS</i>	29
3.4. Observações sobre os estimadores.....	29
CAPÍTULO 4 - Resultados	31
4.1. <i>Balance</i> e avaliação da hipótese de <i>overlap</i>	32
4.2. Análise dos Resultados.....	36
CONSIDERAÇÕES FINAIS	40
ANEXO I – INFORMAÇÃO CENTRALIZADA.....	43
ANEXO II - TABELAS E GRÁFICOS	44
REFERÊNCIAS BIBLIOGRÁFICAS	49

GLOSSÁRIO

CAE – Classificação Portuguesa de Atividades Económicas

CRC – Central de Responsabilidade de Crédito

NNM – *Nearest neighbor matching*

NNMe – *Nearest neighbor exact matching*

NO – Novas Operações (base de dados do Banco de Portugal)

NUTS - Nomenclatura das Unidades Territoriais para Fins Estatísticos

OECD – Organização para a Cooperação e Desenvolvimento Económico

OLS – Mínimos quadrados ordinários

PIB – Produto Interno Bruto

PSM – *Propensity score matching*

SNF – Sociedade não financeira

ÍNDICE DE FIGURAS

Figura 1. Gráfico das densidades do <i>propensity score</i> (amostra completa)	47
Figura 2. Gráfico das densidades do <i>propensity score</i> para regra “percentil 99%”	47
Figura 3. Gráfico das densidades do <i>propensity score</i> para regra “[0,1;0,9]”	48
Figura 4. Gráfico das densidades do <i>propensity score</i> para regra “mínimo e máximo”	48

ÍNDICE DE TABELAS

Tabela I. Composição das classes de ligação banco – empresa relativamente à consulta da CRC (%)	18
Tabela II. Composição das consultas por classes de ligação banco - empresa (%).....	19
Tabela III. Distribuição do montante dos contratos de crédito	20
Tabela IV. Taxa de juro média (% ao ano)	20
Tabela V. Descrição das variáveis.....	44
Tabela VI. Estimativas obtidas por <i>OLS</i> , por <i>matching</i> nas covariáveis e <i>matching</i> no <i>propensity score</i>	45
Tabela VII. Balanceamento das covariáveis (amostra completa)	46
Tabela VIII. Balanceamento das covariáveis (subamostra com regra “percentil 99%”)	46

INTRODUÇÃO

No mercado de crédito as informações são distribuídas assimetricamente e os problemas de seleção adversa (*adverse selection*) e risco moral (*moral hazard*) surgem como dificuldades inerentes aos empréstimos bancários. Isso porque os devedores (empresas ou particulares) geralmente sabem mais sobre seus empreendimentos do que os próprios credores (bancos ou instituições financeiras) – seja em termos dos riscos associados aos seus projetos seja em termos do grau de esforço empenhado em honrar (pagar) suas dívidas.

A literatura sobre assimetria de informação e suas implicações sobre o custo e o financiamento das empresas é extensa¹. No trabalho seminal de Stiglitz e Weiss (1981), os autores mostram que, sob informação assimétrica, a taxa de juro de equilíbrio do mercado de crédito é tal que a procura de crédito excede a oferta. Isso significa que mesmo aqueles clientes que estejam dispostos a pagar taxas de juro mais elevadas não são capazes de obter um empréstimo pois os bancos entendem que o mesmo seria mais arriscado, optando pelo seu racionamento (*credit rationing*). Em tempos mais recentes tem-se desenvolvido a literatura sobre partilha de informação (*information sharing*) e com ela uma gama de estudos, desde os trabalhos seminais de Jappelli e Pagano (1993) e Padilla e Pagano (1997, 2000). Argumenta-se que diante da assimetria de informação uma forma mais eficiente de eliminar-se ou reduzir-se os problemas gerados – de seleção adversa e risco moral –, em contrapartida aos mecanismos mais tradicionais de obtenção de informação individual (*screening*) e monitorização (*monitoring*), seria através de um sistema público ou privado de partilha de informação (Jappelli & Pagano, 2000).

Em geral, segundo a literatura económica, os problemas observados no mercado de crédito podem ser descritos da seguinte forma. Sob a ausência de informação, os bancos reagem de modo a fazer-se sentir na economia uma elevação da taxa de juro média praticada e, no limite, do racionamento do crédito. Ocorre que, por não saberem diferenciar os melhores devedores dos piores, os bancos acabam por utilizar a estratégia de "padronizar" a aplicação da taxa de juro (custo do empréstimo) de modo a compensar as perdas advindas daqueles maus empréstimos, tendo em vista o maior risco associado, penalizando os bons pagadores tanto no acesso quanto no custo do empréstimo. Ainda que no momento da efetivação do contrato de crédito as empresas sejam exigidas a

¹ Ver, por exemplo, Freimer e Gordon (1965); Jaffee e Modigliani (1969); Stiglitz e Weiss (1981); Besanko e Thakor (1987); Berger e Udell (1992) e Ghosh et al (2000).

fornecer informações sobre elas e o projeto a ser financiado, têm sempre a prerrogativa de ditar o quanto é o grau do seu compromisso com a sua dívida.

A bibliografia sobre o tema destaca ainda outros canais pelos quais a introdução de um sistema de partilha de informação afetaria a economia. Segundo Peria e Singh (2014), para além da resolução dos problemas de seleção adversa e risco moral, destacam-se ainda: i. mecanismo disciplinador (*discipline device*) – associado ao problema de risco moral (Padilla & Pagano, 2000); ii. redução do endividamento excessivo por parte das empresas (Bennardo et al, 2015) e; iii. redução dos custos de catividade (*hold-up costs*) (Padilla & Pagano, 1997). No presente trabalho concentrar-nos-emos exclusivamente nos problemas de seleção adversa e risco moral e seus efeitos sobre a taxa de juro.

O estudo divide-se em cinco partes. No primeiro capítulo procuramos apresentar a crescente literatura sobre a partilha de informação. O Capítulo 2 expõe com maiores detalhes os dados utilizados e a estratégia necessária para um melhor entendimento sobre a relação (ligação) entre banco e empresa, fulcrais para a análise aplicada nesta investigação, no contexto da economia portuguesa. No terceiro capítulo apresentamos a metodologia empregue, pautada no conceito de resultado potencial e estimação por procedimento de *matching*. No Capítulo 4, para além da exposição dos resultados alcançados, traçamos as discussões quanto à aplicabilidade das hipóteses adotadas, nomeadamente a hipótese de *overlap*, como também da adequação dos próprios resultados, através da análise do balanceamento (*balance*) das covariáveis. Por fim, o Capítulo 5 tece considerações finais, apontando as suas limitações e evidenciando as principais conclusões do estudo.

CAPÍTULO 1 - Revisão da Literatura

Perante o problema de assimetria de informação e na tentativa de atenuá-lo, os bancos estabelecem processos de análise (*screening*) e monitorização (*monitoring*) dos seus clientes. No entanto, uma alternativa, que se tem mostrado mais eficiente, vem ganhando força nos últimos anos, a saber, partilha de informação (*information sharing*). A partilha de informação pode ocorrer de duas formas, não necessariamente exclusivas. Através de agências de crédito privadas (*credit bureau*), que surgem devido a incentivos de mercado, ou através de registos públicos de crédito (*public credit registry*)

estabelecidos e administrados por Bancos Centrais, onde a participação é tipicamente compulsória e garantida em termos legais (Jappelli & Pagano, 2000, 2002, 2006).

Segundo o Banco Mundial, em janeiro de 2014, mais de 2 mil milhões de indivíduos e 120 milhões de firmas ao redor do mundo eram englobados por um sistema de partilha de informação (agências de crédito privadas ou registos públicos). O número de economias com uma agência de crédito que cobre pelo menos 5% da população adulta cresceu de 49 dentre 145 economias em 2004 para 94 dentre 189 economias em 2014, enquanto para os registos públicos de crédito esse número aumentou de 22 dentre 145 economias para 58 dentre 189 (World Bank, 2014).

A teoria sugere que a partilha de informação pode superar tanto os problemas de seleção adversa no mercado de crédito (Jappelli & Pagano, 1993) quanto o de risco moral (Padilla & Pagano, 2000). Efetivamente, estudos mostram que a existência de tais instituições tende a impactar de forma positiva o mercado de crédito. Jappelli e Pagano (2002), analisando 43 países em 1994 e em 1995, apontam que os países com sistemas ativos de partilha de informações apresentam maiores volumes de empréstimos e menores riscos associados aos créditos.

Beck et al (2004), usando uma base de dados para 74 países, avaliam o efeito da estrutura do mercado bancário no acesso das empresas ao financiamento e concluem, entre outros resultados, que ter um registo de crédito bem desenvolvido reduz os problemas relacionados com a concentração bancária e os obstáculos financeiros. Analisando um painel com 129 países, Djankov et al (2007) demonstram que a presença de sistemas de partilha de informações é positivamente correlacionada com taxas mais elevadas de crédito privado em relação ao PIB, ou seja, associado a uma maior disponibilidade de crédito. O mais interessante é que as agências de crédito públicas estão fortemente associadas ao crédito disponível nos países mais pobres, mas não nos mais ricos, sugerindo um papel destacado dos governos.

Recentemente, houve um aumento no número de estudos com dados a nível individual (tanto a nível das empresas como a nível dos bancos). Brown et al (2009), investigando o mercado de crédito dos países em transição da Europa Oriental e da antiga União Soviética, concluem que a partilha de informações está associada a uma maior disponibilidade de crédito e a um menor custo de financiamento para as empresas. Behr e Sonnekalb (2012) explorando a introdução do registo público de crédito na Albânia em 2008, demonstram que a partilha de informação reduz significativamente a probabilidade

de incumprimento e melhora a qualidade dos empréstimos, mas não encontram efeitos significativos sobre as taxas de juro e nenhum efeito sobre o acesso ao crédito.

Grajz e Laptieva (2016) estudam o efeito da partilha de informações sobre o volume de crédito privado na Ucrânia e concluem que essa partilha de informações está associada a um aumento no volume de empréstimos bancários caso ocorra através de agências de crédito privadas. Por outro lado, ao contrário de Djankov et al (2007) que afirmam que a presença de um registo público de crédito está positivamente correlacionada com o volume de crédito nos países mais pobres, Grajz e Laptieva (2016) não encontram nenhum efeito estatisticamente significativo sobre o volume de crédito provocado pelo registo de crédito público.

Peria e Singh (2014) avaliam o impacto da introdução de sistemas de partilha de informações de crédito (tanto público como privado) no acesso das empresas ao financiamento. A análise utiliza dados plurianuais, ao nível da empresa, para 63 países, cobrindo mais de 75.000 firmas no período de 2002-13 e chegou a resultados similares aos de Grajz e Laptieva (2016) no que diz respeito a um impacto reduzido, ou até mesmo nulo, da existência de um registo público de crédito. Os autores concluem que a introdução de uma agência de crédito privada está associada, entre outros efeitos, a um aumento na probabilidade das empresas terem maior acesso a financiamento e, particularmente, a uma redução nos juro dos empréstimos. No entanto, como referido, tais efeitos não são verificados quando a partilha é processada através de um registo público de crédito.

Utilizando o *matching* como ferramenta analítica, Ioannidou e Ongena (2010) estudam quais as consequências para as condições do financiamento das empresas quando estas trocam os respetivos bancos. Do ponto de vista das empresas, os autores mostram que os devedores que trocam de banco obtêm uma taxa de juro significativamente mais baixa em comparação com a taxa fixada nos créditos concedidos pelo banco antigo (no qual a empresa em questão já era cliente). Por outro lado, o banco novo diminui, inicialmente, a taxa de juro, mas, com o tempo, aumenta-a de forma acentuada, incentivando uma nova mudança por parte das firmas. Esse resultado é consistente com a existência de custos de “catividade” (*hold-up costs*) nas relações entre o banco e a empresa, indo ao encontro da literatura de partilha de informação².

² Ver, por exemplo, Bennardo et al (2015).

Bos et al (2015) estudam o efeito da introdução de um sistema de partilha de informações na Bósnia e Herzegovina, nomeadamente de um registo público de crédito administrado pelo Banco Central do país, e utilizando, também, técnicas de *matching* (em que usam o *propensity score*) para comparar as estratégias dos bancos quanto relativamente às suas respetivas composições de portfólio. Os autores encontraram evidências de que a introdução do registo público levou a um crédito mais caro para as empresas novas, mas reduziu as taxas de juro para clientes repetidos, ou seja, que já possuíam um contrato com o banco subjacente. De facto, os autores mostram como a maior transparência e disponibilidade de informação no mercado de crédito pode beneficiar os bons pagadores no acesso a maior e melhores condições de financiamento, além de permitir uma maior eficiência na alocação dos recursos por parte dos bancos. Esse resultado, assim como o de Ioannidou e Ongena (2010), é consistente com a existência do mecanismo disciplinador dos devedores (*borrower discipline device*) quando da presença de um instrumento de partilha de informação³.

Além daqueles estudos que fazem uma análise comparativa entre diferentes países, os estudos acima apresentados estão, por vezes, interessados na análise dos efeitos da introdução de um sistema de partilha de informação e investigam seu efeito comparando a economia no estado anterior e posterior à criação do sistema (via, por exemplo, estimadores diferença-em-diferença). O presente trabalho procura contribuir para a literatura no que tange à investigação do efeito pontual no mercado de crédito quando os bancos utilizam informação adicional a respeito dos potenciais clientes para tomar decisões de empréstimo.

Em conclusão, o objetivo desta investigação consiste em avaliar se o acesso a um mecanismo singular de partilha de informação por parte das entidades bancárias (consultas à CRC), obtendo conhecimentos que outrora eram privados ao próprio devedor, é capaz de atenuar os problemas inerentes ao mercado de crédito ao proporcionar a possibilidade de taxas de juro médias menos elevadas para a economia portuguesa.

³ Ver Padilla e Pagano (2000).

CAPÍTULO 2 - Dados e Estatística Descritiva

O estudo emprega dados disponibilizados pelo Banco de Portugal, através de informações únicas e exclusivas dessa instituição. Dentre eles, utilizou-se a base de dados das consultas à Central de Responsabilidade de Crédito (CRC) e a base de dados referentes às Novas Operações (NO) desde janeiro até dezembro de 2017. Utilizou-se, também, dados referentes à base das consultas ao segundo semestre de 2016, além da lista de devedor-credor do próprio registo da CRC desde julho até dezembro de 2016.

