

MESTRADO
ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO
TRABALHO DE PROJETO

**IMPACTO DA MOBILIDADE NA PROPAGAÇÃO DA COVID-19 EM
PORTUGAL**

CAROLINA PEREIRA FERREIRA

OUTUBRO – 2020

MESTRADO
ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO
TRABALHO DE PROJETO

**IMPACTO DA MOBILIDADE NA PROPAGAÇÃO DA COVID-19 EM
PORTUGAL**

CAROLINA PEREIRA FERREIRA

ORIENTAÇÃO:
PROF^a DOUTORA ISABEL PROENÇA
PROF^o NUNO SANTOS

OUTUBRO – 2020

RESUMO

Em plena época de pandemia global e ao ver o mundo atravessar aquela que é apontada pelo Fundo Monetário Internacional como a mais grave crise económica desde a *Grande Depressão*, surge o interesse de estudar a importância da mobilidade individual na propagação da doença COVID-19. A análise deste tema, embora relativizado ao âmbito nacional de Portugal continental, pretende obter respostas que sejam o reflexo da realidade que acontece à escala global.

Situando-se dentro do âmbito de projeto profissional aplicado, pretende-se medir a forma como a mobilidade individual observada através do Painel de Mobilidade PSE¹ explica / não explica a progressão espacial da pandemia em Portugal. Para tal, aplicou-se a mobilidade individual observada, medida através de métricas no espaço, a um modelo clássico epidemiológico baseado no agente. Os estudos epidemiológicos existentes na literatura, por serem apenas fundados em mobilidade simulada através de exercícios de simulação das ações diárias dos agentes que promovem o contágio individual, não se alicerçam em mobilidade real e observada, como é o caso do presente estudo. Em suma, com base num número de infetados iniciais selecionados aleatoriamente, foi possível identificar ao longo de todo o período estudado, todas as interações entre agentes, e ao longo de toda a progressão e interação individual, os indivíduos foram classificados como Suscetíveis, Expostos, Infecciosos, Recuperados, bem como colocados em Quarentena e Isolamento, seguindo o modelo SEQIJR (S. Siriprapaiwan *et al.*, 2018) e tendo por base os parâmetros prescritos pelo conhecimento científico para o caso dos surtos SARS.

Trata-se de um projeto conjunto PSE e autoridades sanitárias cujo principal objetivo se prende com o determinar se a forma como os indivíduos se deslocam e interagem está associada à dinâmica espacial de disseminação da doença que é observada e declarada diariamente pelos serviços de saúde. Uma vez que o Painel de Mobilidade PSE acompanha o comportamento de mobilidade dos portugueses, é possível estudar também o impacto de outros temas relevantes, como é o caso da utilização dos transportes públicos e se a mesma pode ser endereçada como um dos fatores disseminadores da pandemia.

¹PSE - Produtos e Serviços de Estatística é uma empresa especialista em *Data Science* e Pesquisa Avançada, responsável pelo Painel de Mobilidade que serviu de base para o presente estudo. Para mais informações consulte www.pse.pt/mobilidade.

Tanto através das várias análises efetuadas aos dados de forma exploratória como através dos resultados experimentais derivados das estimações econométricas, detetou-se a existência de dependência espacial. Como tal, métodos de econometria espacial tornaram possível a introdução de termos espaciais na modelação levando à afirmação de que o número de casos positivos de indivíduos infetados pelo vírus num certo município é influenciado pelo número de casos positivos verificados em municípios vizinhos.

AGRADECIMENTOS

Gratidão é o sentimento que melhor descreve o culminar desta longa e importante jornada. Trilhar este percurso apenas foi possível com o apoio, energia e encorajamento por parte de várias pessoas a quem agradeço e dedico especialmente este projeto.

Em primeiro lugar, gostaria de agradecer à minha orientadora Professora Doutora Isabel Proença pelas suas observações, conhecimentos e sugestões essenciais ao término deste documento.

Ao professor Nuno Santos, um obrigada significativo pelo incansável apoio e transmissão de conhecimentos e ensinamentos muito além da Econometria, bem como a pronta ajuda que me foi prestando ao longo da elaboração deste projeto.

Aos meus pais e às minhas irmãs, um obrigada nunca chegará, por me tornarem quem sou hoje e serem as minhas referências no que toca ao amor que a família aporta.

Ao Miguel, por toda a força transmitida, pela compreensão e, essencialmente, pelo amor e amizade demonstrado diariamente.

Ao Instituto Superior de Economia e Gestão e a todos os professores e companheiros de mestrado com os quais tive o privilégio de me cruzar e apreender conhecimentos e aptidões.

À Produtos e Serviços de Estatística (PSE) pela disponibilização dos materiais imprescindíveis à elaboração deste projeto.

Por fim, um agradecimento a todos os amigos e familiares que estiveram ao meu lado durante a realização deste projeto e em muito contribuem para a minha felicidade.

ÍNDICE

Resumo	i
Agradecimentos	iii
Índice	iv
Índice de Tabelas e Figuras	vi
1. Introdução	8
2. Revisão da Literatura	9
2.1. Modelos Epidemiológicos Clássicos	10
2.2. Modelos de propagação epidémica com base no agente ou multiagente	13
3. Metodologia	15
3.1. Estimacão do modelo epidemiológico com base no agente e mobilidade observada	15
3.2. Abordagem espacial	18
3.2.1. Indicadores de Dependência Espacial	19
3.2.2 Matriz W de ponderacão espacial	19
3.2.3. Modelos de Regressão Espacial para Dados Seccionais	20
3.2.4. Modelos de Regressão Espacial para Dados em Painel	21
4. Análise Exploratória dos Dados	22
4.1. Caracterizacão dos Dados	22
4.2. Caracterizacão das Variáveis	23
5. Modelacão e análise dos resultados	27
5.1. Dinâmica espacial observada explicada pela dinâmica espacial esperada a partir da mobilidade	27
5.2. Dinâmica espacial observada explicada pela dinâmica espacial esperada a partir da mobilidade dos indivíduos superspreaders	30

5.3. Testar o impacto dos indivíduos que andam de transportes públicos na dinâmica da pandemia ao longo do tempo	31
5.4. Testar o impacto dos indivíduos superspreaders na dinâmica da pandemia ao longo do tempo	33
6. Conclusões, Limitações e Trabalho Futuro	34
Referências	38
Anexo I	40
Anexo II	42
Anexo III	43
Anexo IV	45

ÍNDICE DE TABELAS E FIGURAS

Tabela 1 - Características da matriz W a utilizar nos modelos econométricos	20
Tabela 2 - Caracterização das variáveis utilizadas na análise espacial	23
Tabela 3 - Caracterização das variáveis utilizadas na análise macroeconométrica..	26
Tabela 4 - Modelos Pooled para a variável LN_INF_OBS explicada pela variável LN_INF_MOB	28
Tabela 5 - Modelos Pooled para a variável LN_INF_OBS explicada pelas variáveis LN_INF_MOB e LN_INF_SS	30
Tabela 6 - Definição e valores dos parâmetros utilizados na estimação do modelo SEQIJR	40
Tabela 7 - Teste de significância para o Índice de Moran Global da variável LN_INF_OBS sob a hipótese nula de ausência de autocorrelação espacial.	41
Tabela 8 - Estatística Descritiva das variáveis utilizadas na análise espacial	42
Tabela 9 – Estatística Descritiva das variáveis utilizadas na análise macroeconométrica.....	42
Figura 1 - Propagação epidémica com base no agente no espaço e no tempo ao longo de um dia..	14
Figura 2 – Fluxograma do modelo SEQIJR	16
Figura 3 - Diagrama de Dispersão de Moran da variável LN_INF_OBS – Semana 12 a 27, 2020	40
Figura 4 - Estatística I de Moran Global Bivariada entre LN_INF_OBS e LN_INF_MOB	41
Figura 5 - Histograma do número de vizinhos da matriz de contiguidade baseada no critério da rainha.....	41
Figura 6 - Cronograma da variável OBS_DIARIO	45
Figura 7 - Cronograma da variável INF_POR_BUS.....	45

Figura 8 - Cronograma da variável INF_POR_SS	46
Figura 9 - Decomposição da variância (bus)	50
Output 1 - Modelo SARAR (1,1) com variável explicativa LN_INF_MOB	43
Output 2 - Modelo SARAR (1,1) com variáveis explicativas LN_INF_MOB e LN_INF_SS	44
Output 3- Teste de Raiz Unitária para a variável D_OBS_DIARIO	46
Output 4 -Teste de Raiz Unitária para a variável D_INF_POR_BUS	46
Output 5 - Teste de Raiz Unitária para a variável D_INF_POR_SS	47
Output 6 – Seleção da ordem do modelo VAR (bus)	47
Output 7 – Estimação do modelo VAR(6) (bus)	47
Output 8 - Teste de Causalidade de Granger (bus).....	49
Output 9 - Seleção da ordem do modelo VAR (superspreaders)	50
Output 10 - Estimação do modelo VAR(6) (superspreaders).....	51
Output 11 - Teste de Causalidade de Granger (superspreaders)	52

1. INTRODUÇÃO

A partir da segunda metade do século XX tem sido evidente a presença de uma consciencialização para a importância em preparar a humanidade contra doenças infecciosas. Ao vencer o prémio nobel em 1960, o virologista Frank Burnet deu voz às opiniões de vários investigadores afirmando que “a previsão mais provável para o futuro das doenças infecciosas é que elas serão muito monótonas” (Burnet & White, 1972). Ainda que se pensasse que a declaração estaria justificada dado o desenvolvimento de vacinas e antibióticos, não demorou muito tempo até chegar, por exemplo, o VIH, que constitui, atualmente, um dos problemas de saúde pública mais graves presenciados à escala global. No século XXI, em menos de 20 anos, a humanidade viu-se impactada por uma série de doenças infecciosas, como H1N1 (2009), SARS (2002), MERS-CoV (2012), Ebola (2014), Zika (2015) e a mais recente e atual com que todos nos deparamos, a COVID-19, tratando-se da doença provocada pela infeção do Coronavírus SARS-CoV-2 (Síndrome Respiratória Aguda Grave – Coronavírus – 2).

Em finais de 2019, as notícias que relatavam um surto na China de um novo Coronavírus impressionavam todo o mundo. A doença associada, denominada de COVID-19 pela Organização Mundial da Saúde, rapidamente se disseminou pelo planeta, onde, atualmente, afeta uma parte significativa da população mundial. A propagação foi de tal ordem que levou a que, a 11 de março de 2020, a OMS declarasse uma pandemia global.

Em Portugal, o relatório de situação divulgado pela Direção-Geral da Saúde no dia 3 de março de 2020 dava conta dos primeiros 4 casos positivos para a COVID-19 registados em território nacional. A evolução do número de casos diários motivou a que fosse declarado estado de emergência no país entre 18 de março e 2 de maio.

Os modelos e sistemas computacionais tornaram-se uma ferramenta essencial de informação, tanto para os responsáveis pela saúde pública como para o público em geral, na medida em que vêm combater a lacuna proveniente da ausência de sistemas de deteção e de controlo fiéis de epidemias. Uma forma de auxiliar as autoridades a tomarem decisões fundadas passa pelo uso de modelos matemáticos que consigam avaliar o risco de transmissão da infeção.

O que se pretende é aplicar a mobilidade observada pelo Painel de Mobilidade PSE ao modelo de propagação epidémico clássico baseado no agente, ao contrário das investigações de Hackl & Dubernet (2019) e Siriprapaiwan *et al.* (2018), que tal como toda a literatura existente na área da epidemiologia e da saúde, se baseiam puramente em trabalhos de simulação.

O presente trabalho tem como objetivo testar, globalmente e localmente, se a expansão viral decorrente dos resultados obtidos pelo modelo de mobilidade estimado está, ou não, relacionado com aquilo que é observado e declarado pela DGS. Além disso, por ser também possível rastrear a intensidade dos indivíduos que apresentam maior mobilidade ou percorrem mais tempo (indivíduos *superspreaders*), pretende-se testar do ponto de vista econométrico o impacto dos mesmos na propagação da pandemia. Outro ponto importante sobre o qual o presente trabalho pretende obter conclusões prende-se com a determinação do impacto da utilização dos transportes públicos nos números globais de indivíduos infetados.

2. REVISÃO DA LITERATURA

Determinadas hipóteses matemáticas, ao quantificarem aspetos relacionados com a biologia da propagação de epidemias, fundamentam a epidemiologia matemática através de métodos de modelação específicos para a explicação e descrição das infeções cuja transmissão é direta. Esta modelação apresenta uma grande relevância para a área de epidemiologia, na medida em que permite obter um parecer acerca dos mecanismos subjacentes à propagação e ao desenvolvimento da doença na população e, mais do que isso, possibilita a análise em termos do impacto de medidas de controlo e erradicação.

A maioria dos desenvolvimentos primordiais no que à modelação matemática de doenças transmissíveis diz respeito devem-se a médicos de saúde pública. Daniel Bernoulli, em 1760, ficou conhecido por ser o responsável pelo primeiro resultado conhecido em epidemiologia matemática ao defender uma prática de inoculação contra a varíola (Bacaër, 2011).

Relativamente à epidemiologia matemática moderna, as primeiras contribuições surgiram no final do século XIX, como é o caso do trabalho desenvolvido por En'ko (1887) com o seu modelo de tempo discreto ajustado a várias epidemias de sarampo. Por sua vez, os alicerces que corroboram toda a abordagem da epidemiologia sustentada em

modelos compartimentais foram facultados por médicos de saúde pública como Ross (1911), Hamer (1906) e Kermack & McKendrick (1927).