Os dados utilizados são transversais (*cross-section*) quando analisamos o ano de 2017 como um todo, sendo considerado como unidade observacional o contrato de crédito, ou seja, a relação entre banco e empresa. A base de dados das NO contém informação individual sobre os contratos de empréstimos bancários concedidos às sociedades não financeiras (SNF)⁴. Desta base consideramos apenas aqueles contratos estritamente novos, isto é, não foram considerados os contratos renegociados, de renovação automática e de reestruturação da dívida corrente⁵. No que tange às informações individuais, utilizou-se a data da operação, a taxa de juro (em % ao ano), o prazo contratual (em número de dias), o montante do empréstimo de cada operação (em milhões de €), se a operação possui colateral ou não, o setor de atividade a que pertence a SNF subjacente ao contrato (CAE), a região do país na qual se assenta (NUTS) e a sua respetiva dimensão (micro, pequena, média e grande empresa).

Mais precisamente, a variável de tratamento, T_i , é definida como

$$T_i = \begin{cases} 1, & \text{se o contrato } i \text{ envolveu consulta à CRC} \\ 0, & \text{caso contrário} \end{cases} \quad (1)$$

Relativamente à variável dependente, Y , temos: i) Y_i^1 , taxa de juro (potencial) quando ocorre a consulta à CRC e; ii) Y_i^0 , taxa de juro (potencial) quando não ocorre consulta à CRC. Por fim, apesar da unidade observacional ser o contrato de empréstimo, é de extrema importância explicitar, também, além das particularidades dos próprios contratos, as características tanto dos bancos como das empresas envolvidas. Nesse

⁴ No presente estudo as palavras firma, empresa e sociedade não financeira são usadas indistintamente.

⁵ Dado o objetivo do trabalho, a saber, avaliar o impacto das consultas sobre a taxa de juros, é particularmente interessante estudar aqueles contratos estritamente novos onde pouca ou nenhuma informação entre banco e empresa foram permutadas.

sentido, o conjunto de variáveis independentes, covariáveis ou características (X_i) foram agrupados da seguinte maneira:

- i. Características do contrato: Montante (em €), Prazo Contratual (em número de dias), Classe de ligação e Colateral
- ii. Características da empresa: Dimensão (micro, pequena, média e grande), Setor de Atividade (CAE) e Região do País (NUTS)
- iii. Características do banco: o próprio identificador (variável *dummy* – efeito fixo)

As definições da variável dependente e das variáveis independentes podem ser consultadas na Tabela V do Anexo II.

Apesar das informações mais estruturais do trabalho estarem presentes na base das NO, como taxa de juro e as demais variáveis, cabe aqui uma explicação um pouco mais detalhada sobre a CRC por ser ela a provedora das diretrizes utilizadas para um melhor entendimento sobre a relação (ligação) entre banco e empresa, fulcrais para a análise aplicada nesta investigação.

A CRC é um banco de dados gerido pelo Banco de Portugal com base em informações relacionadas com crédito fornecidas pelos seus participantes, ou seja, as instituições que concedem crédito. No que diz respeito ao reporte, os participantes são legalmente obrigados a enviar ao Banco de Portugal as informações relativas aos empréstimos em dívida ao fim de cada mês, tanto para particulares como para pessoa coletiva, residentes e não residentes, sempre que a responsabilidade total seja superior a €50. São contempladas tanto informações positivas (*white information*) como negativas (*black information*), o que significa que todas as situações de crédito são relatadas, sejam elas regulares ou em incumprimento (Banco de Portugal, 2015).

Para além de outras funcionalidades, o objetivo principal da CRC é fornecer apoio às instituições participantes na avaliação do risco da concessão de crédito. Para este fim, estas instituições têm acesso às responsabilidades totais de cada devedor, ou potencial devedor, dentro do sistema financeiro (Banco de Portugal, 2015).

Em particular, destaca-se o serviço disponibilizado através de uma plataforma de consulta *online* que permite às instituições financeiras terem acesso à situação de crédito (centralizada) no sistema financeiro de cada empresa (e indivíduo) constante na CRC. A

informação centralizada⁶ consiste na agregação, para cada cliente, dos saldos de responsabilidades comunicados pelas entidades participantes referentes ao final de cada mês. No caso dos seus clientes de crédito, os bancos recebem mensalmente esta informação, mas para obter acesso a estes dados para potenciais clientes é necessário ter o consentimento do devedor em causa e o pagamento de um valor nominal⁷ (Banco de Portugal, 2017).

Ressalta-se que o facto do Banco de Portugal já disponibilizar automaticamente aos bancos a informação centralizada dos seus clientes é crucial para a análise aqui apresentada. Diante dessa circunstância é necessário caracterizar as ligações entre banco e empresa de modo a conseguir captar os efeitos específicos da consulta – tendo sido feita a distinção entre aqueles clientes antigos e aqueles novos que suscitam maior interesse por parte dos bancos em realizar uma consulta à CRC. Procurou-se definir cada ligação da seguinte forma: i. Nova Ligação Bancária (NLB); ii. Nova Ligação de Mercado (NLM) e; iii. Antiga Ligação (AL).

- i. Significa que a ligação em questão entre a empresa genérica “ α ” e o banco genérico “ β ” não existia nos 6 meses anteriores à efetivação do contrato e passa a existir em 2017, no entanto esse cliente não é novo no mercado de crédito porque já possuía uma relação de empréstimo com outro(s) banco(s), digamos, o banco “ λ ”;
- ii. Significa que tanto a ligação em questão entre a empresa genérica “ α ” e o banco genérico “ β ” não existia nos 6 meses anteriores à realização do contrato (em 2017) assim como também essa empresa estava fora do mercado de crédito (não possuía relação de crédito com qualquer outro banco) no mesmo intervalo de tempo – portanto, a empresa em questão é nova no mercado de crédito;
- iii. Significa que a ligação em questão entre a empresa genérica “ α ” e o banco genérico “ β ” existia nos 6 meses anteriores à efetivação do contrato e por consequência imediata a empresa “ α ” já era participante do mercado de crédito.

⁶ Ver Anexo I para maiores detalhes a respeito da informação centralizada.

⁷ O valor referido é apenas simbólico, ou seja, não traz custo adicional significativo capaz de influenciar na decisão de realizar a consulta à CRC e sua ausência explícita aqui é devido a discricção e reserva necessária.

A motivação dessas definições, como já referido, foi permitir precisar com maior rigor o critério de definição da variável de tratamento que estima o impacto das consultas. Assim, supõe-se que o tratamento corresponde à consulta sobre o (potencial) cliente feita até 6 meses antes da concretização do eventual empréstimo, tempo considerado razoável para o processamento das informações subjacentes à operação. Aqui cabe uma pequena digressão.

É importante ter em mente a limitação desse pressuposto. O facto de se ter registado uma consulta associada à primeira vez em que aparece uma ligação entre banco e empresa não garante que ela seja de facto a consulta subjacente à operação correspondente a esta ligação. Pode acontecer, por exemplo, que a consulta realizada em 2016 seja para uma outra operação de 2016 e não para a operação considerada, realizada em 2017. Portanto, nada garante que as consultas sejam estritamente para a operação associada na base de dados aqui trabalhada. O que é possível afirmar é que para determinada empresa houve uma solicitação de consulta, por parte do banco interessado, no período anterior à contratação do empréstimo – podendo essa consulta ser feita no intervalo de até 6 meses. Apesar de este facto ser uma limitação, no nosso entendimento, não provoca consequências significativas que possam distorcer (ou invalidar) os resultados alcançados⁸.

Perante o que foi mencionado anteriormente pode concluir-se que para as “Antigas Ligações” espera-se que a percentagem de consultas à CRC seja inexistente ou muito baixa uma vez que, para além das trocas de informações que ocorrem naturalmente entre os agentes envolvidos, o Banco de Portugal já fornece, mensalmente, as informações provenientes de uma consulta para os bancos que já tenham regularmente relação com determinado cliente, como foi já referido. Nesse sentido, não é expectável, ainda que possível, a solicitação de consulta ativa por parte do banco.

Por outro lado, sobre as “Novas Ligações Bancárias” esperamos uma taxa de consulta elevada uma vez que, apesar da empresa já estar inserida no mercado de crédito, não possuiu relação com o futuro banco, de modo que é expectável uma consulta por parte deste último com o objetivo de aumentar o seu conhecimento sobre o potencial cliente.

⁸ Como a base de dados sobre as consultas à CRC não é indexada a um contrato específico e sim a uma empresa, a lógica do pressuposto adotado parece-nos bastante razoável. No processo de concessão de crédito existe a etapa de avaliação (*screening*) das características da empresa e do projeto subjacente ao empréstimo e nesse sentido, o que se fez foi considerar que essa etapa, que inclui a solicitação da consulta à CRC como informação adicional, pode ocorrer num intervalo de até 6 meses.

Quanto às “Novas Ligações de Mercado” espera-se que a taxa de consulta seja mais elevada do que a das “Antigas Ligações”, mas é entendido que essas taxas sejam menores do que aquelas observadas nas “Novas Ligações Bancárias” porque as primeiras não possuem informação centralizada subjacente – devido ao facto das empresas desse tipo de ligação estarem fora do mercado nos últimos 6 meses. Nesse caso, se a consulta for solicitada, é informado ao banco que a empresa em questão não possui informação centralizada no último mês de referência⁹.

É possível verificar, na Tabela I incluída de seguida, que de facto é esse o cenário que ocorre. As “Antigas Ligações” possuem as taxas de consulta mais baixas, cerca de 13%, quer dizer, dentre 100 contratos de empréstimos concedidos, sob a qualidade de “Antiga Ligação”, apenas 13 são consultados. Por outro lado, as “Novas Ligações Bancárias” possuem as taxas mais elevadas de consulta, sendo de aproximadamente 66%. Entre essas duas classes, situam-se as “Novas Ligações de Mercado” com uma taxa de consulta em torno dos 37%.

Tabela I. Composição das classes de ligação banco – empresa relativamente à consulta da CRC (%)

Classe de Ligação (%)	Sem Consulta	Consulta	Total
Nova Ligação de Mercado	63,15%	36,85%	100%
Nova Ligação Bancária	34,15%	65,85%	100%
Antiga Ligação	87,31%	12,69%	100%

Fonte: Elaborado pelo autor com base na NO e CRC.

Resumindo, para a definição da base de dados seleccionou-se apenas o primeiro registo da ligação banco-empresa. Para que fique mais claro, se um novo contrato é firmado num determinado mês então qualquer novo contrato, nos meses seguintes, que envolva o mesmo banco e a mesma empresa será ignorado da base de dados. O objetivo é analisar somente a primeira vez em que se realiza a ligação banco-empresa porque, a partir desse momento, conforme mencionado, o banco passa a receber informação sobre o seu cliente de modo automático assim como passa também a conhecê-lo melhor à medida que o histórico das suas relações se vai avolumando, não sendo por isso necessária

⁹ Apesar da informação sobre a empresa ser ou não nova no mercado de crédito (ou pelo menos de se estar ou não ausente nos últimos 6 meses) ser possivelmente declarada na etapa de avaliação, essa situação não é, necessariamente, a regra. Mesmo que essa declaração seja feita, devido, notadamente, aos problemas de assimetria, o banco não sabe, de facto, se essa é a informação verdadeira o que, no nosso entendimento, é o que motivaria a solicitação de uma consulta.

uma consulta ativa à CRC. Em teoria, pelo facto do banco não ter conhecimento prévio da empresa, situação em que a assimetria de informação é mais aguda, espera-se que a motivação para a consulta, assim como a sua necessidade, seja maior.

Na Tabela II, pode observar-se que a prática da consulta à CRC por parte das entidades bancárias varia significativamente com a classificação das ligações entre banco e empresa, o que fundamenta a opção em identificar os três tipos de ligações referidos. Nota-se que das ligações em que não há consulta aproximadamente 80% são “Antigas Ligações”. Por outro lado, das que receberam consulta, mais de 70% correspondem a “Novas Ligações” (55,89% a “Novas Ligações Bancárias” e 15,24% a “Novas Ligações de Mercado”).

Tabela II. Composição das consultas por classes de ligação banco - empresa (%)

Consulta (%)	Sem Consulta	
	Sem Consulta	Consulta
Nova Ligação de Mercado	10,29%	15,24%
Nova Ligação Bancária	11,42%	55,89%
Antiga Ligação	78,29%	28,87%
Total	100%	100%

Fonte: Elaborado pelo autor com base na NO e CRC.

Como ficará claro no Capítulo 3, o perfil da composição da consulta pelos tipos de ligação é relevante para a análise econométrica uma vez que a escolha da quantidade¹⁰ a ser estimada para captar o efeito da consulta levará em conta o facto do grupo de controlo (operações sem consulta) ser composto maioritariamente por “Antigas Ligações” que no fundo são menos interessantes para o objetivo do presente trabalho.

Quando se verificou a existência de contratos diferentes para a mesma ligação (com montante, prazo contratual e juro diferentes, entre outras características) firmados no mesmo mês e no mesmo dia, seleccionámos aleatoriamente apenas um deles. Isto é, supondo que uma ligação em determinado mês aparece, por exemplo, 10 vezes, seleccionámos apenas uma delas de modo aleatório. A seleção aleatória justifica-se para evitar problemas de endogeneidade e por não se saber, precisamente, qual a razão que leva os bancos a fazerem esse tipo de procedimento.

¹⁰ A quantidade referida é o Efeito de Tratamento Médio sobre os Tratados (ATT).

Tabela III. Distribuição do montante dos contratos de crédito

Percentil	10%	20%	30%	40%	50%	60%	70%	80%	90%
Montante (€)	5000	9000	12007,3	15483,2	20321	25000	36289,7	53374,6	125000

Fonte: Elaborada pelo autor com base na NO.

Outro filtro utilizado foi a exclusão de todas as operações abaixo de € 2500. A escolha desse valor de corte reflete a estrutura do mercado de crédito português, como pode ser observado na Tabela III, onde apresentamos a distribuição (frequência) do montante dos contratos de crédito. Por exemplo, se fosse aplicado um limite mínimo de € 25.000, implicaria a exclusão de 60% do total dos contratos de crédito. O último filtro aplicado, com o objetivo de que os dados verifiquem a hipótese de suporte comum¹¹ (*common support* ou *overlap*), foi a eliminação das observações em que sempre ocorre consulta ou nunca ocorre consulta.

Em suma, dadas as explicações acima referidas, a base final dispõe de 48.189 observações, em que cada observação é referente a uma ligação banco-empresa (contrato de crédito). No entanto, vale ressaltar que o número total de observações sofrerá modificações conforme os processos de *matching*¹² adotados e que, em via de regra, resultam em sua redução (ver Capítulo 4).

Tabela IV. Taxa de juro média (% ao ano)

Contratos de crédito	Taxa de juro (média)
Sem consulta	4,47%
Com consulta	4,10%
Total	4,36%

Fonte: Elaborado pelo autor com base na NO e CRC.

Por fim, a partir do resultado da Tabela IV, podemos verificar a existência da diferença entre as taxas de juro (média) dos contratos firmados sob consulta e da taxa de juro média daqueles contratos firmados sob a sua ausência. Os empréstimos que receberam consulta apresentam, em média, uma taxa de juro de 4,10%, inferior aos 4,47% concedidos aos contratos não consultados. Este trabalho tem como objetivo investigar em

¹¹ Ver Capítulo 3.

¹² “(...) *matching* is a procedure that involves dropping, repeating or grouping observations from a observed data set to reduce covariate imbalances between the treated and control groups that were not avoided during data collection” (IMAI et al, 2008, p. 486)

que medida esta diferença pode ser explicada pelo ato da consulta à CRC, ou seja, pelo acesso à informação que de outra maneira seria desconhecida do credor, controlando pelas diferentes características das operações.

CAPÍTULO 3 - Metodologia e Métodos de Estimação

No presente capítulo pretendeu-se expor a metodologia e os métodos de estimação aqui aplicados. Procurou-se condensar os vários e influentes artigos¹³, tanto na área de econometria como na área de estatística, que abordam os temas de inferência e causalidade no contexto de efeitos de tratamento (*treatment effects*) e avaliação de programas (*program evaluation*) com dados não experimentais (*observational data*) aplicados, em sua maioria, a técnicas de estimação por *matching*.

3.1. Efeitos de tratamento e a sua problemática para a Inferência

Considere uma amostra de $i = 1, \dots, N$ unidades. Seja $T_i = \{0, 1\}$ uma variável binária que indica se a unidade participou ou não de um programa (ou de um tratamento), com $T_i = 1$ se participou, e $T_i = 0$ se não participou. Existem N_1 unidades tratadas e N_0 não tratadas, de tal forma que $N_1 + N_0 = N$. Seja X_i o vetor de K variáveis que representam as características observáveis da unidade i e Y_i uma variável de resultado (potencial) sobre a qual se tem interesse em estimar o efeito do programa¹⁴. Neste trabalho a variável de resultado será a taxa de juro de um dado contrato de crédito.

Abordando o resultado potencial (*potential outcome framework*) considerar-se-á que a variável de resultado, Y , toma os seguintes valores: i) Y^1 , o resultado potencial no caso de ocorrência do tratamento e; ii) Y^0 , o resultado potencial no caso de ocorrência do não tratamento (controlo). Podemos, então, escrever para a amostra:

$$Y_i = T_i Y_i^1 + (1 - T_i) Y_i^0 \quad (2)$$

Onde Y_i^1 representa a variável de resultado que a unidade i teria no caso de ser tratada e Y_i^0 no caso de não receber o tratamento (operação sem consulta, no caso

¹³ Para citar alguns, ver Rubin (1973,1974), Rosenbaum e Rubin (1983), Heckman (1997) e Heckman et al (1998a, 1998b), Dehejia e Whaba (1999, 2002), Abadie e Imbens (2002,2006,2008,2011,2016), Hirano et al (2003), Angrist e Hahn (2004), Ho et al (2007), Caliendo e Kopeinig (2008), Cattaneo (2010), Austin (2009,2011), e Imbens (2015).

¹⁴ A exposição desta secção baseia-se, principalmente, nos artigos de Corseuil et al (2017), Imbens (2015), Abadie e Imbens (2002, 2006, 2011) e Imbens e Wooldridge (2009).

específico do presente trabalho). O efeito causal do tratamento T para a unidade i é definido como a diferença nos resultados potenciais, γ_i , isto é:

$$\gamma_i = Y_i^1 - Y_i^0 \quad (3)$$

No entanto, só conseguimos observar, para a mesma unidade i , ou Y_i^1 ou Y_i^0 de forma que Y_i é definido da seguinte maneira:

$$Y_i = \begin{cases} Y_i^1, & \text{se } T_i = 1 \\ Y_i^0, & \text{se } T_i = 0 \end{cases} \quad (4)$$

Portanto, o efeito individual do tratamento, γ_i , não é observável. Ou seja, o resultado é denominado “potencial” exatamente pelo facto da mesma unidade não poder ser submetida simultaneamente ao tratamento e ao não tratamento (a mesma empresa não pode ser consultada e não consultada ao mesmo tempo). Para que fique mais claro, suponha o caso em que a unidade j participe do programa (seja alvo de consulta) então

$$Y_j = T_j Y_j^1 + (1 - T_j) Y_j^0 = 1 \cdot Y_j^1 + (1 - 0) Y_j^0 = Y_j^1 \quad (5)$$

E assim o resultado observado para a unidade j , Y_j , será igual ao resultado sob o tratamento, Y_j^1 . Por outro lado, o resultado potencial (e portanto, não observável) para a unidade j , caso não fosse submetido ao programa, seria Y_j^0 , que, por vezes, é denominado também de contrafactual (*counterfactual*). O objetivo, grosso modo, da literatura que se dedica aos métodos de estimação para o efeito de tratamento é precisamente propor formas de mensurar o contrafactual.

Nesse sentido, apesar de não ser possível observar o efeito individual, podemos estimar o efeito de tratamento médio (*ATE*) comparando os valores esperados de Y_i^1 e Y_i^0 :

$$ATE = \mathbb{E}[\gamma_i] = \mathbb{E}[Y_i^1 - Y_i^0] = \mathbb{E}[Y_i^1] - \mathbb{E}[Y_i^0] \quad (6)$$

No presente trabalho, concentramo-nos noutra quantidade de particular interesse denominado efeito de tratamento médio sobre os tratados (*ATT*). A razão para essa escolha é que a subpopulação de controlo é basicamente formada por “Antigas Ligações” e desse modo, estimar o *ATE* não seria relevante, pois inclui os efeitos sobre as empresas

para as quais o programa – a consulta à CRC –, pelo menos em teoria, não foi destinado¹⁵. O efeito sobre os tratados é definido da seguinte forma,

$$ATT = \mathbb{E}[Y_i | T_i = 1] = \mathbb{E}[Y_i^1 - Y_i^0 | T_i = 1] = \mathbb{E}[Y_i^1 | T_i = 1] - \mathbb{E}[Y_i^0 | T_i = 1] \quad (7)$$

Para determinar o efeito médio, são necessárias estimativas não enviesadas de $\mathbb{E}[Y_i^1 | T_i = 1]$ e $\mathbb{E}[Y_i^0 | T_i = 1]$. Se as variáveis de interesse (Y_i^1 e Y_i^0) fossem independentes de receber ou não o tratamento (T) então,

$$\begin{aligned} \mathbb{E}[Y_i^1 | T_i = 1] &= \mathbb{E}[Y_i^1 | T_i = 0] = \mathbb{E}[Y_i^1] \\ \mathbb{E}[Y_i^0 | T_i = 1] &= \mathbb{E}[Y_i^0 | T_i = 0] = \mathbb{E}[Y_i^0] \end{aligned} \quad (8)$$

Nesse caso o efeito de tratamento médio pode ser estimado pela simples diferença das médias entre as unidades que participaram do programa com as unidades que não participaram, de forma que os dois efeitos médios, ATE e ATT , são iguais,

$$ATT = \mathbb{E}[Y_i^1 - Y_i^0 | T_i = 1] = \mathbb{E}[Y_i^1] - \mathbb{E}[Y_i^0] = ATE \quad (9)$$

Sendo assim, experiências aleatórias resolvem o problema da falta de dados (*missing value*) ou do desconhecimento do contrafactual. Se participar ou não participar do programa é aleatório (e.g, a escolha é feita com um lançamento de uma moeda) então T é independente de Y_i por definição

$$(Y_i^1, Y_i^0) \perp\!\!\!\perp T \quad (10)$$

Entretanto, o problema do contrafactual não é facilmente resolvido quando não estamos perante experiências aleatórias. Como o Banco de Portugal não lança moedas para determinar aquelas empresas que serão consultadas ou não, os dados aqui tratados correspondem, inequivocamente, ao caso observacional e não ao aleatório. Nessa situação, não é possível argumentar que a atribuição da consulta seja independente do resultado potencial (no caso, a taxa de juro).

É possível perceber que o primeiro termo da Equação (7), $\mathbb{E}[Y_i^1 | T_i = 1]$, é diretamente apreendido nos dados e pode ser estimado através de seu análogo amostral: $\bar{Y}^1 = N_1^{-1} \sum_{i:T=1} Y_i$. Todavia, não é possível observar o segundo termo, $\mathbb{E}[Y_i^0 | T_i = 1]$. Poder-se-ia pensar, de maneira intuitiva, que seria válido utilizar como contrafactual a

¹⁵ Como verificado na Tabela I e II, espera-se que a motivação dos bancos para solicitar uma consulta no caso de “Antigas Ligações” seja bastante reduzida - uma vez que as entidades bancárias recebem automaticamente informação equivalente à consulta.

média das unidades de controlo $\mathbb{E}[Y_i^0 | T_i = 0]$ para o lugar de $\mathbb{E}[Y_i^0 | T_i = 1]$. No entanto, esse estimador pode não ser consistente porque em dados observacionais pode acontecer que as covariáveis (X) que determinam a decisão do tratamento também determinem a variável de interesse, levando ao chamado enviesamento de seleção (*selection bias*). Formalmente, a partir da Equação (9) obtém-se:

$$\begin{aligned} ATE &= \mathbb{E}[Y_i^1 | T_i = 1] - \mathbb{E}[Y_i^0 | T_i = 0] \\ &= \mathbb{E}[Y_i^1 | T_i = 1] - \mathbb{E}[Y_i^0 | T_i = 1] + \mathbb{E}[Y_i^0 | T_i = 1] \\ &\quad - \mathbb{E}[Y_i^0 | T_i = 0] = ATT + \text{enviesamento} \end{aligned} \quad (11)$$

O *ATT* é identificado apenas se $\mathbb{E}[Y_i^0 | T_i = 1] - \mathbb{E}[Y_i^0 | T_i = 0] = 0$, ou seja, se os resultados das unidades dos grupos de tratamento e controlo não forem diferentes na ausência de tratamento. Como referido, em experiências aleatórias essa condição é assegurada por definição porque $\mathbb{E}[Y_i^0 | T_i = 1] - \mathbb{E}[Y_i^0 | T_i = 0] = \mathbb{E}[Y_i^0] - \mathbb{E}[Y_i^0] = 0$, mas quando se está lidando com dados observacionais é necessário recorrer a algumas hipóteses para poder identificá-lo.

3.2. Hipóteses

A identificação dos estimadores requer duas hipóteses básicas. A primeira, denominada hipótese de independência condicionada (*conditional independence* ou *unconfoundedness*), requer que a variável de resultado potencial seja independente da participação no programa condicionada às características observáveis das unidades, ou seja:

$$(Y_i^1, Y_i^0) \perp\!\!\!\perp T | X_i \quad (12)$$

Mas, como o interesse é a estimação do *ATT*, seguindo Heckman et al (1998a), essa hipótese reduz-se a:

$$Y_i^0 \perp\!\!\!\perp T | X_i \quad (H1)$$

Como precisamos estimar apenas o contrafactual $\mathbb{E}[Y_i^0 | T_i = 1]$, essa hipótese pode ser ainda menos exigente. Assim, é suficiente assumir, ao controlar-se por X_i , independência condicionada em média que conduz à hipótese:

$$\mathbb{E}[Y_i^0 | T_i, X_i] = \mathbb{E}[Y_i^0 | X_i] \text{ e } \mathbb{E}[Y_i^1 | T_i, X_i] = \mathbb{E}[Y_i^1 | X_i] \quad (H1.1)$$

Esta hipótese tem como implicação que o contrafactual respeitante ao não tratamento de uma unidade com dadas características que tenha sofrido tratamento seja obtido como o resultado observado de uma unidade não tratada com as mesmas características. Dito por outras palavras, sugere que controlando pelas características de cada unidade torna as informações de dados observacionais tão informativas quanto as observações provenientes de experiências aleatórias (Abadie & Imbens, 2011).

Recorrendo a H1 (ou H1.1) é necessário encontrar unidades não tratadas com as mesmas características das unidades tratadas. No entanto, segundo Rosebaum e Rubin (1983), para que se identifique o efeito de tratamento é necessário definir uma hipótese adicional sobre o *propensity score*. Este é uma função que resume as informações contidas no vetor X_i em termos de probabilidade de receber ou não o tratamento, e é definido por:

$$p(X_i) \equiv P(T_i = 1 | X_i) \quad (13)$$

Formalmente, esta segunda hipótese, conhecida como condição de suporte comum (*overlap* ou *common support*), estabelece:

$$0 < p(X_i) < 1 \quad (H2)$$

Novamente, como estamos interessados apenas em *ATT* é suficiente que $p(X_i) < 1$. Esta segunda hipótese (H2) requer que existam valores dos *propensity scores* comuns tanto para os tratados quanto para os não tratados. Então, sob (H1 ou H1.1) e (H2), é possível identificar o efeito do programa sobre os tratados com características $X_i = x$ como:

$$ATT(x) = \mathbb{E}[Y_i^1 | T_i = 1, X_i = x] - \mathbb{E}[Y_i^0 | T_i = 1, X_i = x] \quad (14)$$

Para identificá-lo, precisamos apenas encontrar uma maneira de estimar o segundo termo. Usando a hipótese (H1.1), o segundo termo da Equação (14) pode ser obtido de acordo com:

$$\mathbb{E}[Y_i^0 | T_i = 1, X_i = x] = \mathbb{E}[Y_i^0 | T_i = 0, X_i = x] \quad (15)$$

Finalmente, conclui-se que para se obter o efeito geral do programa sobre os tratados (ou seja, *ATT*), sob H1 e H2, basta calcular a média de $ATT(x)$ na distribuição de X para os tratados (Abadie & Imbens, 2002):

$$\begin{aligned} \mathbb{E}_{X|T_i=1}[\mathbb{E}[Y_i^1 | T_i = 1, X_i = x] - \mathbb{E}[Y_i^0 | T_i = 1, X_i = x]] \\ = \mathbb{E}_{X|T_i=1}[ATT(x)] = ATT \end{aligned} \quad (16)$$

3.3. Métodos de Estimação

A estimação do termo $\mathbb{E}[Y_i^0 | T_i = 1, X_i = x]$ pode ser realizada de várias formas¹⁶. Neste trabalho as estimações são obtidas por *matching*, tanto na forma não paramétrica, *matching* nas covariáveis, como na forma semiparamétrica mediante o *matching* no *propensity score*¹⁷. No entanto, segundo sugere a literatura¹⁸, a estimativa por *OLS* também é incluída como base de comparação.