Os modelos matemáticos desenvolvidos na área da epidemiologia investigam dois parâmetros epidemiológicos: a força da infecção e o número básico de reprodução. Por força da infecção podem ser entendidos dois conceitos: o de incidência, que se entende como o número de novos casos de uma doença por unidade de tempo e, por outro lado, a taxa com que a doença se dissemina pela população. Já o número básico de reprodução, representado por R_0 , consiste no número de casos secundários que um caso primário é capaz de produzir numa população inteiramente suscetível a receber a doença (Hethcote, 2000). Além do R_0 , existe ainda o número de reprodução em função do tempo, denominado por R_t , que constitui o número médio de casos secundários resultantes de um caso infetado, medido em função do tempo. O que distingue estes dois indicadores é que, enquanto que o R_0 deve ser calculado somente no início da epidemia e, portanto, não tem em conta todas as medidas de controlo e contenção realizadas, o R_t é um indicador de transmissibilidade da infecção que deve ser calculado ao longo do curso da epidemia e que, como tal, é um instrumento de auxílio na medição da efetividade das medidas de contenção aplicadas.

2.1. Modelos Epidemiológicos Clássicos

Os modelos epidemiológicos clássicos, normalmente descritos por equações diferenciais ordinárias, como é o caso do *Epidemiological Model Susceptible, Infected and Recovered* (SIR) e do *Epidemiological Model Susceptible, Exposed, Infected and Recovered* (SEIR), assentam numa divisão da população em vários compartimentos através dos quais se modela as transições de uma classe para outra. São modelos bastante empregues no que à análise e compreensão da epidemia dizem respeito e, além disso, pretendem averiguar os efeitos sobre a população aquando da implementação das diferentes estratégias de combate à doença (Hethcote, 2000). Nos últimos anos, uma das abordagens para a conceção de modelos que tem evidenciado maior robustez é a de Sistemas Multiagentes (SMA) ao ter a capacidade de modelar e simular as aplicações para os problemas reais a acontecer na humanidade (Timpka et al., 2009).

Os modelos epidemiológicos tradicionais revelam epidemias relacionadas com doenças transmissíveis através de uma abordagem não espacial e com base na população.

Estes modelos são também apelidados de “modelos compartimentados” visto que uma suposição importante inerente aos mesmos se prende com o facto de a população ser dividida em vários compartimentos distintos que, por sua vez, representam os diferentes estágios da doença. Nestes modelos, as taxas de transferência entre os compartimentos encontram-se expressas de forma matemática como derivadas em relação ao tempo de demora em cada compartimento. Como resultado, o modelo representa a transmissão do vírus através de um conjunto de equações diferenciais ordinárias não lineares que associam uma taxa de transição referente à mobilidade dos indivíduos entre compartimentos.

O modelo SIR, desenvolvido em 1927 por Kermack e McKendrick, é um dos modelos mais empregues para representação de doenças infecciosas. É com base neste que todos os demais modelos são construídos e assentes sob as premissas básicas que dele se podem retirar. Esta abordagem considera que existem 3 estágios possíveis: *Suscetível (S)*, *Infetado (I)* e *Recuperado (R)*. Permite-se que os indivíduos que se encontram na classe *Suscetível* sejam aqueles que ainda não estão infetados e que, como tal, podem contrair a doença a qualquer momento, passando nesse instante para o compartimento *Infetado*, onde se encontram os indivíduos infetados, assumidos também como infecciosos por terem a possibilidade de transmitir a doença para aqueles que se encontram no compartimento *Suscetível*. Designa-se por período infeccioso o tempo que um indivíduo demora até sair do estágio *Infetado*, após o qual se desloca para o compartimento *Recuperado*. Este último estágio representa os indivíduos que recuperaram da doença e, uma vez que neste modelo se está perante a existência de imunidade contra a reinfeção, não voltam a tomar parte ativa na propagação da mesma. Portanto, no compartimento *Recuperado* encontram-se os indivíduos que foram infetados e depois removidos da possibilidade de voltarem a ter o vírus ou de propagarem a infeção.

Uma restrição que é apontada a este modelo reside no facto deste ignorar por completo os efeitos espaciais de propagação das epidemias e, como tal, despreza todas as interações existentes entre os indivíduos por modelar as populações como uma entidade contínua. Esta limitação torna-se importante uma vez que, na presença de uma doença infecciosa, o comportamento individual deve ser levado em consideração na medida em que tem uma grande influência na dinâmica da epidemia. Adicionalmente a esta restrição, o modelo SIR assume que a epidemia se trata de um processo determinístico, o que quer dizer que

o comportamento associado à população é apenas definido pelas regras que determinam o modelo e pela sua própria história, ou por outras palavras, as mesmas condições de entrada darão origem às mesmas condições de saída.

Como já referido, é através deste modelo SIR geral que surgem outros modelos com algumas variações, como é o caso do modelo SEIR, que se diferencia do SIR ao adicionar um estágio intermédio, o *Exposto*. Este compartimento adicional representa os indivíduos que foram expostos à doença (por contacto com indivíduos infetados) mas não apresentam qualquer tipo de sintomas, tratando-se por isso de indivíduos assintomáticos.

A doença COVID-19, tratando-se de um surto de uma nova doença, onde não existe ainda uma vacina disponível, adopta como únicas medidas de controlo o isolamento e a quarentena. Estas ações, associadas tanto à quarentena como ao isolamento, são essenciais no controlo da doença uma vez que se tratam de métodos eficazes de reduzir o contato existente entre populações infetadas e suscetíveis.

O modelo SEQIR é uma generalização do modelo compartimentado SEIR e a diferença perante este reside no acréscimo de duas categorias: *Quarentena* e *Isolamento*. Um indivíduo que se encontre no compartimento *Quarentena* é um indivíduo exposto, com uma certa infecciosidade, que ainda não desenvolveu qualquer sintoma clínico e que é removido do contato com a população em geral. Por sua vez, um indivíduo que se encontre na categoria *Isolamento* é um indivíduo infeccioso que é também removido do contato com a população em geral, muitas das vezes através de internamento em hospitais.

Para que seja possível formular um modelo que inclua estas atuações e que descreva o decorrer da doença assumem-se as seguintes premissas que levam ao modelo SEQIR: os indivíduos na categoria *Exposto* são transferidos para a classe *Quarentena* a uma taxa proporcional γ_1 ou, caso não sejam retirados do contato com a população, são movidos para a classe *Infetado* a uma taxa k_1 ; os indivíduos diagnosticados como infecciosos e presentes no compartimento *Infetado* são isolados e, portanto, transferidos para a categoria *Isolado*, a uma taxa proporcional γ_2 ; os membros em *Quarentena* encontram-se monitorizados e a partir do momento em que desenvolvam sintomas, são movidos para a classe *Isolado* de imediato à taxa k_2 ; os indivíduos do compartimento *Infetado* abandonam o compartimento à taxa d_1 por morte induzida pela doença e uma fração σ_1

recupera; os membros da classe *Isolado* deixam a classe à taxa d_2 por morte associada à doença e uma fração σ_2 tornam-se recuperados. Para uma melhor compreensão do modelo, um fluxograma do mesmo pode ser encontrado em Gumel *et al.* (2004).

O que se verifica na prática é que o período da quarentena é também aplicado a muitos indivíduos presentes na classe *Suscetível* mas essa questão é ignorada no modelo. O modelo antes do início da implementação das medidas de controlo, que variam consoante a gestão que é feita da pandemia, é o caso especial onde: $\gamma_1 = \gamma_2 = k_2 = \sigma_2 = d_2 = 0$, $Q = J = 0$.

2.2. Modelos de propagação epidémica com base no agente ou multiagente

Em absoluto contraste aos modelos compartimentados, os modelos de propagação epidémica com base no agente (*Agent-based models* (ABMs)) admitem a existência de interações entre os indivíduos, que aqui se intitulam de agentes, levando a que muitas das críticas apontadas aos modelos clássicos de expansão epidémica sejam ultrapassadas. Este processo, ao abordar uma modelação que parte do agente, investiga o desenvolvimento da doença através de cada indivíduo e, para tal, uma condição necessária passa pelo rastreamento dos contatos que cada indivíduo toma com outros indivíduos em todas as áreas e redes sociais relevantes, como por exemplo, âmbitos de trabalho e de escola.

O sistema adotará um determinado comportamento dependendo das interações entre os múltiplos agentes dos diferentes tipos, interações essas que são maioritariamente discriminadas através de um conjunto de normas simples e compreensíveis. É permitida uma abstração das ações humanas ao definir-se a autonomia de um agente na tomada de decisão, ou seja, sempre que, por algum motivo, se verificarem variações individuais nas regras comportamentais está-se perante uma questão de heterogeneidade.

Pese embora estes modelos já consigam obter efetivamente o rastreio da progressão da doença através de cada indivíduo, com o objetivo de explicar as interações mais complexas entre os indivíduos suscetíveis e os infetados, vários investigadores começaram a adicionar aos modelos infra-estruturas físicas, como áreas geográficas reais e redes de estradas.

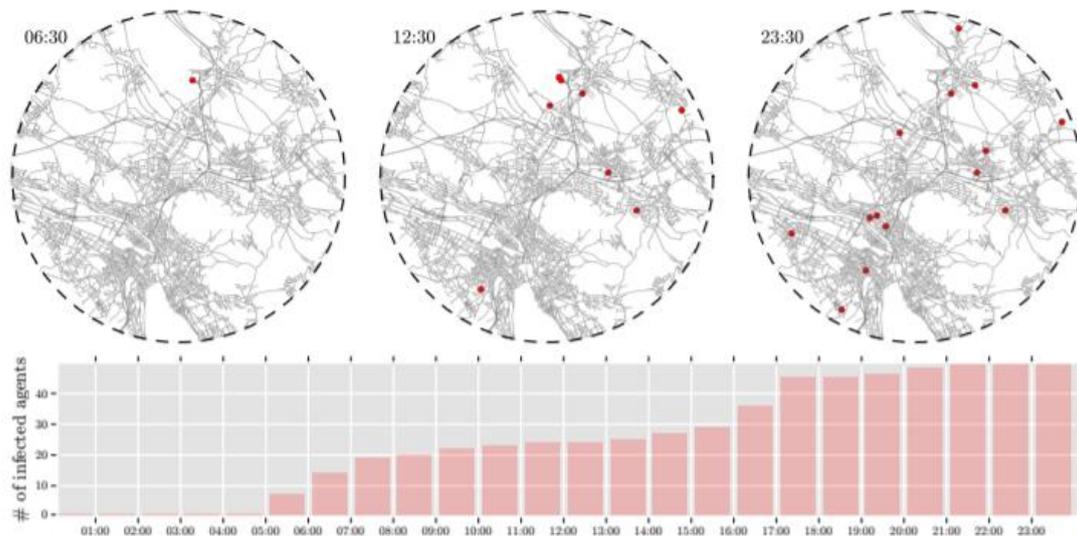


Figura 1 - Propagação epidémica com base no agente no espaço e no tempo ao longo de um dia. Fonte: Hackl, J. & Dubernet, T. (2019).

A Figura 2, presente em Hackl & Dubernet (2019), evidencia a disseminação da infeção ao longo de um dia, tendo por base a história de apenas um agente infetado. É observável que, com o decorrer das várias horas ao longo do dia, o número de agentes infetados (que se encontram assinalados a vermelho na distribuição espacial) aumenta.

Khalil (2010), juntamente com outros investigadores, propôs um modelo multiagente que simulasse a propagação do vírus H1N1 numa comunidade no Egito. O modelo proposto surge como uma extensão dos modelos epidemiológicos compartimentados já analisados anteriormente, no entanto, a escolha por parte dos autores em utilizar um modelo baseado no agente em detrimento de outra abordagem prende-se com a capacidade deste modelo em conseguir representar os fatores espaciais e temporais que tornam a modelação mais próxima da realidade.

Os ABM são qualificados pela presença de agentes heterogéneos, autónomos e potencialmente independentes, que mantêm uma interação coordenada com o intuito de

resolver o problema em questão. Tratando-se de agentes racionais, têm a completa capacidade de se adaptarem ao meio ambiente e de provocar nele as mudanças que achem necessárias de forma a alcançar o que é pretendido (Wooldridge, 2009).

É importante referir que os estudos existentes na literatura, além de serem escassos, apenas são fundados em mobilidade simulada, através de exercícios de simulação que se baseiam no contágio individual de agentes ao simularem as suas ações diárias. Como tal, não se fundamentam em mobilidade real, como é o caso do presente projeto.

3. METODOLOGIA

3.1. Estimação do modelo epidemiológico com base no agente e mobilidade observada

A lógica por detrás dos modelos compartimentados pode também ser implementada através da modelação com base no agente uma vez que, ao se monitorizarem indivíduos de uma população, é possível acompanhar o percurso realizado, indivíduo a indivíduo, nas diferentes horas ao longo do seu dia. Foi exatamente esse o processo que se assumiu na realização deste Trabalho Final de Mestrado, ao se tomar como base o modelo SEQIJR e se aplicar nele a lógica de modelação com base no agente. Quer isto dizer que, com o intuito de se estudar o impacto da mobilidade, aplicaram-se os princípios e parâmetros dos modelos epidemiológicos clássicos aos modelos de propagação epidémica com base no agente. Os agentes geram interações entre si, podem tornar-se infetados, podem ser testados, tratados, hospitalizados e, além de tudo isto, podem ainda ser adicionadas na linha do tempo as políticas levadas a cabo com o objetivo de avaliar os seus possíveis impactos.