3.3.1. *Matching on covariates – Nearest Neighbor Matching (NNM)*

Para além de ser de fácil aplicação, a estimação por *NNM* tem a vantagem de não exigir especificação de forma funcional e não realizar extrapolações significativas, sendo por isso mais robusta. A desvantagem é que não costuma ser, em geral, o método assintoticamente mais eficiente (Abadie & Imbens, 2011). Embora possa parecer uma questão relevante, Abadie e Imbens (2002, 2006, 2011) mostram que no entanto a perda de eficiência é bastante modesta até mesmo perante uma escolha de um número reduzido de *matches* (inferior a 5) e, portanto, na prática as preocupações com a ineficiência dos estimadores por *matching* tornam-se menos graves. Cálculos dos autores referidos sinalizam que, por exemplo, a variância assintótica com um único *match* é maior em até 50% do que a variância assintótica do estimador mais eficiente¹⁹, e esta proporção cai substancialmente para 12,5% quando se escolhe o número de *matches* igual a 4 (Abadie & Imbens, 2002).

Segundo Abadie e Imbens (2006, 2011), para $x \in \mathbb{X}$ e para alguma matriz simétrica definida A , defina $\|x\|_A = (x'Ax)^{1/2}$. Tipicamente escolhe-se A de modo a ser

¹⁶ Ver Frölich et al (2017), Stuart (2010), Imbens e Wooldridge (2009) e Imbens (2004) para versões mais sintéticas. Para algumas abordagens mais aprofundadas sobre métodos de estimação alternativos ver, dentre outros, Hirano et al (2003), Cattaneo (2010) e Tan (2010).

¹⁷ Ver Zhao (2004), para uma maior discussão sobre os dois estimadores.

¹⁸ Ver, por exemplo, Black e Smith (2004) e Abadie e Imbens (2011).

¹⁹ Abadie e Imbens (2002) utilizam os cálculos sobre o limite inferior de Cramer-Rao feitos por Hahn (1998) como base de comparação de eficiência assintótica.

a inversa da matriz de variância-covariância amostral, conduzindo à métrica de *Mahalanobis*²⁰,

$$A_{kxk} = \left(\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X}) \cdot (X_i - \bar{X})' \right)^{-1}, \quad (17)$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

Seja $j_m(i)$ o índice para $j \in \{1, 2, \dots, N\}$ que resolve $T_j = 1 - T_i$ e

$$\sum_{l: T_l = 1 - T_i} \mathbf{1} \{ \|X_l - X_i\|_A \leq \|X_j - X_i\|_A \} = m \quad (18)$$

Onde $\mathbf{1}\{\cdot\}$ é a função indicadora, igual a 1 se a expressão dentro das chavetas for verdadeira e 0 caso contrário. Em outras palavras, $j_m(i)$ é o índice da unidade m -ésima mais próxima da unidade i em termos dos valores de X , dentre as unidades do grupo oposto ao da unidade de i . Em particular, $j_1(i)$ é o *match* mais próximo da unidade i .

Denote $\mathfrak{S}_M(i) = \{j_1(i), \dots, j_M(i)\}$ como o conjunto de índices para as primeiras M correspondências (*matches*) para a unidade i , tal que $M \leq N_0$ e $M \leq N_1$. Por fim, seja $K_m(i)$ o número de vezes que a unidade i é usada como *match* dado que M correspondências foram usadas por unidade sendo igual a:

$$K_m(i) = \sum_{l=1}^N \mathbf{1}\{i \in \mathfrak{S}_M(i)\} \quad (19)$$

Em *matching* sem reposição, $K_m(i) \in \{0, 1\}$, mas com reposição, $K_m(i)$ pode assumir valores inteiros maiores que 1 se a unidade i for utilizada como *match* mais de uma vez. A sua caracterização é importante para a estimação (consistente) da variância desse estimador. No entanto, a sua análise²¹ não será aprofundada por não se encontrar no âmbito deste trabalho.

Como já referido, o efeito individual do tratamento $\gamma_i = Y_i^1 - Y_i^0$ não pode ser diretamente identificado porque na amostra apenas um dos resultados potenciais Y_i^1 ou

²⁰ “The most common metric is the Mahalanobis metric, which is based on the inverse of the full covariance matrix.” (Imbens & Wooldridge, 2009, PG. 38). Além disso, devido ao problema de dimensão (curse of dimensionality) é oportuno a utilização de uma métrica capaz de mapear as múltiplas covariáveis em um escalar (Zhao, 2004).

²¹ O leitor mais interessado pode consultar os vários trabalhos dos autores, como por exemplo, Abadie e Imbens (2002, 2006, 2008, 2011).

Y_i^0 é observado. Portanto, o estimador por *matching* faz a imputação do valor ausente (*missing value*) da seguinte forma,

$$\widehat{Y}_i^1 = \begin{cases} \frac{1}{M} \sum_{j \in \mathfrak{S}_M(i)} Y_j & , se T_i = 0 \\ Y_i & , se T_i = 1 \end{cases} \quad (20)$$

e

$$\widehat{Y}_i^0 = \begin{cases} Y_i & , se T_i = 0 \\ \frac{1}{M} \sum_{j \in \mathfrak{S}_M(i)} Y_j & , se T_i = 1 \end{cases}$$

Sendo o estimador do efeito de tratamento médio, *ATE*, definido por

$$\widehat{ATE} = \frac{1}{N} \sum_{i=1}^N (\widehat{Y}_i^1 - \widehat{Y}_i^0) = \frac{1}{N} \sum_{i=1}^N (2T_i - 1) \cdot \left(1 + \frac{K_m(i)}{M}\right) \cdot Y_i \quad (21)$$

E o efeito de tratamento médio sobre os tratados, *ATT*, definido por

$$\widehat{ATT} = \frac{1}{N_1} \sum_{i=1}^{N_1} (Y_i - \widehat{Y}_i^0) = \frac{1}{N_1} \sum_{i=1}^{N_1} (T_i - (1 - T_i) \cdot \frac{K_m(i)}{M}) \cdot Y_i \quad (22)$$

3.3.2. Matching on Propensity Score (PSM)

Quando estamos perante um estudo em que existem várias covariáveis (X) relevantes, o condicionar por todas pode ser limitativo devido ao problema da “maldição da dimensionalidade” (*curse of dimensionality*). Com a maior dimensão do espaço das covariáveis, os dados tornam-se mais esparsos o que dificulta o próprio processo de *matching* – fica cada vez mais difícil encontrar unidades similares à medida que o tamanho da amostra é fixo e a dimensão do vetor X aumenta. Uma solução é a utilização do *propensity score* já que condensa as informações numa única dimensão. Como vimos, essa questão foi inicialmente sugerida por Rosenbaum e Rubin (1983). Procedimentos de *matching* que fazem uso do *propensity score* para resolver o problema são conhecidos como *propensity score matching* (*PSM*) e serão um dos focos deste trabalho (Caliendo & Kopeinig, 2008).

O estimador de *matching* por *propensity score* é análogo ao apresentado na secção anterior. Considere, ao invés de $j_m(i)$, o índice $\omega_m(i)$ para $j \in \{1, 2, \dots, N\}$ que resolve $T_j = 1 - T_i$ e

$$\sum_{l: T_l = 1 - T_i} \mathbf{1}\{|\widehat{p}(x_l) - \widehat{p}(x_i)| \leq |\widehat{p}(x_j) - \widehat{p}(x_i)|\} = m \quad (23)$$

Onde $\widehat{p}(x)$ é o *propensity score* estimado.

É preciso estimar $\widehat{p}(x_i)$ uma vez que não é conhecido. Usualmente estima-se por procedimentos paramétricos do tipo *probit* ou *logit*. No presente trabalho utilizamos *logit*,

$$p(X_i) = P(T_i = 1|X_i) = \frac{\exp(X_i'\theta)}{1+\exp(X_i'\theta)} \quad (24)$$

Onde θ é um vetor de parâmetros desconhecidos.

3.3.3. OLS

Para além dos dois métodos por *matching*, estimamos também o efeito das consultas por intermédio do método de mínimos quadrados ordinários (*OLS*). Essa abordagem é apresentada para fins comparativos em que se espera a presença de algum grau de enviesamento quando comparado aos demais estimadores. A regressão estimada é a seguinte:

$$y_i = \beta_1 + X_i'\beta_2 + cons_i \cdot \gamma + \varepsilon_i \quad (25)$$

onde y_i é a taxa de juro do contrato i ; X_i é um vetor $k \times 1$ que representa as características do contrato de empréstimo, das empresas e dos bancos conforme definidos na Tabela I do Anexo II; $cons_i$ é variável *dummy* que assume valor 1 quando há consulta e 0 caso contrário para a unidade i ; β_1 é o termo independente; β_2 é um vetor $k \times 1$ com coeficientes desconhecidos; γ é o coeficiente de interesse, nomeadamente, o efeito de tratamento médio; e ε_i é o termo de erro.

3.4. Observações sobre os estimadores

Cabe ressaltar aqui alguns pontos importantes. Em primeiro lugar, Abadie e Imbens (2002, 2006, 2011) mostram que os dois primeiros estimadores acima citados – *NNM* e *PSM* – precisam de correção, por razões distintas. No que tange ao estimador por *NNM*, existem problemas de enviesamento quando se utiliza mais do que uma covariável contínua e para o *PSM*, por ser necessário estimar o *propensity score*, é preciso corrigi-lo, no que tange à variância, de modo a levar em consideração o facto de se ter utilizado estimativas anteriores (procedimento habitual quando se faz estimação em dois passos –

two step estimation)²². Segundo os autores, o estimador por *matching* corrigido (*bias-corrected matching estimator*) apresenta propriedades desejáveis, combinando vantagens e desvantagens dos estimadores por *matching* sem correção e dos estimadores por regressão.

A imputação do contrafactual passa a ser então corrigida em função das diferenças nos valores das covariáveis. De acordo com Abadie e Imbens (2002):

$$\widehat{Y}_i^1 = \begin{cases} \frac{1}{M} \sum_{j \in \mathfrak{S}_M(i)} (Y_j + \hat{\mu}_1(X_i) - \hat{\mu}_1(X_j)) & , se T_i = 0 \\ Y_i & , se T_i = 1 \end{cases}$$

e

$$\widehat{Y}_i^0 = \begin{cases} \frac{1}{M} \sum_{j \in \mathfrak{S}_M(i)} (Y_j + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_j)) & , se T_i = 1 \\ Y_i & , se T_i = 0 \end{cases} \quad (26)$$

Onde $\mu_w(x) = \mathbb{E}[Y_w | X = x]$ é a média condicionada de Y dado $X = x$ e $\hat{\mu}_w(x)$ é um estimador consistente para $\mu_w(x)$ ²³.

A segunda questão relevante que se coloca é a escolha do número de *matching* a ser utilizado. Essa decisão envolve o problema de *trade-off* entre enviesamento e variância. Escolher várias unidades no grupo de controlo ($m > 1$) para servir como contrafactual aumenta o enviesamento já que cada vez menos características similares entre as unidades do grupo de controlo e do grupo dos tratados serão observadas (portanto, a comparação torna-se menos plausível). Por outro lado, incluir várias unidades do grupo de controlo aumenta o número de observações da amostra pareada – derivada do *matching* – reduzindo a variância (Stuart, 2010). Estudos de simulação aconselham selecionar um número baixo, de forma que m esteja entre 1 a 4 (Frölich, 2004; Busso et al, 2011, 2014; Abadie & Imbens, 2011).

Em terceiro lugar, a maior crítica ao método de regressão *OLS* reside em este fazer uma extrapolação demasiado ampla, que pode não ser acertada, por assumir que unidades diferentes são boas substitutas para a estimação dos contrafatuais. A diferença fundamental entre as duas técnicas é que a estimação por *matching* propõe apenas comparar unidades com as mesmas características (estatisticamente falando, significa

²² Para uma maior discussão ver Abadie e Imbens (2016), Hirano et al (2003), Dehejia e Wahba (1999, 2002) e Heckman et al (1998a).

²³ Ver Abadie e Imbens (2002) ou diretamente em Hahn (1998) e Heck et al (1997,1998) para propostas de tal estimador.

comparar unidades com as mesmas distribuições). Na regressão, mesmo tendo sido incluídas as covariáveis (X) de controlo que permitem recorrer a situações *ceteris paribus* para interpretar os efeitos parciais do tratamento, os resultados podem ser falseados na medida em que se utilizam unidades diferentes para obter o contrafactual e essas unidades podem ter um comportamento estruturalmente diferente. Não ter esse cuidado metodológico tende a gerar distorções, como mostram, por exemplo, Black e Smith (2004), Black (2015) e Imbens (2015).

Por fim, argumenta-se que, mesmo sob as hipóteses H1 (ou H1.1) e H2, pode persistir o problema de identificação (*selection bias*) devido a características não observáveis que influenciam a participação ou não num determinado programa (ou tratamento). A introdução de *dummies* para identificar cada banco comercial teve como finalidade precisamente tentar superar essa questão – não considerar a sua presença gera problemas de endogeneidade que exigem técnicas como variáveis instrumentais²⁴ para sua solução, abordagem conhecida na literatura como efeito de tratamento médio local (*LATE*)²⁵. Conforme referimos anteriormente, a solicitação da consulta à CRC é, em última instância, uma decisão monocrática por parte dos bancos e, portanto, incluir o efeito individual pode ser uma estratégia razoável para captar as respetivas possíveis características não observáveis.

CAPÍTULO 4 - Resultados

O presente capítulo dedica-se a traçar as discussões e a apresentar os resultados relacionados com o impacto das consultas realizadas à CRC sobre o custo do crédito (taxa de juro). Assim, como sugerido pela literatura²⁶, antes de introduzirmos propriamente as estimativas obtidas, apresentamos formas de avaliar as hipóteses subjacentes à nossa estratégia de identificação, nomeadamente, a hipótese de *overlap*²⁷, e adequação dos nossos resultados verificando as características *balance* das covariáveis. Em seguida,

²⁴ Ver, principalmente, Heckman (1997) e Imbens e Angrist (1994).

²⁵ Para uma discussão mais recente ver Frölich (2007).