É através do Painel de Mobilidade PSE², tratando-se de um painel representativo da população e de uma mediação contínua que possibilita identificar e medir não só a mobilidade individual, mas também as interações entre indivíduos, que se aplica o modelo clássico epidemiológico SEQIJR, com o objetivo de estimar o número de infetados esperados pela mobilidade. Por sua vez, o modelo não é calculado através de equações

²Através do Painel de Mobilidade PSE, a consultora PSE estuda a mobilidade dos portugueses, em contínuo. Este estudo, que implica uma *APP* instalada nos telemóveis da amostra participante, apresenta um elevado rigor porque regista a localização via *GPS* e simultaneamente porque se baseia numa amostra de 3670 indivíduos estatisticamente representativa do universo em estudo. Deste modo, monitoriza-se a deslocação real da população, 24 horas por dia. Para um universo de 6.996.113 indivíduos residentes nas regiões estudadas a margem de erro imputável ao estudo é de 1.62% para um intervalo de confiança a 95%.

diferenciais ordinárias, como é o caso nos modelos clássicos compartimentais, mas a aplicação através do painel de mobilidade transforma-o num modelo de agente onde se procede à contaminação ao longo do espaço e ao longo do tempo. Como se trata de um ABM, cada agente interage de forma independente e autónoma, permitindo uma alusão mais realista ao dinamismo das relações e interações entre os indivíduos ao longo dos dias.

O modelo aqui adotado apresenta-se como uma generalização do modelo apresentado por Gumel *et al.* (2004) com a diferença de que, como se trata de um painel, a dimensão está definida no momento inicial e, como tal, aumentos na população no que diz respeito a nascimentos ou migrações não são considerados, bem como diminuições relacionadas com mortes naturais para indivíduos livres da doença.

No que toca às condições iniciais necessárias para dar começo ao modelo epidemiológico, foram selecionados 9 indivíduos pertencentes ao painel, de forma aleatória, para se tornarem os infetados iniciais. É com base neste número de infetados iniciais, selecionados aleatoriamente no espaço, que é aplicada a mobilidade com o intuito de estudar se a mesma explica a expansão territorial da infeção. Assim, foram identificadas todas as interações entre agentes, e ao longo de toda a progressão e interação individual, os indivíduos foram classificados como Suscetíveis, Expostos, Infeciosos, Recuperados, bem como colocados em Quarentena e Isolamento para melhor espelhar a realidade, seguindo o modelo SEQIJR, e aplicando as respetivas ponderações dos indivíduos na população.

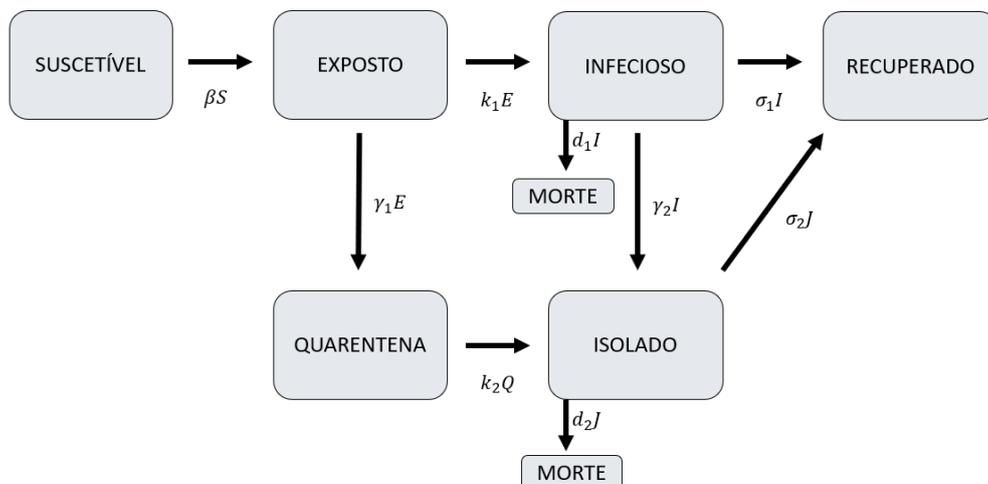


Figura 2 – Fluxograma do modelo SEQIJR. Fonte: Adaptado de Gumel *et al.* (2004)

Atendendo à Figura 2 é possível observar as etapas de transições entre os diferentes compartimentos que compõem o modelo estimado e que podem ser descritos da seguinte forma:

- *Suscetível*: compartimento constituído por todos os indivíduos que se encontram, em determinado dia e hora, na proximidade de algum indivíduo infeccioso e, como tal, têm possibilidade de entrar em contacto com o mesmo e contrair a infeção.
- *Exposto*: classe composta por indivíduos assintomáticos que foram expostos ao vírus, por contacto com indivíduos infecciosos a uma taxa β , mas ainda não desenvolveram sintomas clínicos. A população é reduzida a uma taxa γ_1 por transferência de indivíduos para o compartimento *Quarentena* e a uma taxa de k_1 por desenvolvimento de sintomas clínicos e movimentação para a classe *Infeccioso*.
- *Quarentena*: compartimento formado por indivíduos assintomáticos e colocados em quarentena a uma taxa γ_1 devido ao contato estabelecido com uma fonte de infeção. Assume-se, por questões de simplicidade, que todos os indivíduos em quarentena se tratam de indivíduos infecciosos assintomáticos que irão desenvolver sintomas e, de seguida, serão transferidos para a classe *Isolado* a uma taxa k_2 .
- *Infeccioso*: classe constituída por indivíduos sintomáticos, gerados a partir do desenvolvimento de sintomas clínicos por membros presentes no compartimento *Exposto*. A população é reduzida a uma taxa γ_2 pela transferência para o compartimento *Isolado*, a uma taxa d_1 por morte induzida pela doença e a uma taxa σ_1 por recuperação. É assumido que γ_2 representa a taxa a partir do qual indivíduos sintomáticos procuram atendimento médico e, portanto, são colocados em isolamento.
- *Isolado*: compartimento composto por indivíduos que desenvolveram sintomas clínicos e foram isolados da restante população, por exemplo, por internamento. Podem ser originários, como já foi referido, tanto da classe *Infeccioso* como da classe *Quarentena*. A diminuição da população presente é realizada à taxa σ_2 por recuperação e à taxa d_2 por morte induzida pela doença. Tal como aponta Lee *et al.* (2003), assume-se que $d_1 > d_2$ e $\sigma_2 > \sigma_1$ uma vez que os indivíduos em isolamento têm a possibilidade de receber tratamento durante a hospitalização.

Através da Tabela 6, que se encontra no Anexo I, é dado acesso à descrição e aos valores assumidos para os vários parâmetros a partir dos quais se construiu o modelo. Os valores foram estabelecidos em função da prescrição conhecida e ditada pelo conhecimento científico para o caso dos surtos SARS (Siriprapaiwan *et al.*, 2018) ocorridos em 2002 e 2003 em Pequim, eleitos como sendo os mais idênticos ao SARS-CoV-2 e na falta de informações relativamente à situação atual. Apesar da escassez de informação acerca dos parâmetros da pandemia atual e ainda que possam existir eventuais diferenças de nível, o foco prende-se em determinar se, mesmo assumindo os parâmetros conhecidos de pandemias anteriores, se consegue explicar a propagação espacial da infeção atual no território nacional. No que diz respeito aos parâmetros que estão relacionados com a implementação das políticas de quarentena e isolamento, os mesmos só são incluídos na execução do modelo a partir do dia 18 de março, dia em que foi decretado o estado de emergência em Portugal.

3.2. Abordagem espacial

Com a introdução dos fenómenos espaciais e a possibilidade de incorporar as localizações espaciais das unidades em estudo nas análises que se desejam realizar, motivadas pelo incremento da tecnologia e adequabilidade dos sistemas de informação geográfica, surgem as técnicas de análise de dados espaciais e de econometria espacial.

A metodologia inerente à econometria espacial toma especial interesse quando os problemas centrais requerem modelos e métodos onde o espaço tem uma preponderância importante na determinação desses mesmos modelos.

O que diferencia a econometria espacial da econometria tradicional? LeSage (1999) refere que, ao serem tratados dados que possuem uma componente espacial, surgem dois problemas: a dependência espacial entre as observações e a heterogeneidade espacial. Acontece que a econometria tradicional ignora por completo estas duas componentes.

Ao se permitir a dependência espacial entre as observações admite-se que determinados dados observados num lugar do espaço possam depender de valores observados noutros locais do espaço. Já o termo heterogeneidade espacial, que se traduz nas variações que se verificam nas relações espaciais, pode ser visto, no limite, como sendo a atribuição de relações diferentes para cada ponto do espaço.

3.2.1. Indicadores de Dependência Espacial

As técnicas de análise de dados espaciais foram desenvolvidas com o intuito de procurar quantificar as semelhanças entre os valores associados a uma determinada região e os valores apresentados pelas regiões vizinhas. Estas práticas são exibidas através de indicadores que medem tanto a associação espacial como o padrão específico que os valores associados às unidades geográficas possam apresentar no espaço.

O Índice de Moran Global, apontado indiscutivelmente como o indicador mais utilizado de autocorrelação espacial global, foi sugerido por Moran (1950) e mais tarde generalizado por Cliff & Ord (1972). Esta estatística mede a autocorrelação espacial com base nos valores das variáveis e nas localizações, simultaneamente. Trata-se de um indicador que pode ser sintetizado num único valor para todo o conjunto de dados e é visualizado através de um diagrama de dispersão proposto por Anselin (1996), onde a variável espacialmente desfasada se apresenta no eixo vertical e a variável original no eixo horizontal. A inclinação do ajustamento linear corresponde ao valor do índice e, como tal, valores do coeficiente negativos evidenciam autocorrelação espacial negativa e, em contraste, valores do coeficiente positivos revelam autocorrelação espacial positiva. Para além de exibir a medida global de associação linear espacial, o diagrama de dispersão de Moran disponibiliza ainda a observação de quatro quadrantes que representam os quatro diferentes tipos de associação linear espacial entre o valor de uma dada região e a média dos seus vizinhos.

Quando os dados em mão apresentam um elevado número de regiões é possível que se verifiquem diferentes regimes de associação espacial e que determinados locais apresentem uma dependência espacial mais evidenciada do que outros locais. Para tratar estes fenómenos e identificar possíveis agrupamentos em função da vizinhança de cada região, são empregues os *Local Indicator for Spatial Autocorrelation* (LISA), onde o Índice de Moran Local se considera como a estatística mais usual no conjunto de estatísticas disponíveis.

3.2.2 Matriz W de ponderação espacial

Designada por matriz W , a matriz de ponderação espacial (ou matriz de pesos espaciais) consiste numa matriz quadrada de dimensão n por n , exógena por hipótese e por ser definida *a priori*, e cujos elementos que a compõem, w_{ij} , exibem o grau de

dependência entre as regiões. Assim, seguindo um determinado critério de proximidade, a matriz consegue evidenciar a influência da região j sobre a região i (Almeida, 2012), captando a dependência espacial verificada entre as regiões.

No caso em que se lidam com estruturas geográficas irregulares, como é o caso dos municípios, a definição apropriada da matriz de pesos espaciais é essencial para que a mesma não se torne numa fonte de má especificação do modelo econométrico.

No que toca ao critério de construção da matriz W com base na contiguidade, eleger-se, para elaboração dos modelos, o critério da Rainha uma vez que o mesmo considera como vizinhas não só as regiões que apresentam fronteiras físicas comuns como também as regiões que apresentam vértices em comum. No Anexo I, através da Figura 5, é possível atentar ao histograma do número de vizinhos da matriz de contiguidade construída a partir do critério Rainha.

Tabela 1 - Características da matriz W a utilizar nos modelos econométricos

Número mínimo de vizinhos	1
Número máximo de vizinhos	10
Número médio de vizinhos	5.28

3.2.3. Modelos de Regressão Espacial para Dados Seccionais

No que diz respeito aos modelos de regressão linear espaciais, diferentes especificações podem ser levadas em consideração aquando da violação da hipótese da não existência de autocorrelação nos resíduos do modelo.

Como é sabido, o método de estimação dos Mínimos Quadrados Ordinários (OLS) é referido como sendo o Melhor Estimador Linear Centrado (BLUE). Para que seja possível atender a estas propriedades, é necessário que se cumpra o pressuposto acerca da não existência de autocorrelação no erro aleatório da equação de regressão. Na prática, estes princípios nem sempre se verificam. Se um determinado valor observado num local depende dos valores observados em locais vizinhos, poderá estar-se perante a existência de dependência espacial, que se pode manifestar tanto nas variáveis como nos termos de erro do modelo, como é especificado por Arbia (2014) através das seguintes equações:

$$y = \lambda Wy + X\beta_{(1)} + WX\beta_{(2)} + u \Leftrightarrow y = \lambda Wy + Z\beta + u \quad |\lambda| < 1 \quad (1)$$

$$u = \rho Wu + \varepsilon \quad |\rho| < 1 \quad (2)$$

com W uma matriz de ponderação espacial exógena; X uma matriz de regressores não estocásticos; $\varepsilon|X \sim i. i. d. N(0, \sigma_\varepsilon^2 I_n)$; $\beta_{(1)}, \beta_{(2)}, \lambda$ e ρ os parâmetros a serem estimados; Z a matriz de todos os regressores atuais e espacialmente defasados, definida como $Z = [X, WX]$ e β o vetor de parâmetros da regressão, definido como $\beta = [\beta_{(1)}, \beta_{(2)}]$

Enquanto que (1) aponta como um dos regressores do modelo a variável dependente espacialmente defasada, podendo também conter defasamentos espaciais de outras variáveis exógenas, (2) considera que a dependência espacial se encontra presente no termo de erro. Ao modelo conjunto, constituído por (1) e (2), Kelejian & Prucha (1998) designam de SARAR (1,1) (Modelo Espacial Autorregressivo com estrutura de Erro Autorregressivo), formando aquele que representa o modelo linear mais geral de econometria espacial, através do qual surgem os diversos modelos particulares, impondo restrições nos parâmetros:

- $\beta = 0$ e $\rho = 0$, surge o Modelo Espacial Autorregressivo Puro;
- $\lambda = \rho = 0$, denominado de Modelo de Variável Independente Espacialmente Defasada;
- $\lambda \neq 0$ e $\rho = 0$, permanece o Modelo de Defasamento Espacial (Spatial Lag Model – SLM);
- $\lambda = 0$ e $\rho \neq 0$, surge o Modelo de Erro Espacial (Spatial Error Model – SEM).

De uma forma simples, o que a especificação do modelo SEM permite analisar passa pela quantificação do efeito médio dos erros das unidades vizinhas em relação ao erro de determinada unidade espacial em análise. Por sua vez, a especificação SLM enquadra-se quando o comportamento da variável dependente consegue ser explicado pelos seus valores defasados, ou seja, valores da variável dependente verificados em unidades espaciais vizinhas e ainda por variáveis independentes que permitem caracterizar cada unidade espacial em análise.

3.2.4. Modelos de Regressão Espacial para Dados em Painel

A estrutura de dados inerente aos modelos de dados em painel abrange um conjunto de individualidades, que podem ir desde pessoas, empresas, famílias até a locais

geográficos, observados em vários períodos de tempo. No caso da modelação de dados de painel espaciais, o que se observa são múltiplas observações para os mesmos locais ou áreas geográficas em análise, dando a possibilidade de controlar a influência de determinadas características no tempo específicas às unidades geográficas.

O painel espacial típico e mais comum é caracterizado por apresentar um grande número de unidades espaciais observadas ao longo de séries temporais curtas, como é o caso do painel em estudo. Seja qual for a dimensão existente, a dupla dimensionalidade existente em dados em painel possibilita uma modelação mais rica do que em dados seccionais ou em séries temporais. Ter dados ao longo do tempo para as mesmas unidades geográficas torna-se um exercício útil, não só por permitir obter uma perspetiva das reações dinâmicas, mas também por permitir controlar a heterogeneidade não observada do local em si. Arbia (2014) afirma que, em geral, um conjunto de dados de painel espacial contém mais informações e menos multicolinearidade entre as variáveis em comparação com um conjunto de dados espaciais seccionais.

Sob o pressuposto da não existência de correlação entre os regressores e o termo de erro, a modelação *Pooled OLS* espacial pode ser utilizada para obter um estimador consistente e, ao adicionar dummies temporais como variáveis explicativas do modelo, consegue controlar os efeitos fixos no tempo. Tal como se verifica para os modelos de dados seccionais, a autocorrelação espacial pode ser considerada de várias formas: através das variáveis espaciais desfasadas, endógenas ou exógenas, e/ou através de autocorrelação espacial no erro.

4. ANÁLISE EXPLORATÓRIA DOS DADOS

4.1. Caracterização dos Dados

Como fonte primária de pesquisa e levantamento de dados encontra-se a Direção-Geral da Saúde ao fornecer relatórios de situação diários que dão conta do atual estado da pandemia em Portugal. A partir destes relatórios é possível retirar diferentes tipos de dados, tanto de natureza de séries temporais como de dados em painel. As duas abordagens são tidas em conta.

Ao tomar os valores agregados para o país inteiro, e tendo em conta que os dados são disponibilizados diariamente, está-se perante uma série temporal onde a unidade é o dia. Da mesma forma, é também possível recolher dados referentes a cada município que

registre atividade no que toca a infeções, pelo que se obtém uma estrutura de dados em painel uma vez que existem valores diários para cada município. Optou-se por, na metodologia de dados em painel, agregar a dimensão temporal em semanas de modo a harmonizar ruído e erros de medição conhecidos (devido ao ritmo de recolha de informação por parte das autoridades) que um painel diário aportaria.

Estabeleceu-se como período de recolha dos dados facultados pela DGS entre o dia 3 março (data do primeiro relatório de situação divulgado pela DGS, em que foram declarados os primeiros casos positivos COVID-19 em Portugal) e o dia 2 de julho.

Como forma de medir a mobilidade individual recorreu-se, como já referido, ao Painel de Mobilidade PSE, um painel pioneiro a nível mundial por ter a capacidade de medir e rastrear a mobilidade da população em contínuo. Tratando-se de um painel representativo da população e ao medir a mobilidade todos os dias, todas as horas, ao longo de todo o ano, possibilita a identificação de um conjunto de indicadores que podem ser tidos em conta para o presente estudo, tais como determinar os hábitos e os padrões de mobilidade, identificar quais os meios de transporte utilizados e traçar perfis sociodemográficos. Os resultados da aplicação do modelo epidemiológico à mobilidade observada têm aplicabilidade nas duas metodologias consideradas, ou seja, podem ser abordados e estudados tanto ao nível macroeconómico e de séries temporais, como ao nível de dados em painel com unidades espaciais.

4.2. Caracterização das Variáveis

Abordando primeiramente a metodologia de dados em painel espacial, a base de dados que se encontra em mãos é constituída pelas seguintes variáveis:

Tabela 2 - Caracterização das variáveis utilizadas na análise espacial

Variável	Descrição
INF_OBS	Número de indivíduos infetados, declarados pela DGS por semana, em cada município
INF_MOB	Número de indivíduos suspeitos de infeção, esperados a partir da mobilidade observada por semana, em cada município

INF_SS	Número de indivíduos suspeitos de infeção, esperados a partir da mobilidade observada apenas dos indivíduos <i>superspreaders</i> por semana, em cada município
---------------	---

As variáveis INF_MOB e INF_SS foram obtidas a partir da estimação do modelo SEQIJR, sendo que correspondem ao número de indivíduos estimados suspeitos de infeção, sintomáticos e assintomáticos.

Optou-se por realizar uma transformação logarítmica aos dados de modo a tornar a relação entre as variáveis mais clara e com uma interpretação mais conciliadora. Para tal, não foram tidas em conta as observações nulas.

Foi necessário estabelecer um critério que tornasse possível a classificação dos indivíduos pertencentes ao Painel de Mobilidade PSE como sendo hipermóveis, ou não. Para tal, recorreu-se ao algoritmo *k-means* como ferramenta de agrupamento dos indivíduos. Tendo como variáveis o número de freguesias e concelhos diferentes que cada sujeito contacta durante um dia foi possível obter os diferentes agrupamentos e classificar os indivíduos pertencentes ao agrupamento pretendido. O algoritmo de agrupamento permitiu a identificação deste grupo que representa 3.27% da população total e que se caracteriza por percorrer, em média, um número de freguesias 4 vezes superior e um número de concelhos cerca de 3 vezes superior, comparativamente aos restantes indivíduos. Quando um indivíduo que é hipermóvel contrai a infeção é denominado de *superspreader*. Com base neste processo de classificação, é possível chegar ao número de suspeitos que foram contaminados por indivíduos *superspreaders*.

Procedendo à estimação do modelo SEQIJR de forma global, ou seja, permitindo que todos os indivíduos pertencentes ao painel possam contrair a infeção e interagir com a população, obtém-se a variável INF_MOB. Se, ao invés deste processo, se admitir que apenas os indivíduos *superspreaders* pertencentes ao painel têm a possibilidade de contrair a infeção e interagir com a população, é obtida a variável INF_SS.

No que diz respeito ao processo de extrapolação da amostra para a população, o mesmo é realizado através da atribuição dos pesos que cada indivíduo, respetivamente, representa em função da probabilidade de seleção na população dado o género, idade, distrito, atividade, classe social e meio de transporte predominante (individual/coletivo).

Ao longo da dimensão temporal de recolha dos dados constituiu-se um painel semanal não balanceado, desde a semana 12 do ano 2020 até à semana 27 do ano 2020, composto por 274 municípios pertencentes ao território continental português.

O concelho é identificado através do seu código administrativo proveniente da Carta Administrativa Oficial de Portugal 2018, denominado por DICO e composto por 4 dígitos onde os primeiros 2 dígitos dizem respeito ao distrito (DI) e os últimos 2 ao concelho (CO).

Na Figura 3 , do Anexo I, é possível analisar o diagrama de dispersão de Moran para a variável logaritmizada do número de indivíduos infetados declarados pela DGS por semana em cada município, através do qual se observa o valor do Índice de Moran Global para esta variável. É de notar que os valores da variável foram standardizados (com média zero e desvio padrão um) e são facultados em unidades de desvio padrão. Da mesma forma, também a variável desfasada espacialmente é calculada para esses valores standardizados. Com um Índice de Moran Global de 0.578, declara-se a existência de uma forte autocorrelação positiva entre a variável LN_INF_OBS e a variável W_LN_INF_OBS. Na Tabela 7, presente no Anexo I, encontra-se o teste de significância para o Índice de Moran Global da variável LN_INF_OBS sob a hipótese nula de ausência de autocorrelação espacial entre os municípios que, como é observável, é rejeitada com um valor-p igual a 0.001. Olhando ainda para a Figura 3, do Anexo I, e para a distribuição ao longo dos quatro quadrantes, é visível que existe uma maior densidade de observações no primeiro e no terceiro quadrantes. Isto é, concelhos com valores altos para a variável em análise encontram-se cercados por municípios que também apresentam valores acima da média para a variável e municípios com valores baixos para a variável em estudo estão cercados por municípios que também apresentam baixos valores. Como tal, este resultado, reforçado com o teste de significância de Moran, conduz à necessidade de se incluir a dependência espacial na estimação.

Da mesma forma que se obtiveram os dados apresentados anteriormente que compõem o painel, também é possível obter séries diárias que possibilitam uma análise macroeconómica referente a todo o território, como um só:

Tabela 3 - Caracterização das variáveis utilizadas na análise macroeconométrica

Variável	Descrição
OBS_DIARIO	Número diário de novos casos observados e declarados pela DGS
INF_POR_BUS	Número diário esperado de novos casos suspeitos de infeção por interação com indivíduos que utilizam os transportes públicos
INF_POR_SS	Número diário esperado de novos casos suspeitos de infeção por interação com indivíduos <i>superspreaders</i>

A representação gráfica, referente a cada uma das séries, pode ser visualizada através das Figuras 6, 7 e 8 presentes no Anexo IV.

Uma vez que o Painel de Mobilidade PSE permite a caracterização dos indivíduos nele pertencentes e a identificação dos padrões e tipos de mobilidade de cada um, é possível identificar que indivíduos utilizam no seu dia a dia os transportes públicos ou, mais do que isso, se se tratam de indivíduos *superspreaders*, ou não. Posteriormente a esta classificação individual, torna-se exequível a aplicação do modelo de propagação epidémica com base no agente (sendo que os valores assumidos para os vários parâmetros utilizados na estimação não sofrem alterações) tendo por base estes diferentes critérios. Deste modo, a variável INF_POR_BUS é obtida através do modelo epidémico com base no agente, mas filtrando na categoria dos indivíduos infetados apenas os utilizadores de transportes públicos, pelo que são tidas em conta apenas as interações decorrentes desses indivíduos com todos os demais. Da mesma forma, é igualmente possível obter a variável INF_POR_SS tomando por base apenas as interações decorrentes dos indivíduos *superspreaders*.

As estatísticas descritivas de cada uma das séries em estudo encontram-se na Tabela 8 e 9 (Anexo II). É de fácil constatação que existe uma diferença relativamente ao nível das diferentes séries, onde aquelas que foram ajustadas através do painel apresentam uma ordem de grandeza muito superior comparativamente às séries observadas de casos positivos. Importa referir que esta desigualdade se deve ao facto de se estar a estimar o número de casos suspeitos de infeção totais, como demonstram os resultados apontados pelo Inquérito Serológico Nacional COVID-19 do Instituto Nacional de Saúde Doutor Ricardo Jorge ao referirem que “as diferenças observadas entre a seroprevalência e a incidência acumulada da infeção por SARS-CoV-2 reportada pelo sistema nacional de

vigilância são consistentes com a evidência de sub-captção de casos ligeiros ou assintomáticos pelos sistemas de vigilância”.

Todos os estudos até então realizados apontam para que o número real de infetados seja amplamente superior ao número que os sistemas de vigilância reportam. Esta conclusão pode ser corroborada atendendo ao estudo do Centro de Biomedicina do Algarve e da Fundação Champalimaud, tratando-se do primeiro estudo serológico realizado em Portugal e que revela a existência de uma taxa de infeção 14 vezes superior ao apontado nos testes de diagnóstico. Também os resultados do Inquérito Serológico Nacional vêm dar ênfase a esta questão, ao constatarem que “à semelhança de outros países, só diagnosticámos 15% dos casos”.

5. MODELAÇÃO E ANÁLISE DOS RESULTADOS

O presente capítulo encontra-se dividido em 4 partes, cada uma contendo uma análise particular e atendendo a um objetivo específico. Pretende-se que os resultados obtidos em cada uma das análises, com o conhecimento disponível à data, possam ser vistos como um contributo de mais evidências estatísticas, fomentando a investigação contínua do movimento pandémico.

5.1. Dinâmica espacial observada explicada pela dinâmica espacial esperada a partir da mobilidade

O foco deste primeiro objetivo prende-se com o determinar se a forma como os indivíduos se movem e interagem está associada à dinâmica espacial de disseminação da doença que é observada e declarada diariamente pelos serviços de saúde. Para tal, foi utilizada a variável estimada que designa o número esperado de infeções, resultante na estimação do modelo a partir da aplicação da mobilidade ao modelo epidemiológico SEQIR.