²⁶ Ver, entre outros, Caliendo e Kopeinig (2008) para uma excelente sistematização das etapas envolvidas nas estimativas por *matching*. Rubin (2008) recomenda encontrar um modelo aquedado em termos de *balance* antes de avaliar os resultados estimador para o efeito de tratamento.

²⁷ Vale lembrar que a única hipótese verificável é a de *overlap* uma vez que a hipótese da independência condicionada não é testável, ainda que a avaliação do *balance* possa sugerir algum grau de adequabilidade desta última (ver Imbens 2004, 2014, 2015).

fornecemos as estimativas do efeito da consulta à CRC obtidas através dos modelos de regressão e procedimentos de *matching*.

4.1. *Balance* e avaliação da hipótese de *overlap*

Como já mencionado, a simples comparação entre os valores esperados (médias) de Y_i^1 e Y_i^0 não é capaz de identificar o efeito causal do tratamento. Vimos também que a explicação reside na possibilidade desta comparação estar contaminada pelo efeito de outras variáveis (X_i) correlacionadas tanto com a decisão de participar do tratamento, T_i , bem como com os resultados potenciais, Y_i . É precisamente superar essa questão que motiva a utilização do procedimento por *matching*. Se o *matching* for bem realizado, então é esperado que a possível correlação existente entre T_i e Y_i seja explicada pelas características comuns das unidades, já que serão comparadas apenas aquelas unidades similares, permitindo que se identifique, assim, o efeito causal do tratamento (Abadie & Imbens, 2011).

Nesse sentido, os procedimentos de *matching* devem ser seguidos de uma avaliação do balanceamento das covariáveis entre as unidades dos dois grupos (tratados e controlo), onde o *balance* é definido como a semelhança entre as distribuições empíricas de tais covariáveis nos grupos correspondentes (Stuart, 2010). Portanto, o objetivo de verificar a proximidade das unidades após o *matching* é, precisamente, assegurar-se de que o *selection bias* será significativamente reduzido, ou até mesmo eliminado, na análise final.

Para esse efeito, segundo Imai et al (2008), Austin (2011) e Stuart (2010), realizar um teste estatístico, do tipo *t-test*, sobre a igualdade de médias das covariáveis é, do ponto de vista metodológico, inadequado. Resumidamente, os autores explicam que a (não) significância estatística da diferença das médias pode resultar da dimensão reduzida da amostra utilizada no teste – uma vez que o procedimento de *matching* altera o número de observações da amostra original –, e não porque o *matching* tenha funcionado de forma acertada. Além disso, os testes estatísticos fazem referência ao conceito mais amplo de população ou superpopulação, estatisticamente falando, mas autores como Imai et al

(2008) e Stuart (2010) argumentam que as propriedades do *balance* são particulares a cada amostra²⁸.

Se o método por *matching* estiver bem especificado, então, devemos estar à espera que a distribuição de uma covariável seja a mesma, ou pelo menos muito similar, para todos os níveis de tratamento – os contratos que receberam consulta à CRC e aqueles que não receberam. Nessa situação, a covariável é dita balanceada (*balanced*). No presente trabalho, investigamos as diferenças antes e depois do *matching* através da diferença padronizada (*standardized difference*) e do rácio das variâncias²⁹ (*variance ratio*). Num cenário ideal, uma covariável perfeitamente balanceada teria uma diferença padronizada de zero e uma razão de variâncias de um. Portanto, valores próximos de zero e de um, respetivamente, são desejáveis (Austin 2009, 2011; Abadie & Imbens, 2011).

De forma similar ao utilizado por Austin (2009,2011), a diferença padronizada (d) para uma covariável contínua³⁰ é definida por:

$$d = \frac{(\bar{x}_t - \bar{x}_c)}{\sqrt{\frac{s_t^2 - s_c^2}{2}}} \quad (27)$$

Onde \bar{x}_t e \bar{x}_c denotam a média amostral das covariáveis no grupo de tratamento e controlo, respetivamente e, s_t^2 e s_c^2 denotam a variância amostral das covariáveis no grupo de tratamento e controlo, respetivamente.

A diferença padronizada compara a diferença em unidades de desvio padrão e assim, para além de não ser influenciada pela dimensão da amostra, permite a comparação de variáveis mensuradas em unidades distintas. Somado a isso, a comparação também das variâncias entre os dois grupos produz uma melhor caracterização do grau de semelhança entre as distribuições das covariáveis (Austin, 2009, 2011).

Uma informação adicional como forma de medição da qualidade do processo de *matching*, seguindo de modo similar ao proposto por Busso et al (2011) mesmo que aplicada em outro contexto, é o cálculo da média aritmética simples como um facilitador

²⁸ Argumenta-se ainda que testes de hipóteses não são apropriados devido à relação entre as mudanças no *balance* e mudanças na potência estatística dos testes. Para um maior aprofundamento destas questões, recomenda-se, fortemente, a leitura de Imai et al (2008) e Imbens e Wooldridge (2009).

²⁹ Usualmente esse rácio é expresso na forma (Linden et al, 2013): $VR = \frac{s_t^2}{s_c^2}$ onde s_t^2 e s_c^2 denotam a variância amostral das covariáveis no grupo de tratamento e controlo, respetivamente.

³⁰ Facilmente estendida para o caso discreto utilizando o conceito de proporção - ver Austin (2009, 2011).

na interpretação dos resultados. O facto das medidas de diferença padronizada e rácio de variância não dependerem da dimensão da amostra e também de termos um alto número de covariáveis favorece a utilização dessa ferramenta suplementar.

A observação da Figura 1 mostra que as densidades do *propensity score* de tratados e controlo possuem muito pouca massa de probabilidade em comum, sugerindo que a hipótese de *overlap* revela-se fraca³¹. A fragilidade de tal hipótese – quando o *propensity score* se aproxima dos valores limites 0 e 1 ou quando se tem pouco *overlap* das massas de probabilidade – gera consequências em termos da consistência da estimação, principalmente em relação à variância assintótica, traduzindo-se, por conseguinte, em resultados de inferência estatística menos confiáveis (Imbens, 2004; Crump et al, 2006, 2009).

A relevância e a preocupação com essa questão foi destacada por diversos autores, principalmente a partir dos trabalhos de Heckman et al (1997) e Dehejia e Wahba (1999). Na prática, os investigadores que encontram um conjunto de observações cujos *propensity scores* não apresentam suporte comum muitas vezes reagem a esse problema optando por eliminar as observações com *score* próximo dos valores limites 0 e 1, onde esse valor de corte (*threshold*) é escolhido de maneira *ad hoc* (Imbens & Wooldridge, 2009). Assim, desde o final da década de 1990 e início dos anos 2000 vários métodos de corte (*trimming*) foram propostos na literatura.

Nesse sentido, procurou-se utilizar neste trabalho algumas dessas abordagens para lidar com o dito problema, nomeadamente: a regra “mínimo e máximo” por Caliendo e Kopeinig (2008); a escolha pelo percentil 99% por Lechner e Strittmatter (2017b) e; a regra “[0,1;0,9]” sugerida por Crump et al (2009).

A abordagem de “mínimo e máximo”, apresentada por Caliendo e Kopeinig (2008), consiste na exclusão de todas as observações cujo *propensity score* é menor que o maior dos mínimos e maior que o menor dos máximos no conjunto dos dois grupos. Para que fique mais claro, e seguindo o exemplo dos próprios autores, vamos supor que o *propensity score* esteja dentro do intervalo [0,07; 0,94] no grupo de tratamento e dentro de [0,04; 0,89] no grupo de controlo. Assim, com o critério “mínimos e máximos”, o

³¹ Segundo Lechner (2001, 2008) a questão sobre a falta ou não de suporte comum pode ser avaliada mediante a observação das distribuições do *propensity score*, não sendo necessário, portanto, a adoção de estimadores mais complicados.

suporte comum é dado por $[0,07; 0,89]$. As observações que estão fora desta região são descartadas da análise. (Caliendo & Kopeining, 2008, p. 45).

A partir de simulações com diferentes estimadores e especificações – inclusive de variações do processo gerador dos dados –, Lechner e Strettmatter (2017b) sugerem dois procedimentos de *trimming* que apresentaram os melhores desempenhos, nomeadamente, a exclusão das observações do grupo de tratamento com *propensity score* maior que o valor máximo do *propensity score* do grupo de controlo e o procedimento que propõe a exclusão dos valores acima do percentil 99%. No presente trabalho, o critério escolhido foi o da exclusão de valores acima do percentil 99%, visto que com o outro critério elegido pelos autores mencionados (quanto ao valor máximo) os resultados obtidos foram muito similares aos produzidos pela regra de “mínimo e máximo”.

Finalmente, Crump et al (2006, 2009) desenvolveram uma abordagem mais sistemática, em contraponto às práticas mais *ad hoc* encontradas na literatura, para lidar com a falta de *overlap*. Embora tenham chegado a uma fórmula fechada para decidir o valor de corte limite, simulações feitas pelos autores permitiram a sugestão de um procedimento mais simples, uma espécie de “regra de bolso”, consistindo na eliminação das observações com um *propensity score* estimado fora do intervalo $[0,1;0,9]$ e que mostrou alcançar resultados muito próximos da regra ótima. Seguindo autores como Busso et al (2011, 2014) e Lechner e Strettmatter (2017a, 2017b) iremos implementar também este procedimento no estudo empírico deste trabalho.

Perante os três critérios escolhidos, o critério de “mínimo e máximo” é aquele que confere maior grau de discricionariedade (*ad hoc*). Já as outras duas regras possuem um carácter mais sistemático. A regra “percentil 99%” é resultado de um estudo de simulação com diversos procedimentos de *trimming* existentes na literatura, enquanto o critério “[0,1;0,9]” busca estabelecer um valor de corte ótimo para o *propensity score* no que tange a exclusão de observações. Portanto, as três regras escolhidas são de certa forma representativas e absorvem bem os métodos expostos na literatura para lidar com o problema de *overlap* fraco, seja de modo *ad hoc*, por estudo comparativo (simulação) ou por otimização.

Todavia, chama-se a atenção para o facto de que a adoção de tais regras requer mudanças no objeto de estimação (*estimand*) assim como na sua interpretação. A partir do momento em que se eliminam observações para endereçar o problema de *overlap*, as conclusões resultantes passam a ser cada vez mais restritas e sensíveis à amostra

analisada. Isto é, em consequência da aplicação de tais regras, as estimativas encontradas para o efeito de tratamento médio ficam estritamente relacionadas com a própria amostra, não permitindo uma possível extensão dos resultados para o total da população. Sendo assim, as conclusões passam a ter maior validade interna – consistência – em detrimento de validade externa – extrapolações em termos da generalidade dos resultados encontrados (Crump et al, 2009). Em suma, quanto maior o *trimming*, mais os resultados subjacentes se aplicam estritamente à amostra utilizada, sendo a conclusão específica e não geral.

4.2. Análise dos Resultados

Nessa secção, explicitam-se os resultados obtidos para o *ATT* com os métodos de *matching PSM*, *NNM* e *NNMe* que se encontram sintetizados na Tabela VI do Anexo II. O número de *matches* (m) escolhido contempla quatro situações, baseada em Abadie e Imbens (2011), nomeadamente, $m = 1$, $m = 4$, $m = 16$ e $m = 64$.

O estimador *NNMe*, que significa “*exact nearest neighbor matching*”, é um caso particular do estimador *NNM* com a diferença que o pareamento (*matching*) entre as observações só é realizado após uma pré-classificação de acordo com as covariáveis escolhidas. No cenário ideal, o *NNMe* deve ser utilizado contemplando todas as variáveis discretas presentes no modelo. No entanto, isso nem sempre é possível já que pode não haver observações suficientes para gerar estratificações cada vez mais complexas. No presente trabalho foi dada prioridade à comparação entre contratos de mesma classe de ligação devido à sua relevância para a investigação. Além destas, as variáveis “NUTS” e “Colateral” foram também incluídas no procedimento de “*matching exato*” como o resultado de todas as combinações possíveis entre as variáveis discretas adotadas na tentativa de incluí-las no procedimento de “*matching exato*”.

Na segunda coluna da Tabela VI, encontra-se a estimativa obtida por mínimos quadrados ordinários (*OLS*), apresentados para fins comparativos. O primeiro quadro da Tabela VI apresenta os resultados considerando-se a amostra completa. Em seguida, apresentamos as estimativas derivadas de três subamostras, detalhadas nos três quadros subsequentes. No segundo quadro, referente à regra “mínimo e máximo”, descartamos todas as observações com um *propensity score* estimado fora do intervalo [0,0003;0,9671]. No terceiro, seguindo Lechner e Strettmatter (2017b), excluimos todas as observações do grupo dos tratados com um *propensity score* estimado acima do

percentil 99% do grupo de controlo (maiores que 0,7475). Finalmente, no último quadro calculamos as estimativas com base no critério de Crump et al (2006, 2009), sendo descartadas todas as observações com *propensity score* fora do intervalo [0,1;0,9].

A partir da análise da mesma tabela, alguns aspetos se destacam. O primeiro deles, assumindo a maior adequabilidade e consistência dos estimadores por *matching* para estimar o efeito de tratamento médio com ou sem o problema do suporte comum, diz respeito ao enviesamento dos resultados do *OLS*. Como mencionamos anteriormente, na regressão tradicional, sem nenhum tipo de ajustamento, extrapola-se demasiado os dados ao ignorar-se o conceito de *overlap*. Nesse contexto, recorrer ao *ceteris paribus* não garante a interpretação adequada dos efeitos parciais do tratamento. O efeito de tratamento médio obtido por *OLS* é de -0,1496 p.p enquanto que as estimativas por *matching* (tanto *PSM* como *NNM* e *NNMe*) variam desde -0,1624 p.p a -0,3271 p.p.

O segundo aspeto refere-se ao *trade-off* entre enviesamento e variância. Com o aumento do número de m , fica claro pelo desvio padrão que à medida que as estimativas resultantes vão se distanciando daquelas encontradas para m igual ou inferior a 4, ele reduz-se substancialmente. Observa-se que, por exemplo, no primeiro quadro da Tabela VI, na estimativa por *NNM* com $m = 1$ temos um desvio padrão de 0,0921 e estimativa de -0,1953, enquanto que para $m = 64$ temos um desvio padrão de 0,0291 e estimativa de -0,3271. A escolha de $m = 1$ ou $m = 4$ conduz-nos a uma inferência mais confiável em termos de enviesamento, com o custo de sacrificar alguma precisão, o que vai de acordo com a preferência da literatura diante da decisão a ser tomada entre enviesamento e variância (Imbens & Wooldridge, 2009).