Uma vez que a DGS apenas começou a divulgar os resultados desagregados por concelho a partir do relatório de situação do dia 23 de março, a série para análise foi construída a partir desta data, que corresponde, no calendário anual, à semana 12 do ano 2020. Como tal, teve-se por base um painel semanal cuja unidade espacial é o município, com início na semana 12 e término na semana 27, uma vez que o período de recolha dos dados terminou no dia 3 de julho. Foram apenas duas as variáveis incluídas neste painel: LN_INF_OBS (logaritmo do número total de casos confirmados publicados pela DGS no

relatório de situação diário) e LN_INF_MOB (logaritmo do número de suspeitos de infecção esperados com base na mobilidade aplicada num modelo epidémico de agente). O objetivo é simples e claro: determinar se a progressão espacial da pandemia de COVID-19 em Portugal depende/não depende da mobilidade dos indivíduos.

O GeoDa é um software que, embora vocacionado para dados seccionais, apresenta funcionalidades que possibilitam a realização de análises espaço-temporais. Estas funções permitem que sejam criadas variáveis no espaço e no tempo a partir de um conjunto de dados seccionais onde é possível observar as variáveis em diferentes momentos no tempo. Como não poderia deixar de ser, o software inclui também funcionalidades para estimar modelos de regressão espacial para dados espaço-temporais. Existe ainda um software, denominado de GeoDaSpace, também desenvolvido pelo Centro GeoDa de Análise Geoespacial e Computação, que apresenta como vantagem a possibilidade de estimação de modelos espaciais que controlam tanto a autocorrelação espacial quanto a heteroscedasticidade.

A Tabela 4 apresenta um resumo que permite a comparação dos resultados obtidos com a estimação de diferentes modelos (através do software GeoDaSpace), incluindo estimativas dos coeficientes e respetivas significâncias estatísticas.

Tabela 4 - Modelos Pooled para a variável LN_INF_OBS explicada pela variável LN_INF_MOB

Método	POOLED			
Modelos	OLS (Modelo Não Espacial)	SLM	SEM	SARAR
CONSTANT	0.1215	-0.6129***	-0.1070	-0.5813***
LN_INF_MOB	0.2571***	0.1551***	0.1621***	0.1607***
T_13	0.8622***	0.8837***	0.8511***	0.8917***
T_14	1.2043***	1.2166***	1.1934***	1.2422***
T_15	1.5205***	1.5399***	1.5499***	1.6149***
T_16	1.7623***	1.7683***	1.7099***	1.7807***
T_17	1.8714***	1.8867***	1.8807***	1.9510***
T_18	1.9520***	1.9688***	1.9882***	2.0609***
T_19	2.0027***	2.0133***	2.0123***	2.0948***
T_20	2.0185***	2.0007***	1.9699***	2.0503***

T_21	2.0216***	1.9978***	1.9278***	2.0059***
T_22	2.0102***	1.9966***	1.9968***	2.0822***
T_23	2.0617***	2.0211***	2.0187***	2.1059***
T_24	2.0753***	2.0222***	1.9666***	2.0486***
T_25	2.1500***	2.1038***	2.0439***	2.1355***
T_26	2.2020***	2.1445***	2.0780***	2.1672***
T_27	2.2139***	2.1823***	2.2054***	2.3018***
W_LN_INF_MOB		0.1087***	0.1027***	0.0999***
W_LN_INF_OBS		0.1949***		0.1727***
Wu			0.6630***	0.5970***
R²	0.3757	0.4504	0.4436	0.5368
Nº Observações	4384	4384	4384	4384

Níveis de Significância: *10%, **5%, ***1%.

É de fácil observação que a inclusão de termos espaciais no modelo melhora substancialmente o seu ajustamento. Kelejian & Prucha (2010) desenvolveram, no contexto espacial, um estimador de mínimos quadrados generalizado a dois passos. O modelo SARAR(1,1), estimado tendo em conta esta inferência robusta que não faz suposições acerca da ausência de autocorrelação no termo de erro ou ausência de heteroscedasticidade, torna-se o modelo mais adequado, ao incluir tanto o desfasamento espacial da variável dependente como permitir a existência de autocorrelação espacial no termo de erro, ambos com uma elevada significância estatística. Analisando o coeficiente associado à variável LN_INF_MOB, e sendo o mesmo altamente significativo, afirma-se que se o número de infetados esperados através da mobilidade aumentar 1%, espera-se que o número de infetados observados aumente em 0.1607591%.

O output do modelo eleito (Output 1) pode ser encontrado no Anexo III, com um R^2 de 0.5368, o que quer dizer que cerca de 54% da variação do logaritmo do número de infetados observados consegue ser explicada pelas variáveis independentes incluídas no modelo. Todas as variáveis são estatisticamente significativas a 1% e, além disso, é curioso notar que o coeficiente associado à variável dependente espacialmente desfasada (W_LN_INF_OBS) apresenta um valor muito próximo do coeficiente associado às infeções causadas pela mobilidade (LN_INF_MOB), indicando assim que os municípios vizinhos ajudam na explicação da variável dependente para um determinado município tanto quanto a variável LN_INF_MOB.

Deste modo, há indícios de significância estatística de que a dinâmica espacial esperada pela mobilidade (medida pela variável LN_INF_MOB) ajuda a explicar a dinâmica espacial verificada nos vários municípios portugueses.

5.2. Dinâmica espacial observada explicada pela dinâmica espacial esperada a partir da mobilidade dos indivíduos superspreaders

Enquanto que no ponto anterior desta secção a variável independente era o logaritmo do número de indivíduos infetados pela mobilidade de forma global, independentemente de quem os infetou ser um indivíduo que se mobiliza muito ou não, agora é adicionada uma nova variável independente que tem em conta apenas o número esperado de infeções causadas por interação com indivíduos *superspreaders*. Então, de forma análoga ao estudado no primeiro ponto, o objetivo final prende-se com a determinação do impacto dos indivíduos *superspreaders* na dinâmica espacial observada de disseminação da doença.

Através da Tabela 5 é possível atentar ao valor dos coeficientes e das respetivas significâncias estatísticas de cada um, em cada um dos modelos gerados.

Tabela 5 - Modelos Pooled para a variável LN_INF_OBS explicada pelas variáveis LN_INF_MOB e LN_INF_SS

Método	POOLED			
	Modelos OLS (Modelo Não Espacial)	SLM	SEM	SARAR
CONSTANT	0.0953	-0.7011***	0.1709**	-0.6910***
LN_INF_MOB	0.2272***	0.1566***	0.1705***	0.1649***
LN_INF_SS	0.1076***	0.0716***	0.0590***	0.0639***
T_13	0.8791***	0.9370***	0.8680***	0.9407***
T_14	1.2368***	1.2989***	1.2670***	1.3410***
T_15	1.5554***	1.6307***	1.6050***	1.7043***
T_16	1.7938***	1.8485***	1.7703***	1.8840***
T_17	1.8988***	1.9742***	1.9324***	2.0424***
T_18	1.9935***	2.0749***	2.0415***	2.1577***
T_19	2.0468***	2.1590***	2.1188***	2.2452***
T_20	2.0575***	2.1264***	2.0835***	2.2036***

T_21	2.0482***	2.0690***	1.9804***	2.1064***
T_22	2.0278***	2.0855***	2.0773***	2.2006***
T_23	2.1090***	2.2067***	2.1674***	2.2992***
T_24	2.1145***	2.1777***	2.0955***	2.2238***
T_25	2.1889***	2.2694***	2.1770***	2.3212***
T_26	2.2446***	2.3032***	2.2091***	2.3497***
T_27	2.2794***	2.4174***	2.3800***	2.5275***
W_LN_INF_MOB		0.1048***	0.1034***	0.0993***
W_LN_INF_SS		0.0521***	0.0357***	0.0388***
W_LN_INF_OBS		0.3391***		0.3121***
Wu			0.3962***	0.4699***
R²	0.3847	0.4094	0.3883	0.5676
Nº Observações	4384	4384	4384	4384

Níveis de Significância: *10%, **5%, ***1%.

Tal como verificado para o tópico anterior, também nesta abordagem se nota a importância da inclusão de termos espaciais e, mais uma vez, o modelo SARAR (1,1) apresenta os melhores resultados de entre todos os modelos estimados.

Através da tabela anterior e do Output 2, que se encontra no Anexo III, é observável que o modelo SARAR(1,1), obtido também através do estimador GS2SLS de Kelejian & Prucha (2010), gera um R^2 de 0.5676 e que todas as variáveis são estatisticamente significativas. Assim, a realidade observada no que à dinâmica espacial de propagação da epidemia diz respeito pode ser conduzida pela forma como os indivíduos *superspreaders* se deslocam e interagem com os demais indivíduos. Analisando o coeficiente associado à variável LN_INF_SS, e sendo o mesmo altamente significativo, afirma-se que se o número de infetados esperados através da mobilidade por interação com indivíduos *superspreaders* aumentar 1%, espera-se que o número de infetados observados aumente em 0.0639%.

5.3. Testar o impacto dos indivíduos que andam de transportes públicos na dinâmica da pandemia ao longo do tempo

Ao longo de todo este período de descoberta e de compreensão da forma como o vírus se transmite pela população, muito se tem falado acerca dos eventuais perigos da utilização dos transportes públicos. Por forma a estudar este tema, o presente trabalho

toma por base uma metodologia VAR (Vetorial Autorregressiva) como mote para o estudo da (não) causalidade à Granger entre o número diário observado de casos positivos COVID-19 e o número de indivíduos suspeitos de infeção por contacto com utilizadores de transportes públicos.

Com o intuito de investigar e obter um parecer acerca deste assunto, foram tidas em consideração duas séries diárias: a primeira série, OBS_DIARIO, representa a série diária de evolução do número de novos casos que são observados e declarados diariamente; por sua vez, a segunda série em mãos, INF_POR_BUS, traduz-se numa série diária de novos indivíduos suspeitos de infeção por interação com indivíduos que, na realização dos seus percursos normais do dia a dia, utilizam como meio de transporte os transportes públicos (construída através da aplicação da mobilidade ao modelo de propagação epidémica com base no agente e interações daí decorrentes). O cronograma de cada uma das séries pode ser observado através das Figuras 6 e 7, presentes no Anexo IV.

Uma vez que o objetivo é estudar a causalidade entre o número de casos declarados e o número de casos suspeitos de indivíduos infetados por quem anda de transportes, deve-se ter em conta tanto a possibilidade da existência de efeitos de feedback como a possibilidade da causalidade não ser contemporânea. Com o termo “efeitos de feedback”, entenda-se que, tanto as infeções podem ser determinadas pelo número suspeito de casos positivos infetados por indivíduos que andam de transportes, como o número suspeito de casos positivos infetados por indivíduos que andam de transportes pode ser determinado pelas infeções. Para que tal seja praticável, é seguida uma abordagem que não impõe qualquer tipo de restrição no que toca à dependência ou independência entre ambas as variáveis.

Os testes de raiz unitária, que podem ser encontrados no Output 3 e 4 do Anexo IV, apresentam evidência de que ambas as séries são estacionárias, após serem tomadas as primeiras diferenças (Δ OBS_DIARIO $\sim I(0)$; Δ INF_POR_BUS $\sim I(0)$).

Uma vez que a incorporação de variáveis exógenas se torna importante na anulação de efeitos de causalidades espúrias, como variável exógena e para incluir as medidas de controlo e prevenção implementadas pelo governo, introduziu-se um indicador (MAIO18) que assume o valor 1 após o dia 18 de maio, dia em que se deu início à segunda fase de desconfinamento em todo o país.

Tendo as variáveis endógenas estacionárias, estabilidade do modelo VAR, e resíduos a seguirem um processo de ruído branco (Lütkepohl, 2006), é viável a aplicação da metodologia VAR como veículo para o principal propósito deste tópico: testar as relações de causalidade.

Atendendo aos critérios de seleção da ordem do modelo VAR, que podem ser observados através do Output 6 no Anexo IV, e uma vez que a maioria indica como 6 o número de lags ideal, o modelo foi gerado tendo em conta esta indicação. O output do modelo estimado pode ser encontrado no Anexo IV (Output 7), onde cada coluna corresponde a uma equação no VAR e cada linha diz respeito a um regressor. Através do Output 8, também no Anexo IV, é possível analisar os resultados do Teste de Causalidade de Granger. Como é de fácil constatação, aos níveis de significância de 0.10, encontra-se evidência estatística para rejeitar a hipótese nula de ausência de relação causal entre as variáveis. A análise demonstra que existem indícios de significância estatística de que a variação do número de infetados suspeitos diários provocados por indivíduos que andam de transportes causa *à Granger* a variação do número de casos diários observados e declarados. Pelo contrário, não se encontra evidência estatística para rejeitar a ausência de causalidade inversa a qualquer nível de confiança.

A Figura 9, do Anexo IV, ao exibir uma representação gráfica da decomposição da variância para ambas as variáveis, permite que se retirem algumas informações pertinentes: num intervalo de 14 dias, a variação do número diário de suspeitos de infeção por contacto com indivíduos que utilizam os transportes públicos chega a explicar cerca de 11% da variação do número de casos positivos diários observados; em oposição, e a ir de encontro ao já concluído pela análise ao teste de causalidade de Granger, os choques do número de casos declarados representam uma mínima parte (cerca de 2%) da variação do número de indivíduos infetados por quem se desloca de transportes públicos.

5.4. Testar o impacto dos indivíduos superspreaders na dinâmica da pandemia ao longo do tempo

Tal como a análise precedente, e continuando a trabalhar com dados diários, também o mesmo procedimento é possível de ser aplicado à série diária de novos casos suspeitos de infeção contraída por interação e contacto com indivíduos *superspreaders*. A série, denominada por INF_POR_SS (construída também através da aplicação da mobilidade

ao modelo de propagação epidémica) e cujo cronograma pode ser observável através da Figura 8 do Anexo IV, é transformada num processo estacionário através da aplicação da primeira diferença, como é verificado pelo Output 5, também presente no Anexo IV, ao exibir o teste de raiz unitária.

No Output 10, do Anexo IV, encontra-se a estimação do modelo VAR de ordem 6 eleito tendo em conta os critérios de seleção para a escolha da ordem do modelo (presentes no Output 6, também do Anexo IV). Atentando ao Output 11, onde se encontra o teste de causalidade de Granger, verifica-se a inexistência, a qualquer nível de significância habitual, de uma relação de causalidade *à Granger* entre as variáveis nos dois sentidos. Assim, não existe evidência estatística que os contágios por *superspreaders* tenham uma relação de causalidade direta com a progressão global da pandemia.³

6. CONCLUSÕES, LIMITAÇÕES E TRABALHO FUTURO

O facto de se adicionar a questão da mobilidade individual permite uma maior aproximação à realidade de comportamentos existentes no quotidiano dos indivíduos dentro de uma população, na medida em que os mesmos interagem entre si e se deslocam espacialmente.

Algumas limitações do presente estudo prendem-se, essencialmente, com a simplicidade inerente ao modelo de disseminação epidémica implementado. Por ter em conta apenas os indivíduos presentes em cada um dos compartimentos que o constituem, o modelo não considera efeitos como a vacinação, a imunidade, os períodos de incubação e, por exemplo, características pessoais como a idade e o género. A falta de informação no que concerne a variáveis responsáveis pelo perfil e caracterização dos infetados e suspeitos corresponde a uma lacuna na comunicação efetuada por parte dos sistemas de saúde. Este aspeto alia-se de igual modo à limitação dos dados apurados pelos sistemas de vigilância que se verifica quando ocorrem interrupções de divulgação de algumas variáveis e relatórios. Como já referido, também a falta de informação acerca dos parâmetros da pandemia atual se considera uma limitação do estudo. Todo este enquadramento pode dar origem a eventuais problemas clássicos de especificação, como

³ Procedeu-se a uma nova análise econométrica, utilizando dados semanais e não diários, através da qual se chegaram às mesmas conclusões aqui apresentadas, ou seja, ausência de causalidade *à Granger* nos dois sentidos.

é caso do viés causado por variáveis omitidas, pelo que, em investigações futuras poderá fazer sentido incluir detalhes adicionais que estejam disponíveis por forma a endereçar esta questão.

Comprova-se, através do teste de Moran, a presença de dependência espacial nos dados semanais referentes ao número de indivíduos infetados em cada município. Da análise espacial realizada, tendo em conta os vários modelos econométricos estimados através do software GeoDaSpace e os desfechos que se podem retirar dos mesmos, a principal conclusão consiste na deteção da relevância da inclusão da mobilidade observada dos indivíduos como variável explicativa da dinâmica espacial de propagação da doença observada. Além da importância desta variável, em todos os modelos se obteve evidência de que os coeficientes estimados associados a termos espaciais são estatisticamente significativos, indicando a existência da dependência espacial e corroborando a importância da não omissão destes efeitos.

Através dos testes efetuados acerca do impacto dos indivíduos *superspreaders* na dinâmica espacial de disseminação da doença, verificou-se a significância estatística do coeficiente associado a esta variável, contribuindo para a explicação da subdeclaração de casos de infeção observados e declarados pela DGS semanalmente. Pelo contrário, quando se aferiu acerca do impacto dos *superspreaders* na evolução global da pandemia, não se encontrou a existência de uma relação causal entre as séries. Assim, uma conclusão interessante que se pode recolher desta diferença é que os indivíduos classificados como *superspreaders* apenas têm impacto no espaço e que, ao longo do tempo, o facto de indivíduos *superspreaders* contaminarem outros indivíduos não ajuda a explicar o número de casos diários declarados. Assim dizendo, existem evidências estatísticas de que os *superspreaders* são agentes de disseminação no espaço e, portanto, têm impacto na forma como, espacialmente, a pandemia evolui, mas cujos contágios individuais não têm impacto na progressão global da pandemia. Estas duas conclusões que à primeira vista poderão parecer contraditórias podem revelar pistas da maior importância, sugerindo que os *superspreaders* poderão ser agentes de dispersão que claramente “plantam” a doença em locais no espaço. São agentes de dispersão, na medida em que saem dos grandes centros, onde se encontra o maior número de infeções, e assumem este papel na disseminação da doença, ao passo que na relação com os grandes números globais de indivíduos infetados, são já contágios secundários que promovem o nível.

Como tal, um dos vetores (de entre outros) que podem servir para controlo da disseminação espacial pode passar pela identificação dos indivíduos *superspreaders*, através da identificação dos perfis de mobilidade dos indivíduos aquando dos inquéritos epidemiológicos.

Em oposição ao evidenciado para o caso dos indivíduos *superspreaders*, o número de suspeitos de infeção causadas por interação com indivíduos que utilizam os transportes públicos já representa uma fatia de responsabilidade na dinâmica global da pandemia. Tal como evidencia o estudo de Troko *et al.* (2011), ao mostrar que o risco de infeção é cerca de seis vezes maior quando os agentes utilizam os transportes públicos, também aqui se conclui, através da causalidade de Granger, que o número esperado de novos casos suspeitos de infeção por interação com indivíduos que andam de transportes causam as infeções observadas. Deste modo, e em tom de conclusão, verificam-se indícios de significância estatística de que não é a progressão geral da pandemia que motiva mais ou menos infeções provocadas por indivíduos que utilizam os transportes públicos, mas sim o contrário: os transportes públicos são um dos fatores disseminadores da pandemia.

É de destacar que, a partir de um número inicial de infetados aleatórios no espaço, o modelo conseguiu criar contágios nos municípios que representam 99.5% das infeções reportadas pela DGS, ou seja, a mobilidade explica a progressão da infeção no território quase na sua totalidade.

Tratando-se de um contributo que fomenta a investigação contínua do movimento pandémico, o presente trabalho sugere que se continuem a desenvolver abordagens que permitam o estudo do fenómeno, tanto do ponto de vista quantitativo como do ponto de vista qualitativo. Estas ações em muito podem ajudar os sistemas nacionais a tomar decisões e a repensar formas de ação. Prever a incidência futura em função da mobilidade, bem como conhecer os perfis de mobilidade dos indivíduos, podem tornar-se contributos decisivos para o acompanhamento da situação pandémica e auxiliar para a tomada de decisões.

Uma outra conclusão relevante, que vai de encontro aos estudos do INSA (Instituto Nacional de Saúde Doutor Ricardo Jorge), prende-se com a determinação do valor esperado de indivíduos infetados no país. Segundo os dados apurados pelo INSA relativos ao primeiro inquérito serológico nacional à imunidade da população para a COVID-19, o

valor real de infetados no país é seis vezes superior ao declarado. No modelo aqui construído, o valor esperado encontrado para o número suspeito de infeções foi de 1.646.331, o que representa cerca de 4 vezes mais o número de casos suspeitos declarados pela DGS no dia 3 de julho de 2020. Este aspeto surge, uma vez mais, alinhado e em concordância com as conclusões manifestadas pelos estudos desenvolvidos em Portugal já apresentados.

Um possível desenvolvimento futuro seria a realização de várias simulações com o intuito de determinar se as conclusões se mantêm inalteradas. A realização de vários ensaios onde se poderia alterar, por exemplo, os critérios iniciais utilizados, tanto para determinação dos indivíduos infetados iniciais, como para os parâmetros de infecciosidade que o modelo pressupõe, possibilitaria uma outra robustez à análise. Tendo a estrutura já definida e a metodologia a aplicar, tal como foi realizado para o primeiro ensaio, o mesmo poderá ser aplicado para os demais.

REFERÊNCIAS

- Almeida, E. (2012). *Econometria Espacial Aplicada*, 1ª Ed. Guanabara: Editora Alínea.
- Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial. *Spatial Analytical*, 4, 111.
- Arbia, G. (2014). *A primer for spatial econometrics with applications in R*. Palgrave Macmillan.
- Bacaër, N. (2011). Daniel Bernoulli, d'Alembert and the inoculation of smallpox (1760). In *A short history of mathematical population dynamics* (pp. 21-30). Springer, London.
- Cliff, A.D. & Ord, J.K. (1972) *Spatial Autocorrelation*, London: Pion
- En'ko, P. D. (1887). On the susceptibility to measles and scarlatina. *Vrach. St. Petersburg*, 8, 723-744.
- Gumel, A. B., Ruan, S., Day, T., Watmough, J., Brauer, F., Van den Driessche, P., ... & Wu, J. (2004). Modelling strategies for controlling SARS outbreaks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1554), 2223-2232.
- Hackl, J., & Dubernet, T. (2019). Epidemic spreading in urban areas using agent-based transportation models. *Future Internet*, 11(4), 92.
- Hamer, W. H. (1906). *Epidemic disease in England: the evidence of variability and of persistency of type*. Bedford Press.
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM review*, 42(4), 599-653.
- Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA). (2020). Inquérito Serológico Nacional COVID-19 indica seroprevalência de 2,9% de infeção por SARS-CoV-2 em Portugal. Disponível em: <http://www.insa.min-saude.pt/inquerito-serologico-nacional-covid-19-indica-seroprevalencia-de-29-de-infecao-por-sars-cov-2-em-portugal/>
- Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), 99-121.

- Kelejian, H. H., & Prucha, I. R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140(1), 131-154.
- Khalil, K. M., Abdel-Aziz, M., Nazmy, T. T., & Salem, A. B. M. (2012). An agent-based modeling for pandemic influenza in Egypt. In *Handbook on Decision Making* (pp. 205-218). Springer, Berlin, Heidelberg.
- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772), 700-721.
- Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G. M., ... & Lui, S. F. (2003). A major outbreak of severe acute respiratory syndrome in Hong Kong. *New England Journal of Medicine*, 348(20), 1986-1994.
- LeSage, J. P. (1999). The theory and practice of spatial econometrics. *University of Toledo. Toledo, Ohio*, 28(11).
- Lütkepohl, H. (2006). Structural vector autoregressive analysis for cointegrated variables. *Allgemeines Statistisches Archiv*, 90(1), 75-88.
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17-23.
- Sirirapaiwan, S., Moore, E. J., & Koonprasert, S. (2018). Generalized reproduction numbers, sensitivity analysis and critical immunity levels of an SEQIJR disease model with immunization and varying total population size. *Mathematics and Computers in Simulation*, 146, 70-89.
- Timpka, T., Eriksson, H., Gursky, E. A., Nyce, J. M., Morin, M., Jenvald, J., ... & Ekberg, J. (2009). Population-based simulations of influenza pandemics: validity and significance for public health policy. *Bulletin of the World Health Organization*, 87, 305-311.
- Troko, J. et al. (2011). *Is public transport a risk factor for acute respiratory infection?* 11^aEd. BMC Infect. Dis.
- Wooldridge, M. (2009). *An introduction to multiagent systems*. John Wiley & Sons.

ANEXO I

Tabela 6 - Definição e valores dos parâmetros utilizados na estimação do modelo SEQJR

Parâmetro	Definição	Valor
β	Taxa de infecciosidade e taxa de contato entre um indivíduo suscetível e um indivíduo infecioso	0.23
γ_1	Taxa de quarentena de indivíduos assintomáticos expostos	0.1
γ_2	Taxa de isolamento de indivíduos infecciosos	0.5
σ_1	Taxa de recuperação de indivíduos infecciosos	0.0413
σ_2	Taxa de recuperação de indivíduos isolados	0.0431
k_1	Taxa de desenvolvimento de sintomas em indivíduos assintomáticos expostos	0.1
k_2	Taxa de desenvolvimento de sintomas em indivíduos em quarentena	0.125
d_1	Taxa de mortalidade induzida pela doença em indivíduos infecciosos	0.0055
d_2	Taxa de mortalidade induzida pela doença em indivíduos isolados	0.0041

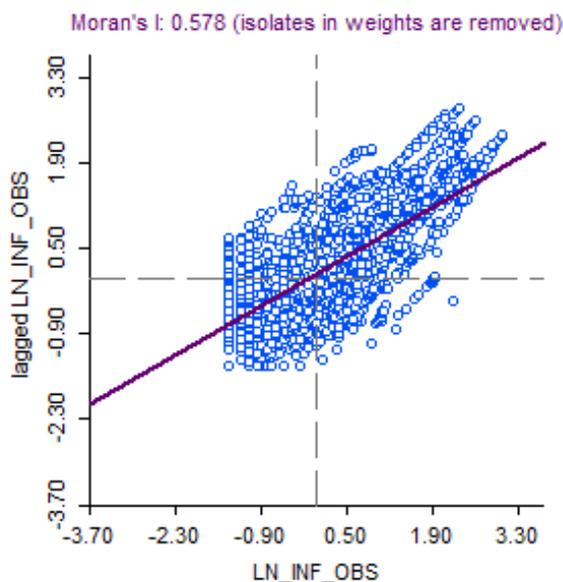


Figura 3 - Diagrama de Dispersão de Moran da variável LN_INF_OBS – Semana 12 a 27, 2020

Fonte: Elaboração própria; Software: GeoDa

Tabela 7 - Teste de significância para o Índice de Moran Global da variável LN_INF_OBS sob a hipótese nula de ausência de autocorrelação espacial.