Em terceiro lugar, tem-se que os resultados por *NNM* são penalizados quando há demasiada redução na dimensão da amostra. Quanto menor o número de observações, mais difícil é a tarefa de encontrar unidades similares e piores são as comparações. Essa é uma razão importante para justificar a disparidade das estimativas deste estimador, em sua maioria estatisticamente não significativas, quando utilizada a regra [0,1;0,9] uma vez que uma grande quantidade de observações foi descartada (22.902 observações).

Ademais, destaca-se que, diante das regras adotadas e a partir do exame das Figuras 2, 3 e 4 do Anexo II, apenas o critério proposto por Lechner e Strettmatter (2017b) teve o resultado esperado, qual seja, a melhora do *overlap* entre as densidades do *propensity score* de ambos os grupos. As figuras mencionadas mostram a densidade

estimada das probabilidades previstas de que uma empresa que não recebeu consulta (controle) receba a consulta e a densidade estimada das probabilidades previstas de que uma empresa que foi consultada (tratamento) não a seja. Quanto à Figura 2, observa-se que não há muita massa de probabilidade perto de 1, e as duas densidades estimadas têm a maioria de suas respectivas massas em regiões nas quais elas se sobrepõem³². Nesse sentido, ao contrário do observado na Figura 1 do Anexo II, não há evidências de que a hipótese de *overlap* seja fraca. Portanto, a avaliação dos resultados e sua interpretação será concentrada nos quadros 1 e 3 da Tabela VI referentes, respectivamente, à amostra completa e à subamostra derivada da regra “percentil 99%”.

No caso mais específico do primeiro quadro da Tabela VI, situação na qual existe fragilidade da hipótese de *overlap*, a análise dos resultados ficará centrada no estimador *NNM* uma vez que estudos de simulação, como por exemplo Busso et al (2011, 2014) e Lechner e Strattmatter (2017a, 2017b), mostram que o estimador por *matching* nas covariáveis é o que tem o melhor desempenho em comparação com outros estimadores na presença de *overlap* fraco.

Os resultados obtidos³³ quanto ao balanceamento das covariáveis entre os grupos dos tratados e o grupo de controle são, de modo geral, muito similares. Com exceção do rácio de variância dos identificadores bancários para os estimadores *NNMe* para os quais resultam em valores por vezes muito discrepantes daquele esperado, o balanceamento é atingido na maioria dos casos de modo satisfatório, com os valores para a diferença padronizada próximos de 0 e o rácio de variância próximo de 1. Utilizando-se a média simples como ferramenta suplementar para a interpretação (Tabela VII e VIII), o estimador com melhor propriedade de balanceamento, tanto para a amostra completa como para a subamostra derivada da regra “percentil 99%”, é o estimador *NNM* com $m = 1$. A principal diferença entre os estimadores *NNM* e *NNMe* é que o primeiro é capaz de comparar mais fielmente os contratos realizados com o mesmo banco, enquanto que o *NNMe* faz a comparação exata entre os contratos com a mesma classe de ligação.

Uma vez apontadas as limitações e as cautelas necessárias, os resultados da Tabela VI de maior credibilidade são, de um modo geral, aquelas estimativas referentes à escolha de $m = 1$ ou $m = 4$, estatisticamente significativas a 5%. Seguindo os mesmo moldes de

³² Devido à estimação do *ATT*, a maior preocupação sobre a violação do *overlap* é no limite superior, quando essa probabilidade fica muito próximo ou igual a 1. O limite inferior é menos importante, nesse caso.

³³ Disponíveis mediante solicitação.

exposição de autores que se utilizam de várias estimações, na tentativa de explorar a sensibilidade dos estimadores, como por exemplo Dehejia e Wahba (1999), Smith e Todd (2005) e Imbens (2015), argumentamos que, para os dados obtidos através das bases da CRC e da NO relativos ao ano de 2017, a solicitação da consulta teve um efeito negativo sobre a taxa de juro média dos contratos de crédito com estimativas pontuais no intervalo -0,1624 p.p e -0,2408 p.p para o total da amostra e -0,1227 p.p e -0,2612 p.p para a subamostra gerada pela regra “percentil 99%”. Mesmo não tendo sido eficiente em termos de resolução do problema de *overlap*, os resultados obtidos através dos quadros 2 e 4 da Tabela VI não se afastam da conclusão mais geral, nomeadamente, do efeito negativo das consultas à CRC sobre a taxa de juro. Podemos considerar, de forma razoável, que essas estimativas situam-se em torno de -0,20 p.p.

Mais precisamente, como a taxa de juro do contrato está mensurada em percentagem ao ano, e tomando como exemplo a estimativa por *NNM(1)* – primeiro quadro da Tabela VI – , a interpretação dos resultados pode ser feita da seguinte forma: dentre os contratos firmados em 2017 que receberam consulta, o efeito de tratamento médio estimado foi de uma redução em 0,1953 p.p na taxa de juro. Ou seja, aqueles contratos que recebem consulta têm, em média, uma taxa de juro associada mais pequena do que caso não tivessem sido consultadas através da CRC. No exemplo citado, o efeito de tratamento estimado de -0,1953 p.p representa uma descida aproximada de 5% em relação à taxa de juro média dos contratos de crédito, que receberam consulta, praticada no ano de 2017 (ver Tabela IV).

A conclusão é que, sob os limites do presente trabalho, os resultados alcançados ratificam a teoria económica, ilustrando, como esperado, que a utilização de uma ferramenta de partilha de informação – consultas à CRC – tem um efeito negativo sobre a taxa de juro (média) dos contratos de empréstimos realizados em 2017. Esse resultado sugere uma atenuação do problema de informação assimétrica, isto é, que a solicitação da consulta à CRC ajuda, por parte dos bancos, na melhor caracterização das empresas, o que de modo geral beneficia os bons pagadores possibilitando uma maior eficiência do mercado de crédito como um todo.

CONSIDERAÇÕES FINAIS

Este estudo investiga o impacto da solicitação de consulta à CRC sobre a taxa de juro média dos contratos de crédito concedidos às empresas residentes em Portugal no ano de 2017. Foram empregues dados disponibilizados pelo Banco de Portugal, dentre eles, a base de dados das consultas à CRC e a base de dados referentes às NO desde janeiro até dezembro de 2017. Utilizou-se, também, dados referentes à base das consultas ao segundo semestre de 2016 além da lista de devedor-credor do próprio registo da CRC desde julho até dezembro de 2016.

A contribuição deste estudo para a literatura reside em dois aspetos maiores. O primeiro deles resulta da investigação do efeito pontual no mercado de crédito quando os bancos consultam a CRC para obter informação adicional a respeito dos potenciais clientes. Geralmente, estudos que investigam os efeitos da partilha de informação sobre o mercado preocupam-se em realizar uma análise comparativa entre o estado anterior e o estado posterior da economia após a introdução de algum sistema de partilha de informação.

O segundo contributo maior diz respeito à utilização de técnicas de *matching*, tanto aquelas não paramétricas – através do estimador *nearest neighbor matching* – quanto as semiparamétricas – , mediante o *matching on propensity score* – no contexto de dados não experimentais e da estimação do efeito de tratamento médio. Além destes, seguindo a literatura econométrica, utilizamos também o estimador convencional por mínimos quadrados ordinários, para servir como base de comparação.

A maioria dos resultados obtidos está de acordo com o que foi referido anteriormente na literatura. A utilização de uma ferramenta de partilha de informação – consultas à CRC – teve um efeito negativo sobre a taxa de juro (média) dos contratos de empréstimos realizados em 2017. Esse resultado sugere uma atenuação do problema de informação assimétrica, isto é, que a solicitação da consulta à CRC ajuda, por parte dos bancos, na melhor caracterização das empresas, o que de modo geral beneficia os bons pagadores possibilitando uma maior eficiência do mercado de crédito como um todo.

Embora o resultado alcançado corrobore com a literatura, algumas reflexões podem ser apontadas, como sugestão para seu melhoramento, em decorrência de possíveis imperfeições da própria CRC como um mecanismo de partilha de informação. Podemos apontar duas delas:

- i. Inflexibilidade do limite temporal de informação – só é possível aceder às informações do mês imediatamente anterior ao mês da consulta e;
- ii. Pouca variabilidade de informação – é disponibilizada, apenas, informação centralizada e algumas variantes.

Num artigo desenvolvido por Vercammen (1995), o autor argumenta que a amplitude do intervalo temporal de informação disponível é decisivo sobre o grau de redução dos problemas associados à assimetria de informação – seleção adversa e risco moral. Tanto um histórico creditício muito curto quanto um muito longo geram consequências negativas sobre o mercado de crédito. No que tange às características das informações disponibilizadas pela CRC, nomeadamente, o curto histórico creditício, é de se esperar que o seu efeito sobre os problemas de seleção adversa e risco moral seja relativamente reduzido. Esse cenário fica evidente com um simples exemplo.

Suponha que o único mecanismo de se obter informação sobre algum devedor seja através da CRC. Pode-se imaginar nesse caso uma empresa mal pagadora em toda a sua vida creditícia, mas que, no entanto, no mês imediatamente anterior ao pedido de empréstimo, tenha resolvido o seu histórico irregular colocando as suas dívidas e compromissos em situação normal. No momento da solicitação da consulta à CRC pelo banco subjacente, a informação que chegará a esse credor é do tipo *white information*, levando a uma possível má identificação. Em consequência, o credor, por entender que tal empresa possui características de bom pagador, reduziria o custo do contrato (taxa de juro associada) quando na verdade deveria aumentá-lo ou até mesmo rejeitá-lo. Desse modo, as informações disponibilizadas pela CRC encontram-se no limite inferior do horizonte temporal estudado por Vercammen (1995).

No que tange ao segundo ponto, apesar de sua cobertura universal quanto às transações financeiras³⁴, a base de dados proveniente de um registo público de crédito está longe de estar completa, porque normalmente falha, por exemplo, em incluir dados mais detalhados e extensos. Por outro lado, agências de crédito privadas fornecem uma gama de informações que podem incluir desde o balanço das empresas assim como dados sobre diretores e executivos. Ademais, essas agências costumam cruzar informações recebidas de devedores com outras bases de dados como, por exemplo, registos criminais

³⁴ Como já referido, a CRC cobre qualquer transação de crédito acima de 50€.

e fiscais. Outro serviço relevante disponibilizado por essas agências privadas é o chamado “*credit score*” ou “*rating*”, um índice que fornece uma análise de risco do potencial cliente, variando numa escalada entre os perfis mais arriscados e os menos arriscados, baseado, normalmente, em modelos estatísticos aplicados (Jappelli & Pagano, 2000).

Este cenário no qual existe apenas uma das duas fontes de partilha de informação penaliza a economia como um todo, já que essas duas fontes costumam ser complementares. Para o funcionamento adequado e mais eficiente do mercado de crédito é esperado que os bancos possam obter a maior quantidade de informação possível de modo a proporcionar uma avaliação mais clara sobre os potenciais clientes, não se restringindo, portanto, somente a um desses dois mecanismos de partilha de informação (Jappelli & Pagano, 2000; OECD, 2012).

De acordo com o *Doing Business* do Banco Mundial essa parece ser a situação na qual se encontra Portugal. Nas estatísticas disponibilizadas³⁵ para o ano de 2018 apenas 7,8% da população adulta é coberta por uma agência de crédito privada enquanto a média para os países membros da Organização para a Cooperação do Desenvolvimento Económico (OECD) é de 63,7%. Assim, a resolução dessas questões pode vir a ser um importante fator para o aprimoramento do ambiente de crédito no país.

Uma extensão desse trabalho, como sugestão para uma investigação futura, seria avaliar quais os impactos que esse valor médio estimado em torno de -0,20 p.p (o que representa uma taxa aproximada de 5% em relação à taxa de juro média dos contratos de crédito praticada no ano de 2017) teria sobre a dinâmica da economia portuguesa, no geral, e as decisões económicas das empresas, em particular.

³⁵ Disponível em: <http://portugues.doingbusiness.org/pt/data/exploreconomies/portugal>.

ANEXO I – INFORMAÇÃO CENTRALIZADA

Suponhamos que uma empresa “Y” obtém crédito de três bancos diferentes totalizando 60.000€. Esses contratos são no valor de, respetivamente: (i) 10.000€ com o banco “A”; (ii) 20.000€ com o banco “B” e; (iii) 30.000€ com o banco “C”. O Banco de Portugal comunica mensalmente ao banco “A”, ao banco “B” e ao banco “C” as responsabilidades totais da empresa “Y” (ou seja, 60.000€) sem mencionar as entidades que concederam o crédito.

Consideremos, agora, que a empresa “Y” pretende contrair um empréstimo com o banco “D” do qual não é cliente. Caso o banco “D” deseje conhecer as responsabilidades totais da empresa “Y” terá de fazer uma solicitação para aceder a essas informações, por intermédio de uma consulta à CRC, com o devido consentimento do devedor em causa (empresa “Y”) e o pagamento de um valor nominal.

Para além da informação centralizada é disponibilizado, também, o tipo de contrato (produto financeiro), o prazo contratual, a situação do crédito (regular ou em incumprimento), o número total de credores e o valor da prestação mensal, mas sempre referente apenas ao mês imediatamente anterior ao da solicitação da consulta (Banco de Portugal, 2015).

ANEXO II - TABELAS E GRÁFICOS

Tabela V. Descrição das variáveis.