Estatística Teste	Valor-p
48.4374	0.001

Fonte: Elaboração própria; Software: GeoDa

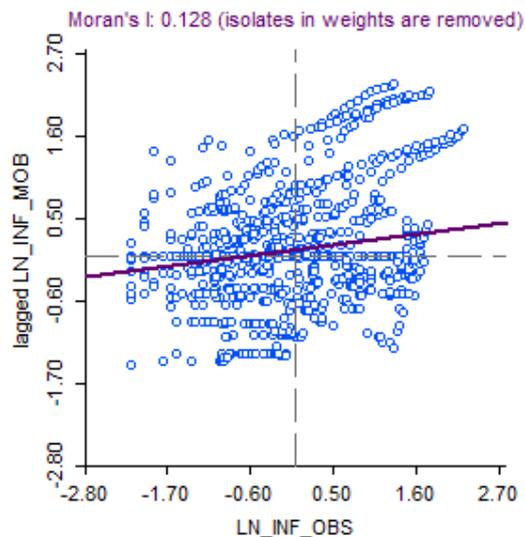


Figura 4 - Estatística I de Moran Global Bivariada entre LN_INF_OBS e LN_INF_MOB

Fonte: Elaboração própria; Software: GeoDa

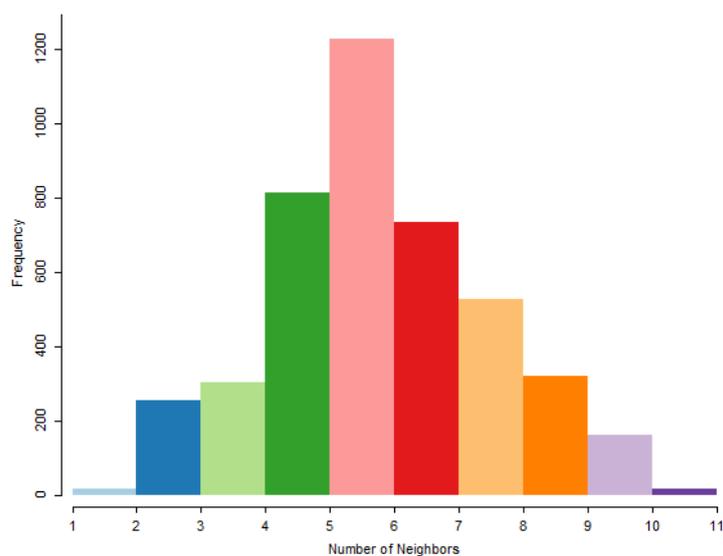


Figura 5 - Histograma do número de vizinhos da matriz de contiguidade baseada no critério da rainha

Fonte: Elaboração própria; Software: GeoDa

ANEXO II

Tabela 8 - Estatística Descritiva das variáveis utilizadas na análise espacial

Variável	Média	Desvio-Padrão	Mínimo	Máximo
INF_OBS	96.91961	268.1274	0	3573
INF_MOB	12074.18	51333.81	0	738345
INF_SS	763.3254	4638.881	0	118143.8

Fonte: Elaboração própria

Tabela 9 – Estatística Descritiva das variáveis utilizadas na análise macroeconômica

Variável	Média	Desvio-Padrão	Mínimo	Máximo
OBS_DIARIO	350.6721	244.889533	2	1516
INF_POR_BUS	14657.85	10062.25938	1339.059	41713.24
INF_POR_SS	4500.717	3170.465059	150.7707	13069.55

Fonte: Elaboração própria

ANEXO III

Output 1 - Modelo SARAR (1,1) com variável explicativa LN_INF_MOB

```

REGRESSION
-----|
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
-----|
Data set           :PANEL SPACE TIME.csv
Weights matrix     :File: WEIGHTS SPACE TIME.gal
Dependent Variable : LN_INF_OBS           Number of Observations: 4384
Mean dependent var : 2.4199              Number of Variables   : 19
S.D. dependent var : 2.0220              Degrees of Freedom    : 4365
Pseudo R-squared   : 0.5368
Spatial Pseudo R-squared: 0.4489
N. of iterations   : 1                   Steploc computed      : Yes
-----|
Variable           Coefficient      Std.Error      z-Statistic     Probability
-----|-----|-----|-----|-----|
CONSTANT           -0.5813640      0.1134755     -5.1232541     0.0000003
LN_INF_MOB         0.1607591      0.0063953     25.1369003     0.0000000
T_13               0.8917905      0.0936534     9.5222442     0.0000000
T_14               1.2422038      0.0934159     13.2975641     0.0000000
T_15               1.6149356      0.0945673     17.0770949     0.0000000
T_16               1.7807539      0.0948449     18.7754237     0.0000000
T_17               1.9510370      0.0950419     20.5281809     0.0000000
T_18               2.0609536      0.0925059     22.2791476     0.0000000
T_19               2.0948631      0.0929716     22.5322879     0.0000000
T_20               2.0503426      0.0910408     22.5211455     0.0000000
T_21               2.0059902      0.0918712     21.8348016     0.0000000
T_22               2.0822279      0.0899901     23.1384121     0.0000000
T_23               2.1059172      0.0907604     23.2030422     0.0000000
T_24               2.0486918      0.0895999     22.8648955     0.0000000
T_25               2.1355550      0.0909822     23.4722264     0.0000000
T_26               2.1672173      0.0921486     23.5187195     0.0000000
T_27               2.3018617      0.0948309     24.2733253     0.0000000
W_LN_INF_MOB      0.0999975      0.0065039     15.3750683     0.0000000
W_LN_INF_OBS      0.1727140      0.0350866     4.9225102     0.0000009
lambda            0.5970857      0.0267039     22.3595261     0.0000000
-----|
Instrumented: W_LN_INF_OBS
Instruments: W_LN_INF_MOB, W_T_13, W_T_14, W_T_15, W_T_16, W_T_17, W_T_18,
             W_T_19, W_T_20, W_T_21, W_T_22, W_T_23, W_T_24, W_T_25, W_T_26,
             W_T_27, W_W_LN_INF_MOB
===== END OF REPORT =====

```

Fonte: Elaboração própria; Software: GeoDaSpace

Output 2 - Modelo SARAR (1,1) com variáveis explicativas LN_INF_MOB e
LN_INF_SS

```

REGRESSION
-----
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
-----
Data set           :PANEL SPACE TIME.csv
Weights matrix     :File: WEIGHTS SPACE TIME.gal
Dependent Variable : LN_INF_OBS           Number of Observations: 4384
Mean dependent var : 2.4199              Number of Variables   : 20
S.D. dependent var : 2.0220              Degrees of Freedom    : 4364
Pseudo R-squared  : 0.5676
Spatial Pseudo R-squared: 0.4062
N. of iterations   : 1                   Stepic computed       : Yes
|
-----
Variable          Coefficient      Std.Error      z-Statistic     Probability
-----
CONSTANT          -0.6910497      0.0995162     -6.9440954     0.0000000
LN_INF_MOB        0.1649329      0.0069730     23.6529328     0.0000000
LN_INF_SS         0.0639002      0.0101336     6.3057547      0.0000000
T_13              0.9407750      0.0966816     9.7306546      0.0000000
T_14              1.3410220      0.0964582     13.9026298     0.0000000
T_15              1.7043398      0.0961252     17.7304146     0.0000000
T_16              1.8840133      0.0967908     19.4647911     0.0000000
T_17              2.0424770      0.0977453     20.8959143     0.0000000
T_18              2.1577485      0.0954130     22.6148346     0.0000000
T_19              2.2452574      0.0961619     23.3487149     0.0000000
T_20              2.2036665      0.0915118     24.0806730     0.0000000
T_21              2.1064186      0.0938604     22.4420394     0.0000000
T_22              2.2006381      0.0915752     24.0309344     0.0000000
T_23              2.2992439      0.0928886     24.7527075     0.0000000
T_24              2.2238135      0.0918616     24.2082977     0.0000000
T_25              2.3212027      0.0929052     24.9846325     0.0000000
T_26              2.3497826      0.0940020     24.9971618     0.0000000
T_27              2.5275988      0.0974207     25.9451932     0.0000000
W_LN_INF_OBS     0.3121935      0.0308521     10.1190488     0.0000000
W_LN_INF_SS      0.0388578      0.0071947     5.4008625      0.0000001
lambda           0.4699610      0.0316570     14.8454202     0.0000000
-----
Instrumented: W_LN_INF_OBS
Instruments: W_LN_INF_MOB, W_LN_INF_SS, W_T_13, W_T_14, W_T_15, W_T_16,
             W_T_17, W_T_18, W_T_19, W_T_20, W_T_21, W_T_22, W_T_23, W_T_24,
             W_T_25, W_T_26, W_T_27, W_W_LN_INF_SS
===== END OF REPORT =====

```

Fonte: Elaboração própria; Software: GeoDaSpace

ANEXO IV



Figura 6 - Cronograma da variável OBS_DIARIO

Fonte: Elaboração própria; Software: EViews

Número diário de novos casos infectados por indivíduos que usam os transportes públicos

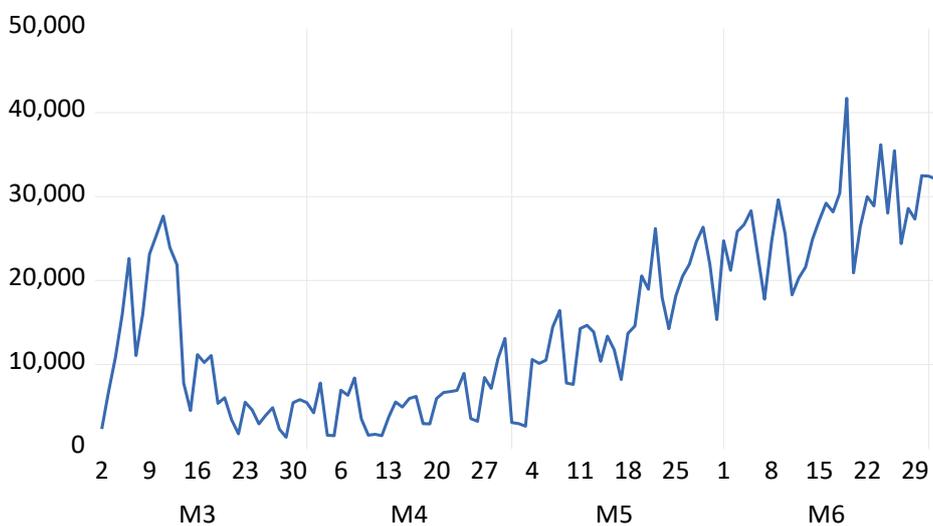


Figura 7 - Cronograma da variável INF_POR_BUS

Fonte: Elaboração própria; Software: EViews

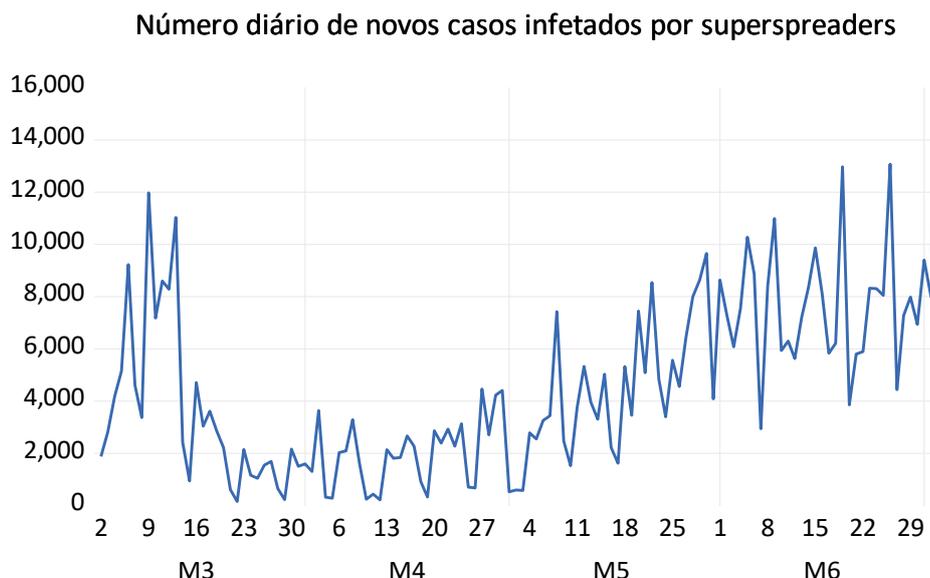


Figura 8 - Cronograma da variável INF_POR_SS

Fonte: Elaboração própria; Software: EViews

Output 3– Teste de Raiz Unitária para a variável D_OBS_DIARIO

Null Hypothesis: D_OBS_DIARIO has a unit root
 Exogenous: None
 Lag Length: 0 (Automatic - based on SIC, maxlag=12)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-18.81047	0.0000
Test critical values:		
1% level	-2.584375	
5% level	-1.943516	
10% level	-1.614956	

*MacKinnon (1996) one-sided p-values.

Fonte: Elaboração própria; Software: EViews

Output 4 -Teste de Raiz Unitária para a variável D_INF_POR_BUS

Null Hypothesis: D_INF_POR_BUS has a unit root
 Exogenous: None
 Lag Length: 6 (Automatic - based on SIC, maxlag=12)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.928108	0.0000
Test critical values:		
1% level	-2.585050	
5% level	-1.943612	
10% level	-1.614897	

*MacKinnon (1996) one-sided p-values.

Fonte: Elaboração própria; Software: EViews

Output 5 - Teste de Raiz Unitária para a variável D_INF_POR_SS

Null Hypothesis: D_INF_POR_SS has a unit root
 Exogenous: None
 Lag Length: 5 (Automatic - based on SIC, maxlag=12)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.271939	0.0000
Test critical values:		
1% level	-2.584877	
5% level	-1.943587	
10% level	-1.614912	

*MacKinnon (1996) one-sided p-values.

Fonte: Elaboração própria; Software: EViews

Output 6 – Seleção da ordem do modelo VAR (bus)

VAR Lag Order Selection Criteria
 Endogenous variables: D(OBS_DIARIO) D(INF_POR_BUS)
 Exogenous variables: C MAIO18
 Sample: 3/01/2020 7/03/2020
 Included observations: 107

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-1770.442	NA	8.70e+11	33.16714	33.26706	33.20765
1	-1747.703	43.77783	6.13e+11	32.81688	33.01672	32.89789
2	-1737.900	18.50599	5.50e+11	32.70842	33.00818*	32.82994*
3	-1732.838	9.367767	5.39e+11	32.68856	33.08824	32.85058
4	-1727.455	9.760029	5.26e+11	32.66271	33.16230	32.86524
5	-1724.033	6.076366	5.32e+11	32.67351	33.27303	32.91655
6	-1713.732	17.90683*	4.73e+11*	32.55573*	33.25517	32.83927
7	-1709.966	6.405969	4.76e+11	32.56010	33.35945	32.88415
8	-1709.057	1.512067	5.05e+11	32.61788	33.51715	32.98243
9	-1708.074	1.597443	5.36e+11	32.67429	33.67347	33.07934
10	-1705.347	4.333221	5.51e+11	32.69807	33.79718	33.14364
11	-1702.187	4.901780	5.61e+11	32.71378	33.91281	33.19985
12	-1699.033	4.776067	5.73e+11	32.72958	34.02853	33.25616
13	-1693.139	8.703404	5.56e+11	32.69418	34.09305	33.26126
14	-1689.908	4.649426	5.67e+11	32.70857	34.20735	33.31615

* indicates lag order selected by the criterion
 LR: sequential modified LR test statistic (each test at 5% level)
 FPE: Final prediction error
 AIC: Akaike information criterion
 SC: Schwarz information criterion
 HQ: Hannan-Quinn information criterion

Fonte: Elaboração própria; Software: EViews

Output 7 – Estimação do modelo VAR(6) (bus)

Vector Autoregression Estimates
 Sample (adjusted): 3/10/2020 7/02/2020
 Included observations: 115 after adjustments
 Standard errors in () & t-statistics in []

	D(OBS_DIARIO)	D(INF_POR_BUS)

D(OBS_DIARIO(-1))	-0.614724 (0.09513) [-6.46197]	-2.150897 (2.62884) [-0.81819]
D(OBS_DIARIO(-2))	-0.238424 (0.11178) [-2.13290]	-4.452354 (3.08907) [-1.44133]
D(OBS_DIARIO(-3))	-0.250613 (0.11327) [-2.21257]	-4.318844 (3.13008) [-1.37979]
D(OBS_DIARIO(-4))	-0.211857 (0.11411) [-1.85658]	-2.320139 (3.15339) [-0.73576]
D(OBS_DIARIO(-5))	-0.052945 (0.11339) [-0.46692]	-1.036024 (3.13349) [-0.33063]
D(OBS_DIARIO(-6))	0.168362 (0.09528) [1.76705]	0.907948 (2.63297) [0.34484]
D(INF_POR_BUS(-1))	0.001886 (0.00339) [0.55700]	-0.464342 (0.09359) [-4.96149]
D(INF_POR_BUS(-2))	0.005864 (0.00365) [1.60595]	-0.350343 (0.10091) [-3.47200]
D(INF_POR_BUS(-3))	-0.002777 (0.00378) [-0.73438]	-0.364080 (0.10451) [-3.48384]
D(INF_POR_BUS(-4))	-0.000536 (0.00374) [-0.14346]	-0.304325 (0.10334) [-2.94484]
D(INF_POR_BUS(-5))	-0.003727 (0.00355) [-1.04969]	-0.203575 (0.09811) [-2.07506]
D(INF_POR_BUS(-6))	-0.008409 (0.00335) [-2.51340]	-0.275356 (0.09245) [-2.97842]
C	6.948013 (18.3773) [0.37808]	-172.3256 (507.846) [-0.33933]
MAIO18	0.407421 (29.9053) [0.01362]	1420.555 (826.414) [1.71894]
R-squared	0.425161	0.286216
Adj. R-squared	0.351171	0.194343
Sum sq. resids	2370189.	1.81E+09
S.E. equation	153.1901	4233.310
F-statistic	5.746251	3.115342
Log likelihood	-734.3570	-1116.049
Akaike AIC	13.01490	19.65302
Schwarz SC	13.34907	19.98719
Mean dependent	2.773913	77.96075

S.D. dependent	190.1804	4716.339
Determinant resid covariance (dof adj.)		4.18E+11
Determinant resid covariance		3.23E+11
Log likelihood		-1850.119
Akaike information criterion		32.66294
Schwarz criterion		33.33127
Number of coefficients		28

Fonte: Elaboração própria; Software: EViews

Output 8 - Teste de Causalidade de Granger (bus)

VAR Granger Causality/Block Exogeneity Wald Tests

Sample: 3/01/2020 7/03/2020

Included observations: 115

Dependent variable: D(OBS_DIARIO)

Excluded	Chi-sq	df	Prob.
D(INF_POR_BUS)	12.47415	6	0.0522
All	12.47415	6	0.0522

Dependent variable: D(INF_POR_BUS)

Excluded	Chi-sq	df	Prob.
D(OBS_DIARIO)	3.351681	6	0.7636
All	3.351681	6	0.7636

Fonte: Elaboração própria; Software: EViews

Variance Decomposition using Cholesky (d.f. adjusted) Factors

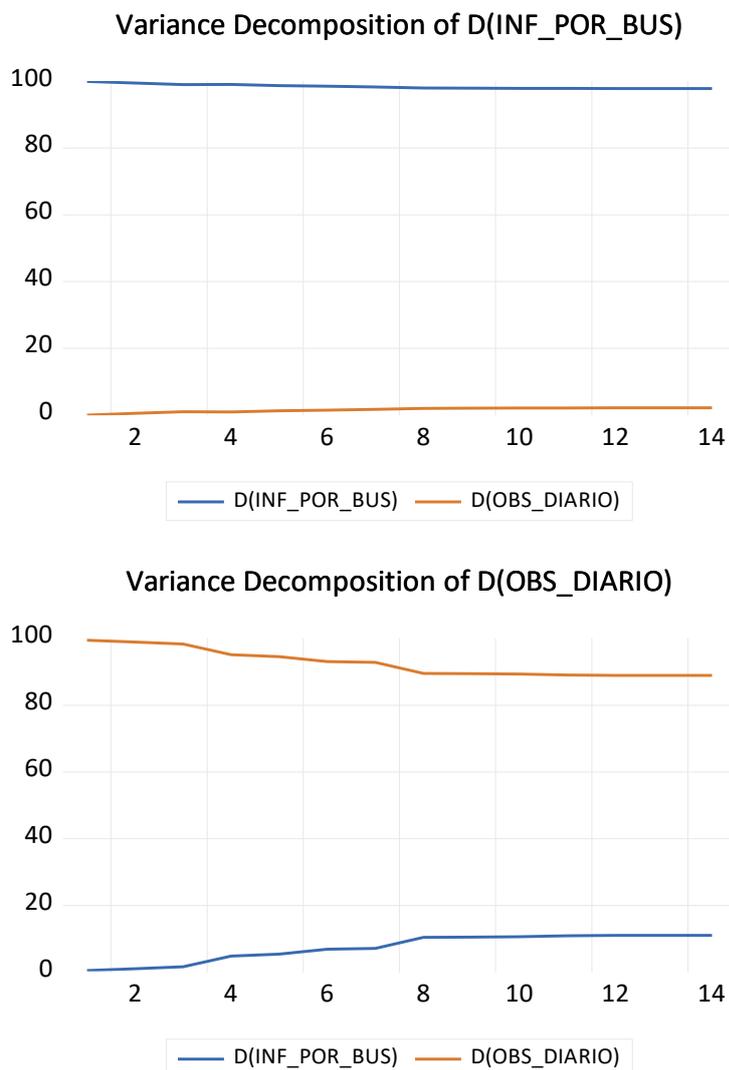


Figura 9 - Decomposição da variância (bus)

Fonte: Elaboração própria; Software: EViews

Output 9 - Seleção da ordem do modelo VAR (superspreaders)

VAR Lag Order Selection Criteria
 Endogenous variables: D(OBS_DIARIO) D(INF_POR_SS)
 Exogenous variables: C MAIO18
 Sample: 3/01/2020 7/03/2020
 Included observations: 107

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-1706.401	NA	2.63e+11	31.97012	32.07004	32.01063
1	-1679.344	52.09166	1.71e+11	31.53914	31.73898	31.62015
2	-1667.356	22.63259	1.47e+11	31.38982	31.68958*	31.51134
3	-1662.591	8.816150	1.45e+11	31.37554	31.77521	31.53756
4	-1656.697	10.68705	1.40e+11	31.34013	31.83972	31.54266
5	-1652.958	6.638221	1.41e+11	31.34502	31.94453	31.58805
6	-1638.413	25.28392*	1.16e+11*	31.14791*	31.84735	31.43146*
7	-1636.756	2.818442	1.21e+11	31.19171	31.99106	31.51576

8	-1634.391	3.935514	1.25e+11	31.22226	32.12153	31.58681
9	-1631.665	4.432717	1.29e+11	31.24607	32.24526	31.65113
10	-1626.284	8.549413	1.26e+11	31.22026	32.31936	31.66582
11	-1624.190	3.247848	1.31e+11	31.25589	32.45492	31.74196
12	-1621.742	3.706150	1.35e+11	31.28490	32.58385	31.81148
13	-1618.439	4.878360	1.38e+11	31.29792	32.69678	31.86500
14	-1614.486	5.688462	1.38e+11	31.29881	32.79759	31.90640

* indicates lag order selected by the criterion

LR: sequential modified LR test statistic (each test at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hannan-Quinn information criterion

Fonte: Elaboração própria; Software: EViews

Output 10 - Estimação do modelo VAR(6) (superspreaders)

Vector Autoregression Estimates

Date: 10/05/20 Time: 19:58

Sample (adjusted): 3/10/2020 7/02/2020

Included observations: 115 after adjustments

Standard errors in () & t-statistics in []

	D(OBS_DIARIO)	D(INF_POR_SS)
D(OBS_DIARIO(-1))	-0.624560 (0.09509) [-6.56786]	-0.705298 (1.23936) [-0.56908]
D(OBS_DIARIO(-2))	-0.243327 (0.11219) [-2.16893]	-1.173785 (1.46215) [-0.80278]
D(OBS_DIARIO(-3))	-0.257490 (0.11225) [-2.29380]	-1.796280 (1.46303) [-1.22778]
D(OBS_DIARIO(-4))	-0.239659 (0.11303) [-2.12032]	-1.818905 (1.47313) [-1.23472]
D(OBS_DIARIO(-5))	-0.059254 (0.11388) [-0.52034]	-0.944341 (1.48416) [-0.63628]
D(OBS_DIARIO(-6))	0.166873 (0.09613) [1.73595]	0.420721 (1.25284) [0.33581]
D(INF_POR_SS(-1))	0.004740 (0.00666) [0.71142]	-0.731789 (0.08683) [-8.42809]
D(INF_POR_SS(-2))	0.010475 (0.00824) [1.27093]	-0.609615 (0.10742) [-5.67485]
D(INF_POR_SS(-3))	-0.003273 (0.00907) [-0.36087]	-0.469184 (0.11820) [-3.96940]
D(INF_POR_SS(-4))	-0.007911 (0.00895)	-0.383075 (0.11667)

		[-0.88369]	[-3.28329]
D(INF_POR_SS(-5))	-0.012196 (0.00804) [-1.51712]		-0.347208 (0.10477) [-3.31404]
D(INF_POR_SS(-6))	-0.017429 (0.00676) [-2.57675]		-0.407504 (0.08815) [-4.62261]
C	6.919567 (18.5893) [0.37223]		-140.0503 (242.277) [-0.57806]
MAIO18	0.650117 (29.8858) [0.02175]		541.0987 (389.505) [1.38919]
R-squared	0.412517		0.488454
Adj. R-squared	0.336900		0.422612
Sum sq. resids	2422323.		4.11E+08
S.E. equation	154.8657		2018.382
F-statistic	5.455368		7.418525
Log likelihood	-735.6080		-1030.870
Akaike AIC	13.03666		18.17165
Schwarz SC	13.37083		18.50581
Mean dependent	2.773913		-34.65706
S.D. dependent	190.1804		2656.254
Determinant resid covariance (dof adj.)			9.70E+10
Determinant resid covariance			7.48E+10
Log likelihood			-1766.076
Akaike information criterion			31.20132
Schwarz criterion			31.86965
Number of coefficients			28

Fonte: Elaboração própria; Software: EViews

Output 11 - Teste de Causalidade de Granger (superspreaders)

VAR Granger Causality/Block Exogeneity Wald Tests

Sample: 3/01/2020 7/03/2020

Included observations: 115

Dependent variable: D(OBS_DIARIO)

Excluded	Chi-sq	df	Prob.
D(INF_POR_SS)	10.03193	6	0.1233
All	10.03193	6	0.1233

Dependent variable: D(INF_POR_SS)

Excluded	Chi-sq	df	Prob.
D(OBS_DIARIO)	2.778010	6	0.8361
All	2.778010	6	0.8361

Fonte: Elaboração própria; Software: EViews