1. **Consulta** – é uma variável *dummy* que assume o valor ‘1’ se a instituição que concede o crédito no contrato de empréstimo solicitou consulta à CRC (nos termos definidos no estudo) sobre a empresa associada a este contrato, assumindo ‘0’ no caso contrário;
2. **Taxa de juro** – Taxa acordada anualizada é a taxa de juro individualmente acordada entre a instituição reportante e a sociedade não financeira em relação a cada empréstimo, convertida numa base anual e cotada como percentagem ao ano;
3. **Montante** – Valor do empréstimo expresso em milhões de euros;
4. **Prazo Contratual** – Prazo contratual corresponde ao prazo de vencimento do empréstimo, definido em número de dias;
5. **Emprego de Colateral** – O campo empréstimo com colateral é uma variável *dummy* a preencher com ‘0’ caso a operação de empréstimo não tenha colateral e/ou garantia associada e com ‘1’ no caso contrário. De notar que apenas é reportada a existência de garantia no caso de esta corresponder a, pelo menos, 100% do valor do empréstimo;
6. **Dimensão** – Código que permite distinguir a dimensão da sociedade não financeira beneficiária do empréstimo. A preencher com ‘1’ no caso de ser uma micro empresa, com ‘2’ no caso de ser uma pequena empresa, com ‘3’ no caso de ser uma média empresa e com ‘4’ no caso de ser uma grande empresa. Estas definições seguem de perto a Recomendação da Comissão Europeia de 6 de Maio de 2003* e resultaram na criação de 3 *dummies*;
7. **NUTS** – Identificação da região do devedor contemplando as seguintes regiões: Norte, Centro, Grande Lisboa, Alentejo, Algarve, Região Autónoma dos Açores e Região Autónoma da Madeira. Essa classificação resultou na criação de 6 *dummies*;
8. **Setor de atividade (CAE)** – Classificação da Atividade Económica a 2 dígitos da sociedade não financeira beneficiária do empréstimo**, tendo sido gerado 17 *dummies* ;
9. **Classe de Ligação** – Código que permite distinguir a classe de ligação do contrato de empréstimo. A preencher com ‘0’ no caso de ser uma “nova ligação de mercado”, com ‘1’ no caso de ser uma “nova ligação bancária” e com ‘2’ no caso de ser uma “antiga ligação”. Essa classificação resultou na criação de 2 *dummies*;
10. **Identificador bancário** – Código que permite distinguir os bancos comerciais participantes no mercado de crédito em Portugal, tendo sido gerado 31 *dummies*.

* Disponível em: <https://publications.europa.eu/en/publication-detail/-/publication/6ca8d655-126b-4a42-ada4-e9058fa45155/language-en>

** Disponível em: https://www.ine.pt/ine_novidades/semin/cae/CAE_REV_3.pdf

Fonte: Elaborado pelo autor com base na NO e CRC.

Tabela VI. Estimativas obtidas por *OLS*, por *matching* nas covariáveis e *matching* no *propensity score*

	<i>Amostra completa</i>	<i>m = 1</i>	<i>m = 4</i>	<i>m = 16</i>	<i>m = 64</i>
<i>OLS</i>	- 0,1496*** (0,0264)				
<i>PSM</i>		-0,1743 (0,1062)	-0,2408*** (0,0890)	-0,3384*** (0,0912)	-0,3245*** (0,0898)
<i>NNM</i>		-0,1953** (0,0921)	-0,2150*** (0,0478)	-0,2746*** (0,0319)	-0,3271*** (0,0291)
<i>NNMe</i>		-0,2283*** (0,0785)	-0,1624*** (0,0563)		
	<i>Regra “mínimo e máximo”</i>	<i>m = 1</i>	<i>m = 4</i>	<i>m = 16</i>	<i>m = 64</i>
<i>PSM</i>		-0,1896* (0,1004)	-0,2611*** (0,0884)	-0,3242*** (0,0906)	-0,3105*** (0,0894)
<i>NNM</i>		-0,1915** (0,0924)	-0,2135*** (0,0479)	-0,2746*** (0,0319)	-0,3253*** (0,0292)
<i>NNMe</i>		-0,2313*** (0,0787)	-0,1615*** (0,0565)		
	<i>Regra “percentil 99%”</i>	<i>m = 1</i>	<i>m = 4</i>	<i>m = 16</i>	<i>m = 64</i>
<i>PSM</i>		-0,1706*** (0,0521)	-0,1374*** (0,0410)	-0,1807*** (0,0355)	-0,2116*** (0,0349)
<i>NNM</i>		-0,1227** (0,0921)	-0,1309*** (0,0374)	-0,2230*** (0,0305)	-0,3576*** (0,0291)
<i>NNMe</i>		-0,1691*** (0,0562)	-0,2612*** (0,0464)		
	<i>Regra “[0,1 ; 0,9]”</i>	<i>m = 1</i>	<i>m = 4</i>	<i>m = 16</i>	<i>m = 64</i>
<i>PSM</i>		-0,2384** (0,1027)	-0,1798** (0,0808)	-0,2115*** (0,0849)	-0,3073*** (0,0914)
<i>NNM</i>		-0,0521 (0,1033)	-0,0368 (0,0461)	-0,0519 (0,0333)	-0,0757** (0,0291)
<i>NNMe</i>		-0,2081* (0,1154)			

Legenda: * $p - \text{valor} < 0,10$; ** $p - \text{valor} < 0,05$; *** $p - \text{valor} < 0,01$; desvio padrão entre parênteses. **Nota 1:** *NNMe* significa “exact nearest neighbor matching” onde foi dado prioridade a comparação entre contratos de mesma classe de ligação devido a sua relevância quanto a decisão de solicitar consulta à CRC. Além destas, as variáveis “NUTS” e “Colateral” foram também incluídas na fórmula de *matching* exato, escolha esta resultante de várias combinações entre as variáveis discretas utilizadas. **Nota 2:** Regra “mínimo e máximo” excluiu 58 observações; Regra “percentil 99%” excluiu 5945 observações e; Regra “[0,1;0,9]” excluiu 22902 observações.

Fonte: Elaborado pelo autor com base na NO e CRC.

Tabela VII. Balanceamento das covariáveis (amostra completa)

		Dados originais		Dados <i>matching</i>	
Número de observações		48.189		27.242	
Grupo de tratamento		13.621		13.621	
Grupo de controlo		34.568		13.621	

Estimadores	Dados originais		$m = 1$		$m = 4$		$m = 16$		$m = 64$	
	d	VR	d	VR	d	VR	d	VR	d	VR
<i>PSM</i>	0,1205	2,1170	0,0425	1,0937	0,0442	1,0405	0,0452	1,0141	0,0423	1,0235
<i>NNM</i>	0,1205	2,1170	0,0232	1,0211	0,0301	1,0517	0,0439	1,2938	0,0610	2,3673
<i>NNM exact</i>	0,1205	2,1170	0,0462	1,9722	0,0586	2,9410				

Legenda: d = média em valores absolutos das diferenças padronizadas ; VR = média simples dos rácios de variância

Fonte: Elaborado pelo autor com base na NO e CRC.

Tabela VIII. Balanceamento das covariáveis (subamostra com regra “percentil 99%”)

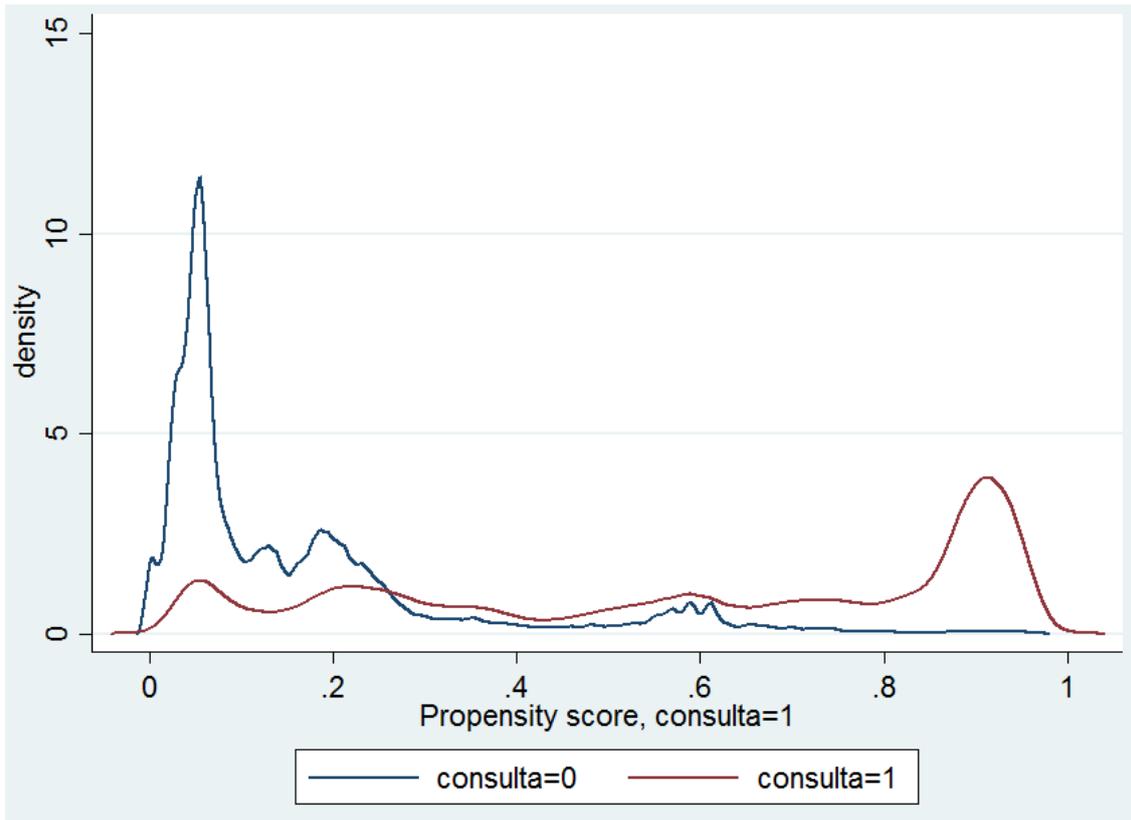
		Dados originais		Dados <i>matching</i>	
Número de observações		42.244		15.352	
Grupo de tratamento		7.676		7.676	
Grupo de controlo		34.568		7.676	

Estimadores	Dados originais		$m = 1$		$m = 4$		$m = 16$		$m = 64$	
	d	VR	d	VR	d	VR	d	VR	d	VR
<i>PSM</i>	0,0759	1,3962	0,0350	0,9066	0,0307	0,9094	0,0288	0,9151	0,0269	0,9363
<i>NNM</i>	0,0759	1,3962	0,0127	1,0187	0,0171	1,0512	0,0258	1,2992	0,0391	2,3745
<i>NNM exact</i>	0,0759	1,3962	0,0296	1,3847	0,0368	1,6891				

Legenda: d = média em valores absolutos das diferenças padronizadas ; VR = média simples dos rácios de variância

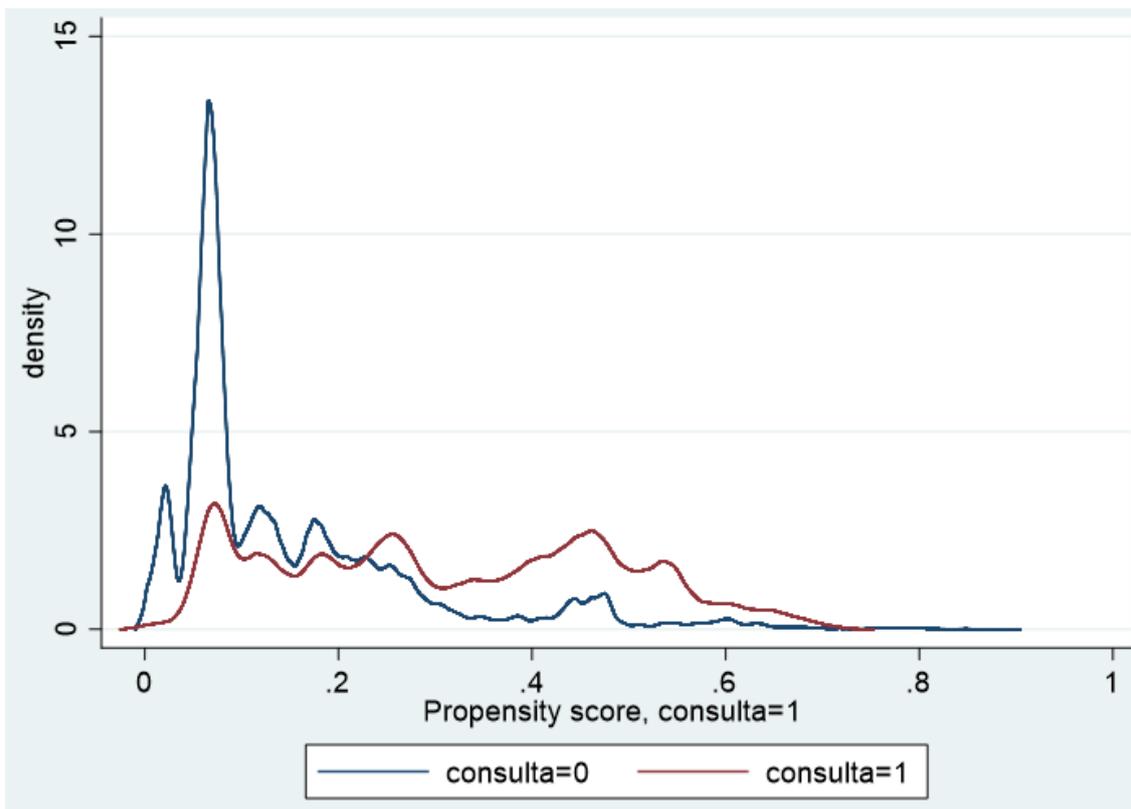
Fonte: Elaborado pelo autor com base na NO e CRC.

Figura 1. Gráfico das densidades do propensity score (amostra completa)



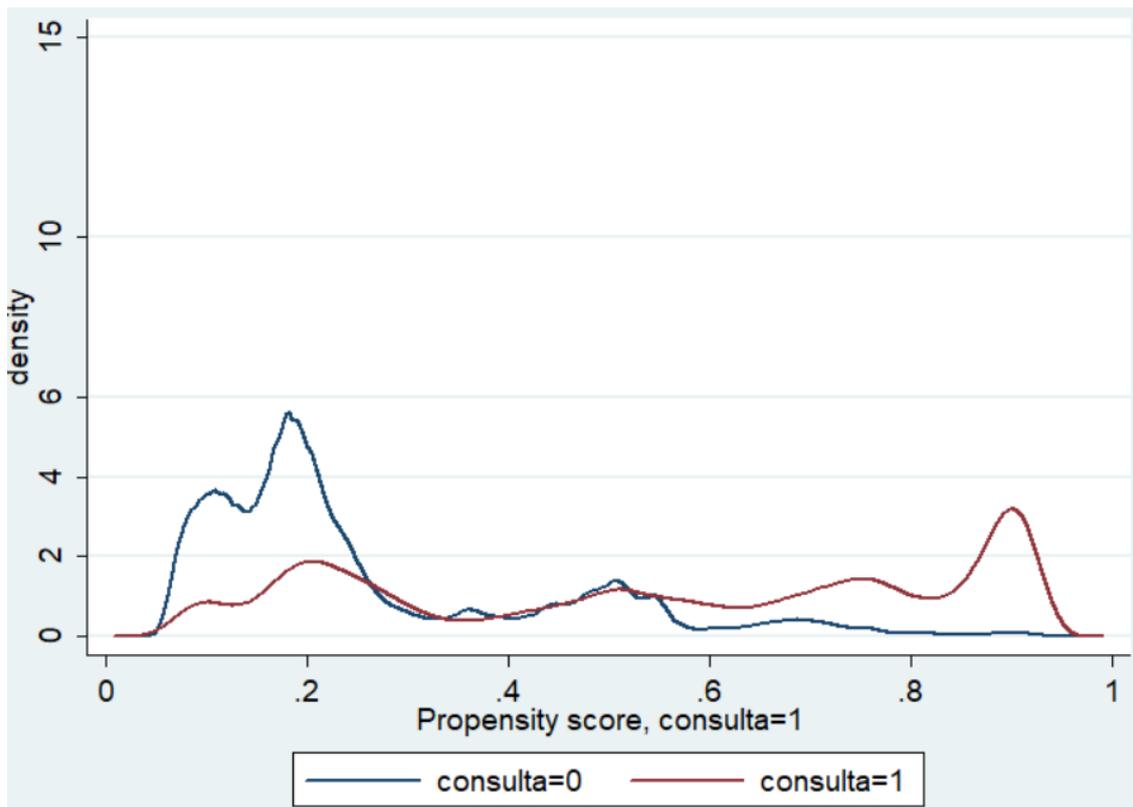
Fonte: Elaborado pelo autor com base na NO e CRC.

Figura 2. Gráfico das densidades do propensity score para regra “percentil 99%”



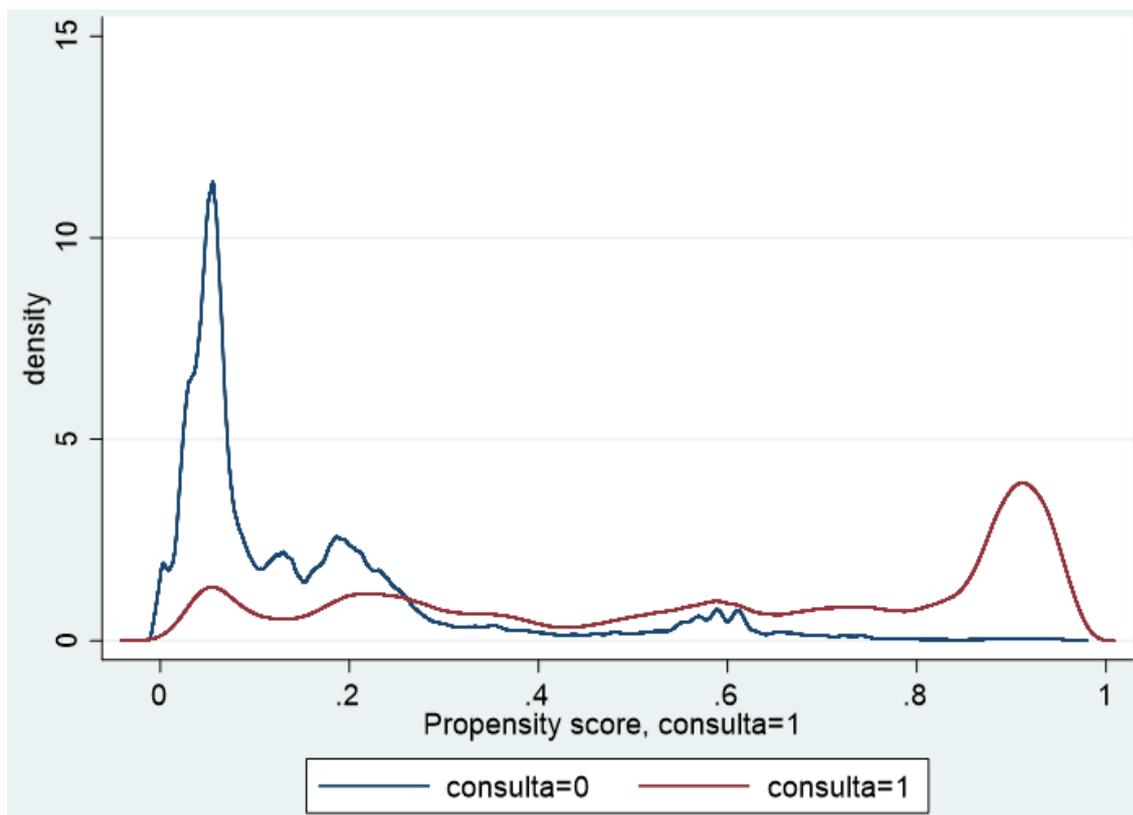
Fonte: Elaborado pelo autor com base na NO e CRC.

Figura 3. Gráfico das densidades do propensity score para regra “[0,1;0,9]”



Fonte: Elaborado pelo autor com base na NO e CRC.

Figura 4. Gráfico das densidades do propensity score para regra “mínimo e máximo”



Fonte: Elaborado pelo autor com base na NO e CRC.

REFERÊNCIAS BIBLIOGRÁFICAS

Abadie, A., e Imbens, G. (2002). Simple and Bias-Corrected Matching Estimators for Average Treatment Effects. *Technical Working Paper T0283*, NBER.

_____. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, 74(1), 235-267.

_____. (2008). On the Failure of the Bootstrap for Matching Estimators. *Econometrica*, 76(6), 1537-1557.

_____. (2011). Bias-Corrected Matching Estimators for Average Treatment Effects. *Journal of Business & Economic Statistics*, 29(1), 1-11.

_____. (2016). Matching on the Estimated Propensity Score. *Econometrica*, 84(2), 781-807.

Angrist, J., e Hahn, J. (2004). When to control for covariates? Panel asymptotics for estimates of treatment effects. *Review of Economics and Statistics*, 86(1), 58-72.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28(25), 3083-3107.

_____. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3), 399-424.

Banco de Portugal. (2015). *Central de Responsabilidades de Crédito – Cadernos do Banco de Portugal*. Disponível em: https://www.bportugal.pt/sites/default/files/anexos/pdfboletim/5_central_responsabilidades_de_credito.pdf

_____. (2017). *Boletim Económico – Outubro 2017*. Disponível em: https://www.bportugal.pt/sites/default/files/anexos/pdf-boletim/be_out2017_p.pdf

Beck, T., Demirgüç-Kunt, A., e Maksimovic, V. (2004). Bank competition and access to finance: International evidence. *Journal of Money, Credit, and Banking*, 36(3), 627-648.

Behr, P., e Sonnekalb, S. (2012). The effect of information sharing between lenders on access to credit, cost of credit, and loan performance – Evidence from a credit registry introduction. *Journal of Banking & Finance*, 36(11), 3017-3032.

Bennardo, A., Pagano, M., e Piccolo S. (2015). Multiple-bank lending, creditor rights and information sharing. *Review of Finance*, 19(2), 519-570.

Berger, A. N., e Udell, G. F. (1992). Some evidence on the empirical significance of credit rationing. *Journal of Political Economy*, 100(5), 1047-1077.

Besanko, D., e Thakor, A. V. (1987). Collateral and rationing: sorting equilibria in monopolistic and competitive credit markets. *International Economic Review*, 28(3), 671-689.

Black, D. (2015). Matching as a regression estimator. *IZA World of Labor*, 186.

Black, D., e Smith, J. (2004). How robust is the evidence on the effects of the college quality? Evidence from matching. *Journal of Econometrics*, 121(1), 99-124.

Bos, J. W.B., Haas, R., e Millone, M. (2015). Show Me Yours and I'll Show You Mine: Sharing Borrower Information in a Competitive Credit Market. In: *CentER Discussion Paper Series No. 2015-027*.

Brown, M., Jappelli, T., e Pagano, M. (2009). Information sharing and credit: Firm-level evidence from transition countries". In: *Journal of Financial Intermediation*, 18(2), 151-172.

Busso, M., DiNardo, J., e McCrary, J. (2011). New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators. Disponível em: <https://eml.berkeley.edu/~jmccrary/BDM2011.pdf>

_____. (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics*, 96(5), 885-897.

Caliendo, M., e Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72.

Cattaneo, M. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, 155(2), 138-154.

Cochran, W., e Rubin, D. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, 417-446.

Corseuil, C. H, Foguel, M, e Tomelin, L. (2017). Uma avaliação de impacto de um programa de qualificação profissional na empresa sobre a inserção dos jovens no mercado de trabalho formal. Rio de Janeiro: IPEA, *Texto para Discussão n. 2347*. Disponível em: http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=31692

Crump, R., Hotz, V. J., Imbens, G., & Mitnik, O. (2006). Moving the goalposts: Addressing limited overlap in the estimation of average treatment effects by changing the estimand. *NBER Technical Working Paper No. 330*.

_____. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1), 187-199.

Dehejia, R., e Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 94(448), 1053-1062.

_____. (2002). Propensity score matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1),151-161.

Djankov, S., McLiesh, C., e Shleifer, A. (2007). Private credit in 129 countries. *Journal of Financial Economics*, 84(2), 299-329.

Freimer, M. e Gordon, M. J. (1965). Why Bankers Ration Credit. *Quarterly Journal of Economics*, 79(3), 397-416.

Frölich, M. (2004). Finite Sample Properties of Propensity-Score Matching And Weighting Estimators. *Review of Economics and Statistics*, 86(1), 77-90.

_____. (2007). Nonparametric IV estimation of local average treatment effects with covariates. *Journal of Econometrics*, 139(1), 35-7.

Frölich, M., Huber, M., e Wiesenfarth, M. (2017). The finite sample performance of semi- and non-parametric estimators for treatment effects and policy evaluation. *Computational Statistics & Data Analysis*, 115, 91-102.

Ghosh, P., Mookherjee, D., e Ray, D. (2000). Credit rationing in developing countries: an overview of the theory. *Readings in the theory of economic development*, 6(2), 383-401.

Grajzl, P., e Laptieva, N. (2016). Information sharing and the volume of private credit in transition: evidence from Ukrainian bank-level panel data. *Journal of Comparative Economics*, 44(2), 434-449.

Hahn, J., (1998). On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. *Econometrica*, 66(2), 315-331.

Heckman, J. (1997). Instrumental variables – a study of the implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources*, 32(3), 441-462.

Heckman, J., Ichimura, H., e Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4), 605-654.

_____. (1998a). Matching As An Econometric Evaluation Estimator. *Review of Economic Studies*, 65(2), 261-294.

_____. (1998b). Characterizing Selection Bias Using Experimental Data. *Econometrica*, 66(5), 1017-1098.

Hirano, K., Imbens, G.W., e Ridder, G. (2003). Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica*, 71(4), 1161-1189.

Ho, D., Imai, K., King, G., e Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15, 199-236.

Imai, K., King, G., e Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society, Series A*, 171,481-501.

Imbens, G. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *Review of Economics and Statistics*, 86(1), 4-30.

_____. (2015). Matching Methods in Practice: Three Examples. *Journal of Human Resources*, 50(2), 373-419.

Imbens, G., e Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475.

Imbens, G., e Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5-86.

Ioannidou, V., e Ongena, S. (2010). Time for a Change: Loan Conditions and Bank Behavior when Firms Switch Banks. In: *The Journal of Finance*, 65(5), 1847-1877.

Jaffee, D.M., e Modigliani, F. (1969). A theory and test of credit rationing. *The American Economic Review*, 59(5), 850-872.

Jappelli, T., e Pagano, M. (1993), Information Sharing in Credit Markets. *The Journal of Finance*, 48(5), 1693-1718.

_____. (2000). Information Sharing in Credit Markets: A Survey. *CSEF Working Papers 36*, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy.

_____. (2002). Information Sharing, Lending and Defaults: Cross-Country Evidence. *Journal of Banking and Finance*, 26(10), 2017-45

_____. (2006). Role and Effects of Credit Information Sharing. In: *The Economics of Consumer Credit*, edited by G. Bertola, R. Disney and C. Grant. Cambridge: MIT Press, 347-371.

Lechner, M. (2001). A note on the common support problem in applied evaluation studies. *Discussion Paper No. 2001-01*, University of St Gallen, SIAW.

_____. (2008). A Note on the Common Support Problem in Applied Evaluation Studies. *Annales d'Economie et de Statistique*, 91/92, 217-234.

Lechner, M., e Strittmatter, A. (2017a). Practical procedures to deal with common support problems in matching estimation. *Econometric Reviews*, DOI: [10.1080/07474938.2017.1318509](https://doi.org/10.1080/07474938.2017.1318509).

_____. (2017b). Practical Procedures to Deal with Common Support Problems in Matching Estimation. *IZA Discussion Papers*, No. 10532, Institute for the Study of Labor (IZA), Bonn.

Linden, A., e Samuels, S. (2013). Using balance statistics to determine the optimal number of controls in matching studies. *Journal of Evaluation in Clinical Practice*, 19(5), 968-975.

OECD (2012). Facilitating access to finance: Discussion paper on credit information sharing. *OECD Publishing*. Disponível em: <http://www.oecd.org/global-relations/45370071.pdf>

Padilla, A. J., e Pagano, M. (1997). Endogenous communication among lenders and entrepreneurial incentives. *The Review of Financial Studies*, 10(1), 205-236.

_____. (2000). Sharing default information as a borrower discipline device. *European Economic Review*, 44(10), 1951-1980.

Peria, M. S. M., e Singh, S. (2014). The impact of credit information sharing reforms on firm financing. *Policy Research Working Paper*, 7013. The World Bank.

Rosenbaum, P.R., e Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41-55.

Rubin, D. (1973). The Use of Matched Sampling and Regression Adjustments to Remove Bias in Observational Studies. *Biometrics*, 29(1), 185-203.

_____. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688-701.

_____. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics*, 2(3), 808-840.

Smith, J., e Todd, P. (2005). Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?. *Journal of Econometrics*, 125(1-2), 305-353.

Stiglitz, J. E., e Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American economic review*, 71(3), 393-410.

Stuart, E. (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science*, 25(1), 1-21.

Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika*, 97(3), 661-682.

Vercammen, J.A. (1995). Credit Bureau Policy and Sustainable Reputation Effects in Credit Markets. *Economica*, 62(248), 461-478.

World Bank, (2014). *Doing Business 2015: Going Beyond Efficiency*. Washington, DC: World Bank. DOI: 10.1596/978-1-4648-0351-2. License: Creative Commons Attribution CC BY 3.0 IGO.

Zhao, Z. (2004). Using matching to estimate treatment effects: data requirements, matching metrics, and Monte Carlo evidence. *Review of Economics and Statistics*, 86(1), 91-107.